OPEN ACCESS
International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-12 Issue-1, May 2023

Machine Translation on Dravidian Languages

B. V. Kiranmayee, Raparthi Sai Priya, Rayapurthi Vijaya, Palthiya Suresh, Regulapati Venkat Goutham

Abstract: ***The Dravidian languages are spoken all over the world. Despite their distinctiveness, Dravidian languages haven't gotten much attention because there aren't enough resources to handle tasks like translation that require language technology. Since Dravidian languages are largely spoken in southern India, machine translation is necessary. For those who speak these regional languages, this would improve information creation and access. It can be challenging to translate between languages, particularly that of Dravidian, because of lexical divergence, ambiguity, and other, lexical, syntactic and semantic issues. This research looks into a number of machine translation models for different languages, conducts a thorough literature review on the various machine translation techniques from earlier studies, and analyses their methodology. The major objective of this research is to evaluate the viability and effectiveness of a machine translation process for Dravidian languages.***

Keywords: Linguistic Rules, Long Term Dependencies, Bilingual Text Corpora, BLEU Metric, Language Embedding, Translation Memory, andParallel Corpora.

I. INTRODUCTION

India is a multilingual nation with a very diversified population. India is the country with the second-highest number of languages in the world since people from different regions speak their own regional tongues (T. Madhavi Kumari et al [33]). Approximately 80 different variations of the Dravidian language, are spoken by 200 million people worldwide (Sai Koneru et al). In southern India, four Dravidian languages that are well-known literary languages are Telugu, Tamil, Kannada, and Malayalam. Each of thesehas a long, well-preserved history and its unique script.

Manuscript received on 15 March 2023 | Revised Manuscript received on 22 March 2023 | Manuscript Accepted on 15 May 2023 | Manuscript published on 30 May 2023.
*Correspondence Author(s)
Dr. B. V. Kiranmayee*, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. E-mail: kiranmayee_bv@vnrvjiet.in, ORCID ID: https://orcid.org/0000-0002-7387-9592
Palthiya Suresh, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: https://orcid.org/0009-0009-3050-5206
Raparthi Sai Priya, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: https://orcid.org/0009-0005-6075-6074
Rayapurthi Vijaya*, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: https://orcid.org/0009-0003-7981-5308
Regulapati Venkat Goutham, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: https://orcid.org/0000-0002-7335-776X

They makeup the vast majority of Dravidian speakers and are the core of the linguistic nations of Andhra Pradesh, Tamil Nadu, Karnataka, and Kerala All have liberally drawn from Sanskrit. Of all the Dravidian languages, Tamil is the most commonly used and has the most extensive literary tradition. Only Sanskrit's literature can be compared to Tamil's in India in terms of its antiquity, wealth, and geographical reach among the Dravidian languages. Dravidian Languages are regional languages of southern states of India where most of the people only speak their regional language.It is essential to interpret the languages of different cultures for understanding their rituals and business activities. Despite being one of the largest language families in the world, the Dravidian languages are not particularly well-known due to a lack of resources for language technology activities like translation. Language translation is important in the realm of text-processing applications, such as information extraction, machine learning, and natural language comprehension.

Machine translation (MT) refers to an automated system that analyses text from a Source Language (SL), conducts calculations on that input, and then outputs similar content in a required Target Language (TL), ideally without any human participation. (Nadeem Jadoon Khan et al [40]) Numerous techniques are used in machine translation, including Neural machine translation (NMT), Trans- former model translation, Hybrid approach model, Example-based translation (EBMT), Multilingual model, Rule-based machine translation (RBMT), Statistical machine translation (SMT), and Long- Short Term Memory model (LSTM). Given the enormous variety of regional languages spoken in India, MT is crucial there. At least30 different languages and over 2000 dialects are recognized as existing (Nithya B et al [11]). The goal of translating across Dravidian languages is to improve information generation and access for the region's monolingual speakers. The selection of the best machine translation model for Dravidian languages from the pool of potential models is the main objective of this research effort.

1.1. Applications Of Machine Translation

Applications for machine translation are being used in more industries as they achieve noticeably high levels of accuracy, leading to the development of new tools and enhanced machine-learning models. Because it saves time, lowers costs, and boosts production, machine translation (MT) is quickly becoming a crucial instrument in the translation industry. Even so, it still needs improvement before it can equal the accuracy of human translation. The ideal option in light of this is to incorporate both machine translation anda post-editor who is a human.

This will produce the best results by combining machine speed and human linguist accuracy. Machine translation software is already a reliable choice for corporations when it comes to translating words. It provides a lot of advantages for diverse businesses with different translation demands and is quick, affordable, and constantly improving.

A. Technical translations: Technical texts benefit greatly from machine translation (MT). Technical texts frequently include short sentences with repeated technical terms, and MT is ideal in this aspect. Since technical writing is frequently straightforward and repetitious, MT software can easily handle it and maintain your terminology consistency. A qualified linguist will assist with further post-editing to improve the grammar and spelling for a composition that is cogent and cohesive. You may then save time and improve consistency across all upcoming projects by using this post-edited information to train specific MT engines for your projects.

B. Legal translations: In order to conduct business, legal documents including contracts, policies, confidentiality agreements, and court transcripts are just a few examples. When it comes to translation in the legal industry, volume is frequently one of the largest obstacles, which is where MT comes into play. Machine translation technology can speed up the understanding and processing of a lot of legal papers for law firms and corporate legal departments. The majority of the volume can be handled by machines, but there is no room for error when translating legal documents. It is crucial that translated documents accurately represent the meaning, organization, and language of the original text when dealing with legal issues. Here, NMT can really excel by offering quick legal translations without compromising quality.

C. E-commerce industry: Due to its digital nature, it is significantly simpler for an e-commerce company to expand into new areas and nations. Any retailer aiming to expand its customer base and enhance sales should now consider international expansion as a basic strategy. To absolutely take benefit of this worldwide potential, each piece of content material on an e-trade internet site ought to be translated into the language of the meant audience. This will boom agree with further to create your services or products extra handy to a much wider audience. E-commerce websites are an ideal choice for automatic translation software due to the enormous amount of online material and the standardized structure of websites. MT is also excellent for developing focused advertising campaigns that have a predetermined format and can increase your exposure in a new area. Businesses may easily over- come language barriers thanks to MT, which will help them build their brand awareness in new regions and boost sales.

D. Fraud investigations: The use of MT technology in fraud investigations is another situation where it excels. Every type of business, in both the public and private sectors, is susceptible to fraud. These companies might be the targets of fraud themselves, or con artist might be exploiting their name to deceive unwary others. A corporation will conduct a fraud investigation when it has suspicions of any fraudulent activity in order to acquire information and safeguard the business from harm. By assisting businesses in identifying and resolving potential fraud and unethical behavior, translation services frequently play significant roles throughout these investigations. By swiftly and efficiently translating emails, MT technology is excellent for spotting fraudulent behavior early on. Businesses can lessen potential financial and reputational harm by adopting this action early on. The translations can then be evaluated and verified by the translation company if any emails are found to be crucial to the investigation.

II. ABOUT MODELS

Machine translation is defined as translating source language into target language without human interventions. There are many approaches for machine translation such as Rule-based machine translation, Example-based machine translation, Neural machine translation, Hybrid machine translation, Transformer model, Statistical machine translation and Long Short-Term memory model which are described below in detail about each model.

A. Rule-Based Machine Translation (RBMT)

The first professional automatic translation system, rule-based ma-chine translation (RBMT), is built on syntactic ideas that enable words to be arranged variously and to have varied meanings ac- cording to the context. Three distinct uses of RBMT technology are tested on a large collection of language norms: analysis, propagation, and generation. Programmers and linguists spent a lot of work understanding and mapping the rules between the two languages while creating the rules. RBMT uses manually compiled translation dictionaries, some of which can be altered and enhanced by the user, to enhance translations. The translations are nonetheless readable because they are provided by rules, but they are frequently less fluid, giving the output style a more "machine-like" reading experience. When a translation’s text is understandable but requires intensive post-editing to be customized for a particular target audience and writing style, this is known as "gist" value. The RBMT developers, who are presenting their products as hybrid machine translation (MT) models in an effort to get around some of RBMT’s shortcomings, have lately added a number of statistical machine translation (SMT) techniques to the basic RBMT technology.

Retrieval Number: 100.1/ijrte.A75380512123
DOI: 10.35940/ijrte.A7538.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

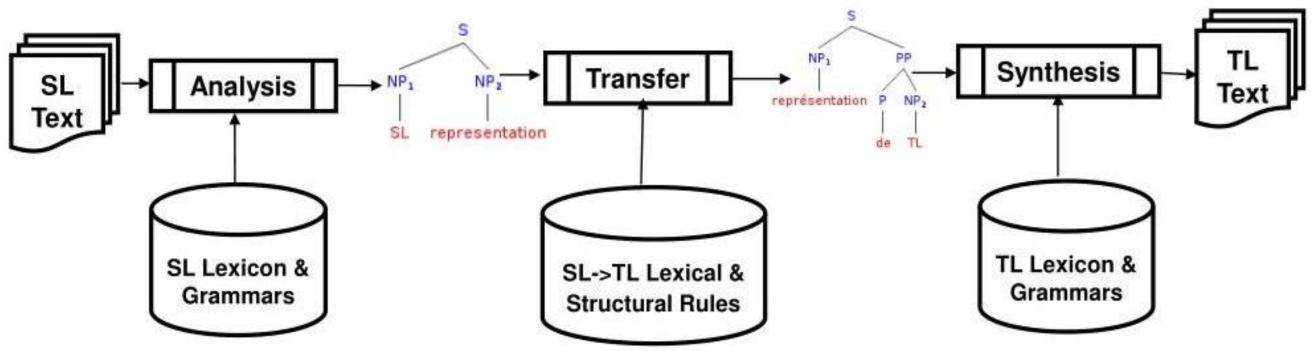

Figure 1: Rule based Machine Translation

B. Neural Machine Translation (NMT)

Neural network technology is used in contemporary machine translation methods like neural machine translation to predict the likelihood of phrase sequences. This could refer to a single word, anentire sentence, or, more recently, an entire document, depending on the context. In order to overcome challenges with language translation and localization, NMT is a fundamentally novel approach that trains neural models using deep neural networks and artificial intelligence. Neural machine translation often produces translations that are more precise, fluid, and relevant than statistical machinetranslation. Neural machine translation requires far less memorythan conventional statistical machine translation models. This NMTapproach varies from traditional translation SMT systems in that the neural translation model is refined from beginning to conclusion to optimize translation efficiency. Neural machine translation creates a solitary substantial learning algorithm that scans a textand provides the accurate translation, in contrast to conventionalphrase-based translation systems that consist of numerous small,individually tuned sub components.

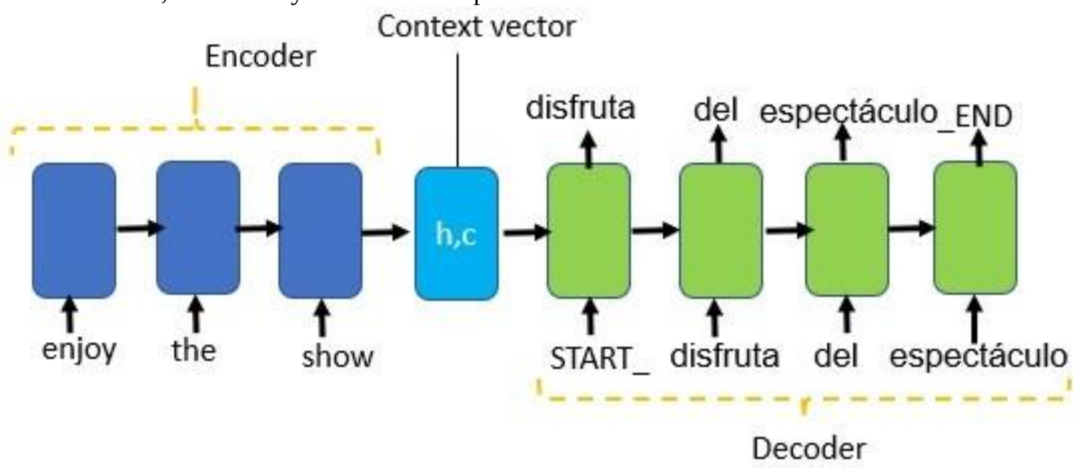

Figure 2: Neural Machine Translation System

C. Example-Based Machine Translation (EBMT)

Example-based machine translation is one of the automatic translation strategies. By using appropriate matching, the method trans- forms the user input into the desired information using a corpus of two languages. The parse tree is where Example-based machine translation is used most frequently. If not, it will search the data- base for the sentence that is closest in meaning, and then adjust thetranslation to account for any discrepancies. Additionally, there is a chance to successfully translate the source sentences. It is crucialto have a vast amount of translated text in order to move forward with Example-based machine translation. In other words, many parallel multilingual documents should be stored in the database. These texts have been translated by professionals. The specialists' skill sets extend beyond linguistic competency. Additionally, they are fluent in both languages. As a result, it is possible to ensure that the database utilized for Example-based machine translation isreliable and capable of consistently producing excellent results. Theapproach of extracting information from bilingual texts is known as example-based machine translation. In this case, it does not appear that the knowledge has any overt formal expression. It is also not connected to any encoding method. Instead, the text coupling method is used to simply encode the knowledge. The translation tool will take a piece of text, compare it to a similar piece of textstored in the database, and then translate it.

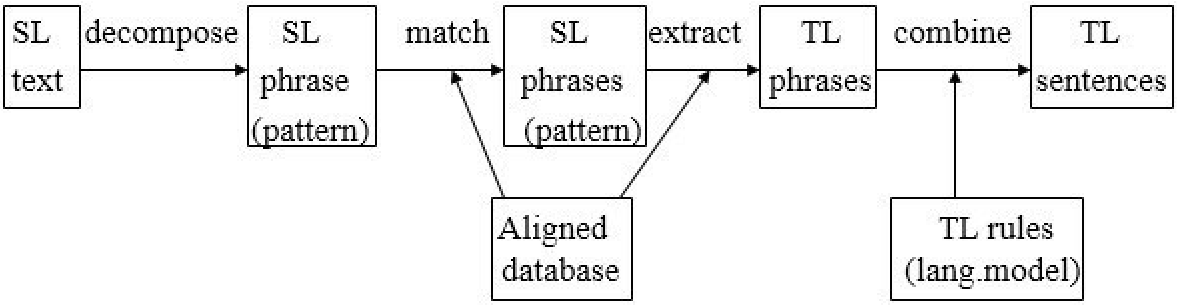

Figure 3: Example based Machine Translation

D. Hybrid Machine Translation

An automated translation system called a hybrid machine translation system incorporates many language translation technologies. The creation of hybrid machine translation systems is driven bythe fact that no single technique has been able to produce resultswith a respectable level of accuracy. The simultaneous operationof numerous machine translation systems is used in this hybrid machine translation technique. The end outcome is produced bycombining the output of each sub-system. Statistical and rule-basedtranslation subsystems are the most typically utilized in these systems, while many combinations have been studied. Merging thesystems at the end of the translation pipeline could offer an im- provement in translation quality that is difficult to find elsewhere because any change made to one system will almost surely resultin an improvement to the combined one. Another reason to adopt acombined system is that it can offer linguistic insight into the areaswhere the RBMT system and SMT system perform worse or better.Hybrid machine translation (HMT), a paradigm, tries to combine the advantages of many methods. Due to the complementary ad-vantages and disadvantages of the different approaches, HMT has emerged as a distinct area of research in machine translation.

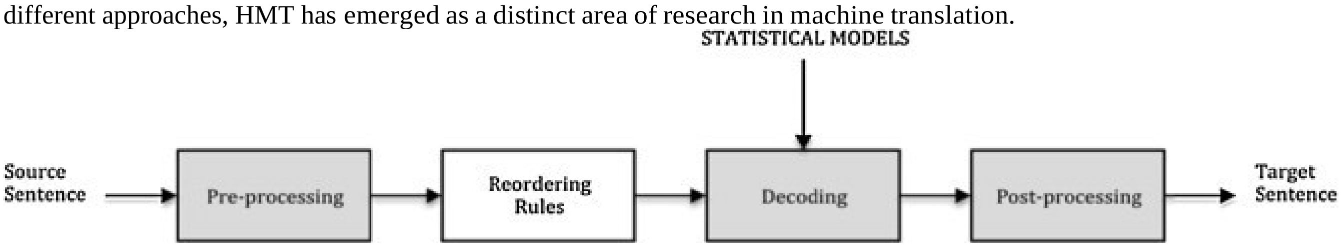

Figure 4: Hybrid Machine Translation

E. Transformer Model

The self-attentional mechanism is used by a deep learning model known as a transformer, which assigns a variable weight to the significance of each incoming data element. Natural language processing (NLP) and computer vision are two fields that heavily employ it (CV). Transformers are designed to examine sequential input data, such as natural language, and have applications in text summarization and translations. They work similarly to recurrent neural networks. Unlike RNNs input is processed by transformers at once.The transformer need not examine each word separately if the in- coming data is in natural language. RNN parallelization is lesser than Transformer which reduces training time. Transformer models are the favored model for NLP difficulties, substituting RNN models like LSTM. Using the added parallelization for training, larger datasets can be trained. Pretrained systems like GPT (Generative Pre-trained Trans-former) and BERT (Bidirectional Encoder Representations from Transformers) were developed as a result. These systems, which may be customized for applications, weretrained using sizable language datasets like the Common Crawland Wikipedia Corpus.

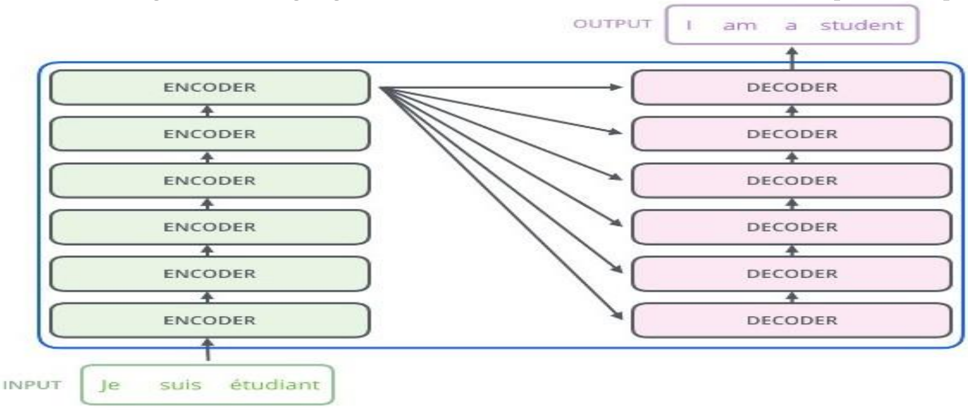

Figure 5: Transformer Model

Retrieval Number: 100.1/ijrte.A75380512123
DOI: 10.35940/ijrte.A7538.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-12 Issue-1, May 2023

F. Statistical Machine Translation

The statistical machine translation (SMT) theory states that proper-ties of numerical methods are verified using bidirectional textualcollections before being translated into some other languages. SMTrefines its transcribing skills by examining past human interpretations. The bulk of SMT methods used today use groups of wordsand combine different cultures or societies on overlapping phrases.By translating full word sequences, where the lengths may vary,phrase-based translation aims to loosen the restrictions of word-based translation. The word pairings are referred to as phrases, but they are frequently statistically rather than linguistically discovered phrases from bilingual text corpora. The examination of bilingual text corpora (source and destination languages) and monolingual corpora is used to develop statistical models for text translation (target language). To choose the most likely translation, these statistical models employ statistical weights.

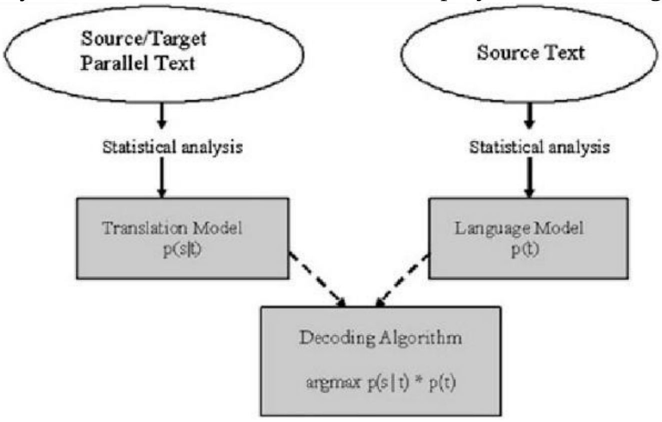

Figure 6: Statistical Machine Translation

G. Long Short-Term Memory Model

Long Short-Term Memory (LSTM) networks, which enable higherretention of information for previous information, are generated bymodifying recurrent neural networks. Gradient descent and long- term dependency are two problems with RNNs that are mostly addressed by LSTM. LSTM is well-suited to classifying, processing, and predicting time series with uncertain time delays. The model is trained via back-propagation. Except for the internal cell content and hidden layer being transmitted forward, the essential functioning of a long short-term memory structure as well as the behavior of a recurrent neural network are similar. The four gates that make up an LSTM are the forget gate, the input gate, the inputmodulation gate, and the output gate. A "memory cell" found in LSTM units can store data in memory for extended periods of time.

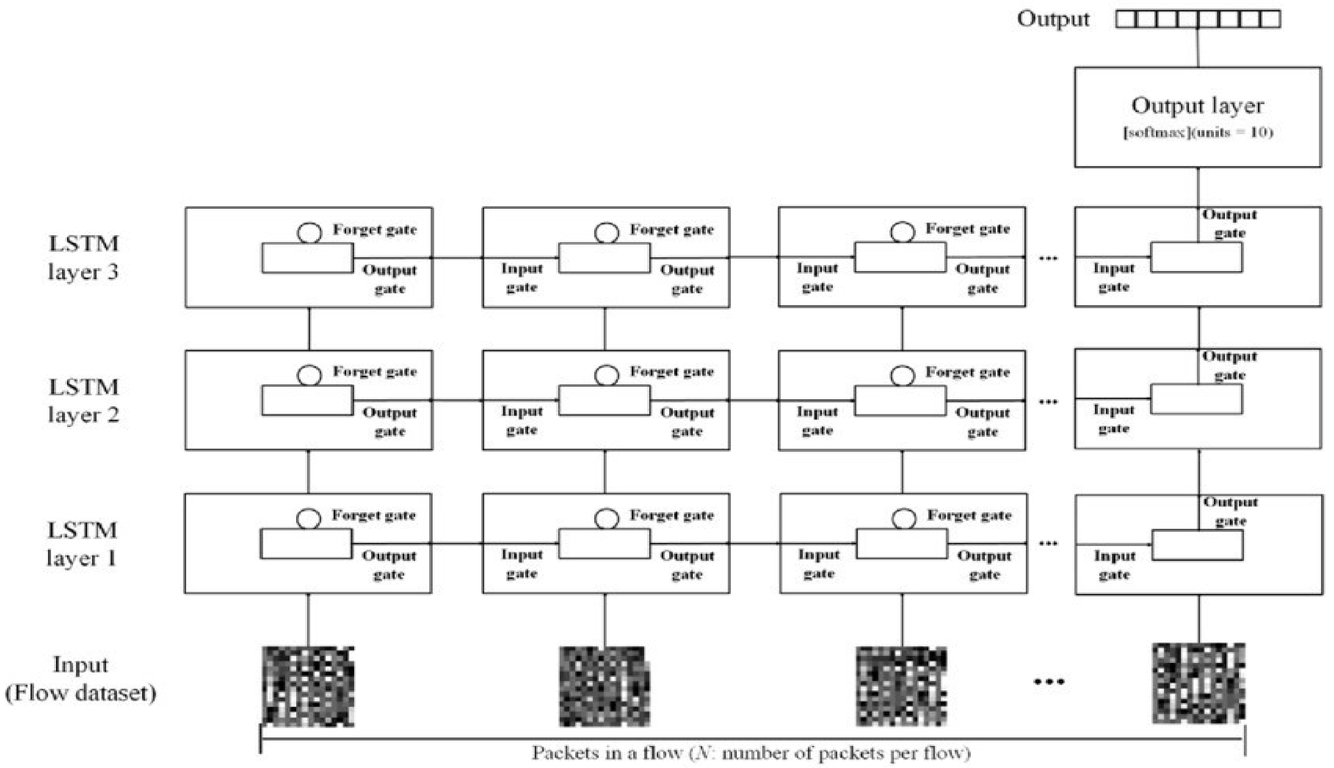

Figure 7: Long short-term Memory Model

Retrieval Number: 100.1/ijrte.A75380512123
DOI: [10.35940/ijrte.A7538.0512123](https://doi.org/10.35940/ijrte.A7538.0512123)
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

III. LITERATURE SURVEY

Peter F. Brown et al [1] proposed A Statistical Approach To Machine Translation. They said translation requires the computation of two kinds of probabilities, language model probabilities, and translation probabilities. They referred to the first kind as monolingual probabilities because they are computed using data in only one language. The second kind they referred to as bilingual probabilities because they are computed using data in two languages. They utilized the language model and translation model. They are most concerned exclusively in this paper with the computation of bilingual probabilities. They began with the monolingual probabilities. They have a large corpus of English text, consisting of about one hundred million words. To compute the probability of any sentence, they broke it into words and multiplied together the probabilities of the individual words and obtained. Finally, they applied a method for deriving a probabilistic phrase structure grammar to construct grammar for both French and English and to base future translation models on the grammatical constructs. Franz Josef Och et al [2] They discuss enhanced alignment models for statistical machine translation in this paper. The statistical translation strategy employs both a language model and a translation model. Either a bigram or a broad m-gram model serves as the language model. A lexical model and an alignment model make up the translation model. The authors propose two strategies for translating statistical information into experimental outcomes. The Verbmobil task, a 6,000-word spoken language exam, was used in the assessments. Both the output of the speech recognizer and the text transcription were put through tests. Philipp Koehn et al [3] contributed the Europarl corpus that was assembled primarily to support research in statistical machine translation, Information extraction, word sense disambiguation, and anaphora resolution are just a few of the numerous natural language problems it has been used to solve. This study addresses both acquiring the corpus and using it to address the issue of statistical machine translation.110 machine translation systems have been created using the corpus to accommodate all conceivable language pairs. The systems that were produced and how well they performed show the various difficulties that statistical machine translation for various language pairs faces. The typical five steps when purchasing a parallel corpus to be used with a statistical ma- chine translation system are obtaining raw data, Segmenting the text into sentences, removing and mapping parallel blocks of text, normalizing and tokenizing the corpus, and mapping sentences from one language to another (sentence alignment) are all steps in preparing the corpus for SMT systems. This represents the first real attempt to develop a system for, say, from Greek to Finnish. It may be the largest number of machine translation systems to be constructed in less than three weeks. SMT research has several challenges, which they have only just begun to solve, as evidenced by the wildly different quality of the numerous SMT systems for the various language pairs. The primary function of the field, which is to translate a few languages into English, ignores many of these challenges. Kishore Papineni et al [4]. The method for automatically assessing machine translation output depends on a number of metrics. These language-independent assessments cover a wide range of topics, including accuracy, fluency, and sufficiency. They also blend evaluations made by humans and machines to get a comprehensive evaluation score. Then it is calculated how well this score matches a human judgment of the same output. The results show that this method can generate results that are well correlated with human review, making it a valuable and efficient tool for assessing machine translation. Chris Callison-Burch et al [5] The blue metric has long served as the benchmark for assessing machine translation systems, but because it relies on automated, quantitative criteria, it may not adequately account for the subtleties of human language. Additionally, its focus on lexical similarity might lead to translations that are accurate but lack depth. They contend that the community should evaluate machine translation systems more comprehensively, taking into account both quantitative and qualitative factors. They provided two case studies that illustrate the flaws in the Bleu metric to back up their claim. In the first, a system's human evaluation was greater for one with a lower Bleu score. In the second, a system with a higher Bleu score nonetheless received a lower rating than a system with a lower Bleu score. These instances demonstrate that the Bleu metric is not necessarily a reliable indicator of quality and that using it as the primary basis for system design might result in accurate but meaningless solutions. In the end, they advocate that the machine translation community concentrate on developing systems that not only provide translations with a high Bleu score but also ones that are meaningful and convey the same meaning as the original text. Alternative evaluation techniques that are better suited to capturing the subtleties of human language should be investigated by the community in order to achieve this. They conclude that while Bleu is a useful tool for monitoring the development of MT systems, it shouldn't be the only indicator of translation quality. When using Bleu, care should be given to use it appropriately and to do the proper manual evaluations as needed.

Rashmi Gangadharaiah et al [6] proposed an Example-Based Ma- chine Translation for Indian Languages. It demonstrates that certain discernible language principles to aid in the process of making automatic translation easier and less expensive when converting from English to low resource languages and the impact of the number of rules on accuracy. All Indian languages have the same structural foundation, hence they all share the same linguistic principles. As a result, the G-EBMT approach applies to all Indian languages. Using the BLEU assessment metric to translate sentences from English to Kannada, a straightforward EBMT system combined with language principles yielded a result of 0.7164. Remya Rajan et al [7], The ROMAN TO UNICODE file, UNI- CODE TO ROMAN file, Word Dictionary File, Morph dictionary file, and Transfer Link Rule File are the essential fundamental elements supporting the RBMT approach for converting English to Malayalam. Each verb, noun, pronoun, determiner, adjective, and other word is included in the Word Dictionary File.

Retrieval Number: 100.1/ijrte.A75380512123
DOI: 10.35940/ijrte.A7538.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Romanized Malayalam is mapped to Unicode format in the ROMANTOUNI- CODE file. The UNICODETOROMAN file is a mapping file that converts Unicode to Romanized Malayalam. The Morph dictionary File contains data about the word's tag, source morphological features, and targets morphological features. The recipient design is changed for a specific reference form according to the Transfer Link Rule File. Due to the differences in the subject-verb-object and subject-object-verb grammatical structures between English and Malayalam, the transfer link rule file is crucial. The specific RBMT system steps are described in this publication.

Unnikrishnan P et al [8] Incorporating syntactic and morphological data, this study provides the creation of an English-to-Dravidian language statistical machine translation (ST) system, encompassing Malayalam and Kannada. With limited technological resources and tools, creating a fully operational MT system for any two natu- ral languages that is bilingual is a challenging and time-consuming task. The creation of a perfectly aligned parallel corpus for the system's training is the first and most crucial stage in SMT. According to an experimental study, the SMT system's English to South Dravidian bilingual parallel corpus-building process is ineffective. This study has implemented a novel method for developing parallel corpora to enhance the effectiveness of the translation system. Separating the root suffix on both English and Dravidian words, the English source sentence is rearranged using Dravidian syntax, and using morphological information are the main concepts that they have put into practice and have proven to be very effective SMT system from English to Dravidian languages in South. The experiment revealed that practically all simple sentences in all twelve tenses, as well as their negatives and question forms, may be correctly translated using the suggested translation approach. Additional transfer regulations that apply to various categories of sentences can be added to the proposed system to boost its performance. Sometimes simple sentences aren't necessary. Long sentences may prevent the parser from providing the correct syntactic structure. This issue can be resolved by creating a clever technique that divides long words into manageable chunks that can then be solved individually Tamil and Telugu can be easily incorporated into the strategy because South Dravidian languages all have a very identical structural foundations. Given that many languages in India use the Subject object Verb order and are very morphologically rich, the methodology described here is generally suitable to SMT from English to Indian languages.

Mary Priya Sebastian et al [9] described the process of utilizing statistical models to translate from English to Malayalam. This study emphasized on incorporating pre-processing methods like suffix separation for better alignments is a better idea because it improved the BLEU score. During the training phase, the translator uses a monolingual Malayalam corpus as well as a bilingual Eng- lish/Malayalam corpus, and it automatically generates the Malayalam translation of an unknown English sentence.

The discussion covers various methods for enhancing the alignment model by adding morphological inputs to the bilingual corpus. With this method, the results of the training have been enhanced by deleting the irrelevant alignments from the sentence pairs. The training step of SMT is made less complicated by the alignment model's knowledge of category tags, name entities, cognates, and predictable Malayalam translations, which reduces inconsequential alignments. This method aids in enhancing the accuracy for Malayalam terms found in the parallel corpus of word translations. WER, F measure, and BLEU metrics are used to assess the SMT's performance, and the findings show that the translations are of average quality. By including the matching bilingual corpus and its order conversion rules, this approach can be further developed and used to translate any language into Malayalam.

Aditya Vyawahare et al [10], It utilizes different translation models, such as Seq to Seq models like long short-term memory, bidirectional long short-term memory, and Conv2Seq, and also trained state-of-the-art transformers from scratch and evaluates their performance. Additionally, it improved the translation models for monolingual data before implementing back-translation. The issues resulting from the lack of data for low-resource languages can be avoided with the aid of back translation. For training the models Fairseq toolkit is used. Secured rank 1 for translation from Kannada to Tamil, Telugu, and Malayalam and placed third and fourth, respectively, for translations from Kannada to Sanskrit and from Kannada to Tulu. Using the sentence bleu function provided by the translated package 7 in the NLTK library, the BLEU scores were determined. Finally, accuracy was calculated using the BLEU evaluation metric.

Nithya B et al [11] proposed and created a hybrid translation process from English to Malayalam. This hybrid strategy adds a translation memory to the basic automatic statistical translation. By utilizing corpus-based machine learning approaches, a statistical machine translator translates text. The translation memory stores recently completed translations in the cache and do not require additional translations to be made. The hybrid technique is found to enhance the translator's performance in tests utilizing the BLEU score and precision measure to implement and evaluate the system. SMT systems need little human effort and little linguistic expertise. As a result, creating an SMT system for a new set of languages is easy and rapid. The availability of training data is the only need. The TM system returns the previously translated output for a sentence that has already been translated. Only those sentences for which there is no translation are given to the machine translation system after the TM evaluates the source text. Thus, using a TM eliminates the need for extra translations. The suggested hybrid approach combines statistical machine translation with TM. The system provides an improvement to a system for phrase-based SMT by integrating with TM. Because the most recent translations are cached, this integration accelerates the translation process. The post-editing feature, which enables human input, aids in improving the translation. More sentences can be added to the parallel corpus to further expand the study.

Retrieval Number: 100.1/ijrte.A75380512123
DOI: 10.35940/ijrte.A7538.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Mallamma V Reddy et al[12] studied machine translation NLP issues and suggested a query translation strategy. The study observed that difficulty in NL understanding originates from lexical, syntactic, and contextual ambiguity. This research made an alterna- tive argument that NL comprehension is significantly harder than NL generation. This study discusses the challenges of statistical ma- chine translation, including sentence alignment, compound words, idioms, morphology, and different word orders, as well as a statistical model approach to the machine transliteration problem. They introduced a query translation-based technique in their system since it is efficient to translate or translate the question in respect to documents. The program makes use of the English stop word list, the Machine-Readable Dictionary, the transliterator, art of Speech Tagger (POST). and Stemmer. This project made a model that can be used to extract proper names and transliterations. The Cross-Language Information Retrieval Tool is constructed using ASP.NET as the front end. Without using pronunciation dictionaries, the suggested method can be simply applied to additional language pairs with various sound systems.

Manoj Kumar et al [13], the distinction between rule-based and corpus-based machine translation is discussed in this study. Example gathering, Matching, and Recombination are the three main phases in example-based machine translation, which is based on the concept of reusing relevant text examples. The process of gathering instances of relevant text sentences and building a corresponding vocabulary for the interpretation system is known as example acquisition. Locating the input sentence's relevant textual segments is the goal of matching and the act of integrating the translated text fragments into the target text is known as recombination or sentence synthesis. The resulting replication produces the intended converted term and boosts the objective statement's legibility. Finally, the model is manually assessed by subject-matter specialists.

Asha Hegde et al [14], The NMT models are applied for converting English to Tamil, English to Telugu, English to Malayalam, and Tamil to Telugu text. In this study, various machine translation models including Rule-Based, Hybrid, Statistical, Encoder to Decoder, and Recurrent Neural Network models are explained. To improve the translation quality of the system, a multi-layer encoder-decoder model can be built. This work also employed a one-hot encoding embedding strategy for the source and target texts. The Tamil to Telugu translation model placed 2nd with 0.43 language translation quality. Additionally, the models showed BLEU scores of 1.66, 0.29, and 0.48 for the corpora of English to Tamil, English to Telugu, and English to Malayalam, respectively. The model is evaluated using expert judgments and a traditional diagnostic criterion, such as the Bilingual Evaluation Understudy. Kyunghyun Cho et al [15] have looked into the characteristics of a new family of machine translation systems that are entirely based on neural networks. They explicitly selected two encoder- decoder models, the newly suggested gated recursive convolutional neural network and RNN with gated hidden units, which differ in the encoder they use. There are many alternative encoder-decoder models. After training those two models on pairs of sentences in both English and French, their performance was examined using BLEU scores in relation to sentence length and the presence of un- known or uncommon words in sentences. The investigation showed that phrase length has a considerable negative impact on how well neural machine translation performs. However, qualitative analysis reveals that both algorithms excel at producing accurate translations. Additionally, this study discovers that the suggested gated recursive convolutional network automatically learns the grammatical structure of a text.

Xing Wang, Zhaopeng Tu et al [16] Neural machine translation (NMT) has gained popularity over the past several years due to its ease of use and cutting-edge features. This paper presents a comprehensive paradigm for combining SMT word knowledge into NMT in order to overcome restrictions above the word-level. The word predictions generated by the NMT decoder in this situation are more precise. They have demonstrated how SMT word knowledge can be used to improve the NMT encoder-architecture. decoder's to put the concept into practice, they employ two model variations. The efficiency of the suggested framework has been shown by experimental results on translation assignments from Chinese to English and from English to German.

Sandeep Saini et al [17] The process of using software to translate text or speech from one natural language to another is known as machine translation (MT). Along with the 22 official languages, India also has more than 1650 dialects.

Minh-Thang Luong et al [18] This study examines two forms of attentional processes for neural machine translation (NMT): local attention, which only considers a small number of source words at once, and global attention, which continuously monitors all source words. They provide as examples of how well both approaches of back-and-forth translation between German and English work. Local attention systems gain 5.0 BLEU points more than non-attentional systems do when using current established tactics like a dropout. In the WMT'15 English to German translation issue, the ensemble model incorporating multiple attention architectures produced a new state-of-the-art result with 25.9 BLEU points, an improvement of 1.0 BLEU points above the previous best system supported by NMT and an n-gram reranked. They also con- ducted experiments with various source lengths, demonstrating that the global model performs better with longer source phrases while the local model can only use a very limited number of source phrases. To comprehend the translation process better, they studied attentional settings and created a state-of-the-art visualization tool. As a result of their investigation, they draw the conclusion that attentional contexts can be considered as sub-subjects and that attentional weights are utilized to align source and target terms as well as pick which themes are communicated in the translation. Hamid Palangi et al[19] described phrase embedding is a hot topic in current natural language processing. It was reported that Recurrent neural networks (RNN) incorporating Long Short-Term Memory (LSTM) cells were utilized to build a model that solves the problem.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The suggested LSTM-RNN model sequentially extracts the information from each word in a phrase and embeds it in a semantic vector. In this study, data on user click throughs collected by a commercial web search engine is used to train the LSTM-RNN in a weakly supervised way. To comprehend how the embedding procedure functions, visualization and analysis are conducted. It has been discovered that the model automatically eliminates un-necessary words and recognizes the significant keywords in the sentence. Additionally, it has been discovered that the LSTMRNN's identified keywords automatically activate various cells, whereas words relating to the same topic activate a different cell. The embedding vector is a semantic representation of the sentence that can be applied in a variety of ways. On a web search task, Several cutting-edge methods are found to be greatly outperformed by the LSTM-RNN embedding. It emphasizes how the proposed method generates sentence embedding vectors that are highly beneficial for web document retrieval tasks. The Paragraph Vector, a well-known broad sentence embedding technique, is compared. Effective translation tasks can also be performed using this model.

Rohit Gupta et al [20] proposed a MT evaluation metric that was submitted for the WMT-15 metrics. Modern Machine Translation (MT) evaluation measures are often complicated, require significant external resources (such as paraphrase), and need to adjust to produce the best results. This study employ a measure based on a kind of recurrent neural networks called Long Short Term Memory (LSTM) networks, and dense vector spaces (RNNs). According to system-level correlation using Spearman and Pearson (Pre-TrueSkill) and Pearson (TrueSkill), this new metric for WMT15 performs the best overall. The suggested metric is straightforward in the sense that, aside from the thick word vectors, it doesn't need a lot of equipment or resources. This cannot be claimed for the majority of cutting-edge MT evaluation measures, which are frequently intricate and demand substantial feature engineering. RNNs, specifically Tree Long Short Term Memory (Tree-LSTM) networks, are the foundation of the metric. A memory cell is used by the sequencelearning method known as LSTM to maintain a state over an extended length of time. As a result scattered word representations, can be used to represent sentences. This dense-vector-space-based metric Reval is competitive with the most sophisticated alternative approaches currently being used, which involve combining different systems, using a lot of external resources, and engineering features, and tuning them. Neeha Ashraf et al [21] compared and investigated MT techniques. The goal of this analysis is to compare the ease of use, reference system quality, and development of an experimental framework for MT technique comparison across multilingual MT techniques. The numerous MT techniques, including EBMT, RBMT, SMT, CBMT, and Hybrid Techniques, are explained in this paper. According to the study, SMT is the most widely used MT technique, while EBMT has made an effort to determine its personality.

The author wanted to draw attention to the fundamental differences between each world: the grammatical correctness of RBMT, the lexical preference of SMT, and SMT's tolerance for irregular structures. As a result, MT system-supported combination of approaches is a fun task and can help a lot with issues with current MT methods. They intended to mix entirely distinct models in a way that makes use of their benefits and helps to address issues with translation procedures in order to provide respectable results. Hybrid MT isinspired by the fact that practitioners are increasingly choosing to combine the most basic elements of very complex pure rule or corpus-based MT systems

Dr. Piyush Kumar Pareek et al [22], To reduce redundant n-grams, most current techniques are limited to monolinguals. To address this, an efficient machine translation model based on ma- chine learning techniques is deployed. The studies carried out on Internet information show exceptional accomplishments in terms of precision and processing difficulty of arrangement when considering different thresholds. The accuracy performance for English to Kannada, English to Telugu, and Kannada to Telugu was 91.63,90.65, and 88.63 percent respectively.

Dr.(Mrs.)V.VIDYAPRIYA et al [23] The aim of the research is to develop an English to Tamil neural machine translation. The NMT system allows users to enter text search queries in their native tongues, which are then translated and used to locate relevant data in other languages. Requests are translated from one language to another using statistical and rule-based methods. The suggested English-to-Tamil translation system delivered effective outcomes when compared to the existing system. The proposed NMT system blends a rule-based approach with a statistical approach. The text is transliterated using statistical MT. Tokenization, preprocessing, and translation are a few of the processes needed. Sainik Kumar Mahata et al. [24] proposed Machine Translation Using Recurrent Neural Network on Statistical Machine Translation. They presented a novel method for machine translation that scores the phrases produced by a statistical machine translation (SMT) system by using a recurrent neural network (RNN) encoder-decoder system. They employed the Moses system, a tool for automatically developing translation models for any language pair, which is trained using a sizable corpus of translated texts (parallel corpus). They employed the RNN system to score the words produced by the SMT system. Once the model has been trained, an effective searchalgorithm quickly discovers the greatest probability translation among the exponential number of options. The translation systemoutput is chosen from the sentences with the greatest scores. They tested their strategy on an English-Hindi machine translation job and demonstrated that it produces better translation quality than the SMT system, receiving a blue score of 3.57. Sai Koneru et al [25] proposed Unsupervised neural machine translation on Dravidian languages. It concentrates on the unsupervised translation of Kannada, a Dravidian language with limited resources, into English. Demonstrate the importance of integrating the writing systems using a small amount of auxiliary data across English and other comparable languages. Additionally, it demonstrates the value of transliteration in UNMT between closely related languages that do not use the same writing system.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Examined several structures to incorporate comparable languages, enabling MUNMT for different language pairs, and offer a metric to assess language resemblance and demonstrate that it serves as an effective guide for choosing the additional languages. Investigate the function of lexical overlap in an unsupervised training pipeline that also includes supervised training and pretraining on monolingual data. In an unsupervised training pipeline, the role of lexical overlap is studied along with pretraining on single data and supervised training on accessory dialect pairings. Evaluate multiple architectures such as Unsupervised, dis- similar languages, Unsupervised similar languages, Unsupervised, Dissimilar languages + cross translation, and Unsupervised, similar languages + cross translation with different reference languages. For Dravidian languages, datasets are taken from the AI4Bharat- IndicNLP corpus, while for English, it is from Wikipedia dumps. For transliteration between Latin script and Indian script, it employs the Indic-transliteration toolbox. The BLEU score for supervised translation between language pairs is En->Kn: 16.4, Kn->En: 12.3, Te->Kn: 20.1, Kn->Te: 11.9, and the BLEU score for unsupervised translation between language pairs is En->Kn: 0.1, Kn->En: 0.3, Te->Kn: 3.5, and Kn->Te: 2.3. Finally, it suggests a metric for lexical over-lap and demonstrate how it may be used to choose the best reference languages.

Xu Tan et al [26], Automated translations from 23 Languagesinto English and from English into 23 Languages were performed. In this study, a framework for grouping languages into clusters is developed, and a single multilingual model is trained for eachcluster. It describes two methods for grouping languages: (1) prior knowledge, in which languages are grouped by language family, and (2) language embedding, in which each language is represented by an embedding vector, and the languages are grouped in the embedding space. In terms of the BLEU score, Language embedding outperforms the prior knowledge strategy. IWSLT datasets from TED lectures that include different languages are used to eval-uate the experiments.

Bharathi Raja Chakravarthi et al [27] provided MMDravi, a Multilingual Multimodal dataset for languages in a Dravidian family with limited resources. It contains thirty thousand sentences generated using a variety of machine translation outputs. Multimodal machinetranslation is the procedure for the conversion of a text between two languages utilizing information from multiple modalities. In order to utilize multilingual corpora and other modalities, for Dravidian languages that are related closely they created a Multilingual Multimodal Neural Machine Translation system (MMNMT) utilizing data from MMDravi and a corpus' phonetic transcriptions. This method enhances the accuracy of language translation with few resources by utilizing phonetic transcription, visual characteristics, and multilingual corpora. The best scores for this model were 52.3 for the En-Ta language pair and 37.6 for the Ml-En language combination. According to the evaluation that used phonetic transcription, multilingual, and multimodal NMT, the recommended system beats the present strategy of multimodal and multilingualin less-resourced neural machine translation across all languagecombinations taken into consideration.

Bharathi Raja Chakravarthi et al[28], suggested Comparing Different Orthographies for Dravidian Languages with Limited Re- sources. In this model, many sources and target languages are trained concurrently, and this approach is known as multilingual neural machine translation. This has been demonstrated to enhance translation quality; however, the use of these multi-way modelswas restricted in this study because they concentrated on languageswith different scripts. The BLEU ratings were calculated to assess if combining them into a single script would allow the system to benefit from the phonetic similarities between these closely relatedlanguages. Finally, the BLEU evaluation metric was used to calculate accuracy.

B. Premjith et al [29], Due to the lack of high-quality corpora and the diverse morphological spectrum of Indian languages, machine translation from English to these languages is typically a challenging procedure. Large corpora are necessary for NMT systems to deliver better translations. The equivalent phrases should also covera variety of topics and express similar sentiments. Using such a cor-pus to model the system can ensure accurate translations while the model is being evaluated. The community of researchers studying machine translation will benefit from our offering of this dataset. The quality of translation is significantly influenced by the size of the sentences as well as the size and breadth of the corpus.

Ashwani Tanwar et al [30], proposed Translating Morphologically Rich Indian Languages Under Zero-Resource Conditions. They came up with a new approach called zero-resource conditions for Machine translations of morphologically-rich Indo-Aryan and Dravidian languages. This process of language transfer learning for monolingual corpora and parallel translation for other languages will be based on Zero-Slot Systems. They used BLEU and TER metrics for comparing direct translations with Zero-Slot Systems. These Systems are trained models used for fine-tuning with real human-generated data. The output of both Indo-Aryan and Dravidian languages are measured according to the complexity of their morphologic. As a result, they found that systems with Dravidian languages performed very well and almost very close to the level of direct human translations because of the morphological and complexity in the language. Due to this, much more room were pro-vided for transfer learning. After tuning Systems with Indo-Aryan languages are also showed much better results.

Ashman Pramodya et al [31] Neural machine translation is currently the most promising machine translation technique (NMT).Transformer performs 2.43 BLEU points better than LSTM in the Tamil to Sinhala direction. The first comparison in this study is between statistical machine translation (SMT) and natural language translation (NMT). NMT provides direct translation from the source language to the destination language. based on SMT, the most effective Sinhala-Tamil open-source translator at the time. They compare the performance of SMT and NMT models on the identical parallel corpus. They compared the recently developed Transformer archi- tecture with recurrent models.

Retrieval Number: 100.1/ijrte.A75380512123
DOI: [10.35940/ijrte.A7538.0512123](https://doi.org/10.35940/ijrte.A7538.0512123)
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

They were contrasted with statisticalmachine translation on low-resource Tamil to Sinhala translation assignments (SMT). They were contrasted with the Statistical Ma- chine Translation on low-resource Tamil to Sinhala translation assignments (SMT).

Pushpalatha Kadavigere Nagaraj et al [32], worked only on the unidirectional translation from Kannada to English using Neural Machine Translation(NMT). But later on, they found Recurrent Neural Network(RNN) has been the most effective method for translation. They used the Sequence to Sequence the dataset of the Encoder-decoder process using Long-short Term Memory (LSTM)as an RNN unit. They measured results with regard to Statistical Ma- chine Translation (SMT) and found the best Bi-Lingual Evaluation study value with an accuracy of 86.32

T. Madhavi Kumari et al [33] proposed a neural machine translation system (NMT) in in this work to effectively translate Indiclanguages like Hindi and Telegu. One of the main causes of this increase is deep neural networks' capacity to learn a logical rep- resentation of words. Encoder-decoder with attention mechanism is part of the proposed NMT paradigm created for the Telegu language. Sentences in two languages from the TDIL database and a corpus reported in this research were used to train the models for the NMT system. The tokenizer module was used to tokenize thesentences provided in the tool kit called Moses before being fed intothe network. The neural network technique (LSTM and Bi-RNN) and the number of hidden layer units (250 and 550) were just a few of the parameters used to train the system. In this work, theyhave created a system that makes use of an attention-based neural model. With evaluation metrics like BLEU, TER and perplexity, the suggested attention-based NMT model is put to the test. When compared to Google Translate, the network translated English toGujarati more accurately by a margin of 6 BLEU points.

Karthik Puranik et al [34] proposed Transformer models for the translation of English ->Marathi and English<=>Irish language pairs. Indic-Trans, a Transformer-4x multilingual NMT model, and Helsinki-NLP Opus-MT models are discussed in this study and shown that the Helsinki-NLP Opus MT model dominated the Irish language task, surpassing other Transformer models. Ranked 1, 1, and 2 in English Marathi, Irish English, and English Irishrespectively with BLEU scores of 24.2, 25.8, and 34.6. Raj Prajapati et al [35], proposed Machine Translation of Dravidian Languages. The final output of this project was translations from English to Telugu, Tamil, and Malayalam. This work majorlyfocuses on storing a good amount of data backed up. They have re- moved all the unnecessary sentences through good pre- processing of it. This helped them to produce effective translations. They havealso done so many experiments on Byte Pair Encoding (BPE) and other hyperparameters. Shubham Dewangan et al [36], studied neural machine translation (NMT) for the languages of India.

Reported work on Indian language Statistical Machine Translation (SMT) shows strong performance within the Indo-Aryan family, but rather poor performance within the Dravidian family as well as between the two families. The general consensus is that NMT produces output that is more fluent than SMT. In NMT, it is currently standard practice to train the models using sub word units. Byte pair encoding (BPE) is a preferred method of sub word. The phrase table injection technique, which successfully integrates SMT and NMT, was also suggested. This study conducted extensive trials with BPE-based NMT models for Indian languages. Improvements in SMT vary from 0.88 (te-bn) to 5.96 (ml-hi), whereas improvements in NMT range from 0.36 (ml-te) to 1.9(ta-hi) BLEU, when Dravidian languages are on the source side Syed Abdul et al [37], the primary contributions of this study are the English to Hindi text processing method using convolutional with learning algorithm, the development of a new synchronous data, and the use of several measures to assess the machine translation model. In this study Long short-term structure, a framework was explained. The effectiveness of the model is examined using several computerized strategic use of data, including BLEU, F-measure, NIST, and WER. After several phases, the model generates a 45.83 overall BLEU score.

Ashalatha Nayak et al [38] The OpenNMT architecture is used to implement the transformer. The encoder and decoder of the model consists of a feed-forward layer and six layers of self-attention mechanisms. The Adam optimizer is used to train the model, with a learning rate of 0.0005. 32 batches are utilized to optimize the model. The IndicNLP library, which was created especially for Indian languages, is used in tokenization. Both the source language and the destination language are processed by the tokenizer. The output of the tokenizer is incorporated into the model during training. The accuracy of the machine translation models is evaluated using the BLEU score, a popular statistic. The score is based on the number of matching n-grams that can be discovered in both the model output and the reference output sentence. The baseline model produced positive results, with an average BLEU score of 0.85 across all language pairs. This shows how successful the suggested strategy is. Using more sophisticated methods like back-translation, ensembling, data augmentation, etc., the model can be improved even more. In conclusion, the Transformer-based method can produce accurate translations of Dravidian languages

Ralf D. Brown et al [39], proposed Example-Based Machine Translation in the Pangloss System. Pangloss is a machine translation system that uses multiple engines to translate input text. He used the PanEBMT model, PanEBMT is one of the translation engines used by Pangloss and EBMT is a type of translation that uses a collection of aligned source/target text pairs to find the best match for a source-language passage. PanEBMT uses a bilingual dictionary, a target-language root/synonym list, and a language-specific configuration file as its main sources of knowledge.

Nadeem Jadoon Khan et al [40], presented the result of the state-of-art phrase-based Statistical Machine Translation on multiple Indian Languages (SMT). Their aim is to publish the development of SMT for the language pairs. Due to the availability of a large parallel corpus, the SMT system was successful. They used this data to predict the accuracy of the translation. The final report was the performance of systems translating Indian languages into English with average 10 percent accuracy.

IV. RESULT

To overcome the drawbacks of traditional models it is necessary to use better models which gives better accuracy that means better translation quality as compared to the traditional models. The type of model that best meets your demands will vary according to the language pairs. The models which are better than traditional models are Long Short-Term Memory, Transformers and BERT models.

V. DISCUSSION

The kind of innovation that best suits your requirements will change contingent upon the language pair. The traditional models require technical expertise in addition to taking too much time to implement from scratch. The implementation of a traditional algorithm from the ground up takes too much time and necessitates technical expertise. It takes some input and some logic in the form of code to generate the output. For example, if RBMT is used, it can be easily customized, handles tags well, and is predictable. However, in order to improve the quality of an RBMT, rules need to be changed, which requires more linguistic knowledge. SMT engines, on the other hand, are really appealing because they don't need linguists to customize them: The engine learns by itself through statistical analysis of translation memory corpora. However, when language pairs have significantly different word orders, the engine typically performs less well. When looking at recent models like LSTM, Transformers, and Bert, it's clear that they are producing good results due to their unique characteristics. For example, LSTM models have long-term dependencies and Transformers take a lot of GPU to implement, but they are easy to use and more efficient. Although these models have some drawbacks, a clear comparison of models' usage, feasibility, and efficiency helps choose which methodology to use based on language and usage.

VI. CONCLUSION

The purpose of this work is to view the trends in Machine Translation studies within the past years and to provide a complete literature analysis to classify and examine the various machine translation techniques and related research. Depending on the application, several MT groups have used various formalisms. Themain objective of this research is to find workable MT techniques for Dravidian languages with few resources. The different Datasetsthat were created for Dravidian languages have been covered inthis paper that explained the significance and nature of the Dravidian family. In this study, a comparative analysis of numerous procedures from various earlier studies and various experimental strategies from earlier work that has been successful is discussed. The significance of this work lies in the fact that the translation quality is improved as a result of the careful examination of the source and target languages, as well as the use of several modelsthat include linguistic formation and syntax analysis. This work has described the research on various NLP techniques that improvestranslation accuracy and the combination of various experimental models that has proven to be efficient. This work provides research about the challenges in Machine Translation, natural language processing and also gives a detailed view of the potential techniquesof MT for Indian languages and mainly for the Dravidian family.

DECLARATION

Funding/ Grants/ Financial Support	No, We did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence
Availability of Data and Material/ Data Access Statement	Yes, Availability of Data and Material and the research data associated with an article is available in References through links.
Authors Contributions	We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. we further confirm that all authors have equally contributed to this manuscript.

REFERENCES

1. Peter F. Brown et al, IBM, Thomas J. Watson Research Center, A Statistical Approach To Machine Translation (1990), Available at: https://aclanthology.org/J90-2002.pdf
2. Franz Josef Och, Christoph Tillmann and Hermann Ney, University of Technology, Aachen, Germany, Improved Alignment Models for Statistical Machine Translation (2000), Available at: https://dl.acm.org/doi/10.3115/1075218.1075274
3. Philipp Koehn, School of Informatics University of Edinburgh, Scotland, Europarl: A Parallel Corpus for Statistical Machine Translation (2001), Available at https://aclanthology.org/2005.mtsummit-papers.11/
4. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, IBM T. J. Watson Research Center, USA, BLEU: a Method for Automatic Evaluation of Machine Translation (2002), Available at: https://aclanthology.org/P02-1040/ [CrossRef]
5. Chris Callison-Burch, Miles Osborne and Philipp Koehn,School on Informatics, University of Edinburgh, Re- evaluating the Role of BLEU in Machine Translation Research (2006), Available at: https://aclanthology.org/E06-1032/
6. Rashmi Gangadharaiah et al, Language Technologies Institute, Carnegie Mellon University, Application of Linguistic Rules to Generalized Example Based Machine Translation for Indian Languages(2006), Available at: http://www.cs.cmu.edu/~rgangadh/mspil06_rashmi_modified. pdf [CrossRef]
7. Remya Rajan et al, National Institute of Advanced Industrial Science and Technology (AIST) Tsukuba, Rule Based Machine Translation from Englis to Malayalam (2009), Available at: https://ieeexplore.ieee.org/document/5376582
8. Unnikrishnan P et al, Computational Engineering and Networking Department, Amrita Vishwa Vidyapeetham, A NovelApproach for English to South Dravidian Language StatisticalMachine Translation System (2010), Available at: http://www.enggjournals.com/
9. Mary Priya Sebastian et al, r, Dept. of Computer Science, Rajagiri School of Engg. & Technology, A framework for translating English text into Malayalam using statistical models (2011), Available at: https://dyuthi.cusat.ac.in/xmlui/handle/purl/4150

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

10. Aditya Vyawahare et al, Pune Institute of ComputerTechnology, Neural Machine Translation On DravidianLanguages(2011), https://aclanthology.org/2022.dravidianlangtech-1.28/ [CrossRef]
11. Nithya B, Shibily Joseph, Dept. of CSE, Government Engineering College, Thrissur, A Hybrid Approach to English to Malayalam Machine Translation (2013), Available at: https://research.ijcaonline.org/volume81/number8/
12. Mallamma V Reddy, Dr. M. Hanumanthappa , Department of Computer Science and Applications, Jnanabharathi Campus, Bangalore University, Nlp Challenges For Machine Translation From English To Indian Languages(2013),Available at: https://www.interscience.in/ijcsi/vol4/iss1/5/
13. Anju E S and Manoj Kumar K V, Dept. of Computer Science & Engineering, Government Engineering College Thrissur,Malayalam to English Machine Translation: An EBMT System(2014)Available at: http://iosrjen.org/Papers/vol4_issue1%20(part- 1)/C04111823.pdf [CrossRef]
14. Asha Hegde, Ibrahim Gashaw and H. L. Shashirekha , Dept ofComputer Science Mangalore University, Machine Translation for Dravidian Languages using Stacked Long Short Term Memory(2014),mAvailable at:https://aclanthology.org/2021.dravidianlangtech-1.50.pdf
15. Kyunghyun Cho et al, Universite de Montreal, On the Properties of Neural Machine Translation: Encoder–Decoder Approaches (2014), Available at: https://arxiv.org/abs/1409.1259/
16. Xing Wang, Zhaopeng Tu, and Min Zhang, School of ComputerScience and Technology, Soochow University, Incorporating Statistical Machine Translation Word Knowledge into Neural Machine Translation(2015), Available at: https://ieeexplore.ieee.org/document/8421063/authors#aut hors
17. Sandeep Saini and Vineet Sahula , Department of Electronics and Communication Engineering , Malviya National Institute of Technology Jaipur, A Survey of Machine Translation Techniques and Systems for Indian Languages (2015), Available at: https://ieeexplore.ieee.org/document/7078789
18. Minh-Thang Luong, Hieu Pham and Christopher D. Manning, Computer Science Department, Stanford University, Effective Approaches to Attention-based Neural Machine Translation (2015), Available at: https://arxiv.org/abs/1508.04025
19. Hamid Palangi, Yelong Shen, Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval (2015), Available at: https://arxiv.org/abs/1502.06922
20. Rohit Gupta et al, Research Group in ComputationalLinguistics, University of Wolverhampton, Machine Translation Evaluation using Recurrent Neural Networks (2015), Available at:https://aclanthology.org/W15-3047/ [CrossRef]
21. Neeha Ashraf et al, Manzoor ahmad, research scholar computer sciences, university of Kashmir, Machine Translation Techniques and their Comparative Study (2015), Available at: https://citeseerx.Ist.Psu.Edu/ [CrossRef]
22. Dr. Piyush kumar Pareek et al, Dept. of CSE, East west Institute of Technology, An Efficient Machine Translation Model for Dravidian Language (2017), Available at: https://ieeexplore.ieee.org/document/8256970
23. Dr. (Mrs) V.Vidyapriya et al, Research Department of Computer Science, Quaid-E-Millath Govt. College for Women,Chennai, Neural Machine Translator for English to Tamil Translation using K-Means Clustering Algorithm (2018), Available at: http://ijream.org/papers/IJREAMV04I0945010.pdf
24. Sainik Kumar Mahata et al, Machine Translation UsingRecurrent Neural Network on Statistical Machine Translation(2018), Available at: https://www.degruyter.com/document/doi/10.1515/jisys- 2018-0016/html/
25. Sai Koneru, Danni Liu and Jan Niehues, Department of DataScience and Knowledge Engineering, Maastricht University, Unsupervised Machine Translation on Dravidian Languages (2018),Available at: https://aclanthology.org/2021.dravidianlangtech-1.7.pdf
26. Xu Tan et al, Microsoft Research, University of Science and Technology of China, Peking University, Multilingual Neural Machine Translation with Language Clustering (2019), Available at: https://arxiv.org/abs/1908.09324
27. Bharathi Raja Chakravarthi et al, Bernardo Stearns, Insight Centre for Data Analytics, National University of Ireland Galway, Multilingual Multimodal Machine Translation for Dravidian Languages utilizing Phonetic Transcription (2019), Available at: https://aclanthology.org/W19-6809.pdf
28. Bharathi Raja Chakravarthi et al,Insight Centre for Data Analytics, Data Science Institute, National University of Ireland Galway, Comparison of Different Orthographies for Machine Translation of Under-Resourced Dravidian Languages(2019), Available at: https://drops.dagstuhl.de/opus/volltexte/2019/10370/
29. B. Premjith et al, Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Coimbatore, M.Anand Kumar and K.P. Soman, Neural Machine Translation System for English to Indian Language Translation Using MTIL Parallel Corpus(2019),Availableat:https://www.degruyter.com/document/doi/10.1515/jisys- 2019-2510/html?lang=en [CrossRef]
30. ASHWANI TANWAR et al, Dhirubhai Ambani Institute of Information and Communication Technology, Translating Morphologically Rich Indian Languages under Zero- Resource Conditions(2020),Available at: https://dl.acm.org/doi/10.1145/3407912/ [CrossRef]
31. Ashman Pramodya, Randil Pushpananda and Ruvan Weerasinghel, Language Technology Research Laboratory, University of Colombo School of Computing, A Comparison of Transformer, Recurrent Neural Networks and SMT in Tamil to Sinhala MT (2020), Available at:https://ieeexplore.ieee.org/document/9325431 [CrossRef]
32. Pushpalatha Kadavigere Nagaraj et al, Kannada to English Machine Translation Using Deep Neural Network (2020), Available at: https://www.iieta.org/journals/isi/paper/10.18280/isi.2 6011
33. T. Madhavi Kumari et al, Associate Professor, Dept. of ECE, JNTUH CEH, Neural Machine Translation System of Dravidian Languages(2021), Available at: https://www.academia.edu/en/49557241/_IJCST_V9I1P2 0
34. Karthik Puranik, Adeep Hande, Attentive fine-tuning of Transformers for Translation of low-resourced languages (2021), Available at https://aclanthology .org /2021 .mtsummit- loresmt.14.pdf
35. Raj Prajapati et al, DA-IICT / Gandhinagar, India, Neural Machine Translation(2021), Available at: https://aclanthology.org/2021.dravidianlangtech-1.36
36. Shubham Dewangan et al, Experience of neural machine translation between Indian languages (2021), Available at: https://doi.org/10.1007/s10590-021-09263- 3/
37. Syed Abdul, Basit Andrabi and Abdul Wahid, Department of Computer Science and Information Technology, Maulana Azad National Urdu University, Machine Translation System Using Deep Learning for English to Urdu (2022), Available at: https://www.hindawi.com/journals/cin/2022/7873012/ [CrossRef]
38. Ashalatha Nayak et al, Department of Computer Science & Engineering, Manipal Institute of Technology, Machine Translation in Dravidian Languages (2022), Available at: https://aclanthology.org/2022.dravidianlangtech-1.19/
39. Ralf D. Brown, Center for Machine Translation, Carncgie Mellon Universityf, Pittsburgh, Example-Based Machine Translation in the Pangloss System (2022), Available at: https://dl.acm.org/doi/10.3115
40. Nadeem Jadoon Khan et al, COMSAT University, Machine Translation Approaches and Survey for Indian Languages (2022), Available at: https://arxiv.org/abs/1701.04290

AUTHORS PROFILE

Dr. B.V.Kiranmayee awarded her Ph.D. in Data Mining area from JNTU-H, Hyderabad. She did her Post graduation, M.Tech in Computer Science and Engineering from JNTUH University, Hyderabad. She has been working in the Department of Computer Science and Engineering, at VNR Vignana Jyothi Institute of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University. Currently, she is a Professor in the Computer Science and Engineering Department. She teaches C programming, Data structures, MFCS, Computer Organization, Design and analysis of Algorithms, UNIX Programming, Human computer interaction, Digital logic Design, Advanced databases, cloud computing and Data warehousing and datamining. She is a life member of ISTE, CSI, and a senior member. Kiranmayee's research interests include Data Mining, Machine learning, Algorithms.

Raparthi Sai Priya, Pursuing B. Tech in Computer Science and Engineering from VNR Vignana Jyothi Institute of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University, Hyderabad, India. She is currently working as a Software Engineering intern at NCR Corporation. She is an active member of the Computer Society of India.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

She has a Certification in Big Data Computing through National Programme on Technology Enhanced Learning (NPTEL) and a Full stack certification through VNR VJIET. Sai Priya's research interest includes Machine Learning, Data Science, and Natural language Processing. She has been an active member in this research work and has a significant contribution.

Rayapurthi Vijaya, Pursuing B. Tech in Computer Science and Engineering from VNR Vignana Jyothi Institute of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University, Hyderabad, India. She is currently working as intern at Teachnook. She is an active member of the Computer Society of India (CSI). She was participated in Web development hackathon which was conducted by DBS (The Development Bank of Singapore Limited) company. She completed Certification courses such as Full Stack Development and Google Data Analytics. She was participated in Web page development workshop which was conducted in VNR Vignana Jyothi Institute of Engineering and Technology.

Palthiya Suresh, Pursuing B.Tech in Computer Science and Engineering from VNR Vignana Jyothi Institute of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University, Hyderabad, India. He is currently working as Enterprise Service Engineer intern at E2open. He is an active member of the Computer Society of India.

Regulapati Venkat Goutham, Pursuing B.Tech in CSE from VNR Vignana Jyothi Institute of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University, Hyderabad, India. He is an active member of the Computer Society of India.

Retrieval Number: 100.1/ijrte.A75380512123
DOI: 10.35940/ijrte.A7538.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)