AssamLegalTrans: A Parallel Corpus, Benchmark and Analysis for English-Assamese Machine Translation of Legal Judgments

Telem Joyson Singh[♢], Hemanta Baruah[♢], Sanasam Ranbir Singh[♢], Anindita Talukdar[♢], Nasrin Shahnaz[♢♣], Okram Jimmy Singh[♢], Priyankoo Sarmah[♢], Pallav Kumar Dutta[♢], Sukumar Nandi[♢], Pranab Duara[♠]

[♢]Indian Institute of Technology Guwahati, Assam, India
[♣]Cotton University, Assam, India
[♠]Gauhati High Court, Assam, India
{tjoyson, hemanta.b, ranbir}@iitg.ac.in
{aninditatalukdar,nasrin}@iisi.iitg.ac.in, okjimmy@rnd.iitg.ac.in,
{priyankoo, pkdutta, sukumar}@iitg.ac.in, pranab.duara@gov.in

Abstract

In India, the official language for writing judgments in higher courts is English, which creates a language barrier for citizens not proficient in English. Machine Translation (MT) provides a scalable solution, but its progress for low-resource languages like Assamese is significantly limited due to the lack of legal domain data. To address this gap, we introduce the first-of-its-kind English-Assamese parallel corpus for the translation of Indian court judgments. This dataset consists of over 55,000 manually translated and validated sentence pairs from over 500 judgments of the Gauhati High Court and the Supreme Court of India. Using this dataset, we perform a comprehensive evaluation of state-of-the-art multilingual models, including NLLB-200 and Sarvam-Translate, in both zero-shot and fine-tuned settings, comparing their performance against commercial systems. Our experiments show that fine-tuning on our legal-domain dataset significantly improves the translation quality. We also conduct a thorough error analysis that points out important issues in legal translation. These include precisely translating legal terms, properly transliterating named entities, expanding abbreviations, and transforming sentence structures, such as changing passive voice to active voice, when translating from English to Assamese. By creating a publicly available dataset and examining the specific challenges, this work offers a reproducible foundation and a clear way to develop more accurate and reliable legal machine translation systems. This will help improve access to justice for Assamese speakers.

Keywords: parallel corpus, legal translation, Assamese

1. Introduction

In India, the official language for writing judgments in the Supreme Court and High Court is English (Agrahari, 2015). Although this ensures uniformity and judicial consistency, it creates language barriers for citizens who are not proficient in English, limiting access to court judgments. To address this gap, legal machine translation (MT) systems can be adopted as an alternative option for a scalable and effective solution. This work attempts to understaend the challenges in building legal MT systems, considering English-to-Assamese judgment translation as a use case.

Translating Indian legal judgments presents unique challenges that set it apart from general-domain translation. Legal texts are not uniform; they contain distinct sections: a **preamble**, which is dense with named entities, acronyms, and abbreviations; and a **main narrative body**, which contains extensive legal terms and complex long sentences. These documents are characterized by specialized legal terminology that requires uncompromising precision, as terms often have non-negotiable official equivalents in the target language that cannot be substituted with mere synonyms. Furthermore, legal text is dense with named entities—such as the names of people, places, and legal counsel—that must be correctly distinguished for either translation or transliteration. A similar dual challenge exists for abbreviations and acronyms, where some must be transliterated to preserve their identity while others must be expanded and then translated (e.g., 'Art.' becomes 'Article' before translation).

These challenges are further compounded for low-resource languages like Assamese, where progress has been critically constrained by the absence of legal domain data. To date, no publicly available domain-specific parallel corpus for English-Assamese legal translation has been created, leaving researchers without the resources to train or reliably evaluate modern MT systems for this task.

To address this resource gap, this paper introduces a first-of-its-kind English–Assamese parallel corpus specifically focused on the legal domain. This new manually translated and validated paral-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 4921–4930
11–16 May 2026.

lel corpus contains over 55,000 sentence pairs curated approximately from 500 court judgments of the Gauhati High Court and the Supreme Court of India, and covers a range of proceedings, including civil and criminal cases. It provides a foundational resource for developing and systematically evaluating MT models for the Assamese legal domain, enabling standardized and reproducible research in this important area.

Building on this resource, we perform a comprehensive benchmark of existing multilingual models, including NLLB-200 (Costa-Jussa et al., 2022) and Sarvam-Translate (Sarvam AI, 2025), both in zero-shot and fine-tuned settings, and compare their performance against commercial systems like Google Translate. Our experiments demonstrate that fine-tuning leads to substantial improvements in translation quality, with a fine-tuned NLLB model achieving a BLEU score of 36.29, significantly outperforming its zero-shot counterpart, in our benchmark dataset.

In addition to automatic translation quality evaluation, we provide in-depth error analysis focusing on critical aspects of legal translation, such as terminology, named entities, and abbreviations. We also analyze the models' ability to handle complex syntactic structures, including the transformation of passive-voice sentences common in English legal text to the active voice preferred in Assamese.

The contributions of this paper can be summarized as follows:

- We introduce a first-of-its-kind English–Assamese parallel corpus for the legal domain, containing authentic judgments from Indian High Courts and the Supreme Court.
- We present a benchmark evaluation of state-of-the-art multilingual large language models on the legal domain translation task for the Assamese language.
- We perform a detailed analysis that identifies key challenges in legal translation, including the handling of legal terminology, named entities, abbreviations and acronyms, and complex syntactic structures like passive-to-active voice transformations.

2. Related Work

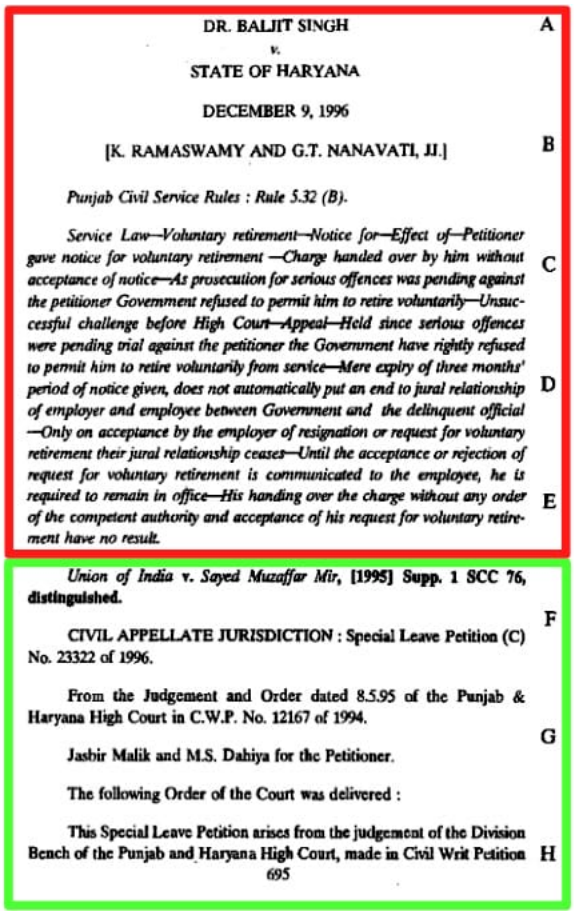
The field of machine translation has made major progress due to neural machine translation (NMT) and large language models (Kocmi et al., 2023, 2024). However, research on machine translation in the Indian legal domain remains in a nascent stage (Mahapatra et al., 2025). This section reviews the relevant literature in the fields of legal translation.

One notable study in Indian legal translation (Mahapatra et al., 2025) constructs a small benchmark corpus and compares the performance of various existing Machine Translation models. Another study (Briva-Iglesias et al., 2024) evaluates the performance of large language models (LLMs) in translating English legal contracts to Portuguese, Turkish, Spanish, and Catalan. It shows that while Google Translate usually performs well on automatic evaluation metrics, human evaluators rated LLMs, particularly GPT-4 (OpenAI Team, 2024), as similarly effective in delivering fluent and contextually accurate translations for legal texts. Another study (Poudel et al., 2024) examines the development of a bidirectional English-Nepali machine translation system for legal documents. It uses a specialized corpus to achieve good translation quality metrics. Additionally, the MT@BZ corpus project (De Camillis et al., 2023) focuses on creating a bilingual corpus of machine-translated legal texts. It provides insights into the types of errors and the quality of translations in less common language pairs. This is important for smaller language communities. A dataset from the EU APE-QUEST project (Ive et al., 2020) shows how important post-editing is for assessing the quality of NMT in the legal field. It reveals that, although neural systems can create high-quality translations, they often do not meet human standards.

Hung et al. (2013) tackled the issue of legal translation by suggesting a way to break down long English legal sentences based on their logical structure. This approach uses the resulting linguistic features to choose translation rules and improve a tree-based statistical machine translation system. Furthermore, customized NMT systems designed for the Swiss legal field show the potential for greater efficiency and a better user experience in legal translation processes (Martinez-Dominguez et al., 2020).

3. Legal Judgments and Translation

Indian court judgments are divided into two main parts: a Preamble and a narrative Main Body (Kalamkar et al., 2022). This is shown in Figure 1, where the preamble is outlined in a red box and the Main Body is outlined in a green box. Each part has its own linguistic features that present unique challenges for machine translation (MT). This is especially true when translating from English, which is an SVO language, to Assamese, which is a morphologically rich SOV language (Singh et al., 2025). This section describes these features and how they affect MT systems.



DR. BALJIT SINGH
v.
STATE OF HARYANA

DECEMBER 9, 1996

[K. RAMASWAMY AND G.T. NANAVATI, JJ.]

Punjab Civil Service Rules : Rule 5.32 (B).

Service Law—Voluntary retirement—Notice for—Effect of—Petitioner gave notice for voluntary retirement —Charge handed over by him without acceptance of notice—As prosecution for serious offences was pending against the petitioner Government refused to permit him to retire voluntarily—Unsuccessful challenge before High Court—Appeal—Held since serious offences were pending trial against the petitioner the Government have rightly refused to permit him to retire voluntarily from service—Mere expiry of three months' period of notice given, does not automatically put an end to jural relationship of employer and employee between Government and the delinquent official —Only on acceptance by the employer of resignation or request for voluntary retirement their jural relationship ceases—Until the acceptance or rejection of request for voluntary retirement is communicated to the employee, he is required to remain in office—His handing over the charge without any order of the competent authority and acceptance of his request for voluntary retirement have no result.

Union of India v. *Sayed Muzaffar Mir*, [1995] Supp. 1 SCC 76, distinguished.

CIVIL APPELLATE JURISDICTION : Special Leave Petition (C) No. 23322 of 1996.

From the Judgement and Order dated 8.5.95 of the Punjab & Haryana High Court in C.W.P. No. 12167 of 1994.

Jasbir Malik and M.S. Dahiya for the Petitioner.

The following Order of the Court was delivered :

This Special Leave Petition arises from the judgement of the Division Bench of the Punjab and Haryana High Court, made in Civil Writ Petition

Figure 1: Illustration of a judgment document showing a dense cluster of named entities in the preamble(red) and the main body(green) text enclosed within two separate bounding boxes in colour.

3.1. Preamble

The preamble of a judgment contains the case's metadata, which consists of entities instead of narrative text. This includes case numbers, court names, the parties involved (such as petitioners and respondents), legal counsel, and judges.

The preamble's primary challenge is the need for different ways of handling named entities, abbreviations, and acronyms:

- **Translation**: Legal designations such as "High Court" need to be translated to their official Assamese equivalent, "উচ্চ ন্যায়ালয়".
- **Transliteration:** Proper nouns, like names of people and places, need to be transliterated to maintain their phonetic identity. For example,"Bhagat Singh" becomes "ভগত সিং".
- **Expansion and Translation**: Common legal abbreviations must be correctly identified, expanded, and then translated. For instance, "Art." must be recognized as "Article" and translated to "অনুচ্ছেদ".
- **Judgment Summary Section Issues**: The title of each court judgment is usually followed by a lengthy summary section. This section describes the kind of judgment, the case's circumstances, any applicable laws or acts, and a synopsis of earlier court cases that preceded the current one. These summaries often consist of several phrases or sentences with em dashes ("—") between them. However, translation and segmentation are challenging since these phrases often deviate from the conventional grammatical sentence structures of English writing. Furthermore, these summaries might occasionally fill a page, with em dashes separating each segment, making it difficult to tell where a phrase ends.

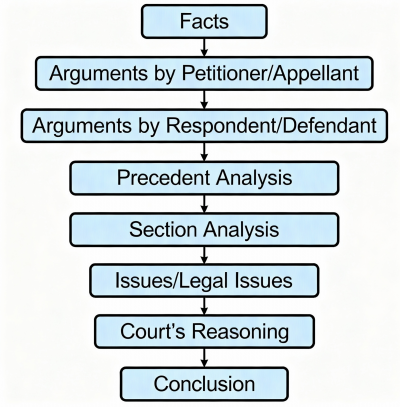


Figure 2: A typical structure of an indian court judgment.

An MT system must be able to distinguish between words to be translated and entities to be transliterated, a non-trivial task that standard models often fail at.

3.2. Main Body

The main body of the judgment presents the court's narrative and decision as shown in Figure 2. It is composed of sections with distinct rhetorical roles (Nigam et al., 2025), each presenting unique linguistic challenges that an effective MT system must navigate.

- **Analysis, Issues and Reasoning of the Court:** These are the core of the judgment, where the court examines evidence, interprets statutes, and applies legal principles, containing the highest density of specialized legal terminology. These terms often have fixed, official equivalents in the target

language that are non-negotiable (e.g., “assessee” must be translated to its official counterpart, “নির্ধাৰিত ব্যক্তি”, not a literal synonym). Our analysis in Section 7 evaluates a model’s ability to adhere to these official glossaries.

- **Facts and Arguments:** These sections contain long and complex English sentences, often with many clauses. To translate them accurately into Assamese, we need to reorder their structure to match the SOV syntax. We assess this challenge quantitatively in Section 8.1.
- **Conclusion (Decision):** This section contains the court’s final decision which requires uncompromising precision in translation. Phrases such as “The appeal is dismissed” must be translated into a precise, standard legal equivalent in Assamese (“আপীলখন খাৰিজ কৰা হ’ল”), as any ambiguity or deviation could fundamentally alter the legal outcome.

4. Dataset

In this section, we describe the creation of a first-of-its-kind parallel corpus in the legal domain for the Assamese language. Compiled from the original English judgments of the Gauhati High Court and the Supreme Court of India, it covers civil, criminal, and other types of cases. It provides a foundational resource for developing and conducting standardized, rigorous evaluations of legal translation for the Assamese language.

4.1. Dataset Construction

Document Translation: The source English judgments are sourced from the Gauhati High Court. The source English documents are translated into Assamese by a team of professional translators, all of whom are native Assamese speakers holding degrees in law, English, or Assamese. We perform document-level translation to overcome the limitations of sentence-level processing, including ambiguity and terminological inconsistency. This approach uses the broader document context to produce accurate, coherent, and fluent translations. Each translated document is then subjected to a peer review by another translator.

Sentence Alignment: Following the translation and review, we performed sentence-level alignment between the source English documents and their Assamese counterparts. We first segmented the text of both languages into sentences using Spacy for English and IndicNLP for Assamese. Subsequently, we used an automatic alignment tool that leverages multilingual sentence embeddings to generate candidate alignments based on cosine similarity scores. Due to the syntactic complexity and length of legal sentences, the sentence segmentation and automatically generated alignments are manually reviewed by annotators to correct errors, such as one-to-many or many-to-one mappings, and to ensure that each sentence pair was semantically equivalent.

Alignment Validation: To ensure the quality of the parallel corpus, a final validation of the sentence alignments was conducted. Two native Assamese speakers, one with a background in English and the other in law, independently reviewed each pair to confirm that the sentences are accurate translations of one another. The inter-annotator agreement, measured using Cohen’ kappa, was 90. In cases of disagreement, a third translator was consulted, and conflicts are resolved through majority voting.

4.2. Dataset Overview

Table 1 summarizes the statistics of the English-Assamese parallel corpus, detailing the distribution of the dataset across training, development, and testing splits. A key observation is that Assamese consistently exhibits a higher Type-Token Ratio (TTR) than English across all data subsets. For example, the training set shows an Assamese TTR of 0.04 compared to English’s 0.02, despite English having a larger overall token count.

This pattern suggests that Assamese is morphologically richer than English. The higher TTR indicates a greater diversity of word forms, which is characteristic of languages with extensive inflectional or agglutinative systems. In contrast, the lower TTR for English reflects its more analytic nature, where grammatical relationships are often conveyed through word order and function words rather than complex morphology. Consequently, while English uses more tokens to express equivalent sentences, Assamese employs a wider vocabulary of unique types.

To ensure the integrity of the splits and minimize data leakage risks in model evaluation, we computed vocabulary overlap statistics between the training set and the development/test sets. Vocabulary overlap is calculated as the percentage of unique types (vocabulary size) in the dev or test set that appear in the training set. For English, the test set shows 94.8% overlap. For Assamese, the overlaps are slightly lower due to its morphological richness: 86.1% for the test.

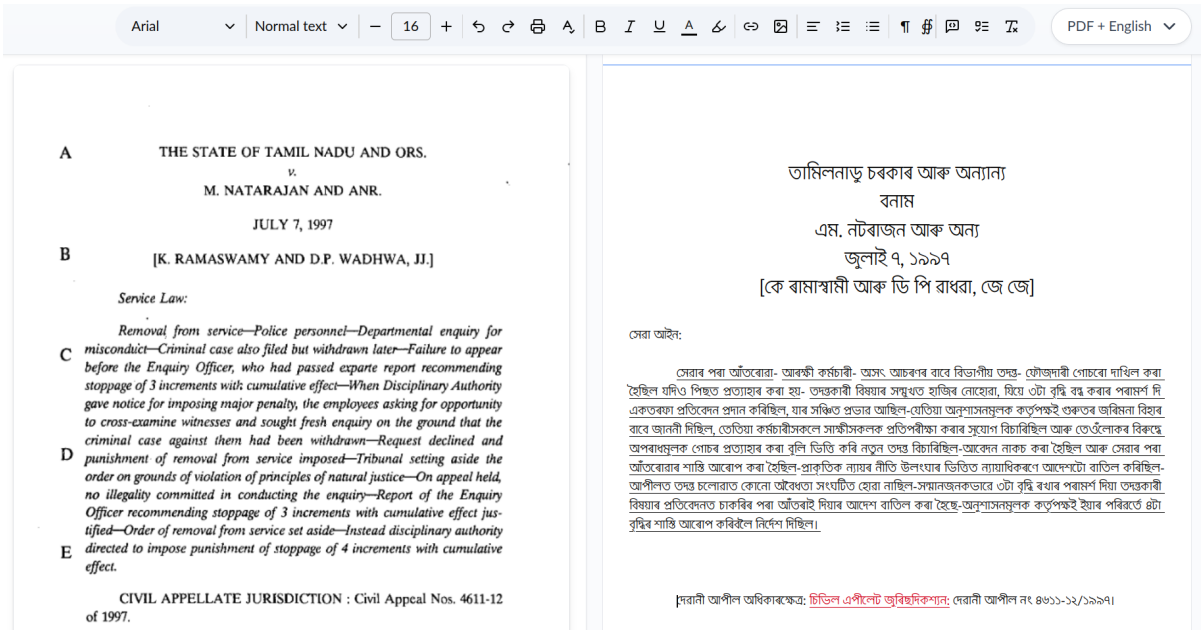


Figure 3: A snapshot of our legal translation system

Dataset	Sentences	Tokens		Types		TTR	
		English	Assamese	English	Assamese	English	Assamese
Train	55626	1480584	1251798	25787	47359	0.02	0.04
Dev	1000	26498	22923	3865	5382	0.15	0.23
Test	1090	36350	31481	3640	5555	0.10	0.18

Table 1: Statistics of the AssamLegalTrans corpus, detailing the number of sentences, tokens, types, and Type-Token Ratio (TTR) for English and Assamese across the train, dev, and test splits.

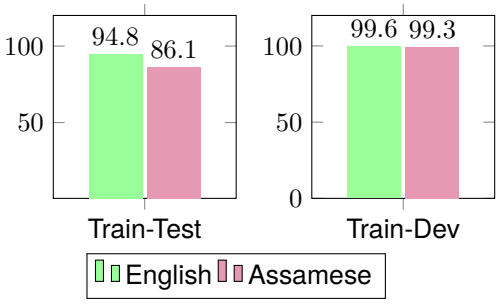


Figure 4: Vocabulary overlap statistics between the training set and the development/test sets.

5. Experimental Setup

After constructing the legal parallel corpus, we evaluate the translation performance of various open-source multilingual MT systems and LLMs. We compare both zero-shot performance and the performance of models fine-tuned on our newly created legal domain corpus.

5.1. Models

We identified models trained on Assamese data. We selected a range of models to establish a comprehensive benchmark, including open-source models known for multilingual capabilities and widely-used commercial systems.

NLLB-200: NLLB (Costa-Jussa et al., 2022) is a Meta AI model covering 200 languages, including Assamese, trained on large mined parallel corpora. It serves as a strong baseline for low-resource translation and broad-coverage evaluation.

Sarvam-Translate: Sarvam-Translate (Sarvam AI, 2025) is an open-source Sarvam AI model focused on Indic languages, including Assamese, built on Gemma3-4B-IT (Gemma Team, 2025). Its specialization for Indian languages motivates evaluation on our legal-domain dataset.

Commercial Baselines: For comparison against state-of-the-art proprietary systems, we include three black-box models: **Google Translate**, **Microsoft Translator**, and **OpenAI's**

ChatGPT (GPT-4). Translations from these systems are obtained via their public APIs in October 2025. These models serve as powerful commercial baselines, but are not fine-tuned in our experiments.

5.2. Fine-tuning and Implementation Details

To specialize open-source models in the legal domain, we fine-tune both NLLB-200 and Sarvam-Translate on the training split of our AssamLegal-Trans corpus. Due to computational constraints, we choose the 1.3 billion parameter variant of NLLB-200. In addition, we perform parameter-efficient fine-tuning using Low-Rank Adaptation (LoRA) (Hu et al., 2022) with a rank of 8.

The fine-tuning is conducted using the Hugging Face transformers library (Wolf et al., 2020) on a single NVIDIA A100 GPU. For both models, we use the AdamW optimizer with a learning rate of 5×10^{-5} and a linear learning rate scheduler with a warm-up period. We train for 5 epochs with a batch size of 4. All other hyperparameters are kept at their default values as provided by the library.

5.3. Metrics

We evaluate translation quality using a set of standard automatic metrics to assess different aspects of performance: lexical accuracy, morphological correctness, and post-editing effort.

- **BLEU**: We report the case-sensitive sacreBLEU, a standard implementation that ensures reproducible scores by managing tokenization and providing a standardized reference (Papineni et al., 2002). It measures n-gram precision between the machine translation and the reference text.
- **CHRF++**: A character-level metric that computes n-gram F-scores, which makes it suitable for morphologically rich languages such as Assamese and more robust to tokenization differences (Popovic, 2017).
- **TER (Translation Edit Rate)**: An error metric that calculates the minimum number of edits (insertions, deletions, substitutions, and shifts) required to change the machine output to match the reference translation, normalized by the reference length (Snover et al., 2006). It is often used to estimate the post-editing effort.

6. Result

Table 2 shows the evaluation of multilingual machine translation (MT) systems and large language models (LLMs) for the English-to-Assamese legal translation task. The performance is measured using BLEU, CHRF++, and TER metrics.

MODELS	BLEU	CHRF++	TER
NLLB-200	24.71	49.13	64.88
Sarvam	25.73	52.29	62.91
Google	27.80	54.50	60.00
Microsoft	29.20	56.20	58.20
ChatGPT	13.02	36.50	78.35
Sarvam-FT	32.49	57.17	60.24
NLLB FT	**36.29**	**59.34**	**52.30**

Table 2: Performance of various MT models on the English-to-Assamese legal translation test set. "-FT" denotes models fine-tuned on the AssamLegalTrans corpus.

The main results demonstrate that fine-tuning pre-trained models with the legal domain-specific corpus yields significant improvements in translation quality. The fine-tuned NLLB model (NLLB-FT) achieved the highest scores in all metrics, with a BLEU score of 36.29, a CHRF++ score of 59.34 and a TER of 52.30. This indicates a substantial improvement over the baseline NLLB and Sarvam models. The relatively lower gain with Sarvam-FT in comparison to NLLB-FT can be attributed to the use of parameter-efficient Low-Rank Adaptation finetuning rather than full finetuning.

7. Error Analysis

Legal translation requires uncompromising precision, as errors in domain-specific elements can alter legal meanings or compromise judicial validity. This section presents a detailed error analysis of the finetuned NLLB model's machine translation output, focusing on three critical areas: **legal terminology**, **named entities**, and **abbreviations/acronyms**.

To ensure the reliability of our analysis, we established a rigorous annotation protocol. Two native Assamese speakers, one of whom has a legal background, independently reviewed the machine-translated output to identify and categorize errors. We measured the inter-annotator agreement for error identification using Cohen's kappa, which yielded a value of 0.85, indicating substantial agreement. Disagreements are resolved through discussion and adjudicated by a third expert to reach a final consensus on each error.

7.1. Legal Terminology Errors

The accurate translation of legal terminology presents a significant challenge, as each legal

Error Types	Percentage
Literal Translation	46.4
Semantic Error	53.6

Table 3: Distribution of Error types in Legal Term Translation

Error Types	Percentage
Transliteration Error	36.3
Translation Error	63.6

Table 4: Percentage of Error types in Named Entities Translation

term must be translated into its single official equivalent, as prescribed by government glossaries, rather than merely a synonym or paraphrase. For this analysis, we evaluated 250 parallel sentences from the test set that contained legal terms.

We identify two primary error categories:

- **Literal Translation Errors**: The model generated a literal or paraphrased translation instead of the mandatory official equivalent.
- **Semantic Errors**: The translation conveyed an incorrect or unintended meaning, deviating from the legal context.

For instance, Example (1) illustrates a literal translation error, where the model fails to use the official Assamese term for "**assessee**." Example (2) shows a semantic error, where "**reference Court**" is incorrectly translated, altering the sentence's legal meaning.

(1) Eng: The judgment and order of a Division Bench of the High Court at Karnataka in writ appeals is challenged, with special leave, by the assessee.
Asm: কর্ণাটক উচ্চ ন্যায়ালয়ৰ বিভাগীয় বিচাৰপীঠৰ লেখ আপীলৰ ৰায় আৰু আদেশ মূল্যায়নকাৰীয়ে বিশেষ ছুটীৰ সৈতে প্রত্যাহ্বান জনায়।

(2) Eng: However, the appellants did not challenge the award of the reference Court which had become final as against them.
Asm: কিন্তু আপীলকাৰীসকলে তেওঁলোকৰ বিৰুদ্ধে চূড়ান্ত হৈ পৰা কোষ আদালতৰ ৰায়ক প্রত্যাহ্বান জনোৱা নাছিল।

Our analysis of the 250 sampled sentences reveals 28 errors in translating legal terms. As shown in Table 3, semantic errors are slightly more prevalent than literal translation errors.

7.2. Named Entity Errors

Named entities (e.g., parties, date, location) require transliteration over translation to retain identity. To evaluate this, we selected 215 parallel sentences containing named entities from the test set. The key challenge is that each named entity must be transliterated correctly, rather than translated.

We identify two key error types:

- **Transliteration Error**: The model generated a phonetically incorrect or inconsistent transliteration.
- **Translation Error**: The model incorrectly translated a named entity that should have been transliterated.

Example (3) shows a transliteration error in the company name, while example (4) illustrates a translation error where "**Consulting Engineers**" is translated into Assamese instead of being transliterated.

(3) Eng: MANGANESE ORE (INDIA) LTD. AND ANR.
Asm: মেঙ্গানিজ আকৰ (ভাৰত) লিমিটেড আৰু অন্যান্য

(4) Eng: Tata Consulting Engineers &Anr. v. Union of India & Ors., (1994) 206 ITR 237, disapproved.
Asm: টাটা কনছালটিং অভিযন্তা আৰু অন্যান্য বনাম ভাৰতীয় সংঘ আৰু অন্যান্য, (১৯৯৪) ২০৬ আইটিআৰ ২৩৭,অস্বীকাৰ কৰা হৈছে।

Our analysis identifies 11 named entity errors in the 215-sentence sample. As detailed in Table 4, translation errors are the more frequent issue, indicating the model's difficulty in distinguishing entities that require transliteration.

7.3. Abbreviation/Acronym Errors

Handling abbreviations and acronyms poses a dual challenge: some must be transliterated, while others need to be expanded and then translated. We analyzed 220 parallel sentences containing abbreviations/acronyms to assess the model's capability.

We identify two categories of error:

- **Transliteration Error**: The model produces an incorrect transliteration of an abbreviation/acronym.
- **Translation Error**: The model failed to correctly expand and translate an abbreviation, resulting in a mistranslation.

Error Types	Percentage
Transliteration Error	70.0
Translation Error	30.0

Table 5: Distribution of error types in abbreviation and acronym translation.

Example (5) contains acronyms (**MISA**, **COFEPOSA**) that should be transliterated. Example (6) includes abbreviations ("**C M**," "**C R**") that require expansion and translation, where the model produced an incorrect rendering.

(5) Eng: On the same day, the order of detention under MISA was revoked and an order of detention was passed under COFEPOSA against Solanki.
Asm: সেইদিনা মিছাৰ অধীনত আটক কৰাৰ আদেশ বাতিল কৰি সোলাংকিৰ বিৰুদ্ধে কোফেপোছাৰ অধীনত আটক কৰাৰ নির্দেশ দিয়া হয়।

(6) Eng: From the Judgment and Order dated 8.1.96 of the Delhi High in C M Number 3825/95 in C R Number 723/1995.
Asm: দিল্লী উচ্চ ন্যায়ালয়ৰ ৮.১.৯৬ তাৰিখৰ ৰায় আৰু আদেশৰ দ্বাৰা দেৱানী গোচৰ নং ৩৮২৫/৯৫, দেৱানী গোচৰ নং ৭২৩/১৯৯৫ত।

In our 220-sentence sample, we found 30 errors related to abbreviations and acronyms. As shown in Table 5, transliteration errors are the dominant issue, accounting for 70% of the mistakes.

8. Syntactic Analysis in English-Assamese Legal MT

Translating English legal judgments to Assamese might present challenges due to significant syntactic differences. English judgments features long, complex sentences and frequent use of the passive voice, while Assamese, as a Subject-Object-Verb (SOV) language, prefers simpler sentences and active voice. This section evaluates the fine-tuned NLLB model's ability to address these syntactic divergences.

8.1. Complex Sentence vs Simple Sentence

English legal texts are known for their long and complex sentences that include many nested clauses. This complexity makes translating legal judgments difficult, especially into languages with different sentence structures. For example, Assamese follows a Subject-Object-Verb (SOV) order, while English uses a Subject-Verb-Object (SVO) structure.

Sent Type	BLEU	CHRF++	TER
Simple	36.50	59.09	52.52
Complex	36.04	59.64	52.03

Table 6: A comparison of translation performance on simple versus complex sentences.

Sent Type	BLEU	CHRF++	TER
Active	36.28	59.15	52.57
Passive	36.30	59.48	52.10

Table 7: A comparison of translation performance on active versus passive voice sentences.

To quantify this challenge, we evaluate our fine-tuned NLLB model on both simple and complex sentences, with the results presented in Table 6. In our analysis, a sentence is considered complex if it has more than 3 clauses, comprising 690 simple and 400 complex sentences.

Contrary to what we expected, the model's performance stays quite stable when translating complex sentences. We see only a small drop in the BLEU score, which decreased by 0.46 points.

8.2. Passive Voice vs Active Voice

Another well-known characteristic of English legal texts is the frequent use of the passive voice, for example, "an enquiry was started.". This is different from typical Assamese writing, which often prefers the active voice. An ideal translation would restructure a passive English sentence into a syntactically natural Assamese equivalent, which may or may not be passive.

Example (7) illustrates this challenge. The English source sentence uses the passive voice twice: "the order... **was revoked**" and "an order... **was passed**.". The model's translation renders these phrases into a standard and natural-sounding active form in Assamese (বাতিল কৰা হয় and জাৰি কৰা হয়).

(7) Eng: On the same day, the order of detention under MISA was revoked and an order of detention was passed under COFEPOSA against Solanki.
Asm: সেইদিনাই এম.আই. এছ. এ. ৰ অধীনত আটক কৰাৰ আদেশ বাতিল কৰা হয় আৰু সোলেনকিৰ বিৰুদ্ধে চি. অ'. এফ. ই. পি. অ'. এছ. এ. ৰ অধীনত আটক কৰাৰ আদেশ জাৰি কৰা হয়।

We tested our model's ability to handle this syntactic transformation with 440 active and 624 passive sentences. Table 7 shows the performance on sentences containing active versus passive voice constructions.

The evaluation metrics demonstrate that the fine-tuned model is exceptionally proficient in handling this linguistic divergence. The BLEU, CHRF++, and TER scores for passive and active sentences are nearly identical. The negligible difference (a 0.02 point increase in BLEU for passive sentences) indicates that the model does not struggle with passive constructions.

9. Conclusion

This paper introduces the first English-Assamese legal parallel corpus and a reproducible benchmark for legal MT, demonstrating that domain-specific fine-tuning substantially improves quality over zero-shot baselines. Beyond reporting gains, the analyzes highlight concrete points: terminology fidelity, named-entity handling, abbreviation expansion, and SVO to SOV restructuring with passive-to-active shifts, thereby clarifying what drives errors and where targeted modeling can pay off in legal translation workflows. In the future, incorporating document-level context and integrating glossary and named entity recognition modules (with principled transliteration and abbreviation resolution) are promising directions to further close the gap to practitioner-ready systems that improve access to judgments for Assamese speakers.

10. Acknowledgements

This work is fully supported by the Gauhati High Court, Guwahati, under the project entitled AI Assisted Legal Translation of Judgments from English to Assamese. The eCourts Division and Translation Wing of the Gauhati High Court, Guwahati, has helped in subjective evaluation of the translation system and providing end users' feedback.

11. Bibliographical References

Gunjan Agrahari. 2015. Languages in administration of justice. Journal of Legal Studies, 3:161.

Sarvam AI. 2025. Sarvam - translate. Sarvam Translate Blog. Accessed: June 29, 2025.

Vicent Briva-Iglesias, Joao Lucas Cavalheiro Camargo, and Gokhan Dogru. 2024. Large language models •ad referendum•: How good are they at machine translation in the legal domain?

Marta R Costa-Jussa, James Cross, Onur Celebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. arXiv preprint arXiv:2207.04672.

Flavia De Camillis, Egon W. Stemle, Elena Chiocchetti, and Francesco Fernicola. 2023. The MT@BZ corpus: machine translation & legal language. In Proceedings of the 24th Annual Conference of the European Association for Machine Translation, pages 171--180, Tampere, Finland. European Association for Machine Translation.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In International Conference on Learning Representations.

Bui Thanh Hung, Nguyen Le Minh, and Akira Shimazu. 2013. Translating legal sentence by segmentation and rule selection. International Journal on Natural Language Computing, 2(4):35--54.

Julia Ive, Lucia Specia, Sara Szoc, Tom Vanallemeersch, Joachim Van den Bogaert, Eduardo Farah, Christine Maroti, Artur Ventura, and Maxim Khalilov. 2020. A post-editing dataset in the legal domain: Do we underestimate neural machine translation quality? In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 3692--3697, Marseille, France. European Language Resources Association.

Prathamesh Kalamkar, Aman Tiwari, Astha Agarwal, Saurabh Karn, Smita Gupta, Vivek Raghavan, and Ashutosh Modi. 2022. Corpus for automatic structuring of legal documents. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 4420--4429, Marseille, France. European Language Resources Association.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondrej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Benjamin Marie, Christof Monz, Kenton Murray, Masaaki Nagata, Martin Popel, Maja Popovic, Mariya Shmatova, Steinthor Steingrimsson, and Vilem Zouhar. 2024. Findings of the WMT24 general machine translation shared task: The LLM era is here but MT is not solved yet. In Proceedings of the Ninth Conference on Machine Translation, pages 1--46, Miami, Florida, USA. Association for Computational Linguistics.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondrej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme

Gowda, Roman Grundkiewicz, Barry Haddow, Philipp Koehn, Benjamin Marie, Christof Monz, Makoto Morishita, Kenton Murray, Masaaki Nagata, Toshiaki Nakazawa, Martin Popel, Maja Popovic, Mariya Shmatova, and Jun Suzuki. 2023. Findings of the 2023 conference on machine translation (WMT23): LLMs are here but not quite there yet. In Proceedings of the Eighth Conference on Machine Translation, pages 1--42, Singapore. Association for Computational Linguistics.

Sayan Mahapatra, Debtanu Datta, Shubham Soni, Adrijit Goswami, and Saptarshi Ghosh. 2025. Milpac: A novel benchmark for evaluating translation of legal text to indian languages. ACM Trans. Asian Low-Resour. Lang. Inf. Process., 24(8).

Ruben Martinez-Dominguez, Matiss Rikters, Arturs Vasilevskis, Marcis Pinnis, and Paula Reichenberg. 2020. Customized neural machine translation systems for the Swiss legal domain. In Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track), pages 217--223, Virtual. Association for Machine Translation in the Americas.

Shubham Kumar Nigam, Tanmay Dubey, Govind Sharma, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. 2025. LegalSeg: Unlocking the structure of Indian legal judgments through rhetorical role classification. In Findings of the Association for Computational Linguistics: NAACL 2025, pages 1129--1144, Albuquerque, New Mexico. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 311--318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Maja Popovic. 2017. chrF++: words helping character n-grams. In Proceedings of the Second Conference on Machine Translation, pages 612--618, Copenhagen, Denmark. Association for Computational Linguistics.

Shabdapurush Poudel, Bal Krishna Bal, and Praveen Acharya. 2024. Bidirectional English-Nepali machine translation(MT) system for legal domain. In Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024, pages 53--58, Torino, Italia. ELRA and ICCL.

Telem Joyson Singh, Sanasam Ranbir Singh, Deepen Naorem, and Priyankoo Sarmah. 2025. Chain-of-morphemes tuning: Injecting morphology in llm-based machine translation. In 2025 International Joint Conference on Neural Networks (IJCNN), pages 1--8.

Matthew Snover, Bonnie Dorr, Rich Schwartz, Linnea Micciulla, and John Makhoul. 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, pages 223--231, Cambridge, Massachusetts, USA. Association for Machine Translation in the Americas.

Gemma Team. 2025. Gemma 3 technical report. ArXiv, abs/2503.19786.

OpenAI Team. 2024. Gpt-4 technical report.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38--45, Online. Association for Computational Linguistics.