Journal Impact Factor (JIF)™ 2023: 0.2 | 5 Year JIF: 0.2 | Quartile: Q2 | *Source: JCR™ 2024 (Clarivate Analytics)*

RUPKATHA JOURNAL
ON INTERDISCIPLINARY STUDIES IN HUMANITIES

AESTHETIX MEDIA

ISSN 0975-2935 | Indexed in the Web of Science Core Collection™ Emerging Sources Citation Index (ESCI)

https://doi.org/10.21659/rupkatha.v17n1.01 | Volume 17, Number 1, 2025 | Dynamic Impact Dimensions

Research article

Error Analysis of Machine Translation for Malayalam Fiction

Hanna C H[1], Sabeerali KP[2] & Vrinda Varma[3]

[1]Research Scholar, Department of Humanities Arts and Social Science, NIT Calicut, India.

[2] Independent Researcher, India.

[3] Assistant Professor, Department of Humanities Arts and Social Science, NIT Calicut, India.

Abstract

Machine translation (MT) has transformed translation studies and linguistics, significantly improving cross-cultural communication and linguistic analysis. The aim of this study is to evaluate and compare the accuracy of Google Translate, AI4Bharat's IndicTrans2 and Bing in handling Malayalam compound nouns, with a particular focus on Named Entity Deviation Errors. This study seeks to identify specific challenges in translating Malayalam noun formations and case markers, and to understand their impact on translation quality. Utilizing a mixed-methods approach, this research involved quantitative and qualitative analyses of three corpora built from a selected fiction text in Malayalam and its human English translation. The findings revealed significant issues in translation accuracy and some common errors were identified, including improper translations of proper nouns, mistranslations of compound nouns, and transliteration issues. Automated metrics used to analyse errors in each MT model revealed that literary-adapted machine translation models produced richer output and showed improved performance compared to general domain models. The study accentuates the necessity of robust linguistic models and larger, high-quality parallel corpora to enhance MT accuracy for low-resource languages such as Malayalam. This study suggests a hybrid approach that develops MT to achieve greater precision and reliability in translation practices, ensuring nuanced and contextually accurate translations.

Keywords: Machine Translation, Google Translate, IndicTrans2, Bing, Malayalam compound nouns, Named entity deviation errors.

Conflicts of Interest: The authors declared no conflicts of interest.

Funding: No funding was received for this research.

Article History: Received: 31 January 2025. Revised: 20 March 2025. Accepted: 21 March 2025. First published: 26 March 2025.

Published by: Aesthetix Media Services, India ↗

Citation: Hanna, C. H., Sabeerali, K.P., & Varma, V. (2025). Error Analysis of Machine Translation for Malayalam Fiction. *Rupkatha Journal,* 17(1). https://doi.org/10.21659/rupkatha.v17n1.01

1. Introduction

Machine translation (MT) has revolutionized translation studies and linguistics. Tools such as Google Translate are increasingly being used, leading to the development of hybrid translation methods. MT's rapid advancements, particularly in neural machine translation, have significantly enhanced translation quality, enabling more effective cross-cultural communication. Linguists have leveraged MT to analyse language patterns across different languages, contributing to understanding language universals and comparative linguistics.

The study aims to evaluate the performance of Google Translate, IndicTrans2 and Bing in translating a Malayalam literary text into English, with a particular focus on compound nouns and Named Entity Deviation Errors (NEDE). By comparing these machine-generated translations with human-translated texts, the research seeks to identify recurring issues in MT, such as improper noun translations, including errors related to case markers in the Malayalam fiction *Anattam Piriyatham* (2021) by Vinoy Thomas, translated into English by Nandakumar K as *Elephantham Misophantham* (2023). This research employs a mixed-methods approach, combining quantitative error analysis with qualitative assessments of the translated text. We manually evaluated Corpus 1 for Named Entity Deviation Errors (NEDE) by comparing a Malayalam parallel corpus with its human translation and three machine translation outputs. To understand the NEDEs, we further analysed corpus 2 and compared how human translations and three MT systems handled Malayalam compound nouns with case markers. We then quantified the error rate using automatic evaluation metrics. Finally, the study adopted a literary corpus 3 with all the passages from the selected fiction to a baseline general domain system to analyse the effectiveness of literary content as data to train MT models.

Through this comparative analysis in accurately translating Malayalam, a low-resource language, the study highlights the limitations of current MT systems. The errors identified in this analysis help focus on areas where further development is needed, particularly for MT systems operating in Malayalam. The study postulates that the inclusion of literary works with unique linguistic and cultural factors, such as complex syntactic structures, idiomatic expressions, and cultural references, in MT training datasets with proper manual annotation might enable systems to better manage the subtleties of language, such as figurative and idiomatic usages. Thus, the research focuses on the need to develop expanded MT training data beyond conventional and non-literary texts that might improve translation accuracy and quality, particularly in the literary domain for low-resource languages.

2. Related Works

2.1 Error Analysis in MT

Accurate evaluation is essential for assessing MT quality. Traditional metrics like BLEU give a swift evaluation (Papineni et al., 2002; Popovic, 2015). Subsequent developments like METEOR (Banerjee & Lavie, 2005) and TER (Snover et al., 2006) incorporated semantic and syntactic information (Snover et al., 2009). Another method is Word Error Rate (WER), one of the earliest MT evaluation metrics (Levenshtein,1966) based on edit-distance, introduced mainly as metrics for speech recognition. Later, the limitations of WER paved the way for the Post Independent

Word Error Rate (PER) (Nießen et al., 2000), which ignores word order in evaluation (Tillmann et al., 1997;). The advent of NMT led to deep learning-based metrics, such as BERT Score and Mover Score that help with capturing semantic nuances more effectively (Zhao et al., 2019; Lee et al., 2023). These studies show a development in error evaluation tools in the case of linguistic feature inclusion, which has led to a deeper, more context-aware understanding of translation quality. Sellamuttu (2024) argues that the integration of graph-based representation and pyramid attention improves the ability of systems to capture difficult linguistic patterns in English, and they proposed the OGDED-PA model with 99.3% accuracy in grammatical error detection. Comprehending the error patterns in MT and their subsequent interpreting is crucial for system refinement. Costa et al. (2015) developed a comprehensive error taxonomy in English to European Portuguese, identifying lexical and semantic errors as the most prevalent and impactful in MT. Mirzaee and Mousavi Razavi (2021) studied error types in English-Persian pairs, finding omission errors most frequent, with no significant directional bias. Costa, Correia, and Coheur (2016) extended their error analysis by correlating error types with translation quality, demonstrating the significant negative impact of semantic errors in rendering the meaning of the source text. The main problems they identified as negatively affecting translation quality are confusion in a sense, wrong choice, and misordering. There is a variety of research on languages like English, Spanish, German, French, Italian, and Chinese suggesting different error taxonomies for MT, which include errors identified at multiple levels like vocabulary, grammar, and discourse, as well as further sub-categorised errors such as concordance errors, stylistic errors, and confusion in word error (Farru´s et al., 2010; Gutie´rrez-Artacho et al., 2019; Krings, 2001; Laurian, 1984; Scha¨fer, 2003; Vilar et al., 2006). Misspelled words are also studied by Gupta et al. (2021), who analysed user-generated reviews. There were studies also focused on including ungrammatical constructions and colloquial expressions. A study on a practical approach for improving MT between English and Chinese through error analysis and reduction revealed that structural errors are predominant. The study reduced these errors by parsing sentences into Naming Telling clauses. (Fang et,al., 2016). The appropriation of parsing as a data preprocessing method is a better step towards increasing accuracy. It helps systems to grasp long sentences. While these studies provide valuable insights, further research is needed to expand error taxonomies to cover a range of language pairs. Research on the correlation between specific error types and language characteristics can contribute to developing tailored error correction strategies for machines. Despite these promising developments, human evaluation continues to be essential, though it is resource-intensive, for assessing fluency and adequacy (Chatzikoumi, 2019). Castilho et al. (2018) discuss the comparison between human and automated analysis, and noting that automated metrics fail in evaluating syntactic and semantic equivalence. Scholars also address the inconvenience of managing the time and the cost of human evaluation (Graham et al., 2013, 2015). Combining automated and human assessments, hybrid evaluation methods are increasingly seen as a promising approach for ensuring accuracy and reliability in machine translation evaluation. (Zhao et al., 2023).

2.2 Error Analysis in Indian Languages

Error analysis in MT systems for Indian languages has revealed significant challenges, especially when dealing with morphologically rich languages such as Hindi and Tamil. The studies in the

Indian context point out the lack of resources in Indian languages that, in turn, hinder the MT quality. Saxena et al. (2022) evaluated the performance of unsupervised statistical MT (USMT) systems for five languages: Hindi, Tamil, Telugu and endangered Kangiri along with English, using monolingual datasets to address the lack of parallel corpora. While they employed evaluation metrics like BLEU, TER, WER, METEOR, and MER, their findings indicated that these metrics often fail to capture the linguistic nuances of Indian languages, leading to inaccurate error scores and assessments. The study by Ramesh et al. (2020) compared phrase-based statistical MT (PB-SMT) and neural MT (NMT) for low-resource language pairs such as English-to-Tamil and Hindi-to-Tamil in a specialized domain and found that PB-SMT generally outperformed NMT in low-resource settings. They highlighted specific errors in translation of terminologies, like lexical selection and word order, which were more frequent in NMT outputs, especially when dealing with data sparsity and polysemy (Ramesh et al., 2020). Moreover, this study noted that standard evaluation metrics like BLEU tend to miscalculate the translation quality for morphologically rich languages due to free word order in languages like Tamil. Saxena et al. (2022) further emphasized that current metrics such as BLEU and METEOR are insufficient for evaluating translations of Indian languages, which often have multiple valid translations due to their agglutinative and complex morphological nature. Still, the most appropriate MT output can be considered as the one which is closer to human translation. While focusing on Malayalam MT error analysis, Sreelekha,S.(2020) explores various approaches to MT between English and Malayalam, focusing on SMT (Statistical Machine Translation) and RBMT(Rule-Based Machine Translation), discussing the challenges of translating between these languages—such as structural, lexical, and semantic ambiguities—while providing a comparative performance analysis that shows SMT generally outperforms RBMT in fluency and naturalness, although RBMT is more effective at handling complex morphological forms in Malayalam, suggesting that incorporating morphological processing into SMT could enhance translation quality. These studies emphasise the need for more targeted error analysis and the development of evaluation metrics that address the linguistic characteristics of Indian languages accurately while also calling for advancements in corpus creation and the integration of linguistic knowledge to enhance MT accuracy.

2.3 Literary Text in MT

Using MT for literary text translation is difficult because it requires more than a word-to-word transfer of meaning; it focuses on artistic integrity and the cultural and emotional needs of that particular genre. One of the early studies in this area, Toral and Way (2014), demonstrated that MT can assist human translators through post-editing works. However, they revealed the limitations of MT in capturing the complex linguistic and emotional subtleties of literary texts. Traditional metrics such as the BLEU score failed to grasp the intricacies of literary texts, limiting the broader applicability of their findings. Later, Matusov (2019) adapted neural machine translation systems specifically for fiction from English to Russian and German to English. Matusov's study contributed an advanced error classification system to address the contextual challenges of literary translation, such as narrative coherence and rhetorical nuance. However, the tiny dataset used in the study and the resource-intensive nature of NMT asks for a larger and more extensive corpus to improve literary MT. In another study by Besacier and Schwartz (2015), it was found that phrase-based statistical machine translation could be used to translate short stories from English to French for better results. They also suggest that MT followed by post-editing is beneficial in producing high quality translation relatively fast. Post-editing can improve translation quality, particularly when translations are produced by non-professional translators. This highlights the need of human intervention in literary translations to maintain cultural and artistic depth.

The complexity of literary texts, rich in metaphors, idiomatic expressions, and rhetorical devices, poses significant challenges for MT systems. Omar and Gomaa (2020) and Vazquez and Mitkov (2023) emphasised that MT systems produce literal translations, failing to understand the contextual and cultural needs of literary texts. Genzel et al. (2010) found that MT struggled to preserve genre-specific elements such as meter and rhyme in poetic translation, leading to translations that lose the artistic qualities of the text. MT also seems to struggle when it comes to discourse-level features such as referential cohesion, leading to less coherent narratives. Voigt and Jurafsky (2012) emphasise considering discourse beyond sentence-level translations to ensure the quality of literary works. More complications in MT's effectiveness is the challenge of cultural adaptation. Bentivogli et al. (2016) and Toral and Way (2018) noted that literal translations often miss cultural nuances, which is especially problematic when translating between languages with different cultural and linguistic frameworks, such as English, Arabic, or Indian languages. This limitation is magnified in low-resource contexts, where MT systems lack the necessary training data to handle the intricacies of culture-specific references. Vazquez and Mitkov (2023) investigated NMT systems across different literary genres and time periods. They used two works from two genres, novels and poetry from different literary periods, to show that better results are obtained from such tailored training data. However, the works from earlier periods faced challenges owing to their complex syntax. Thus, such studies demand better representative corpora across languages from the literary domain in MT development. Omar and Gomaa (2020), identified that students using MT for translating literary texts encountered serious errors and thus advocated incorporating MT into translation education while also emphasising the importance of human post-editing. A study by Guerberof-Arenas and Toral (2024) on the MT of a Kurt Vonnegut story translated into Catalan and Dutch found that human translation had the highest quality and precision for Catalan readers, whereas Dutch readers favoured post-edited MT versions. The highest reader preference was for the original human-translated English version, suggesting that linguistic or cultural factors influence how translations are received by readers. Therefore, beyond literal translation quality, factors such as reader preferences, patterns, and attitudes play an important role in deciding the reception of literary translations.

2.4 In Dravidian languages

MT in literary texts poses complications in different ways, which is more evident when applied to the complex linguistic needs of Dravidian languages. The varied syntactic structures, cultural expressions, and particular literary forms intrinsic to languages such as Tamil, Malayalam, and Kannada pose significant challenges for MT. Ramalingam and Navaneethakrishnan (2024) evaluated the same in their research on Tamil classics, emphasising the distinctive aspects of classical Tamil literary works, which significantly differ from the training data used in NLP systems. This study demonstrated the importance of incorporating discourse parsing to achieve more accurate results in tasks such as text classification and data retrieval. Chakravarti et al.,2021) examine Hindi poetry in MT, which helped study how MT addresses the intricate use of poetic language nuances. Chakravarti et al. 's (2021) study focuses on the challenges MT systems face in conveying the structural and emotional nuances of Hindi poetry when translated into English. The study found that translation, especially from English to a regional language, can be more accurate, whereas the reverse process—in this case, from Hindi to English—presents significantly greater challenges. While such studies highlight the requirement for extensive corpora and enhanced MT models, they also stress the need for translation systems that can maintain the essence of Dravidian languages. Therefore, this evaluation suggests that further improvements in MT are necessary to bridge the gap between Dravidian languages and English. Studies have also focused on the challenge of translating dialects in Dravidian languages and have shown the limitations of MT. Patil et al. (2024) studied this on the dialectal variations in Kannada, which are varied across regions and often fail to be understood by MT systems. This study proposes the use of LSTM models and Automatic Speech Recognition (ASR) to obtain an accurate translation of Kannada dialects into English. Incorporating dialects as data enhanced the system performance, helping achieve a validation accuracy of 86.32%. Tripathi and Rathod's (2024) papers, *Semantic Model for Fragment of Hindi* Part 1 and 2, though not based on literary texts, nevertheless contribute to research in MT by proposing ruled-based syntactic and semantic models for Hindi and have demonstrated effectiveness in different linguistic phenomena. All of these studies show the diverse challenges of MT in Indian languages, especially in the Dravidian context, and further draw attention to the need for specialised systems that can obtain the syntactical, cultural, and emotional richness of Indian literary texts and dialects.

In low-resource languages, the complications faced by MT in literary translations are more evident. Cadotte et al. (2024) explained the use of literary and poetic texts as training data to improve the MT for Innu-Aimun, an Indigenous language in Canada. This study examined how SMT is better than NMT in low-resource language translation, while literary works in a bilingual context were used for training the systems. By integrating literary texts, this approach may improve the translation accuracy and capture the cultural nuances of low-resource languages. Tahseen (2024) examined the complications of translating English literary texts into Arabic using Google Translate and Reverso Translation, highlighting that MT frequently makes errors, especially when translating rhetorical devices, resulting in mistranslations. To explain this, they used Newmark's (1988) translation model, which explains that human translators, who tend to use communicative translation methods, were better at translating the rhetorical impact of the work.

Existing studies reveal that MT in literary texts largely focuses on Indo-European languages like Hindi, French, German, Russian, Dutch and Catalan, Semitic, and Sino-Tibetan, such as Arabic and Chinese, and Dravidian like Tamil and Telugu. Research on Dravidian languages remains limited because of its language roots, which differ from Indo-Aryan and European languages. As only one of the languages in the Dravidian family of languages in which research is limited, research in Malayalam has also received limited scholarly attention. Although the studies examined here have focused on evaluation metrics, error classifications, and corpus limitations, they have not addressed problems specific to the translation of fiction from Malayalam to English.

All these studies conclude that MT's use in literary contexts has significant limitations, and that researchers and translators must exercise prudent caution when employing MT for artistic works.

Nonetheless, these studies also suggest that these issues can be addressed with adequate resources for MT training, highlighting the importance of large datasets, especially for low-resource languages.

2.5 The Case of Malayalam: A Significant Gap in the Literature

Malayalam, a Dravidian language with a rich literary tradition, has been neglected in MT research. The challenges of translating Malayalam literary texts using NMT have not been thoroughly explored, thus highlighting a significant gap in the literature. Malayalam's agglutinative structure, complex morphology, and syntactic features present unique challenges for MT systems. Although not focused on translation in particular, Jayakrishnan et al. 's (2018) research on emotion detection in Malayalam novels highlights Malayalam's linguistic complexity and the challenges it poses for computational processing. The main challenge was to tag and classify sentences using emotion-indicating words. Kiran et al. (2024) addressed the complexities of Malayalam MT and proposed a set of transformation rules to handle various grammatical features in Malayalam, including nouns, case markers, and tenses in their data. Although this study recommends corpus inclusion for better MT systems, current systems are yet to implement this rule-based training in MT. The scarcity of extensive studies on MT in Malayalam literary texts indicates a broader scope of research in Indian MT research. This also necessitates further research that focuses on refining error analysis methodologies and developing more linguistically enhanced corpora.

2.6 Noun formations

The need to understand word formations according to formal criteria is important to know how they are formed in languages. Bauer (1983) categorises word formations in Indo-European languages into compounding, prefixation, suffixation, conversion, backformation, clipping, formation of blends, formation of acronyms, and word manufacturing. Though Malayalam is a Dravidian Language, all of these word formations can be found except backformation and formation of acronyms in Malayalam (Rajendran, 2019). Radford et al (2009) have defined prototypical compounding as a morphologically complex word that has at least two identifiable elements, each capable of independently functioning as separate entities. Among them, compounding is one of the major methods employed in many languages, including Malayalam. Both compounding and suffixing are widely used in the Malayalam language. This study investigates the compounding phenomenon in Malayalam while it is being translated by machines.

2.6.1 Compound Nouns Formations in Malayalam

Scholars have studied the role of compound words in Malayalam as a productive morphological process. In modern Malayalam, compounding is observed largely in the creation of new words (Rajendran, 2019). In such newly coined compounds, simply placing two nouns related to a new concept next to each other can form a compound word. Asher and Kumari (1997), in their *Descriptive Grammar Series*, have classified noun formation using compounding into three categories: noun-noun compounds, verb-noun compounds, and adverb-noun compounds. Ravindran (1975) elaborated on the development of noun-noun combinations, identifying four types of compounds: endocentric, appositional, exocentric (bahuvriihi), and copulative (dvandva).

- * **മഴമേഘങ്ങൾ -**mazhameghangal <mazha “rain” +meghangal “clouds”> “rain clouds
- * **മരക്കൊമ്പ്-**marakkombu <maram “tree” + komb “branch”> “tree branch”[1]

The formations from noun-noun combinations were elaborately developed by Ravindran (1975). Asher (1997) classifies this type of compound noun as hyponymous compounds, which are nothing but endocentric compounds. In which the resultant compound noun is the hyponym of the head noun. Endocentric compounds are also known as head-dependent compounds, which are a specific type of compound word found in many languages, including Malayalam.

The endocentric compound can be formed out of two common nouns, two proper nouns or a common noun and a proper noun.

1. Common noun + common noun is the more productive form in Malayalam noun compounding. Widely used in newspapers, magazines and dictionaries.

2. Proper noun + noun compound-formation is widely used in modern Malayalam. Mainly, names of places and people are used for compounding.

3. Appositional Compound is in which both compound members have the same referent.

This paper investigates the ability of machine translation systems to accurately identify and translate endocentric formations within the selected work, "*Anatham Piriyatham*," with a specific focus on character names representing people and animals, which are also known as Named Entities. Most of the named entities in this fiction are compound nouns where a noun is compounded with an adjective or name of a place or profession.

2.7 Malayalam Noun Case markers

In *A Grammar of Malayalam* (2012), Ravi Shankar explains how Malayalam uses a set of case suffixes to mark grammatical relations and semantic roles, a common feature among Dravidian languages. Because Malayalam relies on suffixes to show roles and relations, the word order does not affect the sentence's meaning. These are the six cases used in Malayalam: nominative, accusative, dative, sociative, instrumental, and locative.

The nominative case indicates the subject of a verb, while the accusative case marks the direct object of a transitive verb. The dative case specifies the indirect object, referring to the recipient of the direct object. The sociative case expresses companionship. The instrumental case denotes how an action is carried out. The genitive case shows possession, and the locative case refers to a location. Lastly, the vocative case is used to address someone or something directly in speech.

In order to deliver effective Malayalam-to-English machine translation, rule-based systems need to include transformation rules that can map case suffixes correctly into their English counterparts,

---

[1] Examples for Noun Compounds are from *Anattam Piriyattam* by Vinoy Thomas (2021).

along with syntactic reordering, prepositional choice, and possessive forms. Without these transformations, the outputs could be grammatically erroneous or convey the wrong meaning.

3. Methodology

This section outlines the methodological framework designed to evaluate the translation accuracy of Google Translate, AI4Bharath's IndicTrans2, and Microsoft Bing when processing Malayalam literary texts into English. Employing a mixed-methods approach, this study focuses on identifying and analysing lexical-level errors, particularly emphasising the translation of compound nouns and their associated case markers—linguistic features that pose significant challenges due to Malayalam's agglutinative nature and complex grammatical structures.

A series of corpora were developed to facilitate a comparative error analysis, each serving a distinct analytical purpose and enabling systematic error analysis across varying linguistic levels. Corpus 1, containing the Malayalam text and its human translation *Elephantam Misophantam* was used for the comparative error analysis. This Corpus 1 comprised texts from a single literary work, with a combined size of 38,481 running words from both Malayalam and English translation, also with its corresponding MT systems outputs: Google, IndicTrans2 and Bing, for an easy understanding of errors in each MT outputs.

For a focused analysis of lexical-level errors in noun translation, Corpus 2 was created. This specialised corpus thus contains 220 sentences extracted from the original Malayalam fiction text, specifically targeting compound nouns, named entities, and their case markers. Corpus 2 includes four parallel versions of each sentence: the original Malayalam text, its human translation, and outputs from the three machine translation systems (Google Translate, IndicTrans2, and Bing). The corpus was structured to facilitate two levels of analysis: first, a human evaluation of errors in case-marked compound nouns, and second, an automated assessment of translation accuracy for each specific case marking found in the compound nouns.

Corpus 3 consists of the entire selected Malayalam fiction and its human translation, comprising 390 passages, each aligned parallelly as Source and Target. This corpus is used as the literary content for the newly trained MT model to test the effectiveness of literary content in training MT models to get better Malayalam translation outputs.

This study employed a targeted sampling strategy to isolate and analyse the behaviour of proper nouns within sentence structures. The sampling process focused exclusively on sentences where proper nouns appear in their nominal form, excluding instances where proper nouns are replaced by pronouns. This methodological decision aimed at facilitating a more straightforward comparative analysis while also acknowledging the prevailing notion that human translation is often the best possible form of translation.

The study employs a mixed-methods approach, combining quantitative and qualitative analyses. The quantitative component focused on determining the frequency of named entity occurrences, specifically character names, within the human-translated corpus and MT output corpora. #Lancsbox was employed to calculate the frequency of named entities, human-translated versions and errors in all three MT outputs. The subsequent qualitative stage entailed a comparative analysis of errors identified within the three outputs. This analysis utilised a purposive sampling strategy, selectively drawing data from both Malayalam text and its human translation. The extracted data was then organised into a tabular format, facilitating a parallel comparison of the original Malayalam sentences, the human translations, and the corresponding three MT outputs. For the next level of error analysis, the study included automated analysis employed using BLEU and METEOR metrics. This analysis was conducted for each selected noun case, for the overall Malayalam literary content translated by MT, and for the literary-adapted MT model. to understand the effectiveness of Malayalam literary texts as training data for MT models.

4. Error Analysis

4.1. Error Analysis of Compound Nouns with their Case Markers

The error analysis of compound nouns with their case markers comprised both human and automated evaluation. Initially, a human evaluation of each case was conducted, followed by an automated analysis using BLEU and METEOR. For the human evaluation, 6 grammatical cases were selected. Subsequently, for the automated analysis, 4 cases—accusative, nominative, genitive and dative—with the highest error frequencies, as identified in the human evaluation, were selected.

4.1.1. Accusative Noun Cases in Translation

In Malayalam, the accusative marks the object of the sentence. In sentences where there is a nominative, accusative, and dative noun, the nominative will be the subject, the accusative the direct object, and the dative the indirect object (Ravi Shankar, 2021). In the text, the compound nouns in their accusative cases are found to be mistranslated. Taking the samples from Google Translate, Bing and IndicTrans2 results for the above situation.

Malayalam sentence: പക്ഷേ, ഇക്കുറി പിടിക്കേണ്ടത് മിന്നൽകൊമ്പനെയാണല്ലോ.

- **Human Translation:** But this time, he had to tangle with none less than Lightning Tusker
- **Google Translate:** But this time it is the lightning bolt that needs to be caught
- **IndicTrans2:** But this time, the lightning rod is the one to be caught.
- **Bing:** But this time it's the lightning horn that needs to be caught.

The proper noun മിന്നൽകൊമ്പൻ (Minnelkomban) exemplifies an endocentric noun compound, a construction where the first element, മിന്നൽ (minnal), functions as an adjective modifying the second element, കൊമ്പൻ (komban), meaning “tusker.”

These instances are the accusative noun cases of മിന്നൽകൊമ്പൻ translated as Lightning Bolt, Lightning Horn, Thunder Bolt, and Minnelkomban by Google Translate. IndicTrans2 results are Lightning Bolt, Lightning Rod, Lightning Horn, Lightning Struck, and Lightning Club. Bing also generated it as Lightning Horn.

4.1.2. Nominative Noun Cases in Translation

In languages with rich inflectional morphology, the components of compound nouns must agree in terms of gender, number, and case with the rest of the sentence. This agreement often includes nominative case markers when the compound noun is used as the subject. The nominative noun denotes the subject of the sentence.

Malayalam sentence: ഒരുതരം ഇരട്ട വേലിക്കളി. തീ ചാടിക്കടന്നാലും ***മിന്നൽകൊമ്പൻ*** കാട്ടിലേക്കു രക്ഷപ്പെടരുത്.

- **Human Translation**: A treacherous double fence. Even if ***Lightning*** managed to run through the fire, he would not be able to escape into the forest.
- **Google Translate**: A kind of double fencing game. Even if the fire jumps, the ***lightning horn*** should not escape into the forest
- **IndicTrans2**: A kind of double fence. The ***lightning bolt*** must not escape into the forest, even if it catches fire.
- **Bing:** A kind of double fence. Even if the fire jumps, don't escape into the ***lightning-horned*** forest.

There are 30 instances of Nominative cases of word ***മിന്നൽകൊമ്പൻ*** being translated incorrectly. There are 10 transliterated results; 20 are “Lightning Horn” in Google translate output. In IndicTrans2, the results are “lightning bolt” (20 times), “lightning horn” (6 times) and “lightning” (7 times), while Bing translated this as “lightning-horned” in most of the instances. It was found that the nominative cases of മിന്നൽകൊമ്പൻ have been mistranslated in all MT models. This suggests a failure of MT algorithms to grasp the context and grammatical subtleties of the Malayalam word "Komban," meaning tusker. Instead of correctly translating it in the context of an elephant's tusks, the systems produce results akin to "horn" or "horned." While semantically related, these terms are inappropriate for describing elephant tusks, highlighting the systems' poor semantic understanding. In some cases, the translation is even "bolt," possibly due to "bolt" being a frequent collocate of "lightning" (minnal) in the IndicTrans2 training data.

4.1.3. Dative Noun Cases in Translation

The dative is used for indirect objects. The function is categorised as a master to denote ‘’independent’

Malayalam sentence: മുത്തിയാനയും കുട്ടികളും കാട്ടിൽതന്നെ നിൽക്കുന്നുണ്ടെന്ന് മിന്നൽകൊമ്പനു മനസ്സിലായി.

- **Human Translation:** ***Lightning*** could sense that Granny and the baby elephants were still in the forest.
- **Google Translate**: ***Minnelakomban*** realized that Muthiyana and the children were standing in the forest.

- **IndicTrans2:** The ***lightning bolt*** realized that Muthiyana and the children were standing in the forest.
- **Bing:** The ***lightning horn*** realized that Muthiyana and the children were still standing in the forest.

In dative noun case instances of മിന്നൽകൊമ്പനു in corpus taken for study are 5, among those the Google translate outputs are transliterated "*Minnelakomban"* in 3 instances while in IndicTrans2 it is "*lightning bolt"* in 4 instances. Bing translated it as "*lightning horn"* in all 5 instances. In case of transliteration we can assume that the system tried to get the idea that a proper noun is better rendered as itself.

4.1.4. Vocative Noun Cases in Translation

The vocative marks the addressee of a statement or invocation. Vocative marker changes according to the phoneme at the end of the word.

Malayalam sentence: "അമ്മേ, ആനമുത്തി, കൂട്ടുകാരേ... ഇന്ന് നേരം വെളുക്കുന്നതിനുമുൻപ് എന്റെ ചേട്ടൻ മിന്നൽകൊമ്പൻ സ്വതന്ത്രനായിരിക്കും." കൊമ്പില്ലാക്കൊമ്പന്റെ ശബ്ദത്തിന് വല്ലാത്തൊരു ഉറപ്പായിരുന്നു.

- **Human Translation**: 'Mother, ***gran***, friends... before this night is over, my brother Lightning will be freed.' His voice was firm and filled with grit.
- **Google translate**: "Mother, ***Anamuthi***, my friends... My brother Lightning Horn will be free before dawn today." Hornhorn's voice was very sure.
- **IndicTrans2**: "Amma, ***Anamuti***, friends... before dawn today, my brother the lightning bolt will be free. The sound of the hornbill was very reassuring.
- **Bing:** "Mother, ***Anamutti***, my friends... My brother will be free from lightning before dawn today." The hornless horn's voice was a certainty.

The vocative case of this compound noun "gran," <ആനമുത്തി> means grandmother Elephant. This is a common endocentric compounding where a common noun and a common noun come together to form a proper noun. All three MT models transliterated this compounding ***Anamuti***. Here again transliteration is better than any literal translation, though it's not sure if the system is able to understand the meaning.

4.1.5. Sociative Noun Cases in Translation

The sociative is grammatically similar to the accusative but semantically different. Like the accusative, the sociative also marks the objective. The sociative nouns do not function in the role of experiencer but only as recipients.

Malayalam sentences: മിന്നാമിനുങ്ങുകൾ കെട്ടുപോയതു കണ്ടപ്പോൾ *കൊമ്പില്ലാക്കൊമ്പനോട്* കടുവ പറഞ്ഞു

- **Human Translation**: “There seems to be some problem there. Walk a little faster,” Failcat said to ***Tuskless Tusker*** when he saw the abrupt snuffing out of the fireflies.
- **Google translate:** When he saw that the ***minnows*** were tied, the tiger said to ***Kombilla Kompan***
- **IndicTrans2**: The tiger said to the ***hornbill*** when he saw that the ***lights*** were out.
- **Bing**: When the tiger saw that the ***fireflies*** had ceased, he said to the ***hornless horn***.

In this sentence, കൊമ്പില്ലാക്കൊമ്പനോട് is in the sociative position. Google Translate transliterated the compounding “***Kombilla Kompan,”*** and IndicTrans2 translated it as “***hornbill***", while Bing output is “Hornless horn” for all instances. In the above sample it is evident that it is not only the compound noun that is mistranslated. ***MinNaminungukal*** <**Fireflies**> has been mistranslated by Google (***minnows***) and IndicTrans2 (***lights***), highlighting the challenges faced by machine translation systems in accurately interpreting fundamental vocabulary within a given sentential context.

4.1.6. Genitive Noun Cases in Translation

The genitive marks the owner in possessive phrases. The genitive noun remains outside the basic sentence structure.

Malayalam sentence: ആ മയിൽപ്പീലിക്കൊമ്പന്റെ മക്കളെല്ലേ അവര് രണ്ടുപേരും?

- **Human Translation:** And aren't the two brothers the sons ***of Peacock Plume Tusker***?' said another little elephant.
- **Google Translate:** Aren't they both the children ***of that peacock horn***?
- **IndicTran2:** Are they not the sons ***of the peacock feather***?
- **Bing:** Aren't they the two children of ***that peacock horn***?

In this sentence “മയിൽപ്പീലിക്കൊമ്പന്റെ” is with the genitive noun case marker, which means “**Peacock Plume Tusker’s**”. This is translated by Google Translate and Bing as “***peacock horn”***, IndicTrans2, it is “***peacock feather”.*** All MT systems failed to convey the original compound meaning, which refers to a tusker with a tusk resembling a peacock plume in its coloration. This suggests a potential failure in capturing colloquial ways of compounding in the Malayalam language, where descriptive adjectives can be incorporated into proper nouns. This difficulty likely contributes to the observed mistranslations.

After thorough human analysis among these grammatical cases, we selected cases with errors in greater numbers, namely, the accusative, nominative, genitive, and dative. We have tested the error rate of these four cases in all three MT models to get a more comprehensive understanding of the errors in a quantitative way. We employed BLEU and METEOR analysis.

Case Marker	System	BLEU	METEOR
Accusative	Google	0.089278	0.269806
	AI4 Bharath	0.089848	0.259181
	Bing	0.085792	0.266399
Nominative	Google	0.087360	0.256748
	AI4 Bharath	0.082120	0.261204
	Bing	0.082865	0.254335
Genitive	Google	0.108782	0.304284
	AI4 Bharath	0.129863	0.326245
	Bing	0.153049	0.349806
Dative	Google	0.134982	0.324381
	AI4 Bharath	0.119426	0.324442
	Bing	0.173449	0.396690

Table 1 Grammatical cases BLEU and METEOR (Accusative, Nominative, Genitive, and Dative).

Table 1 compares the performance of three MT systems—Google, AI4Bharath, and Bing—across four grammatical cases: Accusative, Nominative, Genitive, and Dative cases. The evaluation metrics used are BLEU and METEOR, which measure the quality of translations.

In the accusative case, Google and AI 4 Bharath show similar BLEU scores, while Bing performs slightly lower. METEOR scores show a similar trend, with Google achieving 0.269806, AI4Bharath scoring 0.259181, and Bing slightly behind at 0.266399. In the nominative case, Google leads with a BLEU score of 0.087360, followed by Bing and AI4 Bharath. METEOR scores show for AI4Bharath slightly outperforming Google, while Bing scores the lowest. In the Genitive case, Bing performs better than both Google and AI 4 Bharath, with a METEOR score of 0.349806. AI4Bharath follows with 0.129863 (BLEU) and 0.326245 (METEOR), while Google lags behind at 0.108782 (BLEU) and 0.304284 (METEOR). In the dative case, Bing again demonstrates greater performance, achieving the highest BLEU score (0.173449) and METEOR score (0.396690). Google follows with BLEU and METEOR scores of 0.134982 and 0.324381, respectively. AI4Bharath, even though close to Google in METEOR (0.324442), has a lower BLEU score of 0.119426.

Bing consistently outperforms the other systems in the genitive and dative cases, while Google performs relatively better in the accusative and nominative cases. AI4Bharath's performance is competitive but generally trails behind Bing and Google. These results suggest that Bing excels in translating more complex grammatical cases like genitive and dative, while Google performs reasonably well in simpler cases. But the significantly low metrics in each evaluation proves that all three MT models struggle in identifying nouns in its compound forms with their grammatical cases.

The analysis also reveals significant differences in the accuracy and quality of translations by MT models when handling Malayalam compound nouns and case markers. A high frequency of Named Entity Deviation Errors was observed, particularly with proper nouns and their compound forms in human evaluation. This can be seen in all cases of this compounding. The translation of Malayalam case markers posed substantial challenges for both MT systems. The analysis showed that nominative, accusative, dative, and genitive cases often led to significant errors.

4.2. Adaptation to Literary Content

Literary translation demands a narrative quality; when this comes to MT, the quality criteria have been defined without considering these narrative aspects. (Toral & Way, 2015). Literary translation acts as a mediator between the source and target linguistic culture; that is where the need for more culturally nuanced narrative techniques has been used by human translators.

To address this gap of existing MT models' low accuracy in literary translation we aimed to introduce specially tailored literary data for training. To analyse how the general MT system works for literary texts without the domain-specific data and with literary adaptation, initially we are to test the accuracy level of MT without literary training data and later with the literary content training. For that, the MT model had to be adapted to the style and diverse vocabulary of literary translations. The model we used to train is the M2M100 434M parameter sequence to sequence model (Fan et al. 2020), which offers translation between 9900 directions of 100 languages. In our experiments, we selected 400 sentences from the selected Malayalam fiction and translated them from Malayalam-to-English using the three selected NMT systems. Following the approach of Sennrich et al., (2016), we then used the resulting parallel corpus as synthetic data, mixing it with the data that was used to train the system from Malayalam-to-English. As parallel in-domain data, we used a small corpus of sentence-aligned texts, with a total of 400 sentence pairs and 7702 running words on the English side. We trained the system until convergence in terms of BLEU and METEOR scores on held-out tuning data. These were contiguous sentences from the selected Malayalam fiction and its English translation for Malayalam-to-English.

4.3. Experimental Results

This section reviews the automatic scores for the generated literature translations. We compute case-insensitive BLEU and METEOR scores (Papineni et al., 2002; Lavie & Banarjee, 2005). The texts we used are publicly available.

System	BLEU	METEOR
Google	0.095788	0.275906
IndicTrans2	0.097912	0.279609
Bing	0.105603	0.288863
M2M100+ literary adaptation	0.629857	0.729004

Table:2 BLEU & METEOR before and after literary adaptation.

M2M100 + Literary	**Adaptation**	
Grammatical Case	BLEU	METEOR
Accusative	0.554491	0.664554
Nominative	0.644266	0.741079
Genitive	0.725484	0.797075
Dative	0.541264	0.709779

Table 3 Grammatical cases +literary adaptation

We evaluated the quality of Google's online MT, AI4Bharat IndicTrans2 online and Bing Translator on selected Malayalam fiction *Anattam Piriyattam* and its English Translation *Elephantam Misophantham*. Thus, the test set consisted of 400 sentences and 20046 words on the English side and 3845 sentences on the Malayalam side. The experimental results are summarised in Table 2. First, we see that the BLEU scores are much lower than those of state-of-the-art. This supports the assumption about the particular difficulty of literary translation.

Table 2 shows a comparative analysis of the BLEU and METEOR scores for three systems—Google Translate, IndicTrans2, and Bing Translator—evaluated for Malayalam-to-English literary translation. BLEU score indicates the n-gram overlap between machine-generated translations and human references, emphasising precision. Bing achieved the highest BLEU score (0.1056), followed by IndicTrans2 (0.0979) and Google (0.0958). But the low BLEU scores indicate that all systems struggle with achieving high translation quality.

METEOR, which accounts for precision, recall, word alignment, synonyms, and stemming, is particularly suited for assessing literary translations. Bing also outperformed the others in METEOR, achieving the highest score (0.2888), followed by IndicTrans2 (0.2796) and Google (0.2759). These results suggest that Bing's translations offer better semantic alignment and fluency. IndicTrans2's slightly higher METEOR score than Google highlights its focus on handling Indian languages, including Malayalam, which likely enables it to achieve better results.

Bing shows the most effective outputs for Malayalam-to-English literary translation in terms of both BLEU and METEOR scores. IndicTrans2 shows the potential to handle linguistic intricacies of Malayalam better than Google. Despite all these findings, the generally low scores across all systems indicate limitations in current MT models for literary translation. These challenges are likely due to structural differences between Malayalam and English, as well as the difficulty of capturing the rich cultural and stylistic features of literary texts. Thus, while quantitative metrics provide useful benchmarks, qualitative error analysis is essential to fully understand and improve the quality of literary translations.

The same improvements can be seen in noun-case evaluation after literary adaptation. Table 3 evaluates the performance of the M2M100 machine translation model in handling grammatical cases within the context of literary adaptation. The genitive case performs the best, indicating the model's strong ability to translate possessive relationships, which is crucial for maintaining meaning in literary texts. The nominative case also shows good performance, reflecting the model's effectiveness in translating subjects with simpler syntactic relationships. Conversely, the accusative and dative cases show weaker performance, suggesting challenges in handling objects and indirect objects, which are often more complex in literary syntax. The consistently higher METEOR scores across all cases suggest that, while translations may not always align word-for-word with reference texts, they retain meaning and fluency to a reasonable extent. These results highlight the need for additional fine-tuning of the model with more literary content, particularly tailored for handling grammatical cases, to better handle nuanced syntactic relationships critical for preserving the depth and complexity of literary texts. Though compared to the evaluation metrics for the three selected MT models, the M2M100 trained model shows higher level translation. This indicates the need to develop a better data set with literary content that will enhance the overall quality of MT outputs.

MALAYALAM	HUMAN TRANSLATION	GOOGLE	INDICTRANS 2	BING	TRAINED
മിന്നൽകൊമ്പൻ അങ്ങോട്ടോ ഇങ്ങോട്ടോ നീങ്ങുമ്പോൾ താപ്പാനകൾ ചെന്നു തടയും.	If Lightning made a move in any direction, the kumkis headed him off and blocked his way	As the lightning rod moves this way or that, the heaters go and block it.	When the lightning bolt moves here or there, the arrows go and stop it.	When the lightning horn moves here or there, the thappanas will go and stop it.	If Lightning made a move in any direction, the kumkis would move and block them.
മിന്നൽകൊമ്പനെ തളച്ച കാര്യം കുട്ടിയാനകളാണ് മുത്തിയോടു പറഞ്ഞത്.	The news of Lightning Tusker being corralled was given to her by these baby elephants.	It was the lads who told Muthi about the lightning horn.	It was the children who told Muthy that they had struck the lightning bolt	It was the baby elephants who told Muthi that the lightning had stopped the horn.	The little elephants told her how Lightning Tusker had been chained.
പക്ഷേ, ഇക്കുറി പിടിക്കേണ്ടത് മിന്നൽകൊമ്പനെയാണല്ലോ	But this time, he had to tangle with none less than Lightning Tusker	But this time it is the lightning bolt that needs to be caught	But this time, it's the lightning rod that has to be caught.	But this time it's the lightning horn that needs to be caught.	However, this time, he had to catch Lightning Tusker.

Table 4. Results from the trained M2M100 literary adapted model.

The M2M100+ literary adaptation system clearly shows greater level improvement in performance, with significantly higher BLEU and METEOR scores. The results from trained system are given in Table 4. This indicates that the system, likely augmented with domain-specific adaptations or post-editing for literary translation, is far better at capturing the nuances and stylistic elements of Malayalam literature when translated into English. Though Human annotated and conditioned data will help to achieve a better outcome, human translators are quintessential to making the process more culturally curated and keeping up the aesthetic quality demanded by a literary text.

5. Conclusion

. In conclusion, it is worth highlighting the errors in both machine translation systems and human translation to develop better algorithms and adapt to the linguistic nuances of languages, especially low-resource languages. However, as we have seen through a detailed analysis of Malayalam compound nouns and case markers, there can be serious flaws in machine translation. These systems frequently misinterpret compound nouns and fail to handle case markers accurately, leading to Named Entity Deviation Errors and other translation inaccuracies. The overall translation of literary content also poses serious hindrances to MT models.

Machine translation systems struggle with literary texts, as evidenced by the low BLEU and METEOR scores achieved by Google, IndicTrans2, and Bing. However, if the indications are anything to go by, the M2M100 model, enhanced with literary adaptations, asserts the value of specialised training for this genre. It also suggests that training MT systems in existing literary works is crucial for producing higher-quality literary translations. Furthermore, human intervention and domain-specific training, in addition to literary-focused training, can further improve the accuracy and preserve the artistic nuances of the literature.

The use of human intervention in translation provides the most accurate form of translation, particularly with respect to idioms, grammar, and colloquial expressions. This accuracy is highly dependent on the skill and proficiency of the individual translator. Thus, it is evident that a hybrid approach combining the strengths of both machine and human translation methodologies can achieve the ultimate precision and reliability of MT. While machine translation offers efficiency, human translation ensures the nuance and contextual accuracy necessary for high-quality translation of literary texts. Therefore, the collaboration of both machine translation and human translation is essential for the advancement of translation practices that aim to improve reliability, coherence, and cohesion in translated texts.

For low-resource languages, such as Malayalam, tailoring an error-annotated corpus incorporating multiple levels of linguistic errors can enhance MT research and models. This study focuses mainly on compound nouns and noun case-marker errors in MT. For comprehensive MT models, research should focus on all types of grammatical, morphological, and semantic level error analyses. A detailed error taxonomy to address these errors can be incorporated into the tagging process of the training data. As literary language offers diversified linguistic data, current MT can shift focus from domain/register-specific data augmentation to literary data for training machines. This can be used for literary and non-literary mechanised translations. For the MT of literary texts, a specialised annotated corpus should be developed to ensure reliable translation of stylistic traits, cultural intricacies, and contextual implications of literary writing. The syntactic and semantic variety of literary languages offers a fecund training ground to improve current MT models. Moreover, incorporating parallel literary corpora into MT training will help in linguistic preservation and cross-cultural communication. The findings of this study can be applied to other low-resource languages, and future research in the MT of low-resource languages could adopt a hybrid approach that integrates human post-editing to bridge the existing gap for more efficient translation.

References

Asher, R. E. & Kumari, T. C. 1997. *Malayalam*. London: Routledge.

Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65–72).

Bauer, L. (1983). *English word-formation*. Cambridge university press.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: A case study. *arXiv preprint* arXiv:1608.04631.

Besacier, L., & Schwartz, L. (2015, June). Automated translation of a literary work: A pilot study. In *Fourth Workshop on Computational Linguistics for Literature* (co-located with NAACL 2015).

Cadotte, A., André, N., & Sadat, F. (2024, August). Machine translation through cultural texts: Can verses and prose help low-resource indigenous models? In *Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024)* (pp. 121–127).

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In *Translation Quality Assessment: From Principles to Practice* (pp. 9–38).

Chakrawarti, R. K., Bansal, P., & Bansal, J. (2021). Phrase-Based Statistical Machine Translation of Hindi Poetries into English. In *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2020, Volume 2* (pp. 53-65). Springer Singapore.

Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering, 26*(2), 137–161.

Costa, Â., Correia, R., & Coheur, L. (2016, May). Building a corpus of errors and quality in machine translation: Experiments on error impact. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 288–292).

Costa, Â., Ling, W., Luís, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation, 29*(2), 127–161.

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., ... & Joulin, A. (2021). Beyond english-centric multilingual machine translation. *Journal of Machine Learning Research*, *22*(107), 1-48.

Fang, F., Ge, S., & Song, R. (2016). Error analysis of English-Chinese machine translation. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data: 15th China National Conference, CCL 2016, and 4th International Symposium, NLP-NABD 2016, Yantai, China, October 15–16, 2016, Proceedings* (Vol. 15, pp. 35–49). Springer International Publishing.

Farrús Cabeceran, M., Ruiz Costa-Jussà, M., Mariño Acebal, J. B., & Rodríguez Fonollosa, J. A. (2010). Linguistic-based evaluation criteria to identify statistical machine translation errors. In *14th Annual Conference of the European Association for Machine Translation* (pp. 167–173).

Genzel, D., Uszkoreit, J., & Och, F. J. (2010, October). "Poetic" statistical machine translation: rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 158-166).

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2013, December). Crowd-sourcing of human judgments of machine translation fluency. In *Proceedings of the Australasian Language Technology Association Workshop 2013 (ALTA 2013)* (pp. 16–24).

Graham, Y., Baldwin, T., & Mathur, N. (2015a). Accurate evaluation of segment-level machine translation metrics. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1183–1191).

Guerberof-Arenas, A., & Toral, A. (2024). To be or not to be: A translation reception study of a literary text translated into Dutch and Catalan using machine translation. *Target, 36*(2), 215–244.

Gupta, K. K., Sen, S., Haque, R., Ekbal, A., Bhattacharyya, P., & Way, A. (2021). Augmenting training data with syntactic phrasal-segments in low-resource neural machine translation. *Machine Translation, 35*(4), 661–685.

Gutiérrez-Artacho, J., Olvera-Lobo, M. D., & Rivera-Trigueros, I. (2019, March). Hybrid machine translation oriented to cross-language information retrieval: English-Spanish error analysis. In *World Conference on Information Systems and Technologies* (pp. 185–194). Cham: Springer International Publishing.

Jayakrishnan, R., Gopal, G. N., & Santhikrishna, M. S. (2018, January). Multi-class emotion detection and annotation in Malayalam novels. In *2018 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1–5). IEEE.

Kiran, J., RR, R., & Krishnan, A. S. (2024). Transfer Grammar Rules for Malayalam to English Translation. *Language in India*, *24*(11).

Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes* (Vol. 5). Kent State University Press.

Laurian, A. M. (1984, July). Machine translation: What type of post-editing on what type of documents for what type of users. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics* (pp. 236-238).

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. *arXiv preprint arXiv:1808.07048*.

Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., ... & Lim, H. (2023). A survey on evaluation metrics for machine translation. *Mathematics, 11*(4), 1006.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*(8), 707–710.

Matusov, E. (2019, August). The challenges of using neural machine translation for literature. In *Proceedings of the Qualities of Literary Machine Translation* (pp. 10–19).

Mirzaee, A., & Razavi, M. S. M. (2021). Directionality and error typology in English-Persian simultaneous interpreting: A descriptive-analytic corpus-based study. *New Voices in Translation Studies, 25*(1), 54–80.

Nair, R. S. S. (2012). A grammar of Malayalam. *Language in India*, *12*(11).

Newmark, P. (1988). Pragmatic translation and literalism. *TTR: Traduction, Terminologie, Rédaction, 1*(2), 133–145.

Nießen, S., & Ney, H. (2004). Statistical machine translation with scarce resources using morpho-syntactic information. *Computational Linguistics, 30*(2), 181–204.

Omar, A., & Gomaa, Y. (2020). The machine translation of literature: Implications for translation pedagogy. *International Journal of Emerging Technologies in Learning (iJET), 15*(11), 228–235.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318).

Patil, H. S., Ikshu, K. V., Patil, S. S., Janawad, S. M., Purushotham, U., & Prakash, K. J. (2024, May). Translation of different dialects of Kannada to English using machine learning. In *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)* (pp. 1–6). IEEE.

Popović, M., & Arcan, M. (2015, May). Identifying main obstacles for statistical machine translation of morphologically rich South Slavic languages. In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation* (pp. 97–104).

Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (2009). *Linguistics: an introduction*. Cambridge University Press.

Rajendran, S., & SN, M. R. (2019). Taxonomy of Word Formation in Malayalam. *Journal of Advanced Linguistic Studies 8, 1*, *2*.

Ramalingam, A., & Navaneethakrishnan, S. C. (2022, November). Contextual analysis of Tamil proverbs for automatic meaning extraction. In *International Conference on Speech and Language Technologies for Low-resource Languages* (pp. 231–243). Cham: Springer International Publishing.

Ramesh, A., Parthasarathy, V. B., Haque, R., & Way, A. (2020, December). An error-based investigation of statistical and neural machine translation performance on Hindi-to-Tamil and English-to-Tamil. *Association for Computational Linguistics (ACL)*.

Ravindran, P. N. (1975). *Nominal composition in Malayalam* (PhD thesis). Annamalai University, Annamalainagar.

Saxena, S., Chauhan, S., Arora, P., & Daniel, P. (2022). Explicitly unsupervised statistical machine translation analysis on five Indian languages using automatic evaluation metrics. *Sādhanā, 47*(2), 106.

Sellamuthu, K. (2024). Detection of English Grammatical Errors and Correction using Graph Dual Encoder Decoder with Pyramid Attention Network. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *16*(2).

Schäfer, F. (2003). MT post-editing: How to shed light on the "unknown task" – Experiences at SAP. In *EAMT Workshop: Improving MT through other language technology tools: Resources and tools for building MT*.

Sennrich, R. (2016). Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers* (pp. 223-231).

Snover, M., Madnani, N., Dorr, B., & Schwartz, R. (2009a). Fluency, adequacy, or HTER? Exploring different human judgments with a tunable MT metric. In *Proceedings of the Fourth Workshop on Statistical Machine Translation* (pp. 259–268).

Sreelekha, S. (2020). Machine translation between Malayalam and English. *Linguistics Journal, 14*(2), 7–31.

Tahseen, W. M. (2024). Investigating machine translation errors in rendering English literary texts into Arabic. *Integrated Journal for Research in Arts and Humanities, 4*(1), 68–81.

Tripathi, V., & Rathod, D. (2024). Semantic Model for Fragment of Hindi (Part 1). *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *16*(1).

Tripathi, V., & Rathod, D. (2024). Semantic Model for Fragment of Hindi (Part 2). *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *16*(2).

Thomas, V. (2021). *Anattam Piriyatham*. DC Books.

Thomas, V. (2023a). *Elephantam Misophantham* (N. K., Trans.). Eka

Tillmann, C., Vogel, S., Ney, H., & Zubiaga, A. (1997, July). A DP-based search using monotone alignments in statistical translation. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 289–296).

Toral, A., & Way, A. (2014). Is machine translation ready for literature. In *Proceedings of Translating and the Computer 36*.

Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? In *Translation quality assessment: From principles to practice* (pp. 263–287).

Vázquez, A. I. C., & Mitkov, R. (2023, September). Machine translation of literary texts: Genres, times, and systems. In *Proceedings of the First Workshop on NLP Tools and Resources for Translation and Interpreting Applications* (pp. 48–53).

Vilar, D., Leusch, G., Ney, H., & Banchs, R. E. (2007, June). Human evaluation of machine translation through binary system comparisons. In *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 96–103).

Voigt, R., & Jurafsky, D. (2012, June). Towards a literary machine translation: The role of referential cohesion. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature* (pp. 18–25).

Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C. M., & Eger, S. (2019). MoverScore: Text generation evaluating with contextualized embeddings and earth mover distance. *arXiv preprint arXiv:1909.02622*.

Zhao, A., Huang, K., Yu, H., & Huang, D. (2023, December). DUTNLP system for the WMT2023 Discourse-Level Literary Translation. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 296–301).

Hanna CH is research scholar at the National Institute of Technology Calicut who works on corpus-based translation studies. Her areas of interest are translation studies, corpus linguistics, and machine translation studies.

Sabeerali KP is an independent researcher with a background in computer science and has previously worked as a research assistant at ICFOSS Trivandrum. He is passionate about machine learning and deep learning research, with diverse professional experience as a web and app developer.

Dr. Vrinda Varma is Assistant Professor of English at National Institute of Technology, Calicut, India. Her research areas include gender studies, memory studies, and food studies. She was a Fulbright-Nehru doctoral fellow at Boston University in 2015. She is also a practicing translator and is the recipient of PEN/Heim Translation Grant 2024.