

Findings of WMT 2026 shared task on Low-resource Indic Languages Translation

Partha Pakray¹, Santanu Pal², Advaita Vetagiri³, Kshetrimayum Boynao Singh¹, Sandeep Kumar Dash⁴, Arnab Kumar Maji⁵, Saralin A. Lyngdoh⁵, Lenin Laitonjam⁴, Anupam Jamatia⁶, Ajit Das⁷, Sunita Warjri⁸, Uzzal Sharma⁹, Koj Sambyo¹⁰, Riyanka Manna¹¹

¹NIT Silchar ²TUM, Germany ³SRM University-AP ⁴NIT Mizoram
⁵North-Eastern Hill University ⁶NIT Agartala ⁷Bodoland University
⁸IIIT Manipur ⁹Birangana Sati Sadhani Rajyik Vishwavidyalaya Golaghat
¹⁰NIT Arunachal Pradesh ¹¹ICFAI University Tripura

Abstract

This paper presents the findings of the Low-resource Indic Languages Translation Shared Task organized at the Eleventh Conference on Machine Translation (WMT 2026). We evaluated machine translation systems across ten English–Indic language pairs spanning three major language families of Northeast India, grouped into moderately resourced and extremely low-resource categories. Built upon the expanded INDICNE-CORP 2.0 dataset, this edition introduced a rigorous, bidirectional multi-domain benchmark covering the Healthcare, Political, Travel, Sports, and Entertainment domains to assess real-world generalizability. The task attracted 24 participating teams who deployed diverse methodologies, including parameter-efficient fine-tuning (PEFT) of multilingual foundations, retrieval-augmented generation (RAG) using frontier large language models (LLMs), and hybrid neuro-symbolic decoding constraints. The systems were evaluated using a comprehensive suite of lexical metrics (BLEU, METEOR, TER, ChrF++) and semantic metrics (BERTScore, COMET). Our analysis reveals a persistent performance cliff under extreme data scarcity and a pronounced directional asymmetry, demonstrating that while current architectures readily decode Indic representations into English, they struggle to generate morphologically complex Indic target text. By publicly releasing these datasets, evaluation protocols, and competitive baselines, this shared task establishes a foundation for advancing research in translation technologies for underrepresented and indigenous languages.

1 Introduction

The Indic Machine Translation (MT) Shared Task has become an important benchmark for advancing machine translation research on

low-resource Indic languages. First introduced at the Eighth Conference on Machine Translation (WMT 2023)¹ (Pal et al., 2023), the shared task initially focused on four language pairs; English↔Assamese, English↔Mizo, English↔Khasi, and English↔Manipuri to encourage the development of translation systems for underrepresented languages of Northeast India. The initiative continued at WMT 2024² (Pakray et al., 2024), further strengthening the benchmark datasets and evaluation framework. Building on these efforts, WMT 2025³ (Pakray et al., 2025) expanded the task to seven language pairs by introducing Nyishi, Bodo, and Kokborok, resulting in increased international participation and improved multilingual translation systems.

Despite the remarkable progress achieved by multilingual neural machine translation and large language models, many indigenous languages spoken across Northeast India remain severely underrepresented in terms of digital resources. These languages continue to face challenges including the scarcity of parallel corpora, limited monolingual resources, complex morphological structures, multiple writing systems, and a lack of standardized linguistic resources. Moreover, the effectiveness of existing multilingual and large language models can vary substantially across languages and resource conditions, making it important to examine the strategies adopted to adapt these models to low-resource settings. A systematic analysis of such approaches can provide valuable insights into

¹<https://www2.statmt.org/wmt23/indic-mt-task.html>

²<https://www2.statmt.org/wmt24/indic-mt-task.html>

³<https://www2.statmt.org/wmt25/indic-mt-task.html>

the role of model adaptation, data augmentation, filtering, and inference strategies in improving translation quality for these languages. Consequently, developing robust machine translation systems for these languages remains an active and important research problem.

The Eleventh Conference on Machine Translation (WMT 2026)⁴ further expands the scope of the shared task by introducing ten English-Indic language pairs organized into two categories according to the availability of parallel training data.

- **Category 1 (Moderate Training Data):** English⇔Assamese (en⇔as), English⇔Mizo (en⇔lus), English⇔Khasi (en⇔kha), English⇔Manipuri (en⇔mni-Beng), and English⇔Meitei (en⇔mni-Mtei).
- **Category 2 (Very Limited Training Data):** English⇔Bodo (en⇔bodo), English⇔Kokborok (en⇔trp), English⇔Karbi (en⇔mjw), English⇔Nagamese (en⇔nag), and English⇔Tagin (en⇔tgj).

Compared with the previous edition, the 2026 shared task replaces the Nyishi language pair with the Manipuri state official script known as the Meitei Mayek script and introduces three new language pairs: Karbi, Nagamese, and Tagin. The Mishing language pair, originally planned for this edition, was withdrawn prior to evaluation due to unforeseen circumstances.

The shared task is built upon the updated INDICNE-CORP 2.0 benchmark. A major novelty of this year’s edition is the introduction of a highly balanced, multi-domain evaluation framework spanning Healthcare, Political, Travel, Sports, and Entertainment designed to rigorously assess system generalizability across critical socioeconomic use cases. A total of 24 teams from academia, research laboratories, and industry registered for the shared task, submitting systems based on multilingual transformers, parameter-efficient fine-tuning, retrieval-augmented approaches, and frontier LLMs.

Submitted systems are evaluated using a comprehensive suite of automatic evaluation metrics (Singh et al., 2025) consisting of BLEU (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005),

TER (Snover et al., 2006), ChrF++ (Popović, 2017), BERTScore (Zhang et al., 2020), and COMET (Rei et al., 2020). While BLEU, METEOR, TER, and ChrF++ assess lexical and character-level translation quality, BERTScore and COMET provide semantic evaluation using pretrained neural representations, resulting in a holistic assessment of communicative fidelity.

Through this shared task, our aim is to establish standardized benchmarks, encourage reproducible evaluation, and foster the development of robust multilingual machine translation systems for low-resource Indic languages. We hope that the datasets, evaluation framework, and participant systems presented in this work will facilitate future research on underrepresented languages and contribute to the preservation of the linguistic diversity of Northeast India. The remainder of this paper is organized as follows: Section 2 provides a linguistic overview of the ten target languages. Section 3 details the shared task setup and objectives. Section 4 describes the INDICNE-CORP 2.0 dataset and the multi-domain evaluation benchmark. Section 5 summarizes the participating systems, and Section 6 presents the official evaluation results and comparative analysis. Finally, Section 7 concludes with our key findings.

2 Languages

The WMT 2026 Low-resource Indic Languages Translation shared task covers ten language pairs spoken across Northeast India. These languages belong to three major language families: Indo-Aryan, Tibeto-Burman, and Austroasiatic. They exhibit substantial diversity in morphology, syntax, writing systems, and resource availability, making them representative benchmarks for low-resource machine translation. The shared task includes both moderately resourced and extremely low-resource languages to encourage the development of robust multilingual translation systems under varying levels of data availability.

2.1 Assamese

Assamese is an Indo-Aryan language spoken predominantly in the northeastern Indian state of Assam, where it serves as the primary official language and a vital regional lingua franca. As one of the 22 constitutional languages listed in the Eighth Schedule of India, Assamese plays a key role in facilitating intercultural communication across the

⁴<https://www2.statmt.org/wmt26/indic-mt-task.html>

region's diverse population. The language boasts a rich literary tradition tracing back to the early medieval period, written in an eastern Indic script derived historically from ancient Brahmi. Despite its deep cultural heritage and widespread regional usage, Assamese remains digitally under-resourced. Developing scalable computational infrastructure, parallel corpora, and specialized natural language processing (NLP) tools is imperative to foster its technological integration and ensure its sustained vitality in modern digital environments.

2.2 Mizo

Mizo is a Tibeto-Burman language spoken primarily in Mizoram, India, as well as in adjacent regions of Manipur, Tripura, Assam, and parts of neighboring Myanmar and Bangladesh. Characterized by a distinctive tonal system and rich phonological structure, Mizo forms a central pillar of the cultural and social identity of the Mizo people. While historically preserved through extensive oral traditions, including folk music and narrative poetry, the language transitioned into a written medium in the late 19th century following the adoption of a modified Roman alphabet introduced by Christian missionaries. Despite an active literary domain and widespread regional literacy, Mizo faces substantial barriers in digital representation due to a scarcity of standardized language resources and annotated corpora. Mitigating these technological gaps through neural machine translation and corpus development is essential for its preservation and modern computational accessibility.

2.3 Khasi

The Khasi language (*Ka Ktien Khasi*) is an Austroasiatic language spoken predominantly by the Khasi people in the state of Meghalaya, India (Grierson, 1904). Classified within the Khasic sub-branch of the Mon-Khmer language family, Khasi represents an ethnolinguistic island surrounded by predominantly Tibeto-Burman and Indo-Aryan languages (Rabel, 1961). Although the language exhibits significant dialectal diversity-comprising varieties such as Pnar, Lyngngam, and War-the Sohra (Cherrapunji) dialect is widely recognized as Standard Khasi (Nagaraja et al., 2013). Historically an oral language, Khasi was given a standardized written tradition in 1841 when Welsh missionary Thomas Jones introduced a modified Latin alphabet, establishing an orthography that remains in continuous educational, literary, and administrative

use today (Grierson, 1904).

Structurally, Khasi is an analytic language characterized by a canonical Subject-Verb-Object (SVO) word order (Rabel, 1961). Unlike many of its geographical neighbors in Northeast India, Khasi is a non-tonal, stress-based language with a rich inventory of phonemic vowels and complex onset consonant clusters (Rabel, 1961). Morphologically, the language employs a robust system of prefixation and grammatical gender, classifying nouns using specific markers such as *u* for masculine, *ka* for feminine, and *i* for diminutive (Henderson, 1976). Today, Standard Khasi holds co-official status in Meghalaya and continues to evolve through widespread literary production, digital media, and active ongoing research in natural language processing (Nagaraja et al., 2013).

2.4 Manipuri

Manipuri, also known as *Meiteilon* or *Meeteilon*, is a Sino-Tibetan language primarily spoken in Manipur and serves as the state's principal lingua franca. Despite its official recognition and rich literary tradition, Manipuri remains low-resource for natural language processing and machine translation (MT), with limited parallel corpora, linguistic tools, and pretrained resources (Singh and Bandyopadhyay, 2010; Laitonjam and Singh, 2023).

Manipuri is typologically different from English, following predominantly subject-object-verb (SOV) word order and exhibiting rich agglutinative morphology. Grammatical information such as case, tense, aspect, number, and modality can be expressed through sequences of suffixes, resulting in substantial lexical sparsity and making translation challenging in both directions (Singh and Bandyopadhyay, 2010).

Another important challenge is script variation. Manipuri is written in both the indigenous *Meitei Mayek* and Bengali (Singh et al., 2023) scripts. The coexistence of multiple scripts fragments the already limited resources and introduces issues related to spelling variation, Unicode normalization, and script consistency.

To reflect the two writing systems currently used by the Manipuri-speaking community, the WMT 2026 shared task includes two translation tracks:

- **Manipuri:** written in the Bengali script.
- **Meitei:** written in the indigenous Meitei Mayek script.

Previous studies have explored statistical, unsupervised neural, and corpus-based approaches for English–Manipuri MT (Singh and Singh, 2020; Laitonjam and Singh, 2021). More recently, multilingual pretraining, transfer learning, subword modeling, and synthetic data generation have provided promising directions for addressing data scarcity. Nevertheless, the combination of limited parallel data, morphological complexity, and script variation continues to make Manipuri a challenging language for machine translation.

2.5 Bodo

Bodo is a major Tibeto-Burman language belonging to the Boro-Garo group within the broader Boro-Konyak-Jingphaw sub-branch (Wood, 2008). Spoken predominantly across the Bodoland Territorial Region (BTR) of Assam, India, it is officially recognized under the Eighth Schedule of the Indian Constitution. According to the 2011 Census of India, Bodo has approximately 1.5 million speakers, with extended communities in Arunachal Pradesh, Meghalaya, West Bengal, Nepal (where it is known as Meche), and Bangladesh.

Linguistically, Bodo shares structural similarities with related regional languages such as Kokborok and Dimasa. Bhattacharya (1977) categorizes Bodo into four main dialectal regions: North-West, South-West, North-Central, and South. Although historically written using Roman and Assamese/Bengali scripts, Devanagari was officially adopted as the standardized script in 1975, facilitating modern literary production, administrative use, and Natural Language Processing (NLP) efforts.

2.6 Kokborok

Kokborok is an indigenous language belonging to the Bodo–Garo subgroup of the Tibeto–Burman branch of the Sino–Tibetan language family (Grierson, 1903; Benedict, 1972). It is primarily spoken in the Indian state of Tripura, with smaller speech communities in southern Assam, Mizoram, and the Chittagong Hill Tracts of Bangladesh. Recognized as one of the official languages of Tripura, Kokborok plays a central role in the linguistic and cultural identity of several Indigenous communities, including the Debbarma, Jamatia, Reang (Bru), Noatia, Koloï, Murasing, Rupini, and Uchoi tribes. According to the Census of India 2011, the language has approximately 0.88 million speakers (Office of the Registrar General and Census Commissioner of India, 2011) in India, with additional

speakers residing in neighbouring Bangladesh.

Typologically, Kokborok follows the canonical Subject–Object–Verb (SOV) word order, employs postpositions, and exhibits an agglutinative morphology in which grammatical information is primarily expressed through suffixation. Similar to other Bodo–Garo languages, it also displays productive derivation, compounding, and lexical borrowing from Bengali and English due to prolonged language contact (Karapurkar, 1979; Jacquesson, 2003; Hoque, 2014). Earlier linguistic studies describe Kokborok as a tonal language, although the functional role of lexical tone varies across dialects (Karapurkar, 1972; Jacquesson, 2008). The language generally follows simple consonant–vowel (CV) and consonant–vowel–consonant (CVC) syllable patterns with limited consonant clustering, while grammatical distinctions such as definiteness and case marking contribute to its morphosyntactic complexity (Daladier, 2002).

2.7 Karbi

Karbi is a Tibeto-Burman language spoken primarily in the Karbi Anglong and West Karbi Anglong districts of Assam, India. It serves as the principal language of the Karbi community and exhibits an agglutinative morphological structure together with the canonical Subject–Object–Verb (SOV) word order common among Tibeto-Burman languages. Although Karbi possesses a growing body of written literature and educational materials, computational resources for the language remain scarce. The limited availability of parallel corpora, annotated datasets, and language technologies makes Karbi an important benchmark for evaluating multilingual and low-resource machine translation systems. Its inclusion in WMT 2026 aims to stimulate research on language documentation, corpus creation, and translation technologies for indigenous languages of Northeast India.

2.8 Nagamese

Ngamese is a contact language widely spoken throughout the Indian state of Nagaland. It functions as the primary lingua franca among speakers of different Naga languages and has evolved mainly from Assamese, while incorporating lexical and grammatical influences from English, Hindi, and several indigenous Naga languages. Owing to its extensive use in daily communication, trade, education, and media, Nagamese plays an important role in regional multilingual communication.

Nevertheless, standardized orthography, linguistic resources, and parallel corpora remain limited. The inclusion of Nagamese in the WMT 2026 shared task provides a valuable benchmark for machine translation research on contact languages and extremely low-resource settings.

2.9 Tagin

Tagin is a Tibeto-Burman language spoken predominantly by the Tagin community in the Upper Subansiri district of Arunachal Pradesh, India. The language represents one of the least-resourced languages included in the WMT 2026 shared task, with very limited digital resources, parallel corpora, and computational linguistic tools. Like many indigenous languages of Northeast India, Tagin faces challenges associated with language documentation, standardization, and technological support. The inclusion of Tagin in this shared task aims to encourage the development of multilingual and transfer-learning approaches capable of handling extremely low-resource translation scenarios while contributing to the preservation and digital revitalization of the language.

3 Low-Resource Indic Language Translation 2026 Shared Task

3.1 Overview and Task Description

Following the success of the Low-resource Indic Language Translation shared tasks at WMT 2023, WMT 2024, and WMT 2025, the initiative continues as part of the Eleventh Conference on Machine Translation (WMT 2026). The shared task aims to advance research on machine translation for underrepresented Indic languages by providing standardized benchmark datasets, evaluation protocols, and a common platform for comparing translation systems. Despite significant advances in multilingual neural machine translation and large language models, building effective translation systems for low-resource languages remains challenging due to the scarcity of high-quality parallel corpora, limited linguistic resources, and diverse writing systems.

The WMT 2026 shared task expands the scope of previous editions by covering ten English-Indic language pairs spanning multiple language families and writing systems. Similar to previous years, the language pairs are divided into two categories according to the availability of parallel training data, enabling the evaluation of translation systems under different levels of resource availability.

3.2 Task Categories

The WMT 2026 Low-resource Indic Languages Translation shared task is organized into two categories according to the availability of parallel training data.

3.2.1 Category 1: Moderate Training Data

Category 1 consists of language pairs with a moderate amount of publicly available parallel training data.

- English $\Leftrightarrow$ Assamese (en $\Leftrightarrow$ as)
- English $\Leftrightarrow$ Mizo (en $\Leftrightarrow$ lus)
- English $\Leftrightarrow$ Khasi (en $\Leftrightarrow$ kha)
- English $\Leftrightarrow$ Manipuri (en $\Leftrightarrow$ nni)
- English $\Leftrightarrow$ Meitei (en $\Leftrightarrow$ nni-Mtei)

3.2.2 Category 2: Very Limited Training Data

Category 2 focuses on language pairs with very limited parallel training resources.

- English $\Leftrightarrow$ Bodo (en $\Leftrightarrow$ bodo)
- English $\Leftrightarrow$ Kokborok (en $\Leftrightarrow$ trp)
- English $\Leftrightarrow$ Karbi (en $\Leftrightarrow$ mjlw)
- English $\Leftrightarrow$ Nagamese (en $\Leftrightarrow$ nag)
- English $\Leftrightarrow$ Tagin (en $\Leftrightarrow$ tgi)

Although the Mishng (en $\Leftrightarrow$ mrg) language pair was initially planned for this year’s shared task, it was withdrawn before the evaluation phase due to unforeseen circumstances. Consequently, the official shared task includes ten English-Indic language pairs.

Overall, the shared task covers ten language pairs spanning three major language families and multiple writing systems, providing a challenging benchmark for multilingual and low-resource machine translation.

3.3 Goal

The primary objective of the shared task is to develop robust machine translation systems for low-resource Indic languages despite limited training data. Participants are encouraged to investigate a wide range of approaches, including:

- Leveraging monolingual corpora through semi-supervised and self-supervised learning.

Language	Sentences	English Words	Target Words
Assamese	54,000	1,033,592	878,475
Mizo	50,000	981,535	1,044,121
Khasi	26,000	778,689	948,853
Manipuri	23,687	422,521	357,524
Meitei	17,000	313,175	305,198
Bodo	15,215	228,217	204,921
Tagin	5,343	26,423	25,942
Kokborok	2,269	55,634	51,268
Nagamese	2,000	26,453	27,709
Karbi	1,000	12,787	15,086

Table 1: Statistics of the multilingual parallel corpus used for training and validation.

- Multilingual modeling and cross-lingual knowledge transfer.
- Parameter-efficient fine-tuning of multilingual large language models.
- Retrieval-augmented and data augmentation techniques.
- Novel architectures specifically designed for extremely low-resource translation.

4 Dataset

The WMT 2026 Low-resource Indic Languages Translation shared task provides standardized multilingual benchmark datasets for training, validation, and evaluation of machine translation systems. The dataset covers ten English-Indic language pairs spanning three language families and multiple writing systems. It consists of parallel corpora for supervised machine translation together with manually curated benchmark test sets designed to evaluate translation quality across diverse application domains.

4.1 Training

The training corpus used in the WMT 2026 shared task is based on the updated **IndicNE-Corp 2.0** benchmark dataset, which extends the multilingual resources released in previous editions of the shared task. IndicNE-Corp 2.0 consists of high-quality parallel corpora collected from publicly available resources together with newly curated bilingual datasets for low-resource languages spoken across Northeast India.

The training set covers the following ten language pairs:

The multilingual corpus has been constructed by integrating datasets released in previous research together with newly collected bilingual resources

obtained from government publications, multilingual websites, educational resources, online dictionaries, news articles, religious texts, and manually curated translations. For several extremely low-resource languages, additional parallel corpora were developed specifically for this shared task to improve language coverage and establish standardized benchmarks for machine translation research.

Table 1 summarizes the statistics of the multilingual parallel corpus used for training.

4.2 Testing

To ensure the highest standard of evaluation, the benchmark test sets underwent a rigorous manual curation process. Native speakers and domain-specific language experts validated all translations based on linguistic correctness, cultural nuance, and domain diversity.

Unlike previous editions, the WMT 2026 benchmark challenges translation systems with a significantly larger and more diverse multilingual test suite spanning ten language pairs. Furthermore explicit inclusion of bidirectional evaluation (English $\leftrightarrow$ Indic) allows for a holistic assessment of lexical accuracy, semantic preservation, domain adaptation, and reverse translation capabilities.

The curated benchmark establishes a highly standardized evaluation framework. It not only facilitates the fair comparison of multilingual machine translation systems under varying levels of resource availability but also serves as a robust foundational resource to catalyze future research in low-resource Indic machine translation.

4.2.1 Domain Selection and Motivation

To rigorously assess the generalizability and real-world applicability of machine translation systems across diverse communicative contexts, the WMT 2026 test benchmark is curated across five representative domains: *Entertainment*, *Sports*, *Healthcare*, *Travel*, and *Political*. These domains were deliberately selected to reflect critical socioeconomic use cases in Northeast India while exposing translation models to distinct linguistic, syntactic, and lexical challenges:

- **Healthcare:** Evaluates translation safety, precise handling of medical terminology, anatomical references, and imperative instructional syntax critical for public health advisories in multilingual rural communities.

- **Political & Administrative:** Tests system fidelity on formal register, complex subordinate clauses, institutional nomenclature, and official government communications.
- **Travel & Tourism:** Focuses on geographic named entities, cultural landmarks, directional instructions, and dialogic service interactions vital for regional economic development.
- **Sports:** Introduces dynamic event reporting, highly frequent numerical expressions, temporal markers, and domain-specific action verbs requiring precise lexical disambiguation.
- **Entertainment:** Assesses colloquial expressions, pop-culture terminology, idiomatic phrasing, and informal discourse structures where literal translation often fails.

4.2.2 Comparative Analysis Across Languages

As detailed in Figure 1, the domain distribution is strictly uniform across all ten language pairs evaluated in the shared task. Each language pair—regardless of its categorization into Category 1 (Moderate Training Data) or Category 2 (Very Limited Training Data)—is evaluated against exactly 400 sentences per domain.

This balanced comparative framework ensures that the performance disparities observed in the final results are strictly attributable to the models’ linguistic capabilities and the availability of training data, rather than being skewed by domain overrepresentation in the test sets. Furthermore, the uniform allocation establishes a robust baseline for analyzing cross-lingual transfer learning. By maintaining parity across highly resourced domains (e.g., Entertainment and Travel) and highly specialized domains (e.g., Healthcare and Political), the benchmark allows for precise error analysis regarding how well models generalize specialized vocabulary in extremely low-resource languages like Karbi, Nagamese, and Tagin compared to moderately resourced languages like Assamese and Mizo.

Each domain contributes 400 sentence pairs, resulting in a balanced evaluation dataset covering a broad range of topics.

To evaluate translation performance in both directions, every benchmark is divided equally into:

- **1,000 English → Indic sentences**

- **1,000 Indic → English sentences**

This bidirectional evaluation enables a comprehensive assessment of multilingual translation systems under both forward and reverse translation settings.

The benchmark test sets were manually curated by native speakers and language experts to ensure translation quality, domain diversity, and linguistic correctness. The selected domains represent practical translation scenarios frequently encountered in public communication, healthcare, tourism, media, sports, and government-related content.

Unlike previous editions, the WMT 2026 benchmark evaluates translation systems on a larger and more diverse multilingual test suite spanning ten language pairs. The inclusion of both English→Indic and Indic→English translation directions allows comprehensive evaluation of lexical accuracy, semantic preservation, domain adaptation, and bidirectional translation capability.

Overall, the benchmark provides a standardized evaluation framework for comparing multilingual machine translation systems under varying levels of resource availability and supports future research on low-resource Indic machine translation.

5 Participants and System Descriptions

In the WMT 2026 Low-resource Indic Languages Translation Shared Task, a total of 24 teams from academia, research laboratories, and industry registered for the shared task, as listed in Table 4. These teams submitted machine translation systems for one or more of the ten English-Indic language pairs, resulting in a diverse collection of primary and contrastive submissions. The participating systems employed a wide range of methodologies, including multilingual neural machine translation models, large language models (LLMs), parameter-efficient fine-tuning techniques, retrieval-augmented approaches, synthetic data generation, and multilingual transfer learning. This diversity of approaches provides valuable insights into the current state of low-resource machine translation and highlights the continued progress of multilingual NLP research for underrepresented Indic languages.

The following subsections summarize the approaches adopted by each participating team based on their submitted system description papers.

PANINI: The PANINI (Huidrom et al., 2026) team from the Indraprastha Institute of Informa-

Domain	Primary Linguistic Challenges	Dominant Register / Syntax	Key Evaluation Target
Healthcare	Specialized medical terms, dosage numerals, morphological inflection of verbs	Formal / Imperative & Informative	Semantic fidelity, low hallucination risk
Political	Multi-clausal sentences, institutional named entities, formal honorifics	Highly Formal / Complex Hierarchical	Structural syntactic preservation
Travel	Local toponyms, transliteration across scripts, spatial pre/postpositions	Semi-formal / Descriptive & Directive	Named-entity recognition & script consistency
Sports	Fast-paced action verbs, score/time numerals, dynamic metaphors	Semi-formal to Informal / Active voice	Lexical disambiguation, numeric fidelity
Entertainment	Slang, cultural idioms, loanwords, code-mixing	Informal / Conversational	Lexical adaptation, idiomatic accuracy

Table 2: Comparative analysis of linguistic challenges and evaluation targets across the five selected benchmark domains.

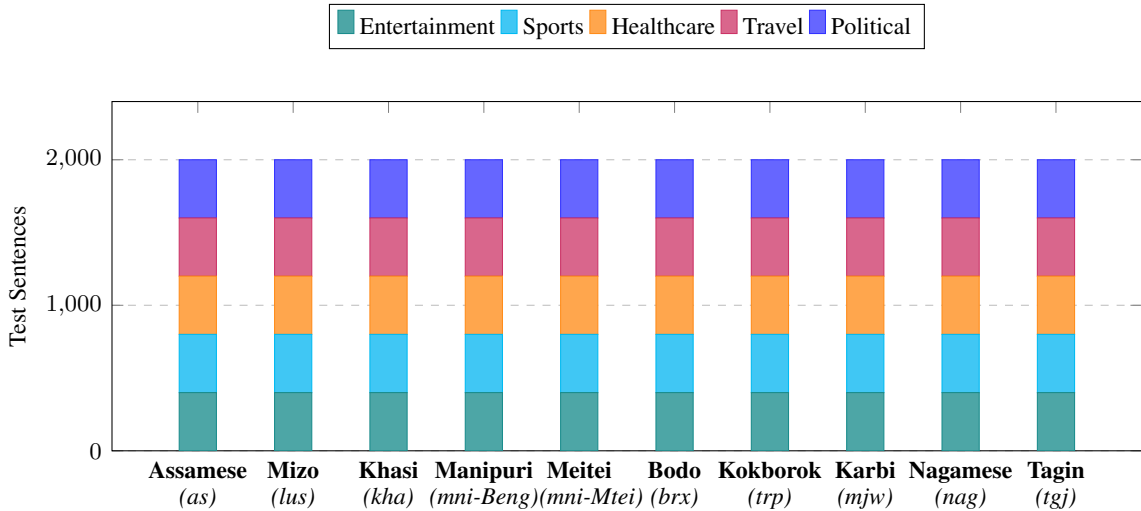


Figure 1: Domain composition of the benchmark test set across all 10 language pairs. Each pair contains exactly 2,000 sentences, distributed uniformly across five domains (400 sentences per domain) and divided equally between the English→Indic and Indic→English translation directions.

Language Pair	Submissions
English - Assamese	15 (primary), 12 (contrastive)
English - Mizo	9 (primary), 9 (contrastive)
English - Khasi	8 (primary), 8 (contrastive)
English - Manipuri	14 (primary), 6 (contrastive)
English - Meitei	9 (primary), 2 (contrastive)
English - Bodo	10 (primary), 10 (contrastive)
English - Kokborok	4 (primary), 6 (contrastive)
English - Karbi	2 (primary), 3 (contrastive)
English - Nagamese	5 (primary), 6 (contrastive)
English - Tagin	4 (primary), 6 (contrastive)

Table 3: Number of submitted systems for the English→Indic translation direction in the WMT 2026 Low-resource Indic Languages Translation shared task.

tion Technology (IIIT) Delhi participated in the WMT 2026 Low-resource Indic Languages Translation Shared Task with English↔Manipuri systems covering both Bengali and Meitei Mayek scripts. Their systems were based on the 1B-parameter IndicTrans2 architecture and adapted us-

ing rank-stabilized LoRA (rsLoRA) for parameter-efficient fine-tuning. The approach combined back-translation, forward translation, and iterative pseudo-labelling, with synthetic data filtered using LaBSE similarity and model-confidence thresholds. Subword regularization and script-aware preprocessing and post-processing were further employed to improve robustness across both Manipuri scripts.

VoxNexus-Translate: The VoxNexus-Translate (Zhou et al., 2026) team from VoxNexus, Shanghai, developed a hybrid machine translation system combining a fine-tuned multilingual decoder-only LLM with Statistical Machine Translation (SMT) word alignment for low-resource Indic languages. Their approach employed IBM Models and fast_align to extract high-confidence bilingual lexicons and phrase mappings, which were incorporated as decoding constraints to reduce hallucinations and improve the translation of named entities, numerals, and domain-specific terminol-

Team Name	Name of University/Lab/Industry/Group
PANINI	Indraprastha Institute of Information Technology, Delhi
VoxNexus-Translate	VoxNexus, Shanghai
ADI-KIIT	Kalinga Institute of Industrial Technology, Odisha
DoDS-IITPKD	Indian Institute of Technology Palakkad, Kerala
gators	University of Florida
BRS-NK	Kathmandu University, Bagmati
TIP-RBG-AI	Kitami Institute of Technology, Hokkaido
EROL	Nalanda University
RozarNLP-cr7	Indian Institute of Information Technology, Design and Manufacturing, Kurnool
Decepticons-IIITH	International Institute of Information Technology, Hyderabad
MTG-LIRA	Centre for Development of Advanced Computing (CDAC), Pune
BVSLP	Banasthali Vidyapith
LTRC_Plural	IIIT Hyderabad
ANVITA	CAIR, Defence Research and Development Organisation (DRDO), Bangalore
Winterpeg	The University of Winnipeg, Manitoba, Canada
KrenTransV0-1	MWire Labs, Shillong, Meghalaya
NLP-MT_Amrita	Amrita Vishwa Vidyapeetham, Bengaluru Campus, Karnataka
DELAB-IIITM	Indian Institute of Information Technology, Senapati, Manipur
JH_NLP_APAShreya2	PES University, Karnataka
星辰之力	China Telecom Corporation Limited, Chongqing Branch
AkibaNLP-TUT	Toyohashi University of Technology / Natural Language Processing Lab.
CIT MT BTECH 2026	Central Institute of Technology Kokrajhar, Assam
SCE-KIIT	Kalinga Institute of Industrial Technology
IndicMT	Jadavpur University, Kolkata

Table 4: Overview of the teams registered for the WMT 2026 Low-resource Indic Languages Translation shared task.

ogy. The system further employed script normalization, transliteration, and rule-based post-processing to improve translation consistency and faithfulness.

ADI-KIIT: The ADI-KIIT (Aditya and Ekbal, 2026) team from Kalinga Institute of Industrial Technology developed bidirectional English–Assamese translation systems using the compact 200M-parameter distilled IndicTrans2 model with LoRA-based parameter-efficient fine-tuning. Their approach explored supervised fine-tuning (SFT) followed by Direct Preference Optimization (DPO), with preference signals derived from GEMBA-style LLM judging, reference-free MQM, and COMET. Based on development-set performance, SFT was selected as the primary system for Assamese-to-English, while the SFT+DPO configuration was used for English-to-Assamese.

DoDS-IITPKD: The DoDS-IITPKD (Salvin et al., 2026) team from the Indian Institute of Technology Palakkad developed bidirectional translation systems for six English–Indic language pairs by adapting two multilingual pretrained architec-

tures. They employed NLLB-200-3.3B for Khasi, Mizo, and Kokborok, and IndicTrans2 for Assamese, Manipuri, and Bodo. The models were adapted using parameter-efficient LoRA together with DoRA and rsLoRA. For Khasi and Kokborok, which are not natively supported by the NLLB tokenizer, the team employed a same-script surrogate language-tag strategy. Their primary systems used only the official shared-task data, while the contrastive systems incorporated additional external corpora.

gators: The gators (Dhawan et al., 2026) team from the University of Florida developed a training-free retrieval-augmented machine translation system using Gemini 2.5 Flash with many-shot in-context learning. Their approach employed BM25 retrieval to select relevant parallel examples from the training corpus and incorporate them into prompts without model fine-tuning. The primary system used only the official WMT 2026 data, while the contrastive system incorporated additional publicly available parallel corpora where

available.

BRS-NK: The BRS-NK (Subedi and Karki, 2026) team from Kathmandu University developed bidirectional English–Assamese and English–Bodo translation systems by fine-tuning the NLLB-200-600M model. While Assamese is natively supported by NLLB, Bodo was handled using the Hindi (hin_Deva) language tag through a proxy-language strategy. Their approach emphasized extensive data preprocessing, including text normalization, numeral conversion, alignment filtering, artifact removal, and deduplication. The primary systems used only the official shared-task data, while the contrastive systems incorporated additional parallel resources.

TIP-RBG-AI: The TIP-RBG-AI (HB et al., 2026) team from Kitami Institute of Technology proposed a zero-shot cross-lingual translation framework using three massively multilingual MT architectures: NLLB-200, MADLAD-400, and IndicTrans2. Rather than fine-tuning the models, their approach focused on inference optimization through script normalization, algorithmic back-transliteration for Romanized text, and phylogenetic script mapping for extremely low-resource languages. The framework aimed to improve semantic fidelity and lexical preservation across both moderate- and very low-resource language settings.

EROL: The EROL (Devi and Lee, 2026) team from Nalanda University developed bidirectional neural machine translation systems for English–Assamese, English–Bodo, and English–Manipuri in both Bengali and Meitei Mayek scripts. Their systems employed the Transformer-Base architecture with six encoder and six decoder layers, shared source–target embeddings, and BPE tokenization. A single bidirectional model was trained for each language pair using standard sequence-to-sequence cross-entropy learning.

RozarNLP-cr7: The RozarNLP-cr7 (Yadav et al., 2026) team from IIITDM Kurnool developed multilingual translation systems based on the NLLB-200 Distilled 1.3B model, adapted using LoRA for parameter-efficient fine-tuning. Their approach employed joint multilingual training to exploit cross-lingual transfer across several low-resource language pairs, including Assamese, Manipuri, Mizo, Khasi, and Nyishi. For Meitei Mayek, where multilingual adaptation remained challenging, the pretrained NLLB-200 model was

retained. IndicTrans2 and pretrained NLLB-200 models without task-specific adaptation were additionally evaluated as contrastive systems.

Decepticons-IIITH: The Decepticons-IIITH (Rao et al., 2026) team from IIIT Hyderabad developed English-to-Indic translation systems for Assamese, Mizo, Nagamese, and Bodo using NLLB-200 for the first three languages and IndicTrans2 for Bodo. Their approach generated synthetic parallel data through back-translation of public monolingual corpora and filtered the resulting pairs using Morphology-Aware Lexicon Matching (MALM). The models followed a two-stage training strategy consisting of continued pretraining on the filtered synthetic data, followed by fine-tuning on the official parallel corpus.

MTG-LIRA: The MTG-LIRA (Hede et al., 2026) team from CDAC Pune developed Transformer-based encoder–decoder NMT systems using the OpenNMT and Eole frameworks for multiple low-resource Indic language pairs, including Assamese, Bodo, Mizo, Nyishi, and Manipuri in both Meitei Mayek and Bengali scripts. Their systems were trained on cleaned and preprocessed WMT 2026 parallel data, with Byte Pair Encoding (BPE) applied for subword tokenization.

BVSLP: The BVSLP (Arora et al., 2026) team from Banasthali Vidyapith developed neural machine translation systems based on an encoder–decoder architecture, augmented with linguistic preprocessing for named-entity handling. Their methodology employed a rule-based Named Entity Recognition (NER) module followed by a Named Entity Translation (NET) component, using bilingual knowledge-base lookup and transliteration to preserve organization, location, and other named entities. The processed data were further segmented using Byte Pair Encoding (BPE) before fine-tuning the NMT models, improving the handling of rare words, morphological variations, and low-resource vocabulary.

LTRC_Plural: The LTRC_Plural (Vishnuraj et al., 2026) team from IIIT Hyderabad developed bidirectional English translation systems for Assamese, Mizo, Khasi, and Tagin using two multilingual foundation models: the encoder–decoder NLLB-1.3B and Sarvam Translate, built on Gemma 3 4B. Their primary systems employed QLoRA-based parameter-efficient fine-tuning, us-

ing NLLB for Assamese and Mizo and a shared multilingual Sarvam adapter with rsLoRA for Khasi and Tagin. The contrastive systems explored additional parallel corpora with temperature-based multilingual sampling and reinforcement learning using RLOO, with BLEU and ChrF++ as reward signals.

ANVITA: The ANVITA (J et al., 2026) team from CAIR, DRDO, Bangalore, developed bidirectional translation systems for Khasi, Nagamese, and Tagin using a transfer-learning approach based on the 220M-parameter T5-base model. The model was fine-tuned using standard NMT and cycle-consistency losses, with data augmentation through English paraphrasing, related-language parallel data, back-translation, and additional external resources. For Nagamese-to-English, the primary system instead employed IndicTrans2 in a zero-shot setting with Assamese transliteration of the Nagamese input.

Winterpeg: The Winterpeg (Rittikar and Rammanna, 2026) team from the University of Winnipeg developed a routed two-model translation system for six low-resource North-East Indian languages. For Assamese, Manipuri (Bengali script), and Mizo, they fine-tuned NLLB-200 on the official parallel data. For Khasi, Nyishi, and Meitei Mayek, they trained a vocabulary-free UTF-8 byte-level decoder-only language model from scratch, augmented with a causal Neural Cellular Automaton (NCA) front-end. This byte-level approach enables translation of previously unseen scripts without requiring explicit vocabulary engineering.

KrenTransV0-1: The KrenTransV0-1 (Nyalang et al., 2026) team from MWire Labs developed a hybrid translation framework covering all 11 low-resource language pairs in the English-to-target direction. Their approach employed fine-tuned IndicTrans2 models for Indic-script languages and NLLB-200 1.3B models for Latin-script languages. For three extremely low-resource languages where NLLB exhibited output collapse, the team instead fine-tuned the M2M-100 418M architecture.

NLP-MT_Amrita: The NLP-MT_Amrita (Sree and Gupta, 2026) team from Amrita Vishwa Vidyapeetham proposed a GRPO-inspired Agentic Retrieval-Augmented Generation (RAG) framework for bidirectional English–Bodo and English–Manipuri translation. Their approach combined IndicTrans2, NLLB-600M, and Sarvam-Translate

with Milvus-based retrieval and multilingual-e5-large embeddings. A dynamic agent explored multiple retrieval and decoding configurations, while a reward function based on SacreBLEU and BERT similarity selected the most suitable translation candidate. The framework operated without task-specific model fine-tuning, providing a scalable approach for low-resource machine translation.

DELAB-IIITM: The DELAB-IIITM (Oinam et al., 2026) team from IIIT Manipur developed English-to-Indic translation systems for Assamese, Bodo, and Manipuri in both Bengali and Meitei Mayek scripts using two pretrained multilingual models: NLLB-200-Distilled-600M and IndicTrans2-1.1B. Their approach employed full fine-tuning of NLLB-200 for English-to-Assamese and English-to-Manipuri, while IndicTrans2 was adapted using LoRA for English-to-Manipuri and English-to-Bodo. All systems were trained exclusively on the official WMT 2026 training data.

JH_NLP_APAShreya2: The JH_NLP_APAShreya2 (Sinha et al., 2026) team from PES University developed an English-to-Manipuri translation system using the NLLB-200-Distilled-600M model with LoRA-based parameter-efficient fine-tuning. The model was trained on the official WMT 2026 English–Manipuri parallel corpus, targeting Manipuri in the Bengali script. Their primary system used the complete training dataset with optimized LoRA parameters to improve translation performance.

星辰之力: The 星辰之力 (Ruan et al., 2026) team from China Telecom Corporation Limited developed English-to-Indic translation systems for Bodo, Karbi, Kokborok, Nagamese, and Tagin using Qwen2.5-32B-Instruct with LoRA-based fine-tuning and Hunyuan-MT-7B with full-parameter fine-tuning. Their approach explored SFT and DPO-based optimization and incorporated automatically generated bilingual word-level alignment tables as “Helpful Vocabulary” to improve lexical guidance in low-resource translation.

AkibaNLP-TUT: AkibaNLP-TUT (Kachi et al., 2026) from Toyohashi University of Technology employed language-specific word-level noise injection to improve robustness and generalization for low-resource translation in the Indic-to-English direction. Their approach augmented Bengali–English or Hindi–English parallel data using lexi-

cal items extracted from monolingual data of the corresponding low-resource language.

CIT MT BTECH 2026: The Central Institute of Technology Kokrajhar team developed English-to-Assamese machine translation systems using a Transformer encoder–decoder model trained from scratch and the pretrained NLLB-200 Distilled 600M model. Their approach compared BPE-based Transformer training with transfer learning through fine-tuning NLLB-200, demonstrating the effectiveness of multilingual pretraining for low-resource Assamese translation.

SCE-KIIT: SCE-KIIT (Maiti et al., 2026b,a) from Kalinga Institute of Industrial Technology developed systems for English to Nagamese and English–Kokborok. For English to Nagamese, they fine-tuned NLLB-200-Distilled-600M. For English–Kokborok, the team proposed KokLLaMA, based on Llama-3.2-3B-Instruct and adapted using QLoRA on 10,856 Kokborok conversational QA pairs. The system employed structured translation prompting and rule-based post-processing to reduce repetitive generation loops. Their approach demonstrates the potential of both multilingual NMT adaptation and cross-task transfer from conversational data for low-resource Indic machine translation.

IndicMT: IndicMT from Jadavpur University developed bidirectional low-resource translation systems using Sarvam-Translate, based on the Gemma 3 architecture, with QLoRA-based parameter-efficient fine-tuning. Their approach employed NLLB-200-based back-translation to generate and filter synthetic parallel data, followed by 4-bit quantized adaptation of the translation model. The framework combined transfer learning, data augmentation, and parameter-efficient fine-tuning to support resource-efficient Indic machine translation.

5.1 Comparative Technological Analytics and Methodological Taxonomy

The submissions to the WMT 2026 shared task reflect a significant paradigm shift in low-resource machine translation. Rather than training models strictly from scratch on bilingual corpora, participating teams converged on five major technical dimensions: foundation model selection, parameter-efficient adaptation, synthetic data filtering, script-handling strategies, and test-time re-

trieval/reasoning frameworks. Table 5 synthesizes the core technical components adopted by all 24 participating teams.

5.1.1 Key Methodological Trends

1. Dominance of Pretrained Multilingual Foundations vs. LLMs The architectural landscape was split into two primary paradigms:

- **Multilingual Encoder–Decoder Models:** NLLB-200 (ranging from 600M to 3.3B) and IndicTrans2 (200M to 1.1B) served as the primary workhorses for over 70% of participating teams, demonstrating that domain-aligned pretraining provides superior parameter efficiency and syntactic stability for agglutinative Indic languages.
- **Autoregressive Large Language Models:** Frontier decoder-only models (e.g., Gemini 2.5 Flash, Qwen2.5-32B, Gemma 3/Sarvam-Translate, and Llama-3.2-3B) were successfully deployed using either in-context learning or parameter-efficient adaptation, particularly excelling in lexical fluency for English target generation.

2. Evolution of Parameter-Efficient Adaptation and Preference Alignment Rather than performing full-parameter updates, teams predominantly leveraged Parameter-Efficient Fine-Tuning (PEFT):

- **Advanced Adapter Variants:** Teams transitioned from standard LoRA to stabilized variants including rank-stabilized LoRA (rsLoRA) and Weight-Decomposed Low-Rank Adaptation (DoRA) to mitigate gradient instability in extremely low-resource settings (e.g., PANINI, DoDS-IITPKD, LTRC_Plural).
- **Post-Training Preference Optimization:** Beyond standard cross-entropy loss, teams explored reinforcement learning and direct alignment—such as Direct Preference Optimization (DPO), Reinforcement Learning from Online Output (RLOO), and Group Relative Policy Optimization (GRPO) guided by COMET, SacreBLEU, and GEMBA-style metric rewards (e.g., ADI-KIIT, LTRC_Plural, NLP-MT_Amrita, 星辰之力).

Team	Base Architecture(s)	Adaptation Paradigm	/ Training	Data Augmentation & Filtering	Inference / Routing Strategy
PANINI (Huidrom et al., 2026)	IndicTrans2 (1B)	PEFT (rsLoRA)		BT, FT, Pseudo-labeling, LaBSE filtering	Script-aware post-processing
VoxNexus-Translate (Zhou et al., 2026)	Decoder-only LLM	SFT + SMT Alignment		IBM Model / fast_align phrase extraction	SMT decoding constraints, Transliteration
ADI-KIIT (Aditya and Ekbal, 2026)	IndicTrans2-Distilled (200M)	LoRA + DPO (GEMBA / COMET)		Preference pair generation	Metric-guided decoding
DoDS-IITPKD (Salvin et al., 2026)	NLLB-200 (3.3B), IndicTrans2	PEFT (LoRA, DoRA, rsLoRA)		External corpora (contrastive)	Surrogate script language-tagging
gators (Dhawan et al., 2026)	Gemini 2.5 Flash	Training-free / Many-shot ICL		External parallel corpora (contrastive)	BM25 dynamic context retrieval
BRS-NK (Subedi and Karki, 2026)	NLLB-200 (600M)	Full Fine-Tuning		Deduplication, artifact / numeral cleaning	Proxy-language tagging (hin_Deva)
TIP-RBG-AI (HB et al., 2026)	NLLB-200, MADLAD-400, IndicTrans2	Zero-Shot Inference		None	Phylogenetic script mapping, Back-transliteration
EROL (Devi and Lee, 2026)	Transformer-Base (from scratch)	Standard Seq2Seq Cross-Entropy		Shared source-target BPE	Standard beam search
RozarNLP-cr7 (Yadav et al., 2026)	NLLB-200 Distilled (1.3B)	Multilingual Joint LoRA		Cross-lingual parameter sharing	Script fallback routing
Deception-IIIITH (Rao et al., 2026)	NLLB-200, IndicTrans2	Continued Pretraining + Fine-Tuning		Back-translation + MALM lexicon filtering	Two-stage progressive adaptation
MTG-LIRA (Hede et al., 2026)	OpenNMT / Eole Transformer	Supervised NMT Training		Cleaned parallel corpus + BPE	Standard beam search
BVSLP (Arora et al., 2026)	Encoder-Decoder Transformer	Supervised Fine-Tuning		Rule-based NER + KB-driven NET	Transliteration and entity preservation
LTRC_Plural (Vishnuraj et al., 2026)	NLLB-200 (1.3B), Sarvam (Gemma 3 4B)	QLoRA, rsLoRA, RL (RLOO)		Multilingual temperature sampling	BLEU / ChrF++ reward optimization
ANVITA (J et al., 2026)	T5-base (220M), IndicTrans2	SFT + Cycle-consistency loss		English paraphrasing, related-pair transfer	Zero-shot script transliteration
Winterpeg (Rittikar and Ramanna, 2026)	NLLB-200, Byte-level LM	SFT + Byte-level causal NCA		Character-level UTF-8 byte modeling	Routed hybrid architecture
KrenTransV0-1 (Nyalang et al., 2026)	IndicTrans2, NLLB-200, M2M-100	Supervised Fine-Tuning		Multi-model script partition	Script/resource-based routing
NLP-MT_Amrita (Sree and Gupta, 2026)	IndicTrans2, NLLB-600M, Sarvam	Training-free Agentic RAG		Dynamic Milvus vector embeddings	Multi-agent decoding + GRPO reward
DELAB-IIIITM (Oinam et al., 2026)	NLLB-200 (600M), IndicTrans2 (1.1B)	Full FT (NLLB) vs. LoRA (IndicTrans2)		In-domain parallel training	Comparative fine-tuning
JH_NLP_APASHreya2 (Sinha et al., 2026)	NLLB-200 Distilled (600M)	LoRA Fine-Tuning		Parameter-optimized adaptation	Constrained generation
星辰之力 (Ruan et al., 2026)	Qwen2.5-32B, Hunyuan-MT-7B	LoRA / Full FT + SFT + DPO		Word alignment "Helpful Vocabulary"	Constrained lexical guidance
AkibaNLP-TUT (Kachi et al., 2026)	Transformer / Multilingual NMT	Supervised Cross-Entropy		Word-level adaptive noise injection	Robust representation transfer
CIT MT BTECH 2026	NLLB-200 (600M), Scratch NMT	Pretrained FT vs. Scratch Training		Subword BPE segmentation	Pretraining transfer benchmark
SCE-KIIT (Maiti et al., 2026b,a)	NLLB-200 (600M), Llama-3.2-3B	QLoRA Fine-Tuning		10.8k Conversational QA transfer	Structured prompting + loop penalty
IndicMT	Sarvam-Translate (Gemma 3)	4-bit QLoRA Adaptation		NLLB-generated synthetic back-translation	Quantized parameter-efficient inference

Table 5: System architecture and methodological taxonomy of participating teams in the WMT 2026 Low-resource Indic Shared Task.

3. Morphological & Script Bottleneck Solutions Given the orthographic challenges presented by Meitei Mayek, Latin-based orthographies (Khasi, Mizo), and unstandardized contact languages (Nagamese), teams implemented novel script handling mechanisms:

- **Surrogate and Proxy Tagging:** For languages unsupported by default tokenizers (e.g., Kokborok, Bodo), teams utilized same-script surrogates or related-language proxy tags (e.g., hin_Deva) to leverage existing embedding representations (DoDS-IITPKD, BRS-NK).
- **Phylogenetic Mapping & Transliteration:** Algorithmic back-transliteration and phyloge-

netic script alignment enabled zero-shot transfer across related Indic orthographies (TIP-RBG-AI, ANVITA).

- **Byte-Level Language Modeling:** Winterpeg addressed subword vocabulary fragmentation by utilizing UTF-8 byte-level decoder models augmented with Neural Cellular Automata (NCA) front-ends.

4. Retrieval-Augmented Generation (RAG) and Hybrid Decoding Several teams demonstrated strong performance without parameter updates by pairing retrieval engines with structured prompting or constrained decoding:

- **In-Context Exemplar Retrieval:** Dynamic retrieval (BM25 or dense embeddings via

Milvus/multilingual-e5) allowed systems like gators and NLP-MT_Amrita to achieve top-tier performance on unseen vocabulary by supplying relevant translation pairs in-prompt.

- **Neuro-Symbolic & SMT Hybridization:** VoxNexus-Translate and 星辰之力 integrated Statistical Machine Translation (IBM Models/fast_align) word-alignment tables as hard decoding constraints or “Helpful Vocabulary” prompts, drastically curtailing hallucination rates in named entities and technical terminology.

6 Results and Discussion

The official evaluation results for all ten English–Indic language pairs across twenty bidirectional translation tracks are reported in Tables 6 through 25. Each table lists participating systems ranked by performance across six standardized lexical and neural evaluation metrics: BLEU, METEOR, TER, ChrF++, BERTScore, and COMET. In accordance with shared task evaluation protocols, system descriptions and discussions focus primarily on verified primary submissions, while contrastive systems and late submissions (highlighted ¶⁵) are included in the tabular benchmarks for completeness and empirical reference.

6.1 Discussion and Comparative Analysis

The empirical outcomes of the WMT 2026 shared task reveal significant heterogeneity across language families, resource tiers, and modeling paradigms. By evaluating systems simultaneously across exact string-matching metrics (BLEU), subword/character overlap (ChrF++), edit distance (TER), and deep cross-lingual embeddings (BERTScore, COMET), the evaluation benchmark exposes key dimensions of modern low-resource translation: (i) the stark divide between moderate and extreme data scarcity, (ii) asymmetric capacity between target-language generation and English decodability, (iii) the behavioral divergence between supervised parameter-efficient fine-tuning (PEFT) and training-free retrieval-augmented generation (RAG), and (iv) script-induced tokenization bottlenecks. The following subsections present a granular comparative analysis of these empirical dynamics.

⁵¶ Late submission indicates systems submitted after the deadline during the late submission period.

6.1.1 Data Regimes and the Resource Divide

The evaluation benchmark demonstrates an unambiguous performance cliff separating Category 1 (Moderate Training Data, $\geq 17k$ parallel sentences) from Category 2 (Very Limited Training Data, $\leq 15k$ down to 1k sentences), as illustrated in Figure 2.

In Category 1 tracks (Assamese, Mizo, Khasi, and Manipuri in both scripts), systems achieved consistent, high-utility translations, with peak English→Indic scores reaching 19.38 BLEU (Assamese, Table 6), 21.11 BLEU (Mizo, Table 8), and 29.30 BLEU (Khasi, Table 10). In these settings, foundational encoder-decoder models (e.g., IndicTrans2 and NLLB-200) successfully generalized across unseen test sets when adapted via parameter-efficient fine-tuning (PEFT).

In contrast, Category 2 languages (Kokborok, Karbi, and Tagin) experienced severe performance degradation. In English→Tagin (Table 24), the highest primary score was limited to 4.81 BLEU (ANVITA), while English→Karbi (Table 20) reached only 5.39 BLEU (gators). Similarly, Kokborok (Table 18) topped at 6.12 BLEU (KrenTransV0-1). Under such extreme data scarcity (1k–5k sentence pairs), supervised parameter updates failed to construct stable cross-lingual representations, leading to high out-of-vocabulary rates and degenerate decoder outputs.

6.1.2 Directional and Typological Asymmetry

A granular examination of the translation directions reveals an important typological distinction:

- **Agglutinative and Indic-Script Tracks:** For Tibeto-Burman languages characterized by complex suffixation (Manipuri and Bodo), Indic→English performance drastically outstripped the English→Indic direction. In Manipuri (Bengali script), the highest primary score into English reached 45.86 BLEU (gators, Table 13), compared to a maximum of only 11.95 BLEU in the reverse direction (Winterpeg, Table 12). A parallel gap occurred in Bodo (35.54 vs. 16.30 BLEU, Tables 17 and 16). This asymmetry demonstrates that pretrained decoders readily generate fluent English from morphologically rich source representations, but struggle to synthesize the intricate case markers, mood inflections, and agglutinative affix chains required on the Indic target side.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	19.38	42.89	63.89	52.34	86.95	83.17
VoxNexus-Translate_primary [¶]	19.23	42.62	60.50	51.78	86.51	83.31
KrenTransV0-1_primary	16.64	41.72	63.59	51.22	86.63	82.90
BVSLP_primary	16.44	38.10	66.71	48.64	85.44	83.23
ADI-KIIT_primary	15.57	39.91	65.52	50.06	86.30	82.36
Winterpeg_primary	15.35	37.99	65.77	48.32	85.64	80.96
EROL_primary	15.31	39.22	65.76	49.42	85.91	81.92
RozarNLP-cr7_primary	15.24	36.91	72.40	49.53	86.10	82.15
BRS-NK_primary	14.96	36.66	66.75	47.81	85.42	80.75
DoDS-IITPKD_primary	14.73	38.19	66.18	49.07	85.74	81.59
Decepticons-IIITH_primary	14.09	35.41	73.78	47.83	85.27	80.59
LTRC_Plural_primary	14.09	35.40	73.21	47.75	85.41	80.87
DELAB-IIITM_primary	13.72	34.79	73.67	47.56	85.55	80.84
MTG-LIRA_primary	12.17	33.74	71.46	45.05	84.51	78.57
CIT MT BTECH 2026_primary [¶]	0.34	5.91	115.43	15.67	71.95	52.00
Contrastive Systems						
gators_contrastive	20.77	44.71	60.06	52.81	86.98	83.27
BRS-NK_contrastive	17.53	40.53	63.22	50.77	86.35	82.18
RozarNLP-cr7_contrastive	15.99	40.84	64.43	50.49	86.43	82.54
DoDS-IITPKD_contrastive	15.08	39.13	65.48	49.77	86.05	81.95
Decepticons-IIITH_contrastive	15.00	36.44	72.29	48.98	85.83	81.65
DELAB-IIITM_contrastive	14.90	36.52	67.03	47.51	85.36	82.01
DoDS-IITPKD_contrastive2	14.73	38.20	66.16	49.07	85.74	81.60
MTG-LIRA_contrastive	14.44	36.58	66.79	47.04	85.32	79.44
LTRC_Plural_contrastive_1	14.01	34.75	69.34	46.63	84.87	81.60
BVSLP_contrastive [¶]	12.86	33.70	76.53	46.73	85.07	80.58
CIT MT BTECH 2026_contrastive1 [¶]	0.78	31.00	81.00	44.42	84.88	80.34
LTRC_Plural_contrastive_2	9.41	28.35	81.81	41.22	83.77	78.57

Table 6: Performance evaluation of primary and contrastive systems for English→Assamese translation across automatic evaluation metrics.

Note: [¶] Late submission indicates systems submitted after the deadline during the late submission period.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	27.26	60.50	64.78	57.43	95.09	86.53
EROL_primary	25.10	59.33	64.72	55.39	94.90	85.86
DoDS-IITPKD_primary	24.37	59.55	67.94	55.63	94.77	85.79
ADI-KIIT_primary	24.08	57.32	67.59	54.28	94.69	85.43
TIP-RBG-AI_primary	23.91	58.24	68.33	54.79	94.66	85.55
RozarNLP-cr7_primary	23.50	57.95	67.20	54.08	94.56	85.49
VoxNexus-Translate_primary [¶]	22.27	51.88	63.62	48.75	94.56	84.34
LTRC_Plural_primary	22.09	56.67	68.60	53.11	94.46	85.31
Winterpeg_primary	21.57	53.91	64.66	51.34	94.29	84.85
BRS-NK_primary	21.27	54.30	69.40	51.33	94.18	84.70
BVSLP_primary [¶]	20.67	54.38	72.70	51.58	94.19	84.74
MTG-LIRA_primary	15.04	43.10	80.33	43.31	92.41	78.77
Akiba-NLP_primary	11.01	39.80	99.69	39.18	90.28	71.81
Contrastive Systems						
gators_contrastive	27.27	60.51	64.77	57.35	95.07	86.48
RozarNLP-cr7_contrastive	25.19	58.89	67.97	55.47	94.72	85.58
ADI-KIIT_contrastive	25.11	58.06	67.09	54.87	94.71	85.51
DoDS-IITPKD_contrastive	24.91	59.70	67.77	55.92	94.81	85.88
DoDS-IITPKD_constrained	24.21	59.44	68.19	55.50	94.76	85.77
LTRC_Plural_contrastive_1	23.27	56.61	68.75	53.48	94.45	85.10
BRS-NK_contrastive	23.13	56.17	68.93	52.97	94.37	84.97
BVSLP_contrastive [¶]	19.74	53.09	80.81	50.71	93.25	82.26
LTRC_Plural_contrastive_2	18.99	53.05	84.35	50.80	93.22	81.98
Akiba-NLP_contrastive2	10.47	40.24	99.71	39.81	90.63	73.08
Akiba-NLP_contrastive1	6.06	30.03	122.98	31.98	88.17	64.46

Table 7: Performance evaluation for primary and contrastive systems on Assamese to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	21.11	47.51	65.94	50.49	81.97	66.77
Winterpeg_primary	19.66	45.78	66.41	47.74	81.06	63.86
Decepticons-IIITH_primary	19.53	46.11	71.37	48.49	81.72	62.45
RozarNLP-cr7_primary	19.33	46.08	67.11	47.93	81.24	64.68
KrenTransV0-1_primary	19.06	45.25	67.81	47.38	80.94	64.08
LTRC_Plural_primary	19.02	45.81	68.81	48.07	81.44	65.29
DoDS-IITPKD_primary	18.75	45.36	69.19	48.56	81.47	64.74
BVSLP_primary [¶]	16.29	42.60	74.20	43.98	79.92	63.26
MTG-LIRA_primary	16.20	39.35	74.95	42.06	78.96	61.21
Contrastive Systems						
LTRC_Plural_contrastive_1	22.85	49.16	63.32	51.02	82.39	65.35
MTG-LIRA_contrastive	22.72	48.94	64.00	51.10	82.02	65.00
gators_contrastive	21.60	47.72	65.31	50.69	81.94	66.83
Decepticons-IIITH_contrastive	19.41	46.23	73.81	48.79	81.74	63.45
DoDS-IITPKD_contrastive2	18.77	45.43	69.08	48.57	81.48	64.75
DoDS-IITPKD_contrastive	18.23	45.09	68.89	47.98	81.56	65.01
RozarNLP-cr7_contrastive	16.66	41.85	70.50	45.39	80.05	64.48
LTRC_Plural_contrastive_2	16.47	41.93	79.44	44.96	79.94	61.91
BVSLP_contrastive [¶]	15.08	42.72	87.24	45.35	80.06	64.12

Table 8: Performance evaluation for primary and contrastive systems on English to Mizo translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	33.46	68.35	55.18	63.36	95.80	81.62
RozarNLP-cr7_primary	31.40	65.18	56.15	60.08	95.35	79.94
TIP-RBG-AI_primary	31.05	64.53	57.51	60.00	95.37	79.67
Winterpeg_primary	30.53	63.20	52.32	58.96	95.29	79.34
LTRC_Plural_primary	28.07	62.31	61.86	58.49	94.98	79.05
DoDS-IITPKD_primary	27.65	63.24	62.39	59.15	95.01	79.43
BVSLP_primary [¶]	24.85	59.44	65.00	55.50	94.74	78.91
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	77.53
Contrastive Systems						
LTRC_Plural_contrastive_1	33.45	66.75	54.63	62.17	95.54	80.30
gators_contrastive	33.28	68.41	55.40	63.36	95.78	81.64
DoDS-IITPKD_contrastive2	31.60	65.29	55.72	60.80	95.72	81.53
DoDS-IITPKD_contrastive	30.18	62.60	57.38	58.74	95.27	79.37
RozarNLP-cr7_contrastive	27.59	60.89	61.34	56.87	94.81	77.54
LTRC_Plural_contrastive_2	20.35	56.23	84.08	53.32	93.44	74.27
BVSLP_contrastive [¶]	9.93	9.55	103.92	40.05	89.78	60.62

Table 9: Performance evaluation for primary and contrastive systems on Mizo to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
KrenTransV0-1_primary	29.30	54.21	50.44	53.01	83.89	66.98
gators_primary	28.66	54.18	51.71	52.96	83.82	68.28
DoDS-IITPKD_primary	20.85	44.57	63.08	44.04	79.54	65.58
LTRC_Plural_primary	14.47	36.23	85.95	36.28	76.95	62.51
ANVITA_primary	10.08	28.94	77.15	30.46	74.13	57.43
RozarNLP-cr7_primary	9.04	32.34	144.02	29.25	72.90	55.24
Winterpeg_primary	7.40	25.68	84.59	30.62	73.60	61.71
IndicMT_primary	0.00	7.91	311.41	5.44	64.98	49.57
Contrastive Systems						
ANVITA_contrastive_2	32.89	56.75	48.53	54.84	84.51	69.69
ANVITA_contrastive	32.46	57.50	48.65	55.54	84.66	69.76
LTRC_Plural_contrastive_1	32.46	57.16	48.70	54.77	84.44	68.06
gators_contrastive	29.14	54.42	51.54	53.16	83.95	68.34
DoDS-IITPKD_contrastive2	21.05	45.03	61.27	44.46	79.81	65.85
LTRC_Plural_contrastive_2	20.80	44.14	68.63	43.35	79.66	61.68
DoDS-IITPKD_contrastive	18.99	44.27	58.60	43.30	80.53	62.31
BVSLP_contrastive [¶]	15.69	36.73	76.82	37.16	76.12	59.16

Table 10: Performance evaluation for primary and contrastive systems on English to Khasi translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	26.28	55.99	64.75	53.15	94.72	77.04
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	77.03
DoDS-IITPKD_primary	16.99	40.92	82.49	41.48	91.91	65.66
TIP-RBG-AI_primary	11.17	36.32	98.74	37.65	91.32	66.23
LTRC_Plural_primary	7.94	28.75	124.56	31.03	89.03	56.03
ANVITA_primary	5.17	22.98	119.93	25.78	88.01	50.70
RozarNLP-cr7_primary	4.54	25.05	189.51	25.75	87.84	51.38
Winterpeg_primary	0.51	10.47	112.93	18.14	84.22	40.72
IndicMT_primary	0.00	4.21	179.90	12.08	80.90	35.42
Contrastive Systems						
ANVITA_contrastive	26.48	55.26	62.48	52.31	94.55	75.62
gators_contrastive	25.76	55.81	65.66	52.92	94.69	76.85
DoDS-IITPKD_contrastive2	24.63	55.01	66.92	51.83	94.55	76.93
LTRC_Plural_contrastive_1	23.72	52.53	67.45	49.83	94.03	74.07
ANVITA_contrastive_2	22.90	52.63	64.05	49.91	93.96	74.15
DoDS-IITPKD_contrastive	19.14	43.25	77.40	43.29	92.63	68.15
LTRC_Plural_contrastive_2	11.82	35.74	111.46	36.81	90.26	59.78
BVSLP_contrastive [¶]	8.68	30.84	130.64	32.05	89.26	55.42

Table 11: Performance evaluation for primary and contrastive systems on Khasi to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
Winterpeg_primary	11.95	25.67	74.93	44.67	87.12	70.36
MTG-LIRA_primary	10.61	17.58	86.37	37.24	82.41	67.00
NLP-MT_Amrita_Primary_IndicTrans2	10.29	17.38	88.27	38.88	84.57	68.01
EROL_primary	10.21	18.40	85.13	38.83	84.42	67.04
NLP-MT_Amrita_Primary_NLLB	9.65	18.50	85.03	38.67	83.98	67.44
BVSLP_primary [¶]	9.18	16.71	89.27	35.99	82.97	66.76
DoDS-IITPKD_primary	8.65	18.43	86.47	39.63	84.67	67.61
KrenTransV0-1_primary	8.16	17.42	105.67	37.92	83.98	67.19
gators_primary	8.04	17.79	90.64	39.43	84.64	67.15
JH_NLP_APAShreya2_primary	8.03	20.15	82.56	40.93	84.86	68.18
PANINI_primary	7.48	20.02	83.00	40.03	84.48	67.71
DELAB-IITM_primary	6.61	20.72	81.40	41.16	85.03	68.18
IndicMT_primary	0.00	0.36	117.33	7.89	67.71	48.23
RozarNLP-cr7_primary	0.00	0.17	100.00	0.06	3.56	31.03
Contrastive Systems						
RozarNLP-cr7_contrastive	10.18	17.81	86.81	39.03	84.26	67.60
DELAB-IITM_contrastive	9.75	18.72	82.59	38.88	84.01	67.61
DoDS-IITPKD_contrastive2	8.66	18.43	84.51	39.75	84.72	67.65
gators_contrastive	8.15	18.28	89.62	39.89	85.00	67.86
BVSLP_contrastive [¶]	7.21	19.17	84.36	39.60	84.54	67.85
DoDS-IITPKD_contrastive	5.23	18.45	86.74	39.60	84.72	67.78

Table 12: Performance evaluation for primary and contrastive systems on English to Manipuri (Bengali script) translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	45.86	77.29	41.23	72.44	96.84	85.93
DoDS-IITPKD_primary	38.84	73.61	47.90	68.52	95.69	82.74
Winterpeg_primary	37.81	70.27	44.56	65.55	95.77	82.68
TIP-RBG-AI_primary	36.39	70.30	50.15	65.70	95.67	82.44
PANINI_primary	35.79	71.98	49.72	66.08	95.74	82.90
RozarNLP-cr7_primary	33.52	70.51	50.03	63.81	95.38	81.69
NLP-MT_Amrta_Primary_NLLB	32.36	67.01	54.87	62.68	95.04	80.51
BVSLP_primary [¶]	26.56	62.47	63.97	57.79	94.69	80.74
NLP-MT_Amrta_Primary_Sarvam	23.61	59.02	68.42	55.59	93.39	75.02
EROL_primary	22.47	56.60	69.41	53.04	94.27	80.15
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	79.97
Akiba-NLP_primary	0.51	2.54	292.36	18.09	83.19	47.14
NLP-MT_Amrta_Primary_IndicTrans2	0.28	8.81	223.41	16.21	84.31	47.85
IndicMT_primary	0.10	3.77	177.57	12.08	82.32	41.94
Contrastive Systems						
gators_contrastive	45.93	77.32	40.27	72.37	96.85	85.97
DoDS-IITPKD_contrastive	40.84	74.57	45.79	69.49	96.33	84.09
DoDS-IITPKD_constrained	37.75	73.49	49.01	68.00	95.75	82.77
BVSLP_contrastive [¶]	30.59	68.09	59.01	63.12	94.95	81.14
RozarNLP-cr7_contrastive	22.64	61.99	71.33	59.07	92.69	72.13

Table 13: Performance evaluation for primary and contrastive systems on Manipuri (Bengali script) to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	16.24	38.02	68.90	52.65	98.50	91.28
MTG-LIRA_primary	12.63	30.38	73.03	46.00	97.15	90.85
PANINI_primary	9.60	27.55	81.18	45.87	98.85	92.31
DELAB-IIITM_primary	5.10	20.04	86.42	38.36	97.67	88.58
EROL_primary	3.96	19.81	118.02	41.39	96.76	84.75
BVSLP_primary [¶]	3.82	12.37	102.30	27.68	96.23	87.54
KrenTransV0-1_primary	3.56	18.86	122.00	38.99	96.81	85.13
RozarNLP-cr7_primary	3.46	17.98	119.20	38.60	96.67	84.57
Winterpeg_primary	0.00	2.27	101.91	12.49	94.16	77.39
Contrastive Systems						
gators_contrastive-Mtei	16.62	38.35	68.95	52.85	98.53	91.34
DELAB-IIITM_contrastive	3.55	18.94	121.96	39.02	96.81	85.09

Table 14: Performance evaluation for primary and contrastive systems on English to Manipuri (Meitei Mayek script) translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	38.61	70.96	47.73	66.38	96.25	81.16
TIP-RBG-AI_primary	37.53	70.38	49.70	65.91	95.93	79.99
PANINI_primary	36.56	70.29	49.12	65.12	96.02	80.20
EROL_primary	24.71	58.97	65.96	55.00	94.69	77.82
BVSLP_primary [¶]	24.61	57.99	65.78	54.60	94.48	76.37
MTG-LIRA_primary	24.43	56.46	67.35	53.66	93.46	71.73
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	76.26
RozarNLP-cr7_primary	0.13	6.92	137.67	13.42	84.65	42.06
Winterpeg_primary	0.00	9.75	122.02	17.33	83.49	40.19
Contrastive Systems						
gators_contrastive	38.13	70.68	48.48	66.16	96.28	81.00
RozarNLP-cr7_contrastive	33.75	67.20	53.67	62.93	95.45	78.44
BVSLP_meit_to_en_contrastive [¶]	15.30	48.61	88.37	46.69	91.40	66.20

Table 15: Performance evaluation for primary and contrastive systems on Manipuri (Meitei Mayek script) to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
Decepticons-IIITH_primary	16.30	37.56	70.20	50.77	85.25	62.38
DoDS-IITPKD_primary	16.07	36.23	72.30	49.88	84.95	62.30
MTG-LIRA_primary	15.91	34.92	71.25	48.32	84.13	61.75
gators_primary	14.87	35.29	73.04	49.06	84.58	61.72
EROL_primary	14.25	35.31	75.40	48.68	84.47	61.38
NLP-MT_Amrita_Primary_IndicTrans2	14.01	34.40	76.09	48.00	84.32	61.58
BRS-NK_primary	12.00	28.74	77.71	43.67	82.90	59.94
DELAB-IIITM_primary	9.74	27.11	86.88	42.03	80.86	58.19
KrenTransV0-1_primary	9.20	30.78	79.10	44.06	82.70	59.64
NLP-MT_Amrita_Primary_NLLB	0.00	0.73	114.63	12.39	68.95	47.73
Contrastive Systems						
BRS-NK_contrastive	17.34	37.95	68.50	50.98	85.10	62.79
DoDS-IITPKD_contrastive	16.78	36.80	72.54	50.26	84.91	62.22
DoDS-IITPKD_contrastive2	16.33	36.52	71.95	49.95	84.95	62.33
MTG-LIRA_contrastive	15.87	34.63	71.02	48.11	84.24	61.72
Decepticons-IIITH_contrastive	15.64	36.18	71.19	49.56	84.78	62.05
gators_contrastive	14.61	34.43	72.19	48.21	84.38	61.86
DELAB-IIITM_contrastive	12.63	34.21	81.32	47.30	84.03	61.27
星辰之力_contrastive2	12.46	31.06	75.49	45.12	83.15	60.27
星辰之力_contrastive1	12.12	30.74	74.20	44.98	83.15	60.29
BVSLP_en_to_bod_contrastive [¶]	5.20	12.73	95.77	17.52	66.91	39.58

Table 16: Performance evaluation for primary and contrastive systems on English to Bodo translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	35.54	68.26	49.91	63.04	95.82	84.23
TIP-RBG-AI_primary	32.15	68.04	55.43	62.23	95.26	82.62
DoDS-IITPKD_primary	31.17	68.24	52.78	60.23	95.00	81.22
EROL_primary	23.50	58.14	66.75	54.02	94.55	81.87
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	81.08
BRS-NK_primary	19.39	52.43	67.71	48.12	92.83	74.86
Akiba-NLP_primary	12.19	43.21	90.33	40.38	89.91	65.16
NLP-MT_Amrita_Primary_NLLB	3.04	18.55	116.70	24.08	81.63	42.99
NLP-MT_Amrita_Primary_Sarvam	2.10	11.11	105.66	16.71	78.50	35.83
NLP-MT_Amrita_Primary_IndicTrans2	0.45	8.45	106.92	16.15	85.37	51.28
Contrastive Systems						
DoDS-IITPKD_contrastive	36.10	71.51	47.47	63.93	95.88	83.78
gators_contrastive	34.16	67.57	51.48	62.19	95.68	83.79
DoDS-IITPKD_constrained	31.41	67.26	52.90	60.00	95.02	81.39
BRS-NK_contrastive	24.29	60.62	62.26	54.23	94.09	78.89
BVSLP_contrastive [¶]	20.56	55.64	67.96	50.81	93.36	76.79
Akiba-NLP_contrastive	11.41	12.58	135.09	17.87	82.76	43.43

Table 17: Performance evaluation for primary and contrastive systems on Bodo to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
KrenTransV0-1_primary	6.12	26.17	80.58	34.91	74.73	60.84
gators_primary	5.88	23.43	99.99	33.90	73.94	60.96
SCE-KIIT_primary	5.11	21.04	98.67	27.09	71.17	54.35
DoDS-IITPKD_primary	3.86	17.44	95.34	27.33	71.19	58.99
Contrastive Systems						
DoDS-IITPKD_contrastive	7.36	28.40	81.52	35.04	75.14	60.72
DoDS-IITPKD_contrastive2	7.35	28.37	81.55	35.03	75.14	60.73
星辰之力_contrastive1	6.51	24.55	85.91	31.32	73.82	58.92
gators_contrastive	6.02	23.29	98.36	33.70	73.90	60.93
星辰之力_contrastive2	3.44	15.86	103.06	26.44	70.76	57.11
SCE-KIIT_contrastive	1.98	11.94	108.65	20.32	68.62	49.89

Table 18: Performance evaluation for primary and contrastive systems on English to Kokborok translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
VoxNexus-Translate_primary [¶]	22.43	52.01	63.42	48.89	94.58	78.25
gators_primary	16.67	41.26	84.06	41.58	92.04	68.81
SCE-KIIT_primary	3.58	23.59	200.86	26.14	86.99	53.49
TIP-RBG-AI_primary	3.35	17.69	113.33	23.17	81.44	40.78
DoDS-IITPKD_primary	3.04	17.32	151.93	22.78	84.99	44.42
Contrastive Systems						
DoDS-IITPKD_contrastive2	20.28	50.42	75.11	48.47	93.67	76.25
gators_contrastive	16.67	41.26	84.06	41.58	92.04	68.81
DoDS-IITPKD_contrastive1	12.08	35.70	94.12	36.65	90.86	65.47
SCE-KIIT_contrastive	1.45	16.72	137.37	22.13	87.00	53.31

Table 19: Performance evaluation for primary and contrastive systems on Kokborok to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
gators_primary	5.39	26.95	114.02	33.38	72.14	57.76
KrenTransV0-1_primary	3.47	15.03	92.41	22.06	62.44	40.45
Contrastive Systems						
gators_contrastive	5.74	27.24	107.55	33.85	72.17	57.94
星辰之力_contrastive1	3.03	19.57	93.37	28.45	71.00	56.49
星辰之力_contrastive2	2.99	18.38	94.11	27.30	70.36	55.79

Table 20: Performance evaluation for primary and contrastive systems on English to Karbi translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
VoxNexus-Translate_primary [¶]	22.42	51.86	63.49	48.80	94.55	76.95
gators_primary	9.62	24.26	92.01	28.93	89.21	57.54
TIP-RBG-AI_primary	3.32	15.92	129.90	23.01	84.24	42.94
Contrastive Systems						
gators_contrastive	9.62	24.26	92.01	28.93	89.21	57.54

Table 21: Performance evaluation for primary and contrastive systems on Karbi to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
Decepticons-IIITH_primary	20.34	59.95	76.40	57.66	82.28	69.55
gators_primary	17.44	56.50	84.77	54.93	81.05	66.34
ANVITA_primary	16.37	52.05	81.65	51.81	80.25	65.16
SCE-KIIT_primary	12.51	42.54	91.82	43.81	77.45	59.04
KrenTransV0-1_primary	0.67	10.95	108.24	16.40	67.56	44.16
Contrastive Systems						
gators_contrastive	16.62	55.61	86.83	54.14	80.66	65.13
ANVITA_contrastive_2	16.14	50.13	82.28	50.22	79.63	64.14
ANVITA_contrastive	15.88	50.25	83.73	50.48	79.66	63.81
星辰之力_contrastive2	12.90	47.30	97.04	47.31	78.81	62.58
星辰之力_contrastive1	12.35	46.94	98.15	47.51	78.43	62.04
Decepticons-IIITH_contrastive	12.01	47.84	105.06	47.78	76.48	62.47

Table 22: Performance evaluation for primary and contrastive systems on English to Nagamese translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
VoxNexus-Translate_primary [¶]	21.77	47.26	65.51	46.35	93.88	77.44
gators_primary	19.80	45.06	67.43	44.88	93.63	77.25
ANVITA_primary	10.91	37.91	80.09	36.10	90.96	66.47
BVSLP_primary [¶]	8.18	30.68	89.11	32.13	88.95	58.93
TIP-RBG-AI_primary	5.61	29.55	88.46	32.06	86.86	56.64
Contrastive Systems						
gators_contrastive	19.48	45.03	67.93	44.66	93.61	77.03
ANVITA_contrastive_2	13.82	39.43	73.78	39.37	92.52	72.95
ANVITA_contrastive	14.72	40.78	72.56	40.48	92.70	73.57

Table 23: Performance evaluation for primary and contrastive systems on Nagamese to English translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
ANVITA_primary	4.81	15.04	92.58	22.39	70.49	52.61
gators_primary	2.41	7.99	220.26	20.48	69.02	52.15
LTRC_Plural_primary	2.25	15.46	195.42	19.17	67.91	51.52
KrenTransV0-1_primary	0.00	3.73	98.38	8.98	68.57	47.32
Contrastive Systems						
星辰之力_contrastive1	4.84	16.77	95.42	23.95	71.77	56.45
星辰之力_contrastive2	4.75	16.67	96.85	23.85	71.68	56.42
LTRC_Plural_contrastive_2	3.88	20.91	44.21	26.14	70.83	54.86
ANVITA_contrastive	3.71	20.12	99.11	28.00	72.76	58.65
gators_contrastive	3.31	19.87	162.83	24.13	71.01	55.51
LTRC_Plural_contrastive_1	1.65	13.63	176.50	21.76	69.46	53.37

Table 24: Performance evaluation for primary and contrastive systems on the English to Tagin translation task.

System	BLEU↑	METEOR↑	TER↓	ChrF++↑	BERTScore↑	COMET↑
Primary Systems						
VoxNexus-Translate_primary	22.38	51.61	63.59	48.69	94.53	78.19
gators_primary	3.89	14.64	97.48	17.46	86.84	48.57
LTRC_Plural_primary	2.62	12.33	126.16	16.91	86.25	47.39
ANVITA_primary	2.43	12.59	100.71	16.72	86.52	47.01
TIP-RBG-AI_primary	1.63	7.89	111.07	12.86	79.63	33.16
Contrastive Systems						
gators_contrastive	4.01	15.12	94.81	17.73	86.88	49.33
ANVITA_contrastive	3.52	15.46	91.48	17.96	87.10	49.56
LTRC_Plural_contrastive_2	2.28	11.87	126.74	17.64	85.79	47.26
LTRC_Plural_contrastive_1	1.30	10.65	121.78	16.70	85.34	46.28

Table 25: Performance evaluation for primary and contrastive systems on Tagin to English translation task.

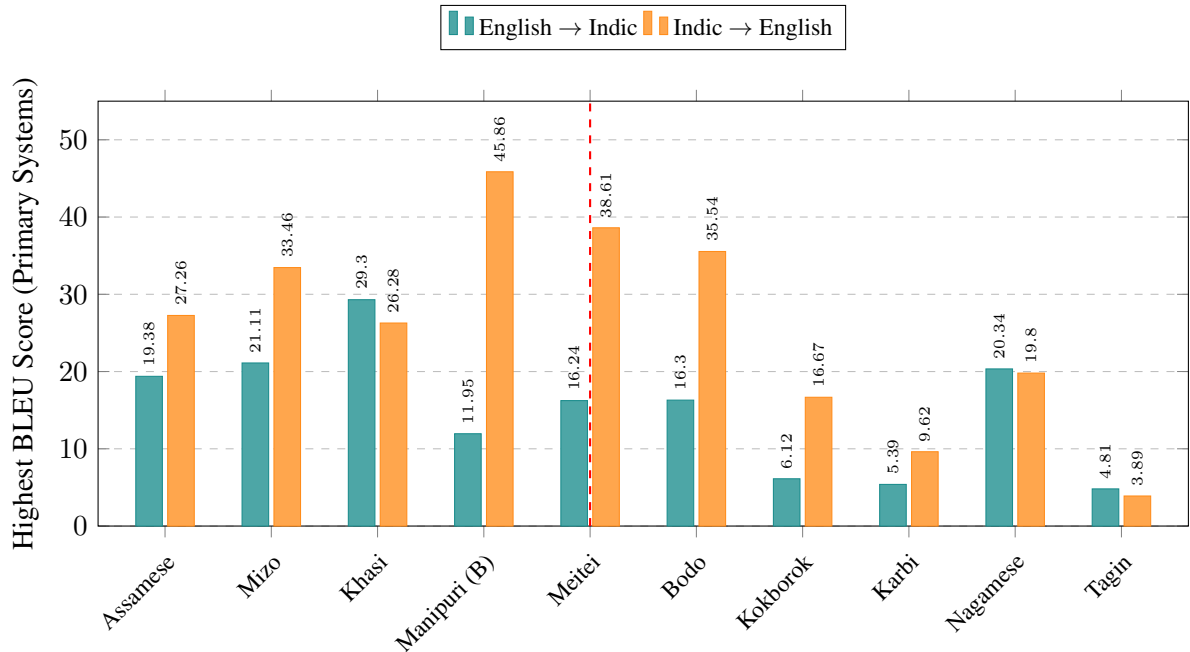


Figure 2: Maximum BLEU scores achieved by primary systems across all ten language pairs. The vertical dashed line separates Category 1 (Moderate Data, left) from Category 2 (Very Limited Data, right).

- **Latin-Script and Lingua Franca Tracks:** Conversely, languages employing Latin orthography (Khasi) or functioning as regional contact Creoles (Nagamese) exhibited near-directional parity or reversed asym-

metry. English→Khasi achieved 29.30 BLEU compared to 26.28 BLEU for Khasi→English (Tables 10 and 11). Similarly, English→Nagamese reached 20.34 BLEU (Decepticons-IIIITH, Table 22) against 19.80

BLEU for Nagamese→English (gators, Table 23). Because Khasi follows an analytic Subject-Verb-Object (SVO) syntax and Nagamese employs simplified contact morphology derived from Assamese, target generation suffered far less from morphological compounding errors.

6.1.3 Metric Divergence: Exact Match vs. Semantic Preservation

The evaluation results highlight a marked divergence between surface-level n-gram metrics (BLEU, ChrF++) and neural semantic representations (COMET, BERTScore). Figure 3 illustrates the correlation between BLEU and COMET for primary submissions in the Manipuri (Bengali script)→English task.

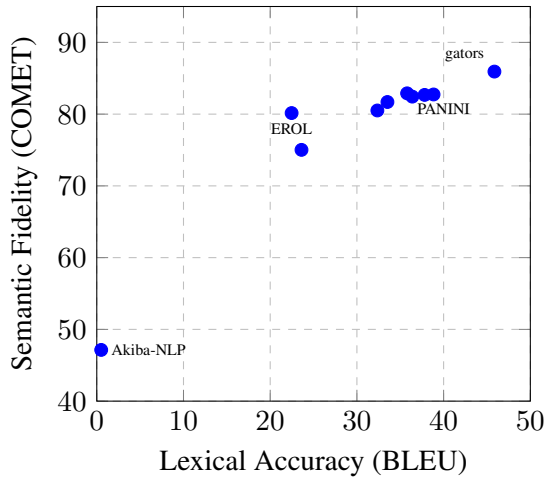


Figure 3: Lexical Overlap (BLEU) versus Semantic Fidelity (COMET) for primary submissions on the Manipuri (Bengali script)→English benchmark.

While frontier LLM-based prompting (**gators**) achieved top rankings on both axes (45.86 BLEU, 85.93 COMET), the mid-tier cluster formed by parameter-efficient transformer models (**PANINI**, **TIP-RBG-AI**, **DoDS-IITPKD**, and **EROL**) exhibits an important trend. Despite a lexical drop of up to 23.39 BLEU points relative to the top system, these models maintained high COMET scores (> 80.0). This demonstrates that adapted sequence-to-sequence models preserve core cross-lingual semantics and contextual discourse even when generating lexical variations or synonymous phrasings penalized by rigid n-gram matching.

Conversely, string edit-distance analysis revealed acute degeneration patterns in under-resourced tracks. In English→Tagin (Table 24) and Khasi→English (Table 11), several unconstrained

generative submissions exhibited Translation Edit Rates (TER) exceeding 120.0, peaking at 311.41 (IndicMT). Inspection indicates that when tokenizers encounter unseen vocabularies in extreme low-resource regimes, autoregressive decoders enter repetitive token loops or hallucinate ungrounded text sequences. This finding demonstrates the necessity of pairing semantic evaluation with length-sensitive edit penalties (TER). While embedding-based metrics like COMET may mask structural collapse by generously scoring topically relevant vocabulary, severe TER spikes act as a critical diagnostic tool to expose catastrophic over-generation and repetitive token loops.

6.1.4 Architectural Paradigms and Methodological Insights

The variance across task tracks reveals structural trade-offs between general massively multilingual pretraining and region-specific architectures. As visualized in Figure 4, we extracted the exact COMET scores from primary submissions utilizing strictly **NLLB-200** (e.g., **BRS-NK**, **RozarNLP-cr7**) versus strictly **IndicTrans2** (e.g., **PANINI**, **ADI-KIIT**) to evaluate foundation robustness.

- **IndicTrans2 vs. NLLB-200 Pretraining:** **IndicTrans2** demonstrated a notably higher median COMET score (74.11 vs. 68.18) and a tighter interquartile range (IQR 21.32 vs. 30.31) compared to **NLLB-200**. It benefited significantly from pre-allocated subword capacity tailored to South Asian linguistic typologies. In contrast, **NLLB-200** exhibited wider variance across Category 2 languages, performing strongly when language tags aligned, but suffering representation collapse when encountering unsupported vocabularies.
- **Retrieval-Augmentation vs. Supervised Fine-Tuning:** In-context exemplar retrieval paired with frontier LLMs (**gators**, **NLP-MT_Amrita**) proved exceptionally competitive in Category 2 tracks. Dynamic exemplar injection bypasses parameter updates entirely, effectively grounding the LLM decoder without succumbing to the catastrophic over-fitting seen in supervised fine-tuning under small corpus regimes.
- **Neuro-Symbolic and Lexical Alignment Constraints:** Teams integrating statistical

phrase tables and word mappings into generative decoding (**VoxNexus-Translate**, 星辰之力) substantially constrained hallucination rates. This hybrid approach ensured higher lexical precision on named entities, numeric expressions, and specialized domain vocabulary.

- **Tokenization-Free and Script-Agnostic Frontiers:** For orthographies unsupported by standard subword vocabularies (e.g., Meitei Mayek), byte-level language modeling (**Winterpeg**) and phylogenetic zero-shot transliteration (**TIP-RBG-AI**) provided promising alternatives, preventing subword fragmentation and enabling cross-script transfer.

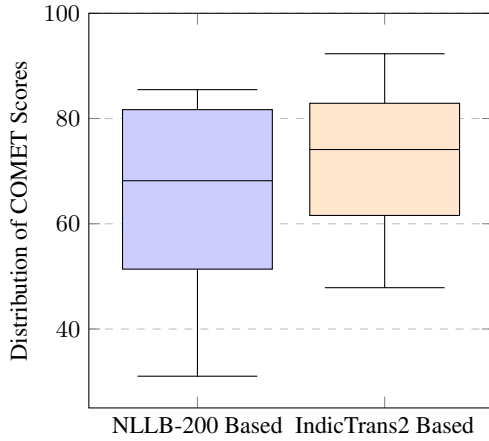


Figure 4: Distribution of COMET scores across shared task tracks grouped by foundational pretrained architecture. Metrics are drawn from primary submissions utilizing strictly NLLB-200 or IndicTrans2 frameworks.

6.2 Linguistic Bottlenecks and Future Directions

Despite major advances in parameter-efficient adaptation, agentic retrieval, and hybrid alignment, several fundamental linguistic challenges remain unsolved:

1. **Subword Fragmentation in Agglutinative Languages:** Fixed-vocabulary BPE tokenizers continue to fragment morphologically rich roots and suffixes in Bodo, Kokborok, and Manipuri, compounding decoding errors across multi-morpheme tokens.
2. **Orthographic and Tonal Ambiguity:** Tonal systems in languages such as Mizo remain

unrepresented in standard orthography, introducing semantic ambiguities that cannot be resolved without broader discourse-level modeling.

3. Contact-Language Code-Mixing:

Nagamese exhibits high dialectal variation and unsystematized phonetic orthography, degrading standard sequence-to-sequence translation pipelines.

Addressing these persistent bottlenecks will require dedicated regional pretraining corpora, linguistically informed morphological segmenters, and evaluation metrics tailored to the phonological and syntactic properties of Northeast Indian languages.

7 Conclusion

This paper presented the findings of the WMT 2026 Low-resource Indic Languages Translation Shared Task. Building upon previous editions, we expanded the evaluation benchmark to ten English-Indic language pairs by introducing Meitei, Karbi, Nagamese, and Tagin alongside Assamese, Mizo, Khasi, Manipuri, Bodo, and Kokborok. The task utilized the enriched INDICNE-CORP 2.0 dataset and introduced newly curated, bidirectional test sets across five distinct domains (Entertainment, Sports, Healthcare, Travel, and Political) to ensure robust evaluation across diverse communicative contexts. Systems submitted by 24 participating teams were evaluated using a complementary suite of lexical (BLEU, METEOR, TER, ChrF++) and semantic (BERTScore, COMET) metrics.

Based on the comparative evaluation of the submitted systems, we report the following principal findings:

- A stark performance dichotomy persists between moderately resourced (Category 1) and extremely low-resource (Category 2) languages. While parameter-efficient fine-tuning (PEFT) on multilingual baselines yielded competitive results for Category 1, the same architectures suffered precipitous performance drops on Category 2 languages, frequently failing to surpass 10 BLEU points. This indicates that low-resource transfer learning requires a critical mass of parallel anchors to be effective.

- Translation quality is consistently higher in the Indic→English direction than in the English→Indic direction. Multilingual models successfully map morphologically rich Indic source text into English, but struggle to generate the complex agglutinative suffixes, tonal markers, and rich morphological inflections required on the Indic target side.
- The results demonstrate a clear architectural divergence based on data availability. While PEFT (e.g., rsLoRA, DoRA) remains the dominant strategy for moderately resourced languages, dynamic in-context retrieval and LLM prompting achieved superior semantic fidelity in data-scarce settings without requiring parameter updates.
- Teams that constrained decoding using neuro-symbolic approaches, such as integrating SMT word alignments, successfully reduced entity hallucination rates. Furthermore, byte-level modeling and phylogenetic script mapping emerged as viable strategies to process unstandardized orthographies and bypass the subword tokenizer bottleneck entirely.

Despite continued progress in multilingual pre-training and hybrid decoding, translating morphologically complex and underrepresented Indic languages remains a challenging open problem. We expect that the standardized training corpora, multi-domain test sets, and comprehensive evaluation framework released through this shared task will facilitate targeted future research in low-resource machine translation.

References

- Aditya and Asif Ekbal. 2026. ADI-KIIT: Reward-guided preference optimization for low-resource english–assamese machine translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Palak Arora, Adhikarimayum Meerajita Sharma, Mrityunjaya Indoria, Keneiwenuo Lhoungu, Lalthafamkimi, and Nisheeth Joshi. 2026. BVSLP — enhancing low-resource neural machine translation with nmt and llm-assisted synthetic data augmentation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Satanjeev Banerjee and Alon Lavie. 2005. **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*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.
- Paul K. Benedict. 1972. *Sino-Tibetan: A Conspectus*. Cambridge University Press.
- Pramod Chandra Bhattacharya. 1977. *A Descriptive Analysis of Boro Language*. Gauhati University Press, Guwahati, Assam, India.
- Anne Daladier. 2002. Definiteness in kokborok. *Recherches Linguistiques de Vincennes*, 31:61–78.
- Laishram Thoibisana Devi and Grace Lee. 2026. EROL: Transformer-based multilingual translation for north-eastern indian languages pair. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Aashish Dhawan, Christopher William Driggers-Ellis, Dzmitry Y Kasinets, Christan Earl Grant, and Daisy Zhe Wang. 2026. gators: Bm25-augmented many-shot translation for low-resource north-eastern indian languages. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- George A. Grierson. 1903. *Linguistic Survey of India, Volume III, Part II*. Government of India.
- George A. Grierson. 1904. *Linguistic Survey of India: Vol. II, Mon-Khmer and Siamese-Chinese Families (Including Khasi and Tai)*. Office of the Superintendent of Government Printing, Calcutta, India.
- Barathi Ganesh HB, Michal Ptaszynski, Meenakshi, and Jairam R. 2026. TIP-RBG-AI: Overcoming orthographic discrepancies via algorithmic translinear pipelines and phylogenetic script mapping. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Vedarth Hede, Neha Gupta, Harish Bapat, and Harsh Ekbote. 2026. MTG-LIRA: Scaling neural machine translation for low-resource indian languages through multilingual learning and synthetic data. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Eugénie J. A. Henderson. 1976. Vestiges of morphology in modern standard khasi. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta, editors, *Austroasiatic Studies, Part I*, pages 481–522. The University Press of Hawaii, Honolulu, HI.
- F. Hoque. 2014. Kokborok: A major tribal language of tripura. *IOSR Journal of Humanities and Social Science*, 19(10):30–35.

- Rudali Huidrom, Vikas Kumar, Hoomexsun Pangsatbam, Kabi Konthoujam, Nongmaithem Zeno, Nicholas Khumukcham, Mangalton Okram, Justice Konjengbam, Sai Varun Jamalpoor, Alex D Nelson Khangembam, Anand Konjengbam, and Vikram Goyal. 2026. PANINI: Improving low-resource english $\leftrightarrow$ manipuri translation in two scripts via rslora and three-stream self-augmentation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Siva Bhavani J, Daneshwari kankanwadi, Abhinav G, and Biswajit Paul. 2026. ANVITA: Machine translation system for low-resource indian languages - khasi, nagamese, and tagin. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- François Jacquesson. 2003. Kokborok: A short analysis. In *Hukumu: 10th Anniversary Volume*, pages 109–122. Kokborok Tei Hukumu Mission.
- François Jacquesson. 2008. *A Kokborok Grammar: Agartala Dialect*. Kokborok Tei Hukumu Mission, Agartala, India.
- Takumi Kachi, Tomoyoshi Akiba, and Hajime Tsukada. 2026. AkibaNLP-TUT: Improving low-resource machine translation via language-specific and length-adaptive noise injection. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Pushpa P. Karapurkar. 1972. *Tripuri Phonetic Reader*. Central Institute of Indian Languages, Mysore.
- Pushpa P. Karapurkar. 1979. *Kokborok Grammar*. Central Institute of Indian Languages, Mysore.
- Lenin Laitonjam and Sanasam Ranbir Singh. 2021. Manipuri-english machine translation using comparable corpus. In *Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021)*, pages 78–88.
- Lenin Laitonjam and Sanasam Ranbir Singh. 2023. Manipuri-english comparable corpus for cross-lingual studies. *Language Resources and Evaluation*, 57(1):377–413.
- Agniva Maiti, Aarsh Muley, and Sovan Kumar Sahoo. 2026a. SCE-KIIT: Fine-tuning nllb-200 for english-nagamese translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Agniva Maiti, Aarsh Muley, and Sovan Kumar Sahoo. 2026b. SCE-KIIT: Kokllama: Cross-task adaptation for low-resource english-kokborok translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- K. S. Nagaraja, Paul Sidwell, and Simon Greenhill. 2013. A lexicostatistical study of the khasian languages: Khasi, pnar, lyngngam, and war. *Mon-Khmer Studies*, 42:1–11.
- Badal Nyalang, Riya Tekcham, and Biman Debbarma. 2026. KrenTransV0-1: Multi-model routing and reranking for low-resource northeast indian translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Office of the Registrar General and Census Commissioner of India. 2011. Census of india 2011: Language tables.
- Dingku Singh Oinam, Kishorjit Nongmeikapam, and Navanath Saharia. 2026. DELAB-IIITM wmt26: Full fine-tuning and lora-based adaptation for english-to-low-resource indic translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Partha Pakray, Reddi Mohana Krishna, Santanu Pal, Advaita Vetagiri, Sandeep Kumar Dash, Arnab Kumar Maji, Saralin A. Lyngdoh, Lenin Laitonjam, Anupam Jamatia, Koj Sambyo, Ajit Das, and Riyanka Manna. 2025. Findings of the WMT 2025 shared task on low-resource Indic languages translation. In *Proceedings of the Tenth Conference on Machine Translation*, Suzhou, China. Association for Computational Linguistics.
- Partha Pakray, Santanu Pal, Advaita Vetagiri, Reddi Krishna, Arnab Kumar Maji, Sandeep Dash, Lenin Laitonjam, Lyngdoh Sarah, and Riyanka Manna. 2024. [Findings of WMT 2024 shared task on low-resource Indic languages translation](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 654–668, Miami, Florida, USA. Association for Computational Linguistics.
- Santanu Pal, Partha Pakray, Sahinur Rahman Laskar, Lenin Laitonjam, Vanlalmuansangi Khenglawt, Sunita Warjri, Pankaj Kundan Dadure, and Sandeep Kumar Dash. 2023. [Findings of the WMT 2023 shared task on low-resource Indic language translation](#). In *Proceedings of the Eighth Conference on Machine Translation*, pages 682–694, Singapore. 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 Popović. 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.

- Lili Rabel. 1961. *Khasi, a Language of Assam*. Louisiana State University Press, Baton Rouge, LA.
- Soujanya Rao, Sakhil Taneja, Krishna Chaitanya, and Dr. Radhika Mamidi. 2026. Decepticons-IITH: Morphology-aware lexicon filtering of back-translated data for low-resource english-to-indic translations. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. **COMET: A neural framework for MT evaluation**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.
- Sujay Uday Rittikar and Sheela Ramanna. 2026. Winterpeg: Do neural cellular automata help where pretraining ends? In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Lu Ruan, Shaoying Jia, Fan Ning, Wei Wang, Zhengzhen Cai, Fei Hu, Heng Hu, and Chenzi Wang. 2026. : An llm-driven pipeline for extremely low-resource english-to-indic translation — from vocabulary injection to post-training and decoding. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- G.L.John Salvin, Amisha Ramesan, Shrikant tryambak budde, Abigail Nyasha Chigweddedza, and Swapnil Hingmire. 2026. DoDS-IITPKD: Lora fine-tuning and llm post-editing for low-resource indic machine translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Kshetrimayum Boynao Singh, Deepak Kumar, and Asif Ekbal. 2025. Evaluation of llm for english to hindi legal domain machine translation systems. In *Proceedings of the Tenth Conference on Machine Translation*, pages 823–833.
- Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Loitongbam Sanayai Meetei, Ningthoujam Justwant Singh, Thoudam Doren Singh, and Sivaji Bandyopadhyay. 2023. A comparative study of transformer and transfer learning mt models for english-manipuri. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 791–796.
- Salam Michael Singh and Thoudam Doren Singh. 2020. Unsupervised neural machine translation for english and manipuri. In *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, pages 69–78.
- Thoudam Doren Singh and Sivaji Bandyopadhyay. 2010. Manipuri-english bidirectional statistical machine translation systems using morphology and dependency relations. In *Proceedings of the 4th Workshop on Syntax and Structure in Statistical Translation*, pages 83–91.
- Aparajita Sinha, Monika Agarwal, Shreya Bonal Shreya Narayana Bhat, and Aryan. 2026. JH_NLP_APAShreya2: Lora fine-tuning of nllb-200 for english–manipuri machine translation at wmt 2026. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- 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.
- Kuruva Divya Sree and Deepa Gupta. 2026. An rl-inspired agenticrag-grpo based framework for low-resource indic machine translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Bipesh Subedi and Nischal Karki. 2026. BRS-NK: Bidirectional finetuning of nllb-200 for low-resource indic languages: Assamese-english and bodo-english machine translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Vishnuraj, Yuvrajsinh Bodana, Dasari Priyanka, Heli Hingrajiya, and Parameswari Krishnamurthy. 2026. LTRC_Plural: Towards enhancing low resource machine translation with supervised fine tuning and reinforcement learning. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Daniel C. Wood. 2008. An initial reconstruction of proto-boro-garo. Master’s thesis, University of Oregon, Eugene, Oregon, U.S.A.
- Suyash Yadav, Utkarsh Chaudhary, and Alapan Kuila. 2026. RozarNLP-cr7 at wmt 2026: Lora fine-tuning of nllb for low-resource machine translation in north-eastern indian languages. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations (ICLR)*.
- Zebiao Zhou, ZiHao Wang, haoran.wang, and xuemin.zhao. 2026. VoxNexus: A large language model based system for low-resource indic language translation. In *Proceedings of the Eleventh Conference on Machine Translation (WMT 2026)*. Association for Computational Linguistics.