ChakmaNMT: Machine Translation for a Low-Resource and Endangered Language via Transliteration

Aunabil Chakma[1], Aditya Chakma[5], Masum Hasan[3], Soham Khisa[6], Chumui Tripura[4], Rifat Shahriyar[2]

[1]University of Arizona, [2]Bangladesh University of Engineering and Technology, [3]University of Rochester, [4]Chittagong University of Engineering and Technology, [5]University of Newcastle, [6]University of Maryland, Baltimore County

aunabilchakma@arizona.edu, m.hasan@rochester.edu, aditya.chakma@uon.edu.au, skhisa1@umbc.edu, u1604131@student.cuet.ac.bd, rifat@cse.buet.ac.bd

Abstract

We present a systematic study of machine translation for Chakma, an endangered and extremely low-resource Indo-Aryan language, with the goal of supporting language access and preservation. We introduce the first and, to our knowledge, largest Chakma–Bangla MT resource: 15,021 parallel translation pairs comprising 9,548 sentence pairs and 5,473 dictionary word pairs, plus 42,783 Chakma monolingual sentences and a trilingual benchmark with 500 unique sources and 600 Chakma references. To address script mismatch and data scarcity, we propose a character-level transliteration framework designed to preserve pronunciation and lexical identity while enabling transfer from Bangla and multilingual pretrained models. We benchmark from-scratch MT, fine-tuned pretrained models, and large language models via in-context learning. For the tested pretrained models, transliteration is crucial. ICL is strongest for CCP→BN; for BN→CCP, BanglaT5 has the highest BLEU while GPT-4.1 with n-gram retrieval has the highest chrF.

Src.	১৯৬১৯৯৯৫ ৬৭৬৬৬৬৫ ৩৭৬৫ ৯৯৬৫ ৮৯৬৫৩৫ ৯৬৫৩৫ ১৯৬৫৩৫ ৬৬৫৬৫ ৯৮৫৬৫ ৯৮৫৬৫ ৩৩৬৫ ।
Ref.	দেশকে সামনের দিকে এগিয়ে নিয়ে যেতে বেসরকারি সেক্টর খুবই গুরুত্বপূর্ণ । (To move the country forward, private sector is very important.)
NMT	দেশটিকে সামনে এগিয়ে নিয়ে যাওয়ার জন্য বেসিরকারি সেক্টরের যথেষ্ট ভূমিকা রয়েছে । (To move the country forward, there is a significant importance of the private sector.)
ICL	দেশগুলোকে বিভিন্নভাবে আরও বেশি করে বেসরকারি খাতে যুক্ত করতে হবে যেখানে সম্ভব। (Countries need to engage the private sector more and more in various ways where possible.)

Figure 1: Illustrative Chakma→Bangla translation example comparing our two best-performing approaches: fine-tuned NMT (BanglaT5) and GPT-4.1 with 400 demonstration pairs selected by n-gram similarity, our best-performing ICL variant. Despite similar automatic scores, the outputs differ in lexical choice and interpretation.

1 Introduction

The Chakma language is spoken by the indigenous Chakma people across Bangladesh, the easternmost regions of India, and western Myanmar, and belongs to the Indo-Aryan language family (Mohsin, 2013). It is spoken by over 700,000 people across the region (Office of the Registrar General & Census Commissioner, India, 2011; Bangladesh Bureau of Statistics, 2022). Despite this population, Chakma remains predominantly oral, with limited use of its writing system, leaving the language critically under-resourced in digital form. As noted by Saikia and Haokip (2023), Chakma is classified as "Definitely Endangered," and continued language loss poses risks to cultural identity and community continuity. Although revitalization efforts exist-such as pre-primary materials produced by the National Curriculum and Textbook Board of Bangladesh (National Curriculum and Textbook Board, 2019) and limited school textbooks in India (Chakma Autonomous District Council, 2024)-they remain largely confined to education, while everyday communication and public discourse increasingly rely on dominant regional languages such as Bangla, underscoring the lack of computational support for Chakma in broader communication settings.

Chakma remains underrepresented in NLP, although prior and recent work includes character recognition and handwriting data (Podder et al., 2023; Ferdeous et al., 2026), speech language identification (Pratap et al., 2023), and parallel-corpus MT (Rahman et al., 2025). Although commercial LLMs can recognize Chakma script, preliminary inspection suggests that their Chakma outputs can be semantically unreliable; our experiments use

Proceedings of the Eleventh Conference on Machine Translation, pages 64–81
October 28-29, 2026 ©2026 Association for Computational Linguistics

GPT-4.1 variants through the API (Section 7). As a result, foundational text-based capabilities-most notably machine translation (MT), which is critical for cross-lingual communication and access to public resources-remain largely unexplored.

Motivated by the remaining gap, we present a systematic study of machine translation for Chakma, an extremely low-resource and endangered language. We investigate how different modeling paradigms, including classical machine translation, neural models, and large language models via in-context learning, behave under severe data scarcity and cross-script conditions. Our goal is to establish practical baselines for Chakma MT while highlighting challenges that arise in extremely low-resource and non-standardized languages. Figure 1 illustrates qualitative differences between our best-performing systems. A fine-tuned BanglaT5 model produces higher-scoring translations with limited parallel data. A GPT-based in-context learning approach, using 400 demonstration examples, produces outputs that appear more faithful in the inspected examples.

Our contributions are summarized as follows:

(1) **First and largest Chakma–Bangla MT resource to date.** Originally released in 2024, ours was the first Chakma–Bangla MT dataset and remains, to our knowledge, the largest documented Chakma MT corpus, with 15,021 parallel translation pairs comprising 9,548 sentence pairs and 5,473 dictionary word pairs. Of the sentence pairs, 8,647 additionally include aligned English; we also provide 42,783 Chakma monolingual sentences and a trilingual benchmark with 500 unique sources and 600 Chakma references. These resources provide a foundation for machine translation and downstream NLP research on Chakma.

(2) **Script-bridging transliteration framework.** We propose a simple, rule-based character-level transliteration system that leverages Chakma–Bangla script similarity to provide near one-to-one mappings designed to preserve pronunciation and lexical identity. This enables effective cross-script transfer and allows pretrained and large language models to be applied in this extremely low-resource setting.

(3) **Comprehensive benchmarking in an extremely low-resource setting.** We systematically benchmark statistical and neural MT, fine-tuned pretrained models (e.g., BanglaT5, mT5), and GPT-based in-context learning, establishing initial baselines for Chakma–Bangla translation.

(4) **Analysis of orthographic variability.** We analyze orthographic inconsistencies in Chakma arising from non-standardized spelling and script usage. We examine how this variability may affect automatic evaluation, particularly BLEU, when multiple valid spellings occur.

2 Related Work

Machine translation has historically relied on abundant parallel data, from statistical systems to neural encoder–decoders (Koehn et al., 2003; Bahdanau et al., 2014; Koehn and Knowles, 2017), leaving languages with scarce digital resources underserved.

Work on low-resource NMT has explored a range of strategies to mitigate data scarcity, including semi-supervised learning with monolingual data (Gulcehre et al., 2015), back-translation (Sennrich et al., 2016), multilingual neural machine translation for cross-lingual transfer (Kocmi and Bojar, 2018), and transliteration for closely related languages with different scripts (Durrani et al., 2010). In extremely low-resource settings, prior work reports very low translation quality overall (often obtaining BLEU scores below 10) (Guzmán et al., 2019), with several MT approaches yielding BLEU scores in the low single digits or around 1–2 under out-of-domain evaluation, highlighting the difficulty of generalization (Zhang et al., 2020).

Following the initial 2024 release of this work, recent Chakma resources include ChakmaBridge, a complementary 807-sentence five-way corpus with MT baselines (Rahman et al., 2025); MELD, which represents Chakma, Garo, and Marma in Bengali script with English and standard Bengali translations (Mahi et al., 2025); a Bangla-transliterated Chakma corpus for studying cross-lingual transfer through MLM fine-tuning (Khisa et al., 2025); and a Chakma handwriting dataset (Ferdeous et al., 2026). LLMs also offer in-context learning (ICL) as an alternative to fine-tuning (Brown et al., 2020), but its direction-dependent performance motivates our study of extremely low-resource, cross-script translation.

3 ChakmaNMT Dataset

Table 1 summarizes the parallel, monolingual, and evaluation data collected in this work, with details discussed below.

Category	Source	Samples	Avg #Tok	Total
Parallel	UN-Disabilities (BN-CCP)	610	16.86	15,021
	UN-Child Rights (BN-CCP)	291	37.66	
	Dictionary (word pairs)	5,473	1.14	
	Crowdsourced (BN-EN-CCP)	3,444	4.51	
	Community translation + expert review (BN-EN-CCP)	5,203	3.60	
Monolingual	Local Chakma sources (CCP)	42,783	5.81	42,783
Evaluation	RisingNews Benchmark Extension (BN-EN-CCP)	500/600	14.8	500/600

Table 1: Overview of the Chakma–Bangla datasets introduced in this work. The parallel total comprises 9,548 sentence pairs and 5,473 dictionary word pairs. Evaluation contains 500 unique source sentences and 600 Chakma references. Avg #Tok counts space-separated Chakma (CCP) tokens; BN and EN denote Bangla and English.

3.1 Parallel Corpus

Parallel Documents After extensive searching, we identified two publicly available documents that contain aligned Bangla (BN) and Chakma (CCP) translations: *UN Convention on the Rights of Persons with Disabilities* (United Nations, 2007) and *UN Convention on the Rights of the Child* (United Nations General Assembly, 1989). The Bangla versions of these documents were available only as scanned PDF files containing images of the original printed documents. We applied Tesseract OCR[1] to extract Bangla text from these scans.

Automatic sentence alignment methods such as Hualign (Varga et al., 2005) were not applicable in our setting due to the lack of a sufficiently rich Chakma lexicon. Consequently, all sentence alignments were performed manually. This process resulted in 610 and 291 CCP-BN sentence pairs from the two documents, respectively. In addition, we incorporated word-level translation pairs from the only available Chakma dictionary, provided to us in JSON format.[2] This yielded 5,473 additional parallel samples.

Community Translation with Expert Review To obtain high-quality sentence-level translations, we organized a manual data collection effort involving native Chakma speakers with strong literacy skills. We prepared paper-based forms containing 10,000 Bangla sentences randomly sampled from the BN-EN corpus of Hasan et al. (2020). A three-day voluntary translation program was conducted in Dighinala, Khagrachari (Bangladesh). Between 7 and 10 proficient Chakma speakers participated each day.

All collected translations were subsequently reviewed by senior community-recognized Chakma writers or teachers for linguistic quality and consistency. After filtering, this process produced a final set of 5,203 reviewed CCP-BN-EN parallel sentence triples. Further details of this collection process are provided in Appendix A.1.

Manual Translation via Crowdsourcing To further expand the dataset, we collected additional translations through a crowdsourcing approach involving non-expert Chakma speakers. We developed a web-based platform that displayed Bangla sentences and allowed users to submit their Chakma translations. The platform link was distributed through social media channels.

The source sentences primarily consisted of common conversational expressions collected from publicly available resources[3]. Contributors translated only the displayed Bangla sentences into Chakma; English was retained when already available as an aligned reference. Since most contributors were unfamiliar with the Chakma script, they were instructed to write Chakma using Bangla characters. We later converted these submissions into Chakma script using a custom transliteration system (see Section 4.2) developed for CCP-BN conversion. After manual verification and filtering by Chakma language experts, this effort yielded 3,444 additional CCP-BN-EN sentence triples.

Overall, we collected 15,021 BN-CCP parallel translation pairs, comprising 9,548 sentence pairs and 5,473 dictionary word pairs. Of the sentence pairs, 8,647 additionally include aligned English translations.

[1] https://github.com/tesseract-ocr/tesseract

[2] The dictionary data were provided directly by the dictionary's owner for research use.

[3] https://www.learnenglishfrombangla.com/2021/07/easily-learn-english-in-bangla-beginner.html, https://www.omniglot.com/language/phrases/bengali.php, and https://en.wikibooks.org/wiki/Bengali/Common_phrases

3.2 Monolingual Data

Figure 2 shows the distribution of the collected Chakma monolingual data by content type. We collected a substantial amount of Chakma monolingual data relative to the available parallel resources. Due to the scarcity of digitally available Chakma texts,[4] we conducted in-person visits to Chakma language scholars to obtain soft copies of Chakma script materials. These materials primarily consist of poems, articles, short stories, and a small number of national-level textbooks. In addition, we collected Indian Chakma textbooks and a Chakma folktale mobile application, and we reused the Chakma dictionary introduced in the parallel data collection, which contains numerous high-quality example sentences accompanying lexical entries and is therefore well suited for monolingual data extraction.

To process these sources, all materials were first transcribed into separate `.docx` files, which preserved the original Chakma fonts used in the documents. However, these were legacy fonts that mapped Chakma glyphs onto ASCII code points using distinct encodings. To address this, we developed a conversion program that maps all source fonts to RibengUni, a Unicode Chakma font. This enabled consistent normalization across heterogeneous sources.

Finally, we applied a simple rule-based segmentation procedure, splitting text at sentence boundaries defined by three punctuation markers: '?', '!', and '।' (full stop). After preprocessing and normalization, we obtained a total of 42,783 Chakma monolingual samples. Tables 3 and 4 provide detailed descriptions of the monolingual data sources, while the font conversion code is available in our normalization repository[5] and the list of supported ASCII fonts is provided in Appendix Table 5.

3.3 Evaluation Data

To evaluate our models, we constructed a carefully curated benchmark dataset (see Table 1). We randomly selected 500 unique Bangla–English sentence pairs from the RisingNews Benchmark of online news articles (Hasan et al., 2020). We gave only the Bangla sentences to three Chakma language researchers for translation into Chakma; English remained an aligned reference. Each annotator translated 200 sentences, including a common subset of 50 sentences translated by all three annotators. The shared subset was included to allow analysis of translation variability and orthographic inconsistency across gold references, which we further discuss in Section 7. This produced 600 Chakma reference translations corresponding to 500 unique source sentences. All 600 references are retained as separate evaluation rows, so each shared source occurs once per annotator/reference. The resulting *RisingNewsChakma* benchmark is used exclusively for evaluation and differs in domain from the UN documents, dictionary entries, conversational expressions, and short general-domain sentences used for training.

4 Machine Translation Approaches

Machine Translation Task This task is formulated as a sequence-to-sequence learning problem at the sentence level. Given a source sentence, the model generates a target sentence token by token. The objective is to learn this mapping from limited parallel data. We study machine translation between Chakma and Bangla.

We compare three complementary approaches to establish strong baselines and understand what works best for this language pair. Our methods differ primarily in how they leverage prior knowledge and handle the script mismatch between Chakma and Bangla. First, we train conventional MT systems from scratch using only our parallel data, which are limited by data scarcity. Second, we fine-tune pretrained sequence-to-sequence models by transferring knowledge from Bangla via script-bridging transliteration and monolingual augmentation. Third, we evaluate large language models using few-shot in-context learning, adapting them to Chakma translation without parameter updates.

4.1 From-Scratch MT Baselines

We train both statistical and neural baselines from scratch on our parallel corpus. We use phrase-based SMT (Koehn et al., 2003) as a classical baseline that remains competitive in low-resource scenarios (Koehn and Knowles, 2017). We also train a GRU-based RNN with attention (Bahdanau et al., 2014; Luong et al., 2015) as a lightweight neural model. Finally, we train a Transformer (Vaswani et al., 2017) as a stronger but more data-demanding neural baseline.

[4] Chakma digital texts and scripts are rarely available online, and most existing materials are accessible only through local scholars or printed sources.

[5] https://github.com/Aunabil4602/chakma-nmt-normalizer

4.2 Script-Bridging Transliteration

To enable the use of pretrained sequence-to-sequence and large language models, we develop a rule-based, character-level transliteration system that bridges the Chakma and Bangla scripts in a near one-to-one manner.[6] The mapping is designed to preserve pronunciation and lexical identity while mapping Chakma text into the Bangla Unicode range, allowing Chakma data to be processed by Bangla-pretrained models. This transliteration step serves as a core foundation for both fine-tuning pretrained models and in-context learning experiments.

Exploiting the close relationship between Chakma and Bangla. Chakma and Bangla exhibit an unusually high degree of phonetic and orthographic similarity among Indo-Aryan languages, making transliteration a natural and low-effort strategy to bridge script mismatch and enable effective transfer from Bangla-pretrained models. Beyond script, the languages share a largely similar subject–object–verb (SOV) word order, although Chakma has systematic morphosyntactic differences (e.g., placing negation before the verb) that may introduce local reordering effects. The transliteration is largely based on straightforward one-to-one character mappings in both directions, with a small number of deterministic phonetic substitutions where exact script-level equivalence is unavailable. The mapping is not bijective and may neutralize some orthographic distinctions; thus, round-trip BLEU and chrF measure surface-form retention and convergence rather than independently establishing semantic preservation (Section 7).

In practice, transliteration is used exclusively as preprocessing and postprocessing: if Chakma is the source, input is transliterated into Bangla before translation; if Chakma is the target, Bangla-script model output is transliterated back into Chakma for evaluation. All pretrained sequence-to-sequence models described in the following subsection are trained and evaluated using this transliterated input–output representation.

We provide full mapping statistics, handling of non-direct characters, and a summary of all non-direct rules (Figure 3) in Appendix A.3.

[6] https://github.com/Aunabil4602/chakma-nmt-normalizer

4.3 Fine-Tuning Pretrained Sequence-to-Sequence Models

We fine-tune pretrained text-to-text models on transliterated Chakma–Bangla parallel data. We experiment with BanglaT5 (Bhattacharjee et al., 2023), mT5-small (Xue et al., 2021), and mBART (Liu et al., 2020). These models allow us to transfer linguistic knowledge from Bangla and multilingual pretraining into the Chakma setting.

We further evaluate two data-centric extensions to improve robustness under scarcity. We apply iterative back-translation (IBT) (Hoang et al., 2018) to generate synthetic parallel data from monolingual corpora. In IBT, we start with the forward direction CCP→BN trained on the original parallel data, the backward direction (BN→CCP) is trained with synthetic data, and in later iterations both directional models are trained with additional synthetic data. We also evaluate multilingual training (MNMT) following Johnson et al. (2017). For MNMT, we add 10k Bangla–English sentence pairs from Hasan et al. (2020) to the training data to encourage cross-lingual transfer across Bangla, Chakma, and English.

4.4 Large Language Models via In-Context Learning

We evaluate few-shot in-context learning as an alternative to fine-tuning using state-of-the-art large language models. Specifically, we test GPT-4.1, GPT-4.1-mini, and o4-mini with a fixed prompting template and a limited number of demonstration translation pairs. This setting assesses their ability to perform Chakma translation using only in-context examples. The full prompting template and example format are shown in Figure 4.

Example Selection Demonstration examples are selected using two retrieval strategies: uniform random sampling from the parallel data and character-level n-gram similarity with each 20-sentence query batch. We use character-level matching to handle orthographic variation in Chakma, where multiple valid spellings make word-level retrieval unreliable. Comparing these strategies allows us to examine how retrieval strategy and demonstration count affect in-context translation quality.

5 Experimental Setup

Data Splits and Evaluation We split the parallel Chakma–Bangla corpus into training and develop-

ment sets.[7] The training set contains 12,016 parallel pairs and the development set contains 3,005 parallel pairs. We evaluate all models on the held-out RisingNewsChakma online-news benchmark. We report BLEU (Post, 2018) and chrF (Popović, 2015) as our primary automatic metrics; following common practice, BLEU is used for model selection on the development set, and chrF is reported alongside BLEU for all experimental results.

From-Scratch SMT and NMT We use the Moses toolkit[8] for phrase-based SMT and PyTorch for neural models. All neural models are trained on Google Colab using NVIDIA V100/A100 GPUs[9]. Data preprocessing follows the normalization scheme of Hasan et al. (2020) with minor language-specific adjustments.[10] We apply SentencePiece (Kudo and Richardson, 2018) for tokenization across SMT and NMT systems. Decoding uses beam search with width 5 and a maximum sequence length of 128 tokens; coverage statistics are reported in Appendix A.4. We train the GRU-based and Transformer models described in Section 4.1.

Fine-Tuning and In-Context Learning We fine-tune BanglaT5, mT5-small, and mBART for Chakma–Bangla translation. For iterative back-translation, we use the full Chakma monolingual corpus and 50k Bangla monolingual sentences.

We also evaluate large language models using few-shot in-context learning without parameter updates, including GPT-4.1, GPT-4.1-mini, and o4-mini. We retained the default temperature setting and did not tune decoding parameters because of API-budget constraints. Each prompt contains 100–400 demonstration pairs and batches 20 query sentences to reduce API cost; we did not compare single-query prompting. One demonstration set is selected for the full 20-query batch and held fixed across the three repeated generations, so reported variation reflects decoding rather than demonstration-selection variability. Demonstrations are drawn from the training split either randomly or by Jaccard similarity over character 6-grams; Appendix A.4 gives the full retrieval procedure.

We additionally conduct ablation experiments by (i) removing transliteration for fine-tuned and in-context models, and (ii) evaluating a zero-shot in-context learning configuration without demonstrations.

Additional experimental details necessary for replication, including normalization, training hyperparameters, model architectures and initialization, multilingual data formatting and oversampling, in-context learning prompt construction, and multiple runs and randomness, are provided in Appendix A.4. The complete experimental code is available on GitHub.[11]

6 Results

Transliteration is crucial for the tested pre-trained models Removing transliteration causes near-zero BLEU and chrF for the selected pre-trained and in-context models under our experimental setup (Table 6). Thus, these models cannot effectively use native Chakma-script data in this configuration; the metrics do not isolate whether the cause is tokenization, representation, or another model-specific factor.

The relative advantage of ICL is direction- and metric-dependent ICL is strongest for CCP→BN. For BN→CCP, BanglaT5 achieves the highest BLEU, whereas GPT-4.1 with n-gram retrieval achieves the highest chrF (Tables 2 and 8). Thus, the relative advantage of ICL and fine-tuning depends on translation direction and evaluation metric.

From-scratch models generalize poorly under extreme data scarcity From-scratch SMT, RNN, and Transformer models achieve modest development performance but near-zero test BLEU in both directions (Table 2). This large development–test gap is consistent with distributional differences between the training sources and the RisingNewsChakma benchmark, although our results do not isolate domain shift from other causes.

Back-translation effects depend on model and direction The effect of iterative back-translation is model- and direction-dependent. It consistently benefits mT5-small, while the largest improvements occur for CCP→BN after two iterations (Table 2). For BanglaT5 BN→CCP, iterative back-translation does not improve over ordinary bilingual fine-tuning. Notably, CCP→BN scores for

[7] https://huggingface.co/datasets/amlan107/chakma-nmt-complete-dataset

[8] https://www2.statmt.org/moses/

[9] https://colab.research.google.com/

[10] https://github.com/Aunabil4602/chakma-nmt-normalizer

[11] https://github.com/Aunabil4602/chakmanmt

System	BN→CCP				CCP→BN			
	Dev		Test		Dev		Test	
	BLEU	chrF	BLEU	chrF	BLEU	chrF	BLEU	chrF
From-scratch trained								
SMT	4.6 ± −−	**26.3** ± −−	0.1 ± −−	20.2 ± −−	3.2 ± −−	24.9 ± −−	0.1 ± −−	**20.4** ± −−
RNN	**5.3** ± 0.91	22.2 ± 3.71	0.1 ± 0.03	10.4 ± 2.45	**8.0** ± 2.51	20.1 ± 4.37	0.1 ± 0.07	8.5 ± 2.43
Transformer	1.6 ± 0.03	25.8 ± 0.21	**0.2** ± 0.04	**23.2** ± 0.30	3.0 ± 0.12	**27.2** ± 0.28	**0.4** ± 0.05	20.0 ± 0.51
Fine-tuned with transliteration enabled								
mBART	3.5 ± 0.35	20.0 ± 1.45	0.2 ± 0.03	11.1 ± 1.09	10.4 ± 6.74	22.6 ± 9.80	0.8 ± 0.51	12.8 ± 3.49
mT5-small	8.5 ± 0.11	32.2 ± 0.20	2.0 ± 0.04	25.8 ± 0.12	14.1 ± 0.33	33.9 ± 0.35	4.6 ± 0.07	27.2 ± 0.25
+IBT-1it	8.4 ± 0.03	33.2 ± 0.04	2.5 ± 0.02	29.4 ± 0.09	14.1 ± 0.33	33.9 ± 0.35	4.6 ± 0.07	27.2 ± 0.25
+IBT-2it	8.3 ± 0.09	33.9 ± 0.14	**2.7** ± 0.09	30.1 ± 0.09	15.1 ± 0.11	37.4 ± 0.14	7.8 ± 0.10	37.8 ± 0.20
+MNMT	6.1 ± 0.14	27.5 ± 0.35	1.5 ± 0.05	24.1 ± 0.73	9.2 ± 0.28	28.1 ± 0.30	3.0 ± 0.10	23.5 ± 0.34
BanglaT5	**10.9** ± 0.06	36.3 ± 0.22	**2.7** ± 0.05	**30.6** ± 0.21	**24.4** ± 0.11	44.6 ± 0.08	11.7 ± 0.19	37.9 ± 0.09
+IBT-1it	10.4 ± 0.19	36.5 ± 0.11	2.2 ± 0.06	27.7 ± 0.21	**24.4** ± 0.11	44.6 ± 0.08	11.7 ± 0.19	37.9 ± 0.09
+IBT-2it	10.5 ± 0.18	**36.8** ± 0.07	2.5 ± 0.18	28.5 ± 0.35	24.2 ± 0.11	**46.8** ± 0.06	**13.9** ± 0.20	**46.3** ± 0.15
+MNMT	8.2 ± 0.26	32.1 ± 0.22	2.4 ± 0.11	29.2 ± 0.32	20.9 ± 0.62	40.1 ± 0.39	12.8 ± 0.12	38.5 ± 0.61
In-context learning with transliteration enabled								
GPT-4.1-mini(R)	-	-	1.1 ± 0.06	28.4 ± 0.19	-	-	10.9 ± 0.19	40.0 ± 0.58
GPT-4.1(R)	-	-	**1.6** ± 0.13	30.4 ± 0.12	-	-	16.5 ± 0.12	48.2 ± 0.35
o4-mini(R)	-	-	1.5 ± 0.03	29.7 ± 0.03	-	-	12.5 ± 0.47	42.9 ± 0.08
GPT-4.1-mini(N)	-	-	1.2 ± 0.01	28.7 ± 0.25	-	-	10.5 ± 0.54	40.4 ± 0.47
GPT-4.1(N)	-	-	1.5 ± 0.10	**31.3** ± 0.53	-	-	**17.8** ± 0.12	**49.6** ± 0.12
o4-mini(N)	-	-	1.2 ± 0.07	29.7 ± 0.14	-	-	12.2 ± 0.27	42.7 ± 0.48

Table 2: Translation performance on BN↔CCP across development and test sets. We report mean ± standard deviation for BLEU and chrF. Results are shown for from-scratch models, fine-tuned pretrained models (including iterative back-translation (IBT), up to two iterations, and multilingual fine-tuning (MNMT)), and GPT-based in-context learning (ICL) with 400 examples using random (R) or n-gram (N) similarity-based sampling. ICL is strongest for CCP→BN. For BN→CCP, BanglaT5 has the highest BLEU, while GPT-4.1 with n-gram retrieval has the highest chrF. Dashes denote unavailable ICL development results or unreported SMT standard deviations, as applicable.

BanglaT5 and mT5 remain unchanged after the first IBT iteration, since the backward model is initially trained on the same parallel data.

Multilingual training generally underperforms bilingual fine-tuning Adding MNMT reduces test BLEU and chrF for both directions with mT5-small and for BN→CCP with BanglaT5 (Table 2). However, BanglaT5 shows a modest improvement for CCP→BN, from 11.7 BLEU / 37.9 chrF to 12.8 BLEU / 38.5 chrF. Thus, the effect of multilingual training is model- and direction-dependent. Table 9 additionally reports EN↔CCP and BN↔CCP performance under the MNMT setup.

ICL retrieval and scaling effects are non-uniform N-gram retrieval generally benefits GPT-4.1, but improvements are not consistent across all model variants or demonstration counts (Table 8). Performance is also non-monotonic as the number of demonstrations increases.

Performance-cost trade-offs between ICL and fine-tuning ICL achieves strong CCP→BN performance on both BLEU and chrF with minimal data, but requires large commercial models and costly prompts. Fine-tuning may be more economical under repeated use, although we do not perform a formal cost analysis. The preferred approach therefore depends on translation direction, evaluation metric, and deployment frequency.

7 Qualitative Analysis

BN→CCP translation is substantially harder than CCP→BN Across all systems, BN→CCP performance is substantially lower than CCP→BN on both BLEU and chrF (Table 2). Even the best BN→CCP results reach only about 2–3 BLEU, while CCP→BN attains 13–18 depending on the method. We hypothesize that this asymmetry arises because pretrained models encode Bangla more effectively, making translation into Bangla easier

than into Chakma. The chrF gap suggests that the difficulty extends beyond a BLEU-specific artifact. Figure 5 provides representative BN→CCP outputs across model families.

BLEU and lexical or orthographic variation in Chakma The BLEU–chrF divergence is consistent with lexical and orthographic variation, although without exhaustive variant annotation or human evaluation we cannot isolate its contribution from semantic errors. BLEU is particularly sensitive to re-transliterated Chakma outputs, where spelling differences introduced by script conversion lower n-gram overlap. This is evident in inter-annotator agreement on 50 shared benchmark sentences (Section 3.3), which is only 4.48 BLEU but 38.82 chrF, indicating stable character overlap despite differing word forms. Figure 6 further illustrates multiple valid spellings for common words, strongly penalizing BLEU. These patterns, driven by the lack of standardized Chakma orthography, motivate treating chrF as a co-primary metric alongside BLEU.

Round-trip transliteration measures surface-form retention The transliteration mapping is designed to preserve pronunciation and lexical identity, but it is not bijective and may neutralize some orthographic distinctions. After one cycle, scores are 41.55 BLEU / 79.32 chrF for BN→CCP→BN and 38.37 BLEU / 79.69 chrF for CCP→BN→CCP, showing substantial character overlap alongside nontrivial word-form differences. Later rounds converge because the deterministic substitutions have stabilized (Table 10). These round-trip metrics measure surface-form retention and convergence; they do not independently establish semantic preservation.

Zero-shot ICL ablation highlights the need for demonstrations As an ablation, we evaluate zero-shot ICL without any demonstrations. For BN→CCP, BLEU remains below 1 across models (Table 11), indicating that zero-shot ICL is ineffective in this direction. CCP→BN performs better even without examples, but still lags behind few-shot ICL. For the tested models and prompts, these results show that demonstrations are important for generating Chakma outputs in this extremely low-resource setting. chrF is similarly low in zero-shot BN→CCP, underscoring that models fail to recover even partial character overlaps without examples.

LLM variants show different effectiveness under ICL Under identical in-context learning (ICL) setups, GPT-4.1 achieves the strongest overall performance, while o4-mini often outperforms GPT-4.1-mini in BLEU and chrF (Table 8). As these models differ in architecture, capacity, and intended design, we do not attribute the observed differences to any single factor. Instead, we report them as an empirical comparison of LLM variants under the same ICL conditions. The chrF ranking matches BLEU, suggesting that model differences affect both token-level and character-level fidelity.

8 Conclusion

This work introduces new Chakma–Bangla resources and evaluates transliteration-based transfer across several machine translation paradigms. Pretrained models and in-context learning substantially outperform from-scratch systems, but performance remains strongly direction-dependent. ICL performs best for CCP→BN, while BN→CCP remains difficult and shows metric-dependent differences between fine-tuning and ICL. The results provide initial baselines rather than deployment-ready systems.

Ethics

This work involves data collection for an endangered and low-resource language with the goal of supporting language preservation and accessibility. All human-generated data were collected with informed consent from contributors, who voluntarily participated and expressed support for this research. No personally identifiable information about dataset contributors is included in the released data. Potential risks include linguistic misrepresentation, reinforcement of contested orthographic conventions, and reuse outside the intended community context. We mitigate these risks through community review and explicit documentation of orthographic variation. Third-party materials retain their original rights; we will redistribute processed text only where permission or source licenses allow, and otherwise provide metadata, document identifiers, conversion code, and access instructions.

Limitations

This work is constrained by the extremely low-resource nature of the Chakma language, which limits the size and diversity of available training

data and leads to relatively low automatic evaluation scores, a common challenge in low-resource machine translation. Our rule-based transliteration framework enables effective cross-script transfer for the tested models, but relies on manually designed mappings and is not character-bijective; round-trip metrics do not independently establish semantic preservation. Automatic metrics such as BLEU may further underestimate translation quality due to orthographic variation and multiple valid spellings in Chakma. Our contribution is resource- and evidence-oriented rather than a new MT architecture, and the findings most directly apply to cross-script pairs with a related higher-resource language. The benchmark contains 500 unique sources and 600 Chakma references, and the low absolute scores make these initial baselines rather than deployment-ready systems. Ultimately, the primary bottleneck remains data scarcity: future work could benefit substantially from automated web-based data crawling and collection systems to expand Chakma textual resources, as well as from more data-driven transliteration and translation approaches tailored to extremely low-resource and non-standardized languages.

Acknowledgments

Our work would not have been possible without the help of several individuals and organizations. Particularly, we would like to thank Injeb Chakma, who assisted us with any issues related to the Chakma language. We are thankful to Arjya Mitra Chakma, Shanti Chakma, and The Zabarang Foundation for providing important documents. We would also like to express our gratitude to Bivuti Chakma for helping us with the RibengUni font and additional documents. For translating the benchmark data, we are highly grateful to Sujan Chakma, Shanti Chakma, and Binoy Bikash Talukdar. Moreover, we extend our sincere thanks to all the people who supported our work and voluntarily participated in the translation process both online and offline.

"Beyond the Hill" group assisted us in data collection and enabled us to compete in a competition hosted by the ICT Division of Bangladesh, where we won 10,000 BDT.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Y. Bengio. 2014. Neural machine translation by jointly learning to align and translate. *ArXiv*, 1409.

Bangladesh Bureau of Statistics. 2022. Population and housing census 2022: Preliminary report. Table A-1.4: Ethnic population by group and sex.

Abhik Bhattacharjee, Tahmid Hasan, Wasi Uddin Ahmad, and Rifat Shahriyar. 2023. BanglaNLG and BanglaT5: Benchmarks and resources for evaluating low-resource natural language generation in Bangla. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 726–735, Dubrovnik, Croatia. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Chakma Autonomous District Council. 2024. Chakma textbooks.

Nadir Durrani, Hassan Sajjad, Alexander Fraser, and Helmut Schmid. 2010. Hindi-to-Urdu machine translation through transliteration. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 465–474, Uppsala, Sweden. Association for Computational Linguistics.

Jannatul Ferdeous, Md. Abdul Kayum, Ahmed Islam, Md Roman Nihal, ASM Shihavuddin, and Muhammad Abul Hasan. 2026. A 48-class handwritten dataset for the endangered Chakma language. *Data in Brief*, 66:112796.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy. PMLR.

Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loïc Barrault, Huei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2015. On using monolingual corpora in neural machine translation. *arXiv preprint arXiv:1503.03535*.

Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. 2019. The FLORES evaluation datasets for low-resource machine translation: Nepali–English and Sinhala–English. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6098–6111, Hong Kong, China. Association for Computational Linguistics.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online. Association for Computational Linguistics.

Vu Cong Duy Hoang, Philipp Koehn, Gholamreza Haffari, and Trevor Cohn. 2018. Iterative back-translation for neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 18–24, Melbourne, Australia. Association for Computational Linguistics.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Adity Khisa, Nusrat Jahan Lia, Tasnim Mahfuz Nafis, Zarif Masud, Tanzir Pial, Shebuti Rayana, and Ahmedul Kabir. 2025. Exploring cross-lingual knowledge transfer via transliteration-based MLM fine-tuning for critically low-resource Chakma language. In *Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025)*, pages 280–291, Mumbai, India. Association for Computational Linguistics.

Tom Kocmi and Ondřej Bojar. 2018. Trivial transfer learning for low-resource neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 244–252, Brussels, Belgium. Association for Computational Linguistics.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver. Association for Computational Linguistics.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 48–54. Association for Computational Linguistics. 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003) ; Conference date: 27-05-2003 Through 01-06-2003.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Preprint*, arXiv:2001.08210.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

Mehraj Hossain Mahi, Anzir Rahman Khan, Mobashsher Hasan Anik, Sheak Rashed Haider Noori, Arif Mahmud, and Mayen Uddin Mojumdar. 2025. MELD: A multilingual ethnic dataset of Chakma, Garo, and Marma in Bengali script with English and standard Bengali translation. *Data in Brief*, 61:111745.

Amena Mohsin. 2013. Language, identity and state. In Naeem Mohaiemen, editor, *Between Ashes and Hope: Chittagong Hill Tracts in the Blind Spot of Bangladesh Nationalism*, page 158. Drishtipat Writers' Collective.

National Curriculum and Textbook Board. 2019. Textbook of small ethnic group (Chakma) for pre-primary. Government of Bangladesh.

Office of the Registrar General & Census Commissioner, India. 2011. District census handbook: Lawngtlai.

Kanchon Kanti Podder, Ludmila Emdad Khan, Jyoti Chakma, Muhammad E.H. Chowdhury, Proma Dutta, Khan Md Anwarus Salam, Amith Khandakar, Mohamed Arselene Ayari, Bikash Kumar Bhawmick, S M Arafin Islam, and Serkan Kiranyaz. 2023. Self-ChakmaNet: A deep learning framework for indigenous language learning using handwritten characters. *Egyptian Informatics Journal*, 24(4):100413.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Mamdouh Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2023. Scaling speech technology to 1,000+ languages. *ArXiv*, abs/2305.13516.

Md. Abdur Rahman, Md. Tofael Ahmed Bhuiyan, and Abdul Kadar Muhammad Masum. 2025. ChakmaBridge: A five-way parallel corpus for navigating the script divide in an endangered language. In *Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025)*, pages 259–265, Mumbai, India. Association for Computational Linguistics.

Jonali Saikia and D. Mary Kim Haokip. 2023. Language endangerment with special reference to Chakma. *Journal of English Language and Literature*, 10(3):86–100.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. *Preprint*, arXiv:1511.06709.

United Nations. 2007. Convention on the rights of persons with disabilities and optional protocol.

United Nations General Assembly. 1989. Convention on the rights of the child. General Assembly Resolution 44/25, 20 November 1989.

Dániel Varga, Péter Halácsy, András Kornai, Viktor Nagy, László Németh, and Viktor Trón. 2005. Parallel corpora for medium density languages. *In Proceedings of the Recent Advances in Natural Language Processing (RANLP 2005)*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Shiyue Zhang, Benjamin Frey, and Mohit Bansal. 2020. ChrEn: Cherokee-English machine translation for endangered language revitalization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 577–595, Online. Association for Computational Linguistics.

A Appendix

A.1 Additional Details on Community Translation with Expert Review

Prior to data collection, we conducted a pre-assessment to evaluate the feasibility of manual translation by volunteer Chakma speakers. The assessment revealed that participation and translation quality were highly sensitive to task complexity and time requirements. In particular, longer sentences significantly increased cognitive load and annotation time, frequently resulting in incomplete or inaccurate translations. These difficulties were exacerbated by the fact that many Bangla or English lexical items are either rarely used in Chakma or lack direct lexical equivalents, often requiring paraphrasing or explanatory reformulation. Based on these findings, we constrained source sentences to a length of 2 to 8 words to reduce annotation burden while maintaining sufficient linguistic content.

To mitigate these challenges and reduce annotator fatigue, we deliberately selected short sentences, with the probability of sentence selection peaking at 4-5 words and gradually decreasing toward both extremes. This distribution reflects an optimal trade-off between linguistic informativeness and annotation feasibility. Very short sentences (e.g., one word) were avoided due to limited contextual value, while longer sentences were excluded to minimize translation difficulty and error propagation.

The participants were predominantly young native Chakma speakers with functional bilingual proficiency in Bangla and English but limited formal training in translation. Senior reviewers were community-recognized Chakma writers or teachers, and translators were instructed to produce natural translations faithful to the Bangla source. Despite their linguistic competence, many participants reported difficulty translating abstract or institutional terms, as Chakma remains primarily an oral language with limited standardized written usage. Some experienced contributors provided handwritten translations, which were later digitized by trained typists due to limited familiarity with Chakma script typing.

A.2 Origins of Orthographic Variation in Chakma

Chakma lacks a widely accepted standardized grammar, resulting in substantial variation in spelling and syntactic structure across written sources. Historically, language experts and local shamans have documented the language using personal conventions without publishing formal grammatical guidelines. As a result, the same lexical items are often written using multiple valid spellings, leading to pervasive orthographic inconsistency. Disagreements among scholars have further hindered consensus on standard grammatical rules. These disagreements are also reflected in dif-

ferences between Indian and Bangladeshi Chakma scholarly traditions. Such variability complicates data normalization and automatic evaluation in downstream NLP tasks.

A.3 Script-Bridging Transliteration: Mapping Coverage and Non-Direct Rules

This appendix provides the full mapping statistics and the handling rules for characters that do not admit an exact one-to-one correspondence between Chakma (CCP) and Bangla (BN). A summary of all non-direct substitutions in both directions is shown in Figure 3.

CCP→BN coverage. In the CCP→BN direction, all 10 Chakma digits have direct mappings. Among the core Chakma letters (vowels and consonants), 36 out of 37 characters map directly to Bangla equivalents. For diacritics, 18 out of 21 Chakma diacritical marks have direct mappings. The remaining Chakma characters correspond to prosodic or orthographic distinctions that do not have explicit representations in Bangla. Notably, one such character functions as a lengthening/extension marker: it modifies the pronunciation of an adjacent letter rather than introducing independent lexical content. Since Bangla lacks an equivalent graphemic mechanism for this feature, we normalize it during transliteration by retaining the base character without adding a distinct symbol in the Bangla rendering. The four Chakma characters without direct Bangla equivalents are rare in the collected data, but their semantic effect is not independently evaluated.

BN→CCP coverage. In the reverse BN→CCP direction, all 10 Bangla digits have direct mappings. Out of 50 Bangla letters, 44 map directly to Chakma characters. Similarly, 10 out of 11 Bangla diacritics have corresponding Chakma equivalents. The remaining Bangla characters are mapped to the closest Chakma counterparts according to pronunciation and lexical identity.

Deterministic handling of non-direct characters. For characters without direct one-to-one correspondences in either direction, we apply deterministic substitutions based on closest phonetic similarity in the target script (Figure 3). Consequently, the transliteration system is not strictly reversible and may neutralize some orthographic distinctions. Round-trip metrics quantify surface-form retention rather than semantic preservation (Section 7). Overall, the system remains predominantly a straightforward character mapping scheme, with a small number of rule-based phonetic normalizations applied only when exact script-level equivalence is unavailable.

A.4 Additional Experimental Details

Multiple Runs and Randomness To assess robustness to training stochasticity, we run all experiments three times with different random seeds (affecting initialization and minibatch order) and report mean and standard deviation. For in-context learning, we fix the demonstration set and repeat generation three times to quantify decoding variability. Therefore, reruns under the same configuration are expected to yield scores consistent with the reported mean ± standard deviation.

Normalization details On top of the normalization described by (Hasan et al., 2020), we apply a minimal and conservative normalization step uniformly to all text, including training data, model inputs, and output labels, across all experiments. For Chakma script, we merge a small number of rarely used, phonetically similar vowel variants into a common representation to reduce superficial spelling variation. This is analogous to collapsing long and short vowel variants in low-resource settings and is intended to simplify orthographic variation. We also normalize all bracket symbols to parentheses in order to reduce sparsity in the data. These normalization steps are applied symmetrically; their semantic effect is not independently evaluated.

Additional details on metrics We report BLEU and chrF scores using `sacreBLEU` via the Hugging Face `evaluate` library with default settings: BLEU uses a maximum word n-gram order of 4 (BLEU-4), and chrF uses character n-grams of order 6.

Sequence length and augmentation overlap Table 1 reports space-separated words rather than model subwords; in the tokenized training data, 3.04% of sequences exceed 32 tokens, 0.73% exceed 64, and 0.02% exceed 128, so the 128-token limit covers over 99.9% of samples. Monolingual augmentation samples were kept disjoint from the parallel splits to avoid overlap and add new content.

SentencePiece Vocabulary Search We use SentencePiece (Kudo and Richardson, 2018) both for (i) vocabulary building and (ii) tokenization for

SMT and NMT. As part of hyper-parameter optimization, we evaluate vocabulary sizes of 1,000, 2,000, 5,000, 10,000, and 20,000.

Training Hyper-Parameters and Optimization Settings We apply gradient clipping with max norm 1.0. We tune learning rates in {0.001, 0.005, 0.0001, 0.0005}, batch sizes in {8, 16, 32}, and training steps in {10,000, 15,000, 20,000}. Warmup steps are varied in {0, 2000, 4000}. We use linear learning-rate decay after warmup. We also tune label smoothing over {0.1, 0.2, 0.3, 0.4, 0.5}. The final tuned hyper-parameter configurations are reported in Table 7.

RNN with Attention: Architecture and Initialization For the RNN baseline, we use a public PyTorch seq2seq implementation.[12] The model incorporates Luong-style attention (Luong et al., 2015). We experiment with 1, 2, and 4 stacked recurrent layers, and consider hidden size and embedding size in {512, 1024}. Dropout is tuned in {0.1, 0.2, 0.3}. All RNN parameters are initialized from a normal distribution with mean 0 and standard deviation 0.1.

Transformer: Model Variants and Initialization For Transformer training, we follow the standard Transformer formulation (Vaswani et al., 2017). We explore MarianNMT-style Transformer[13] configurations available through HuggingFace implementations, and initialize weights with Glorot initialization (Glorot and Bengio, 2010). We vary the number of layers in {1, 2, 6}, attention heads in {1, 2, 6}, dropout in {0.1, 0.2, 0.3}, and feed-forward hidden dimensions in {512, 1024}.

Multilingual Formatting, and Oversampling In multilingual training (MNMT), we prepend a target-language prefix tag to each input sentence: <BN> for Bangla, <EN> for English, and <CCP> for Chakma. To mitigate imbalance, we oversample Chakma-involving parallel pairs to better balance all translation directions, following practices shown to improve multilingual performance (Johnson et al., 2017).

In-Context Learning Prompt Construction For in-context learning (ICL), each prompt contains one batch of 20 query sentences and one shared demonstration set. Random retrieval samples from the training pairs without replacement. For n-gram retrieval, we form the union of character 6-grams across all 20 queries and rank each training source by Jaccard similarity to that union. Equal scores retain the original training-data order. We apply no additional deduplication, so duplicated training rows, if present, can yield repeated demonstrations. The prompt requests 20 numbered outputs aligned by index to the 20 inputs. The selected demonstration set is fixed across the three repeated generations, so the reported variation reflects decoding variability rather than demonstration selection. Character-level matching is used instead of word-level matching due to the absence of standardized spelling. The n-gram size was selected through limited manual experimentation using a small subset of the development data, due to computational budget constraints. This subset was used only for preliminary testing of retrieval behavior, and not for model selection or final evaluation. We evaluated a narrow range of values and fixed $n = 6$, which provided the most stable retrieval behavior in these tests.

mBART: Additional Details We do not apply iterative back-translation (IBT) or multilingual fine-tuning (MNMT) to mBART, as its plain fine-tuning performance is substantially lower than other models (Table 2, Section 6), making these extensions unlikely to provide meaningful improvements.

A.5 Interviews with Chakma Language Experts

We interviewed several scholars in Bangladesh to discuss the variants, for example, the number of characters, diacritics, rules, spelling patterns, etc. The scholars include Arjya Mitra, Injeb Chakma, Ananda Mohon Chakma, and Sugata Chakma. However, almost all of them suggested following the rules maintained by the members of the National Curriculum and Textbook Board of Bangladesh involved in writing the Chakma books for the pre-primary levels because their rules will be followed eventually. The most important rule from them that we followed in our transliteration codes from Bangla to Chakma, is that the core grapheme cannot have more than one diacritic attached to a consonant or a vowel. However, in India, this restriction is not maintained, rather more than one diacritic is seen frequently in their documents.

[12] https://github.com/bentrevett/pytorch-seq2seq/tree/main

[13] https://huggingface.co/docs/transformers/model_doc/marian

Title	Content	Samples
Ajanir dajan firana.docx	Story	206
Amader-Bari-2.pdf	Story	12
Amader-Bari-3.pdf	Story	23
Amader-gaye-dewar-pinon.pdf	Story	10
Amar-Charar-Boi.pdf	Poem	123
Amlokir-Gach.pdf	Story	27
Article 3rd Jamachug.docx	Story	194
Article 4th Furamon.docx	Story	194
Article 5th Pawr Murah.docx	Story	191
Bang-O-Puti-mach.pdf	Story	11
Banor-Berate-Eseche.pdf	Story	35
Banorer-Marfa-khaowa.pdf	Story	10
Bashir-soor.pdf	Story	9
Bie-Bari.pdf	Story	28
Bijhu.pdf	Story	28
Binoy Bikash Talukder20.docx	Poem	647
Binoy Dewan.docx	Poem	2004
Bizute-Berano.pdf	Story	12
Bone-Gie-Gach-Kata.pdf	Story	30
Boner-Mama.pdf	Story	11
Chader-Buri.pdf	Story	28
Chakma Dictionary app	Other	14928
Chakma Folktales app	Story	3765
Chakma Love song Uvagit.docx	Story	13
Chakma Text Book For Class-IV 2010 (IN Govt).docx	Textbook	1088
Chakma Text Book for Class-II 2010 (IN Govt).docx	Textbook	490
Chakma Text Book for Class-III 2010 (IN Govt).docx	Textbook	561
Chakma Text Book for Class-V 2010 (IN Govt).docx	Textbook	940
Chakma Text Book for Class-VI 2010 (IN Govt).docx	Textbook	1543
Chakma Text Book for Class-VII 2010 (IN Govt).docx	Textbook	1858
Chakma.docx	Article	136
Charar Boi-Chakma-Pages.pdf	Poem	31
Cijir Orago Boi-Chakma-Pages.pdf	Other	71
Cijir Talmiloni Kodatara-Chakma-Pages.pdf	Other	45
Cycle-e-Bazare-Jawa.pdf	Story	33
Dhanpudi.doc	Story	1278
Dudur-Kanna.pdf	Story	40
Dui-Bandhobir-Kotha.pdf	Story	16
Ghara Poja pire-Chakma-Pages.pdf	Other	4
H.F.Miller's Rangakura.docx	Story	90
Hotat-Agun.pdf	Story	12
Iskulo Akto-Chakma-Pages.pdf	Other	5
Jhimit-Ekhon-Bhalo.pdf	Story	42
Jhogra-Kora-Valo-Noi.pdf	Story	42
Kalo-and-Forshar-Kotha-1.pdf	Story	22
Kanamachi-Khela.pdf	Story	13
Karo-bipode-hasa-thik-na.pdf	Story	15
Kolar-Kotha-1.pdf	Story	11
Korgosher-sobji-bagan.pdf	Story	12
Lairang-er-nodi-par-howa.pdf	Story	13
Lao-er-Desh-Vromon.pdf	Story	44
Laz-kata-Banor.pdf	Story	12
Lobh-kora-valo-na.pdf	Story	16

Table 3: Names of the sources of our Chakma monolingual data with details (Part 1).

Title	Content	Samples
Mamar-Bari.pdf	Story	19
Mayer-Upadesh-1.pdf	Story	19
Meghla-Akash.pdf	Story	22
Mitar-Fuler-Bagan-1.pdf	Story	10
Moina-Pakhi-1.pdf	Story	16
Monar Sabon-Chakma-Pages.pdf	Story	36
Moni-Malar-Kotha-.pdf	Story	22
Monir-shopno-dekha.pdf	Story	14
Morog-Jhuti-Fool.pdf	Story	25
My Legha by Injeb Chakma.doc	Story	727
Nada-bhet-math for class I (IN Govt Tripura).docx	Textbook	878
Nanarakam-ghor.pdf	Story	14
Nirapod-pani-pan-korbo.pdf	Story	13
Ojhapador Chora-Chakma-Pages.pdf	Poem	30
Paka-Lichu.pdf	Story	19
Porichoy.pdf	Story	16
Projapoti-Ronger-Kotha.pdf	Story	12
Puti-Macher-Fal.pdf	Story	13
Rangdhanu.pdf	Story	20
Ranjuni for Class I (IN Govt) Tripura.docx	Textbook	1459
SRM 1st P. Bargang.docx	Poem	156
SRM 1st R. Krisnachura.docx	Poem	149
SRM 2nd P. Belwa Pawr.docx	Poem	259
SRM 2nd R. Chadarok.docx	Poem	76
Sanye-Pidhe-.pdf	Story	6
Shikkha Boi2017.docx	Poem	722
Shing-Macher-Kata.pdf	Story	36
Shiyal-er-Khang-Garang-Bazano.pdf	Story	19
Shrout.pdf	Story	8
Sial-mamar-school.pdf	Story	14
Sukorer-pat-batha-1.pdf	Story	12
Surjyer-Manush.pdf	Story	21
Tanybi.doc	Story	79
Tarum A Ranjuni-Chakma-Pages.pdf	Other	16
Teen-bondhur-golpo.pdf	Story	13
Text-Book-Chakma-pdf.pdf	Story	1405
Thurong-Barite-Raja.pdf	Story	43
Tin-bondhur-gacher-kotha.pdf	Story	15
Tiya-Pakhi-1.pdf	Story	23
chakma novel hlachinu.docx	Novel	1571
chedon akkan(10).pdf	Article	103
diarrhea-hole-ki-Korbo.pdf	Article	18
ghila khara class 3 p. 62.docx	Story	133
kajer-Kotha.pdf	Story	11
kochpanar rubo rega.docx	Story	151
mle- 2 ananda babu.docx	Poem	174
tin fagala-1.docx	Novel	1765
Changma Ekbacchya Kodha2.doc	Other	170
Chadi 2 Pojhodhe.docx	Novel	1209

Table 4: Names of the sources of our Chakma monolingual data with details (Part 2).

ASCII Font list of Chakma
BivunabaKhamaC
BijoygiriDPC
Udoy Giri
Alaam
Arjyaban
Chakma(SuJoyan)
Punong Jun

Table 5: Legacy fonts identified in our data sources, which map Chakma glyphs onto ASCII code points and are converted to the Unicode Chakma font RibengUni during corpus normalization.

System	BN→CCP				CCP→BN			
	Dev		Test		Dev		Test	
	BLEU	chrF	BLEU	chrF	BLEU	chrF	BLEU	chrF
BanglaT5	**10.9** ± 0.06	**36.3** ± 0.22	**2.7** ± 0.05	**30.6** ± 0.21	**24.4** ± 0.11	**44.6** ± 0.08	**11.7** ± 0.19	**37.9** ± 0.09
BanglaT5†	0.0 ± 0.02	0.5 ± 0.04	0.0 ± 0.00	0.5 ± 0.18	2.3 ± 0.04	14.5 ± 0.09	0.1 ± 0.02	14.3 ± 0.37
mT5-small	**8.5** ± 0.11	**32.2** ± 0.20	**2.0** ± 0.04	**25.8** ± 0.12	**14.1** ± 0.33	**33.9** ± 0.35	**4.6** ± 0.07	**27.2** ± 0.25
mT5-small†	0.2 ± 0.17	9.1 ± 0.91	0.0 ± 0.00	7.7 ± 0.23	0.8 ± 0.04	11.9 ± 0.99	0.1 ± 0.00	11.7 ± 0.37
GPT-4.1(R)	-	-	**1.6** ± 0.13	**30.4** ± 0.12	-	-	**16.5** ± 0.12	**48.2** ± 0.35
GPT-4.1(R)†	-	-	0.4 ± 0.12	18.3 ± 0.12	-	-	0.3 ± 0.10	18.2 ± 0.21

Table 6: Ablation study comparing models evaluated *with* and *without* transliteration on BN→CCP and CCP→BN translation. Rows marked with † indicate evaluation **without transliteration** (native Chakma script as input and output). Comparison includes BanglaT5 and mT5-small fine-tuned models, as well as GPT-based in-context learning (ICL) models with random sampling of 400 examples. Metrics report mean ± std for BLEU and chrF on the development and test sets. Removing transliteration yields near-zero performance for the selected pretrained and ICL models under this experimental setup.

Parameter	RNN	Trans.	T5
Max Epochs	-	-	5
Max Train Steps	20000	20000	-
Warmup Steps/Ratio	4000	4000	0.1
Learning Rate	0.0005	0.0001	0.0005
Batch Size	16	32	16
Max Length	128	128	128
Optimizer	adam	adam	adam
Vocab size	2000	10000	-
Beam width	5	5	5
Clip gradient	1.0	1.0	-
Label Smoothing	0.2	0.5	0.3
d_model	-	512	-
dropout	-	0.2	-
layer_dropout	-	0.1	-
att_heads	-	1	-
ffn_dim	-	512	-
blocks	-	6	-
rnn_dropout	0.3	-	-
layer_normalization	True	-	-
layers	1	6	-
word_embedding	512	-	-
hidden_embedding	1024	-	-
weight_decay	-	-	0.01

Table 7: Final training hyperparameters selected based on validation performance for from-scratch RNN and Transformer models, and for fine-tuning pretrained T5-based models (BanglaT5 and mT5-small). mBART was fine-tuned using the same hyperparameter settings as T5.

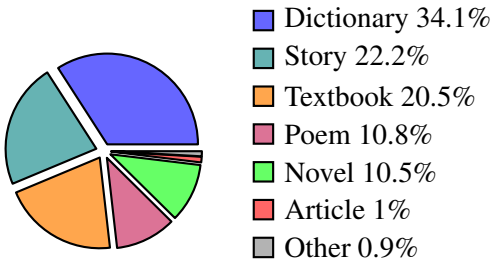

Figure 2: Distribution of Chakma monolingual data by content type. The corpus contains 42,783 monolingual samples collected from diverse sources, including dictionaries, stories, textbooks, poems, novels, and articles.

Bangla Characters	Direction	Chakma Characters
শ, ষ	→	𑄛
ড়, ঢ়	→	𑄜
ঋ	→	𑄟
ৎ	→	𑄠
ওআ	←	𑄋

Figure 3: Nearest-character substitutions used in the Chakma–Bangla transliteration system for characters without direct one-to-one mappings. These substitutions approximate pronunciation and lexical identity while potentially neutralizing orthographic distinctions. The only entry marked with a dash (–) in the Bangla column corresponds to a rare Chakma prosodic lengthening marker that lacks an explicit Bangla graphemic equivalent and is normalized during transliteration. All four Chakma characters without direct Bangla counterparts are rare in the collected data; their semantic effect is not independently evaluated.

You are given translation examples from Chakma to Bangla below:

Chakma Example 1: মরে নুয়া গরি পিজোর্ ন গরিবে
Bangla Example 1: আমাকে পুনরায় জিজ্ঞাসা করবে না ।

Chakma Example 2: গিগির্রানা
Bangla Example 2: কাঁপা

...
...

Chakma Example K: এ সভাগুনৎ কোরাম্ পুরেবাৎয়াই সরিক্ রাস্ট্রআনির্ তিনভাগর্ দিভাগ্ হাজির্ থা -পরিব ।
Bangla Example K: অংশগ্রহণকারী রাষ্ট্রের দুই -তৃতীয়াংশ উপস্থিতি ফোরাম হিসাবে বিবেচিত হবে ।

Provide the Bangla Translation of the Chakma provided below. Only provide the translation and do not output anything else.

Chakma Test 1: বলানয়ায়ান্ গুঝিটৈলো সুনানু মন্ত্রি কোহেয়দে , ধয় কত্তেদি সুনজুক্কানেদয়ই তে কাতার্ বেরা জেব ।
Chakma Test 2: সুনানু মন্ত্রা সেখ্ খাসিনা জধা ভিচ্চে আরব্ আমিরাদৎ (ইউএই) তিন্ দিনর্ সরাকারি পর্ভাচ্ বিদি এচ্চে রেদোৎ দেঝৎ লুহ্গেগি ।
...
...
Chakma Test 20: এ আলহটয়্ সুক্রবার্ পুলিঝর্ বেগ দাঙর্ চোক্ দিয়েবো (আইজিপি) জাবেদ্ পাটোয়ারি কোয়াহেদ , ভুক্ক গরিয়ে কিঐো জনি তুহ্বিচ্ ন চাহ্না , সালেন্ পুলিসসুনে নিজে গিরোহচ্ ঠৈহোনে তুহ্বিচ্ চাহ্নাক্ ।

Figure 4: Prompt format used for our few-shot in-context learning (ICL) experiments, illustrating the structure of source-target examples, task instructions, and test-time input.

System	#Ex.	Random Sampling				N-gram Similarity Sampling			
		BN→CCP		CCP→BN		BN→CCP		CCP→BN	
		BLEU	chrF	BLEU	chrF	BLEU	chrF	BLEU	chrF
GPT-4.1	100	1.3 ± 0.13	28.7 ± 0.40	16.2 ± 0.23	47.8 ± 0.12	1.4 ± 0.17	30.4 ± 0.30	16.9 ± 0.31	48.7 ± 0.28
	200	1.3 ± 0.07	29.5 ± 0.20	**16.8** ± 0.53	47.8 ± 0.14	1.4 ± 0.15	30.5 ± 0.35	17.5 ± 0.28	49.1 ± 0.34
	400	**1.6** ± 0.13	**30.4** ± 0.12	16.5 ± 0.12	**48.2** ± 0.35	**1.5** ± 0.10	**31.3** ± 0.53	**17.8** ± 0.12	**49.6** ± 0.12
GPT-4.1-mini	100	0.9 ± 0.05	27.6 ± 0.26	10.3 ± 0.37	40.2 ± 0.27	**1.2** ± 0.10	28.6 ± 0.35	**11.1** ± 0.50	**40.7** ± 0.30
	200	**1.1** ± 0.14	28.1 ± 0.31	**10.9** ± 0.20	**40.6** ± 0.12	1.1 ± 0.02	28.5 ± 0.24	**11.1** ± 0.03	**40.7** ± 0.10
	400	**1.1** ± 0.06	**28.4** ± 0.19	**10.9** ± 0.19	40.0 ± 0.58	**1.2** ± 0.01	**28.7** ± 0.25	10.5 ± 0.54	40.4 ± 0.47
o4-mini	100	1.3 ± 0.04	28.6 ± 0.29	**12.8** ± 0.06	42.7 ± 0.03	1.1 ± 0.05	29.2 ± 0.14	**12.6** ± 0.42	**42.8** ± 0.29
	200	1.5 ± 0.02	28.3 ± 0.51	**12.8** ± 0.21	**42.9** ± 0.16	1.4 ± 0.04	**29.8** ± 0.18	12.4 ± 0.43	42.2 ± 0.60
	400	1.5 ± 0.03	**29.7** ± 0.03	12.5 ± 0.47	**42.9** ± 0.08	1.2 ± 0.07	29.7 ± 0.14	12.2 ± 0.27	42.7 ± 0.48

Table 8: In-context learning (ICL) performance of different GPT variants under identical experimental settings. Results compare random and n-gram similarity-based sampling of in-context examples across varying numbers of demonstrations (#Ex.). BLEU and chrF are reported as mean ± standard deviation for BN→CCP and CCP→BN translation. The table enables a controlled comparison of LLM variants under the same ICL conditions. Performance is non-monotonic as the number of demonstrations increases, and n-gram retrieval does not improve every model and demonstration count.

System	EN→CCP		CCP→EN		BN→CCP		CCP→BN	
	BLEU	**chrF**	**BLEU**	**chrF**	**BLEU**	**chrF**	**BLEU**	**chrF**
BanglaT5	1.2 ± 0.06	23.0 ± 0.28	6.5 ± 0.36	28.5 ± 0.32	2.4 ± 0.11	29.2 ± 0.32	12.8 ± 0.12	38.5 ± 0.61
mT5-small	0.2 ± 0.01	11.1 ± 0.88	1.0 ± 0.18	15.6 ± 0.34	1.5 ± 0.05	24.1 ± 0.73	3.0 ± 0.10	23.5 ± 0.34

Table 9: Test-set performance of multilingual fine-tuned models on EN↔CCP translation using BanglaT5 and mT5-small. BLEU and chrF are reported as mean ± standard deviation. The BN↔CCP columns show the corresponding MNMT results and are included for comparison with EN↔CCP.

	Src. 1	জাতিসংঘের সংস্থাগুলো এবং প্রযুক্তি বিশেষজ্ঞদের কাছ থেকে জরুরি প্রস্তুতি
	Ref. 1	[illegible] (United Nations agencies and technical experts on emergency preparedness)
	Src. 2	জাতিসংঘের সংস্থাগুলো এবং প্রযুক্তি বিশেষজ্ঞদের কাছ থেকে জরুরি প্রস্তুতি , বসবাসযোগ্যতা এবং দ্বীপের সুরক্ষার বিষয়ে কোনো ছাড়পত্রের জন্য অপেক্ষা না করে রোহিঙ্গাদের সেখানে স্থানান্তরিত করায় সরকার তার বারবার করা প্রতিশ্রুতি ভঙ্গ করেছে ।
	Ref. 2	[illegible] (The government has also reneged on repeated promises to await clearance from United Nations agencies and technical experts on emergency preparedness, habitability, and safety of the island before relocating Rohingya there.)
SMT	Pred. 1	[illegible] (United Nations ...) [Correctly generates a named entity, but the output is not semantically interpretable.]
	Pred. 2	[illegible] (United Nations' General Secretary ...) [Correctly identifies a named entity, but fails to preserve sentence-level semantics.]
RNN	Pred. 1	[illegible] (United Nations' children and others) [Literal lexical translation is produced, but the output is not semantically interpretable.]
	Pred. 2	[illegible] (United Nations' all and others and why) [Literal lexical translation is produced, but the output is not semantically interpretable.]
Trans -former	Pred. 1	[illegible] (United Nations' United Nations' United Nations' United Nations' help help) [Degenerate repetitive generation with severe semantic degradation.]
	Pred. 2	[illegible] (United Nations' United Nations' all country country country country autistic person's rights rights rights and and and and their their their own own own country country country) [Degenerate repetitive generation with severe semantic degradation.]
BanglaT5	Pred. 1	[illegible] (Preparedness required from United Nations agency and technical expert persons) [Semantically faithful with minor orthographic variation.]
	Pred. 2	[illegible] (The government has reneged on repeated promises to await clearance from United Nations agencies and technical experts on emergency preparedness, habitability, and safety of the island relocating Rohingya there.) [Semantically faithful to the reference with minor orthographic variation.]
GPT 4.1 (ICL)	Pred. 1	[illegible] (United Nations agencies and technical experts on emergency preparedness) [Semantically faithful to the reference, with lexical choices that differ slightly from BanglaT5 while preserving meaning.]
	Pred. 2	[illegible] (The government has reneged on repeated promises to await clearance from United Nations agencies and technical experts on emergency preparedness, habitability, and safety of the island before relocating Rohingya there.) [Semantically faithful to the reference, with lexical choices that differ slightly from BanglaT5 while preserving meaning.]

Figure 5: Qualitative examples of BN→CCP translations generated by different models on short and long sentences, illustrating differences across approaches. Parentheses () provide a literal English gloss of each model output for readability, while square brackets [] give a brief qualitative comparison. In these inspected examples, BanglaT5 and GPT models using in-context learning (ICL) appear more faithful than earlier SMT, RNN, and baseline Transformer approaches, particularly for longer sentences.

Round	BN→CCP→BN		CCP→BN→CCP	
	BLEU	**chrF**	**BLEU**	**chrF**
1	41.55	79.32	38.37	79.69
2	99.97	99.99	97.61	99.46
3	100.00	100.00	100.00	100.00

Table 10: Round-trip transliteration quality up to the third iteration on the Benchmark set. Scores are reported as BLEU and chrF on benchmark sentences and show convergence after two rounds.

Src.	ইফতারের আগে বিশেষ মোনাজাতে দেশ ও জাতির অব্যাহত শান্তি, অগ্রগতি এবং সমৃদ্ধি কামনা করা হয়। (Before the iftar, a special munajat was offered seeking continued peace, progress and prosperity of the nation.)
T1.	𑄌𑄢𑄨𑄟𑄴𑄝𑄴 𑄃𑄢𑄬 𑄌𑄨𑄛𑄳𑄚 𑄌𑄧𑄢𑄚𑄴-𑄃𑄚𑄴 𑄎𑄨𑄢 𑄃𑄃 𑄢𑄇𑄴𑄝 𑄐𑄬 𑄝𑄌𑄧𑄬𑄢 𑄃𑄚, 𑄃𑄧𑄟𑄨𑄢𑄇 𑄃𑄃 𑄝𑄬𑄢𑄇𑄨 𑄇𑄝𑄟𑄴 𑄌𑄧𑄟𑄬𑄢।
T2.	𑄌𑄢𑄨𑄟𑄴𑄝 𑄃𑄢𑄇𑄨 𑄌𑄧𑄢𑄚𑄴𑄚 𑄎𑄨𑄢 𑄌𑄟 𑄢𑄇𑄴𑄝 𑄇𑄝𑄇𑄨𑄌𑄬𑄨 𑄎𑄢𑄚𑄬𑄴, 𑄌𑄧𑄟𑄬𑄢𑄴 𑄃 𑄝𑄬𑄢𑄇𑄨 𑄝𑄢𑄬𑄢 𑄇𑄝𑄟𑄴 𑄌𑄬𑄢।
T3.	𑄌𑄢𑄨𑄟𑄴𑄝𑄴 𑄃𑄬𑄟𑄚𑄴 𑄎𑄨𑄢 𑄌𑄧𑄢𑄚𑄴𑄚 𑄎𑄨𑄢 𑄌𑄟 𑄢𑄇𑄴𑄝 𑄝𑄎𑄢𑄇𑄴 𑄎𑄢𑄬𑄴 𑄌𑄟 𑄝𑄬𑄢𑄇𑄨𑄝𑄴 𑄝𑄢𑄬𑄢 𑄇𑄝𑄟𑄴 𑄌𑄧𑄢।

Figure 6: Illustration of orthographic variation in Chakma, where multiple valid spellings of the same word appear across translations produced independently by three language experts from the same Bangla source. Identical colors highlight variant spellings of the same lexical item.

System	BN→CCP		CCP→BN	
	BLEU	chrF	BLEU	chrF
GPT-4.1	0.5 ± 0.09	21.1 ± 1.44	15.5 ± 0.49	45.9 ± 0.45
GPT-4.1-mini	0.6 ± 0.06	24.2 ± 0.71	9.2 ± 0.13	38.9 ± 0.02
o4-mini	0.6 ± 0.03	22.6 ± 1.41	12.2 ± 0.26	42.2 ± 0.23

Table 11: Zero-shot in-context learning ablation showing translation performance with no in-context examples for GPT-4.1, GPT-4.1-mini, and o4-mini on BN→CCP and CCP→BN. Results are reported as mean ± standard deviation for BLEU and chrF.