

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 primarily 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 those of the Dravidian family, due to lexical divergence, ambiguity, and other lexical, syntactic, and semantic issues. This research examines several machine translation models for various languages, conducts a comprehensive literature review on the diverse machine translation techniques from earlier studies, and analyses their methodologies. The primary 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, and Parallel 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, as people from different regions speak their regional languages (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 are well-known for their literary heritage: 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.

Regulapati Venkat Goutham, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: <u>https://orcid.org/0000-0002-7335-776X</u>

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an <u>open access</u> article under the CC-BY-NC-ND license <u>http://creativecommons.org/licenses/by-nc-nd/4.0/</u>

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 drawn liberally from Sanskrit. Among all the Dravidian languages, Tamil is the most widely spoken and has the most extensive literary tradition. Only the literature of Sanskrit can be compared to that of Tamil in India in terms of its antiquity, wealth, and geographical reach among the Dravidian languages. Dravidian Languages are regional languages of the southern states of India, where most people speak their regional language. It is essential to understand the languages of different cultures to comprehend their rituals and business activities. Despite being one of the most prominent language families in the world, the Dravidian languages are not particularly well-known due to a lack of resources for language technology activities, such as translation and interpretation. Language translation is crucial in the realm of text-processing applications, including information extraction, machine learning, and language natural 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 this case is to incorporate both machine translation and a human post-editor.

Published By: Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) © Copyright: All rights reserved.



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

1

^{*}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: <u>kiranmayee bv@vnrvjiet.in</u>, ORCID ID: <u>https://orcid.org/0000-0002-7387-9592</u>

Palthiya Suresh, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: <u>https://orcid.org/0009-0009-3050-5206</u>

Raparthi Sai Priya, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: <u>https://orcid.org/0009-0005-6075-6074</u>

Rayapurthi Vijaya*, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad (Telangana), India. ORCID ID: <u>https://orcid.org/0009-0003-7981-5308</u>

This will produce the best results by combining machine speed and human linguistic accuracy. Machine translation software is already a reliable choice for corporations when it comes to translating words. It offers numerous advantages for diverse businesses with varying translation needs, is quick, affordable, and continually improving.

A. Technical translations: Technical texts benefit greatly from machine translation (MT). Technical texts frequently include shortsentences with repeated technical terms, and MT is ideal in this aspect. Since technical writing is often straightforward and repetitive, MT software can easily handle it and maintain consistency in your terminology. 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 postedited information to train specific MT engines for your projects.

B. Legal translations: 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 most significant obstacles, which is where MT comes into play. Machine translation technology can accelerate the understanding and processing of numerous legal documents for law firms and corporate legal departments. Machines can handle the majority of the volume, but there is no room for error when translating legal documents. It is crucial that translated documents accurately represent the meaning, organisation, and language of the original text when dealing with legal issues. Here, NMT can 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 newareas and nations. Any retailer aiming to expand its customer base and enhance sales should now consider international expansion as a basic strategy. To take advantage of this worldwide potential, each piece of content on an e-trade website ought to be translated into the language of the intended audience. This will help you determine further steps to make your services or products more accessible to a much wider audience. E-commerce websites are an ideal choice for automatic translation software due to the vast amount of online content and the standardised structure of these websites. MT is also excellent for developing focused advertising campaigns that have a predetermined format and canincrease your exposure in a new area. Businesses can easily overcome language barriers thanks to MT, which helps them build 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 tofraud. These companies might be the targets of fraud themselves, or a 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 to acquire information and safeguard the business from harm. By assisting businesses in identifying and resolving potential fraud and unethical behaviour, translation services often play a significant role in these investigations. By swiftly and efficiently translating emails, MT technology is well-suited for detecting fraudulent behaviour early on. Businesses can lessen potential financial and reputational harmby 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 the process of translating a source language into a target language without human intervention. There are several approaches to machine translation, including rule-based machine translation, Example-based machine translation, Neural machine translation, Hybrid machine translation, the Transformer model, Statistical machine translation, and the Long Short-Term Memory model, which are described in detail below for each model.

A. Rule-Based Machine Translation (RBMT)

The first professional automatic translation system, rulebased machine translation (RBMT), is built on syntactic ideas that enable words to be arranged in various ways and to have varied meanings according to the context. Three distinct uses of RBMT technology aretested on an extensive 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 utilises manually compiled translation dictionaries, some of which can be modified and enhanced by the user, to improve translations. The translations are nonetheless readable because they adhere to rules, but they are frequently less fluid, resulting in a more "machine-like" reading experience for the output style. 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 to get around some of RBMT's shortcomings, have lately added several statistical machine translation (SMT) techniques to the basic RBMT technology.



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

2

Published By:

© Copyright: All rights reserved.





Figure 1: Rule-based Machine Translation

B. Neural Machine Translation (NMT)

Neural network technology is utilised in contemporary machine translation methods, such as 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. To overcome challenges in language translation and localisation, 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 NMT approach differs from traditional translation SMT systems in that the neural translation model is refined from start to finish to optimise translation efficiency. Neural machine translation creates a solitary substantial learning algorithm that scans a text and provides an accurate translation, in contrast to conventional phrase-based translation systems that consist of numerous small, individually tuned subcomponents.



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 Examplebased machine translation is used most frequently. If not, it will search the database for the sentence that is closest in meaning and then adjust the translation to account for any discrepancies. Additionally, there is a chance to successfully translate the source sentences. It is crucial to have a vast amount of translated text to move forward with Example-based machine translation. In other words, many parallel multilingual documents should be stored in the database. Professionals have translated these texts. 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 used for Example-based machine translation is reliable and consistently produces excellent results. The approach 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 encode the knowledge simply. The translation tool will take a piece of text, compare it to a similar piece of textstored in the database, and then translate it.

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



Machine Translation on Dravidian Languages



Figure 3: Example-based Machine Translation

D. Hybrid Machine Translation

A hybrid machine translation system, which incorporates various language translation technologies, is an automated translation system. The creation of hybrid machine translation systems is driven bythe fact that no single technique has been able to produce results with a respectable level of accuracy. This hybrid machine translation technique utilises the simultaneous operation of multiple machine translation systems. The end outcome is created 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 the systems at the end of the translation pipeline could offer an improvement in translation quality that is difficult to find elsewhere, because any change made to one system will almost surely result in an improvement to the combined system. Another reason to adopt acombined system is that it can provide linguistic insight into areas where the RBMT system and SMT system perform better or worse. Hybrid machine translation (HMT) is a paradigm that attempts to combine the advantages of various methods. Due to the complementary advantages 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. Transformers are designed to examine sequential input data, such as natural language, and have applications in text summarisation and machine translation. They work similarly to recurrent neural networks. Unlike RNNs, input is processed by transformers at once. The transformer does not need to examine each word separately if the incoming data is in natural language. RNN parallelisation is less than that of the Transformer, which reduces training time. Transformer models are the preferred models for NLP challenges, replacing RNN models such as LSTM. Using the added parallelisation for training, larger datasets can be trained more efficiently. Pretrained systems like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers) were developed as a result. These systems, which can be customised for specific applications, were trained using substantial language datasets, such as the Common Crawl and Wikipedia Corpus.







F. Statistical Machine Translation

The statistical machine translation (SMT) theory states that the properties of numerical methods are verified using bidirectional textual collections before being translated into other languages. SMT refines its transcribing skills by examining past human interpretations. The bulk of SMT methods used today use groups of words and combine different cultures or societies on overlapping phrases.By translating complete 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 the higher retention of information from previous periods, are generated bymodifying recurrent neural networks. Gradient descent and long-term dependency are two problems with RNNs that LSTM addresses. 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 behaviour of a recurrent neural network, is similar. The four gates that make up an LSTM are the forget gate, the input gate, the input modulation gate, and the output gate. A "memory cell" found in LSTM units can store data in memory for extended periods of time.



and Sciences Publication (BEIESP)

© Copyright: All rights reserved.

.iirte.org

oring Innova

III. LITERATURE SURVEY

Peter F. Brown et al [1] proposed A Statistical Approach To Machine Translation. They stated that translation requires the computation of two types 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 utilised both the language model and the 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, comprising approximately one hundred million words. To compute the likelihood of any sentence, they broke it into words and multiplied together the probabilities of the individual words. Finally, they applied a method for deriving a probabilistic phrase structure grammar to construct a 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 recogniser 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 the acquisition of the corpus and its use in addressing the issue of statistical machine translation. One hundred ten machine translation systems have been created using the corpus to accommodate all conceivable language pairs. The systems that were produced and their performance demonstrate the various difficulties that statistical machine translation faces for different language pairs. 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 genuine 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 faces several challenges, which it has only just begun to address, as evidenced by the wildly different quality of the numerous SMT systems for 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 relies on several 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 demonstrate that this method can produce results that are highly correlated with human review, making it a valuable and efficient tool for evaluating 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 argue that the community should evaluate machine translation systems more comprehensively, considering 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 case, 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. Ultimately, they advocate that the machine translation community focus on developing systems that not only provide translations with a high BLEU score but also ones that convey the same meaning as the original text. Alternative evaluation techniques that are better suited to capturing the nuances of human language should be investigated by the community to achieve this. They conclude that while Bleu is a valuable 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 Machine Translation for Indian Languages. It demonstrates that certain discernible language principles can aid in making automatic translation easier and less expensive when converting from English to low-resource languages, as well as the impact of the number of rules on accuracy. All Indian languages share the same structural foundation; hence, they all adhere to 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,

determiner, adjective, and other is included in the Word Dictionary File.





Romanised Malayalam is mapped to the Unicode format **ROMANTOUNI-Code** file. in the 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 target 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-objectverb grammatical structures between English and Malayalam, the transfer link rule file plays a crucial role in facilitating the translation process. 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 natural languages that are bilingual is a challenging and timeconsuming 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 suffixes in both English and Dravidian words, the English source sentence is rearranged using Dravidian syntax, and utilising morphological information is the central concept that they have put into practice and has proven to be very effective in SMT systems from English to Dravidian languages in South India. 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 applicable to various categories of sentences can be incorporated into the proposed system to enhance 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, which can then be solved individually. Tamil and Telugu can be easily incorporated into the strategy because the South Dravidian languages share a very similar structural foundation. Given that many languages in India use the Subject-Object-Verb order and are very morphologically rich, the methodology described here is generally suitable for 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 emphasised the incorporation of preprocessing methods, such as suffix separation, for better alignments, as it improved the BLEU score. During the training phase, the translator utilises a monolingual Malayalam corpus as well as a bilingual English/Malayalam

Retrieval Number: 100.1/ijrte.A75380512123 DOI: <u>10.35940/ijrte.A7538.0512123</u> Journal Website: <u>www.ijrte.org</u> 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 simplified by the alignment model's knowledge of category tags, named entities, cognates, and predictable Malayalam translations, which reduces inconsequential alignments. This method helps enhance the accuracy of Malayalam terms 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, the 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 does 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 require minimal human effort and linguistic expertise. As a result, creating an SMT system for a new set of languages is straightforward 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 enhances a phrase-based SMT system

by integrating it 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 expand the study further.

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 presents an alternative argument that NL comprehension is significantly more complex than NL generation. This study discusses the challenges of statistical machine translation, including sentence alignment, compound words, idioms, morphology, and different word orders, as well as a statistical model approach to the machine translation problem. They introduced a query translationbased technique in their system, as it is efficient for translating questions about the documents. The program utilises the English stop word list, the Machine-Readable Dictionary, the transliterator, the Art of Speech Tagger (POST), and the 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 applied simply to additional language pairs with diverse sound systems.

Manoj Kumar et al [13], the distinction between rulebased and corpus-based machine translation is discussed in study. this gathering, Matching, Example 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 second with a language translation quality of 0.43. 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 traditional diagnostic criteria, such as the Bilingual Evaluation of Study Abroad. 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 encoderdecoder models: the newly suggested gated recursive

convolutional neural network and an 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 known or uncommon words in sentences. The investigation revealed that phrase length has a significant negative impact on the performance of neural machine translation. However, qualitative analysis shows 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 to overcome restrictions beyond 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. To put the concept into practice, the decoder employs 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 considers only a small number of source words at once, and global attention, which continuously monitors all source words. They provide 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 nonattentional 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 conducted experiments with various source lengths, demonstrating that the global model performs better with longer source phrases. In contrast, the local model can only use a very limited number of source phrases. To better comprehend the translation process, they studied attentional settings and developed a state-of-the-art visualisation tool. As a result of their investigation, they conclude that attentional contexts can be considered as subsubjects and that attentional weights are utilised to align source and target terms, as well as to select which themes are communicated in the translation.

Hamid Palangi et al. [19] described phrase embedding as a hot topic in current natural language processing. It was

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





reported that Recurrent neural networks (RNNs) incorporating Long Short-Term Memory (LSTM) cells were utilised to build a model that solves the problem.

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 are used to train the LSTM-RNN in a weakly supervised manner. To comprehend how the embedding procedure functions, visualization and analysis are conducted. It has been discovered that the model automatically eliminates unnecessary words and recognises the significant keywords in the sentence. Additionally, it has been found 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 significantly 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. Practical translation tasks can also be performed using this model.

Rohit Gupta et al [20] proposed an MT evaluation metric that was submitted for the WMT-15 metrics. Modern Machine Translation (MT) evaluation measures are often complex, require significant external resources (such as paraphrasing), and necessitate adjustments to produce the best results. This study employs a measure based on a type of recurrent neural network called Long Short-Term Memory (LSTM) networks and dense vector spaces (DVSs). 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 often complex and require 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 sequence learning 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-spacebased metric, Reval, is competitive with the most sophisticated alternative approaches currently being used, which involve combining different systems, utilising numerous external resources, and engineering features, as well as 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, an MT system-supported combination of approaches is a fun task and can significantly help address issues associated with current MT methods. They intended to combine entirely distinct models in a way that leverages their benefits and helps address the problems with translation procedures, providing reliable 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. ems

Dr.. Piyush Kumar Pareek et al [22], To reduce redundant ngrams, most current techniques are limited to monolinguals. To address this, an efficient machine translation model based on machine learning techniques is deployed. The studies conducted on Internet information demonstrate exceptional achievements in terms of precision and processing difficulty of arrangement when considering various thresholds. The accuracy performance for English to Kannada, English to Telugu, and Kannada to Telugu was 91.63%, 90.65%, and 88.63%, respectively.

Dr.(Mrs.)V.VIDYAPRIYA et al. [23] The research aims 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 proposed English-to-Tamil translation system yielded more effective outcomes compared to the existing system. The proposed NMT system blends a rule-based approach with a statistical approach. The textis transliterated using statistical MT. Tokenisation, preprocessing, and translation are among the processes required. 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 istrained 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 search algorithm quickly discovers the most excellent probability translation among the exponential number of options. The translation systemoutput is chosen from the sentences with the most significant 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.

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



Additionally, it demonstrates the value of transliteration in UNMT between closely related languages that do not use the same writing system.

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 a practical 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 examined, along with pretraining on single data and supervised training on paired dialects. Evaluate multiple architectures such as Unsupervised, dissimilar languages, Unsupervised similar languages, Unsupervised, Dissimilar languages + cross translation, and Unsupervised, similar languages + cross translation with different reference languages. For Dravidian languages, datasets are sourced from the AI4Bharat-IndicNLP corpus, while for English, they are 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 language family groups languages, and (2) language embedding, in which an embedding vector represents each language, 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, which encompass various languages, are utilised to evaluate the experiments.

Bharathi Raja Chakravarthi et al [27] provided MMDravi, a Multilingual Multimodal dataset for languages in the Dravidian family with limited resources. It contains thirty thousand sentences generated using a variety of machine translation outputs. Multimodal machine translation is the process of converting text between two languages using information from multiple modalities. To utilise multilingual corpora and other modalities, for closely related Dravidian languages, they created a Multilingual Multimodal Neural Machine Translation system (MMNMT) that leverages data from MMDravi and the phonetic transcriptions of a corpus. 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 MI-En language combination.

According to the evaluation that utilised phonetic transcription, multilingual, and multimodal NMT, the recommended system outperforms the present strategy of multimodal and multilingual NMT in less-resourced neural

Retrieval Number: 100.1/ijrte.A75380512123 DOI: <u>10.35940/ijrte.A7538.0512123</u> Journal Website: <u>www.ijrte.org</u> machine translation across all considered language combinations.

Bharathi Raja Chakravarthi et al[28], suggested Comparing Different Orthographies for Dravidian Languages with Limited Resources. In this model, multiple sources and target languages are trained concurrently, a technique known as multilingualneural machine translation. This has been demonstrated to enhancetranslation quality; however, the use of these multi-way models was restricted in this study because they concentrated on languages with different scripts. The BLEU ratings were calculated to assess if combining them into a single script would allow the system tobenefit from the phonetic similarities between these closely related languages. 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 essential for NMT systems to produce more accurate translations. The equivalent phrases should also encompass a range of topics and convey similar sentiments. Using such a corpus to model the system can ensure accurate translations during model evaluation. 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 developed 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 finetuning with real human-generated data. The output of both Indo-Aryan and Dravidian languages is measured according to the complexity of their morphology. As a result, they found that systems with Dravidian languages performed very well, almost reaching the level of direct human translations due to the morphological and syntactic complexity of the language. As a result, significantly more room was provided for transfer learning. After tuning, Systems with Indo-Aryan languages also show 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

parallelcorpus. They compared the recently developed Transformer architecture with recurrent models.





They were contrasted with statistical machine translation on low-resource Tamil to Sinhala translation assignments (SMT). They were contrasted with the Statistical Machine 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). However, later on, they found that the Recurrent Neural Network (RNN) has proven to be the most effective method for translation. They used the Sequence-to-Sequence approach on the dataset of the Encoder-Decoder process, utilising Long Short-Term Memory (LSTM)as an RNN unit. They measured results with regard to Statistical Machine 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 leading causes of this increase is the deep neural networks' capacity to learn a logical representation of words. The encoder-decoder with an attention mechanism is part of the proposed NMT paradigm created for the Telugu 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 tokeniser module was used to tokenise the sentences provided in the toolkit called Moses before being fed into he 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, they have developed a system that utilises an attention-based neural model. With evaluation metrics such as BLEU, TER, and perplexity, the proposed attention-based NMT model is evaluated. When compared to Google Translate, the network translated English toGujarati with greater accuracy, 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, which shows that the Helsinki-NLP Opus MT model dominates the Irishlanguage task, surpassing other Transformer models. Ranked 1, 1, and 2 in English, Marathi, Irish English, and English Irish, respectively, 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 primarily focuses on storing a substantial amount of data in a backup. They have removed all the unnecessary sentences through good preprocessing. This helped them to produce effective translations. They have also 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 relatively 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 models using subword units. Byte pair encoding (BPE) is a preferred method of subword segmentation. 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, the Long Short-Term Memory (LSTM) framework was explained. The effectiveness of the model is examined using several computerised metrics for strategic data use, including BLEU, F-measure, NIST, and WER. After several phases, the model generates an overall BLEU score of 45.83.

Ashalatha Nayak et al [38] The OpenNMT architecture is used to implement the transformer. The encoder and decoder of the model consist 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. Thirty-two batches are utilised to optimise the model. The IndicNLP library, specifically designed for Indian languages, is utilised in tokenisation. The tokeniser processes both the source language and the destination language. The output of the tokeniser is incorporated into the model during training. The accuracy of the machine translation models is evaluated using the BLEU score, a widely used metric for evaluating machine translation models. 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 demonstrates the effectiveness of the suggested strategy. Using more sophisticated methods, such as back-translation, ensembling, and data augmentation, the model can be further improved. In conclusion, the Transformer-based method can produce accurate translations of Dravidian languages. ges

Ra.lf 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, one of the translation engines employed by Pangloss. EBMT is a type of translation that utilises 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 primary sources of knowledge.

Nadeem Jadoon Khan et al [40] presented the result of the state-of-the-art phrase-based Statistical Machine Translation on multiple Indian Languages (SMT). They aim 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 evaluated the performance of



systems translating Indian languages into English, achieving an average accuracy of 10%.

IV. RESULT

To overcome the drawbacks of traditional models, it is necessary to utilise more effective models that offer improved accuracy, resulting in enhanced translation quality compared to conventional 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 type of innovation that best suits your requirements will vary depending on the language pair. The traditional models require technical expertise and take too much time to implement from scratch. Implementing a conventional algorithm from scratch requires significant time and technical knowledge. It needs some input and 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, to improve the quality of an RBMT, rules need to be changed, which requires more linguistic knowledge. SMT engines, on the other hand, are 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 examining recent models, such as LSTM, Transformers, and BERT, it's clear that they produce good results due to their unique characteristics. For example, LSTM models have long-term dependencies, and Transformers require a significant amount of GPU resources to implement, but they are easier 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. This paper covers the multiple datasets created for Dravidian languages, explaining the significance and nature of the Dravidian family. In this study, a comparative analysis of various procedures from earlier studies and successful experimental strategies from previous work 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 models that include linguistic formation and syntax analysis. This work describes research on various NLP techniques that improve translation accuracy, as well as the combination of experimental models that has proven to be efficient. This work provides research on the challenges in Machine Translation and natural language processing. Also,

it offers a detailed view of the potential techniques of MT for Indian languages, with a focus on 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 or consent to participate, as it presents evidence. nce
Availability of Data and Material/ Data Access Statement	Yes, the Availability of Data and materials, as well as the research data associated with an article, is available in the 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

- Peter F. Brown et al, IBM, Thomas J. Watson Research Centre, A Statistical Approach To Machine Translation (1990), Available at: https://aclanthology.org/J90-2002.pdf
- 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/
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, IBM T. J. Watson Research Centre, USA, BLEU: a Method for Automatic Evaluation of Machine Translation (2002), Available at: https://aclanthology.org/P02-1040/ [CrossRef]
- Chris Callison-Burch, Miles Osborne and Philipp Koehn, School of Informatics, University of Edinburgh, Re-evaluating the Role of BLEU in Machine Translation Research (2006), Available at: <u>https://aclanthology.org/E06-1032/</u>
- Rashmi Gangadharaiah et al, Language Technologies Institute, Carnegie Mellon University, Application of Linguistic Rules to Generalised Example-Based Machine Translation for Indian Languages(2006), Available http://www.cs.cmu.edu/~rgangadh/mspil06_rashmi_modified.
- [CrossRef]
 7. Remya Rajan et al, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Rule-Based Machine Translation from English to Malayalam (2009), Available at: https://ieeexplore.ieee.org/document/5376582
- Unnikrishnan P et al, Computational Engineering and Networking Department, Amrita Vishwa Vidyapeetham, A Novel Approach for English to South Dravidian Language Statistical Machine Translation System (2010), Available at: <u>http://www.enggjournals.com/</u>
- 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
- 10. Aditya Vyawahare et al, Pune Institute of
- Computer Technology, Neural Machine Translation On





Dravidian Languages (2011), https://aclanthology.org/2022.dravidianlangtech-1.28/ [CrossRef]

- 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/
- 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-1-1)/C04111823.pdf [CrossRef]
- Asha Hegde, Ibrahim Gashaw and H. L. Shashirekha, Dept of Computer Science, Mangalore University, Machine Translation for Dravidian Languages using Stacked Long Short Term Memory (2014), Available at:https://aclanthology.org/2021.dravidianlangtech-1.50.pdf
- Kyunghyun Cho et al, Université de Montréal, 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 Computer Science 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
- 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
- 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
- Rohit Gupta et al, Research Group in Computational Linguistics, University of Wolverhampton, Machine Translation Evaluation using Recurrent Neural Networks (2015), Available at:https://aclanthology.org/W15-3047/ [CrossRef]
- 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.lst.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
- 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/jisyji sys-2018018 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
- 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/

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

- B. Premjith et al, Centre 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), Available at:https://www.degruyter.com/document/doi/10.1515/jisys-2019/2510/html?lang=en [CrossRef]
- 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 Weerasinghe, 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
- 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/2021 mtsummit.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]
- Ashalatha Nayak et al, Department of Computer Science & Engineering, Manipal Institute of Technology, Machine Translation in Dravidian Languages (2022), Available at: <u>https://aclanthology.org/2022.dravidianlangtech-1.19/</u>
- Ralf D. Brown, Centre for Machine Translation, Carnegie Mellon University, Pittsburgh, Example-Based Machine Translation in the Pangloss System (2022), Available at: https://dl.acm.org/doi/10.3115
- Nadeem Jadoon Khan et al, COMSAT University, Machine Translation Approaches and Survey for Indian Languages (2022), Available at: <u>https://arxiv.org/abs/1701.04290</u>

AUTHORS PROFILE



Dr. B.V.Kiranmayee was awarded her Ph.D. in Data Mining from JNTU-H, Hyderabad. She completed 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, which is

affiliated with Jawaharlal Nehru Technological University. Currently, she is a Professor in the Computer Science and Engineering Department. She teaches C programming, Data structures, MFCS, computer organisation, Design and analysis of Algorithms, UNIX Programming, human-computer interaction, Digital logic Design, Advanced databases, cloud computing, and Data warehousing and data mining. She is a life member of ISTE and a senior member of CSI. Kiranmayee's research interests include Data Mining, Machine learning, and 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.

She holds a Certification in Big Data Computing through the National Programme on Technology Enhanced Learning (NPTEL) and a full-stack

certification from VNR VJIET. Sai Priya's research interests include Machine Learning, Data Science, and Natural Language Processing. She has been an active member in this research work



Machine Translation on Dravidian Languages

and has made 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 an intern at Teachnook. She is an active member of the Computer Society of India (CSI). She participated in a Web

development hackathon conducted by DBS (The Development Bank of Singapore Limited). She completed Certification courses, including Full Stack Development and Google Data Analytics. She participated in a Web page development workshop conducted at 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 an 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.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



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

Published By:

© Copyright: All rights reserved.