Tackling Low-Resource NMT with Instruction-Tuned LLaMA: A Study on Kokborok and Bodo

Deepak Kumar, Kshetrimayum Boynao Singh, Asif Ekbal
Indian Institute of Technology, Patna
{deepakkumar1538, boynfrancis, asif.ekbal}@gmail.com

Abstract

This paper presents a new neural machine translation (NMT) system aimed at low-resource language pairs: English to Kokborok and English to Bodo. The framework leverages the LLaMA3-8B-Instruct model along with LoRA-based parameter-efficient fine-tuning. For translating into Kokborok, the model undergoes an initial continued pre-training phase on a dataset containing 75,000 Kokborok and 25,000 English monolingual sentences, followed by instruction-tuning. This tuning uses a reformulated version of WMT25 dataset, adapted to the Alpaca format to support instructional goals. In the Bodo translation, the model is pre-trained on a more extensive dataset of 350,000 Bodo and 125,000 English sentences, using a similar instruction-tuning approach. LoRA adapters are used to modify the large LLaMA3 model for these low-resource settings. Testing with the WMT25 test dataset reveals modest translation results, highlighting the difficulties in translating for low-resource languages. Translating English to Bodo, the model achieved a BLEU score of 4.38, a TER of 92.5, and a chrF score of 35.4. For English to Kokborok, it yielded scores of 5.59 in chrF, 105.4 in TER, and 0.17 in BLEU. These results underscore the intricacies of the task and highlight the critical need for further data collection, domain-specific adaptations, and improvements in model design to better support underrepresented languages.

1 Introduction

Despite significant advancements in neural machine translation (NMT) and the instruction-tuning of large language models (LLMs), the predominant focus of research and datasets remains on English and high-resource languages. Instruction-tuning, a potent method to align LLMs with human preferences, has demonstrated efficacy primarily in contexts where English data are plentiful. Conversely, low-resource languages frequently experience a dearth of robust foundational models due to the scarcity of both monolingual and parallel corpora. In mitigating this limitation, recent studies investigate cross-lingual instruction-tuning by integrating translation-following demonstrations, which enable English-centric LLMs to generalize to novel languages. Currently, evidence suggests that zero-shot cross-lingual transfer is achievable when instruction-tuning is meticulously calibrated, even in scenarios where only English instructions are employed, although factuality and fluency may be compromised in the target language.

More comprehensively, multistage architectures, such as LinguaLIFT, further exemplify how code-switched translation data and task alignment can enhance reasoning in low-resource environments without dependence on extensive multilingual corpora. In light of these insights, this paper introduces an innovative NMT framework targeting pairs of English to Kokborok and English to Bodo language, utilizing the LLaMA3-8B-Instruct model, fine-tuned with LoRA-based adapters for efficient parameter adaptation. The model is initially pre-trained on monolingual corpora of 75,000 Kokborok and 25,000 English sentences for Kokborok and 350,000 Bodo and 125,000 English sentences for Bodo, subsequently undergoing instruction-tuning on WMT25-supplied English–target language parallel data reformatted into an Alpaca style structure. This process aligns translation objectives with instruction follow-up behaviors within the LLM.

Evaluation in WMT25 test sets reveals modest yet meaningful translation performance: for English to Bodo, BLEU = 4.37, TER = 92.5, and chrF = 35.4; for English to Kokborok, chrF = 5.59, TER = 105.4, and BLEU = 0.17. These findings highlight the considerable challenges inherent in low-resource machine translation and emphasize the crucial need for persistent data collection, domain adaptation, and alignment strategies.

Proceedings of the Tenth Workshop on Machine Translation (WMT), pages 1215–1221
November 8–9, 2025. ©2025 Association for Computational Linguistics

Our key contributions are:

- We present a **LoRA-adapted fine-tuning methodology** to enhance English centric, instruction-following models (LLaMA) for translation tasks in low-resource languages such as Kokborok and Bodo.

- We propose a **sequential pipeline** combining monolingual *pre-training* followed by *instruction-tuning*, aimed at strengthening language representations for underrepresented Indian languages.

- We conduct an **empirical evaluation** that highlights both the limitations and capabilities of instruction-tuned LLaMA models in extremely low-resource translation scenarios.

- We issue a **call for future work** aiming at the expansion of high-quality instruction-tuning datasets, domain adaptation techniques, and architecture-level innovations to support the broader inclusion of under-resourced languages in LLM pipelines.

2 Linguistic Background

In this study, we focus on two low-resource Indian languages, Bodo and Kokborok, both belonging to the Tibeto-Burman branch of the Sino-Tibetan language family. These languages are spoken primarily in the northeastern region of India and share several linguistic characteristics such as tonality, agglutinative morphology, and subject object verb (SOV) word order. Despite their socio-cultural and political significance, both languages remain significantly underrepresented in NLP research and lack sufficient digital resources for computational modeling.

2.1 Bodo Language

Bodo[1], also spelled Boro, is a low-resource language spoken predominantly in the Bodoland Territorial Region of Assam, India. According to the 2011 Indian Census, it has more than 1.4 million speakers and is officially recognized as one of the 22 scheduled languages of India. Bodo serves as a medium of instruction in schools throughout the region and has official status in the local administration. The language has historically used Latin and Bengali scripts, but now primarily uses Devanagari. Linguistically, Bodo follows an order of words SOV and exhibits features such as tonality, agglutination, and rich verbal morphology, including case marking and verb inflections. Despite its cultural significance, Bodo lacks substantial computational resources, making it an important target for low-resource NLP, particularly in areas such as neural machine translation (MT), language modeling, and morphological analysis. Its inclusion in multilingual NLP initiatives contributes both to linguistic inclusivity and to technological equity.

2.2 Kokborok Language

Kokborok[2], also known as Tripuri, is another Tibeto-Burman language of the Sino-Tibetan family, spoken mainly in the state of Tripura and parts of Bangladesh. It is the mother tongue of the Tripuri people and one of the most spoken dialects among the various linguistic varieties of the Tripuri. According to the 2011 Census, it has more than 1 million speakers. Kokborok lacks a standardized orthography; however, the Roman script is increasingly used in digital communication and education, along with some continued use of Bengali and Devanagari scripts based on socio-political and institutional preferences. Like Bodo, Kokborok follows an SOV syntactic structure, is tonal, and agglutinative, with a complex system of verbal inflection and case marking. Despite its status as one of the official languages of Tripura, Kokborok is largely lacking digital resources, with very limited availability of annotated corpora, linguistic tools, or parallel datasets. This presents significant challenges for NLP system development, but also offers an opportunity to explore transfer learning, few-shot adaptation, and cross-lingual methods in the context of extremely low-resource settings.

3 Related Work

3.1 Low-Resource MT for Indic Languages

The domain of low-resource neural machine translation (NMT) has been extensively studied with respect to underrepresented Indic languages through the use of a variety of methodologies (Kakwani et al., 2020). In this context, English to Bodo translation systems constructed using approximately 92K parallel corpus sentences with Transformer-based architectures, thereby achieving BLEU scores reaching 11.01.(Boruah et al.,

[1] https://en.wikipedia.org/wiki/Boro_language_(India)

[2] https://en.wikipedia.org/wiki/Kokborok

2023). Similar efforts for languages such as Manipuri (Singh et al., 2023b, 2024), Assamese, Mizo, and Khasi have employed techniques such as transfer learning (Singh et al., 2023a) and tokenization of sub-words to address challenges associated with data sparsity.

3.2 Pre-training and Instruction-Tuning of LLMs

The adaptation of large language models (LLMs) for resource-constrained languages, achieved (Gao et al., 2024) via continued monolingual pre-training (Sennrich et al., 2016) followed by instruction-tuning, has shown favorable results. Specifically, presents an Estonian instruction follow-up LLM derived from LLaMA-2, showcasing that the integration of constrained monolingual pre-training with cross-lingual instruction-tuning considerably improves performance on Estonian tasks. Furthermore, they have introduced Alpaca-est, which represents the first general-task instruction dataset for the Estonian language.

3.3 Cross-Lingual Instruction-Tuning

Cross-lingual instruction-tuning where instructions and translation tasks appear in both English and the target language has been proposed as a cost efficient adaptation method for English centric LLMs (Gao et al., 2024). Such tuning is often implemented in Alpaca style formats to align translation and general task objectives. The inclusion of translation-style instructions during tuning has been empirically shown to boost performance in target languages, especially when paired with monolingual pre-training.

3.4 Parameter Efficient Fine-Tuning (PEFT)

Parameter efficient fine-tuning techniques like LoRA have become increasingly popular for adapting massive LLMs (Hu et al., 2021) to new tasks and languages under compute constraints. PEFT methods enable efficient adaptation without needing to update all model weights, making them especially appealing for low-resource settings.

3.5 Data Augmentation and Synthetic Parallel Corpora

Data augmentation methods especially back-translation remain vital for enhancing low-resource MT performance. These methods generate synthetic parallel data from monolingual corpora (Raja and Vats, 2025), though the quality depends heavily on the strength of the reverse translation model. Controlled generation techniques (e.g., using target sentence length tokens) have also been explored to manage translation output fidelity.

3.6 Evaluation Metrics and Limitations in Low-Resource MT

Evaluation for low-resource MT frequently relies on automatic metrics such as BLEU (Papineni et al., 2002), chrF (Popović, 2017), and TER (Snover et al., 2006). While COMET (Rei et al., 2020), a neural based metric, often correlates better with human judgments, its availability remains limited for languages like Bodo and Kokborok. Human evaluation is crucial to assess fluency and adequacy, particularly for morphologically complex languages.

3.7 Gap and Our Contribution

In the domain of low-resource Indic machine translation (MT), despite developments in parameter efficient fine-tuning (PEFT) techniques, instructional tuning frameworks, and data augmentation methodologies, the translation of Kokborok and Bodo employing instruction following large language models (LLMs) such as LLaMA3 with LoRA adapters remains unexplored. This investigation integrates monolingual pre-training for Bodo and Kokborok with Alpaca style instruction-tuning and LoRA fine-tuning. The study evaluates the performance of the constructed models using the WMT25 (Pakray et al., 2025, 2024; Pal et al., 2023; Kakum et al., 2023) English to Bodo and English to Kokborok datasets and notes limited BLEU (≈ 0.17 for Kokborok), chrF, and TER scores. These results highlight the considerable challenges that persist and stress the necessity for ongoing research specific to these languages.

4 Dataset Preparation

For this study, we collected a substantial amount of monolingual data in Bodo, Kokborok, and English. The dataset comprises approximately 350,000 sentences in Bodo, 75,000 sentences in Kokborok, and 450,000 sentences in English. All data was legally compliant for research use.

For pre-training, we constructed two datasets, one for English to Bodo and another for English to Kokborok, by mixing monolingual data in the following ratios:

- **English to Bodo:** 75% Bodo, 25% English

- **English to Kokborok:** 70% Kokborok, 30% English

These specific ratios were chosen to ensure that the base model becomes more familiar with Bodo or Kokborok while preserving sufficient English fluency to prevent catastrophic forgetting.

We also prepared an instruction-tuning dataset in Alpaca format. We collected WMT25 [3] parallel training data for English to Bodo and English to Kokborok and converted it into Alpaca style instruction response pairs. To enhance the robustness of the model, we maintained a variety of instruction types so that the model learned to follow multiple instruction formats. Finally, we trained two LoRA-adapted LLaMA models using these instruction-tuned datasets for English→Bodo and English→Kokborok translation.

Stage	Language Pair	Sentences
Pre-training	English (eng)	450,000
	Bodo (bodo)	350,000
	Kokborok (trp)	75,000
Fine-tuning	English to Bodo	15,215
	English to Kokborok	2,269
Testing	English to Bodo	1,000
	English to Kokborok	1,000

Table 1: Dataset statistics used for pre-training, fine-tuning, and testing. All monolingual datasets are open-source and licensed under CC BY 4.0. Fine-tuning and testing datasets were converted into instruction response format following Alpaca style instruction-tuning.

5 Methodology

5.1 Monolingual pre-training

We first continued pre-training the LLaMA3-8B-Instruct model on a mixed monolingual corpus for each target language. For English to Kokborok, we use a mix of 75,000 Kokborok sentences 75% and 25,000 English sentences 25% and for English to Bodo, 350,000 Bodo sentences and 125,000 English sentences 75%–25%. These data sets are concatenated and shuffled to increase exposure to the target language while retaining English fluency and minimizing catastrophic forgetting.

Pre-training is conducted using the standard autoregressive language modeling objective, with next token prediction as the training objective. This stage helps the model internalize the vocabulary, grammar, and patterns of the low-resource language before instruction-tuning. As demonstrated in earlier studies (Kuulmets et al., 2024), this step significantly boosts translation performance in low-resource settings.

Following monolingual adaptation, we perform instruction fine-tuning using the WMT25 parallel corpora, consisting of 15,215 sentence pairs for English to Bodo and 2,269 pairs for English to Kokborok. Each sentence pair is reformatted into an Alpaca style instruction response format:

> **Instruction:** Translate the following sentence into [Target Language].
> **Input:** [English Sentence]
> **Output:** [Reference Translation]

This prompt format aligns with the instruction following capabilities of LLaMA-style models and encourages better alignment between English inputs and target-language outputs. No auxiliary tasks or instructions (e.g., summarization or question answering) are included; the dataset is entirely translation-focused.

5.2 LoRA Fine-Tuning

To efficiently adapt the large model to these instruction-tuned datasets, we use Low-Rank Adaptation LoRA (Hu et al., 2021). LoRA adapters are injected into each transformer layer of the model, enabling parameter-efficient fine-tuning. Only low-rank update matrices are trained, while the original model weights remain frozen.

This significantly reduces computational cost and memory usage while maintaining high translation quality. The training objective is the standard cross-entropy loss between predicted tokens and the reference translation, without any post-processing or auxiliary losses.

5.3 Experimental Infrastructure

All pre-training and fine-tuning experiments were conducted on NVIDIA A100 80 GB PCIe GPUs, deployed in a dual-GPU configuration. Each GPU offers up to 80 GB of HBM2e memory with a maximum memory bandwidth of approximately 1.9 - 2.0 TB, enabling rapid data movement, essential for training large-scale models.

This powerful GPU setup provides the computational and memory resources necessary to efficiently pretrain and finetune large language models for low-resource machine translation scenarios.

[3] https://www2.statmt.org/wmt25/indic-mt-task.html

Direction	Setting	BLEU (↑)	METEOR (↑)	ROUGE-L (↑)	chrF (↑)	TER (↓)
en-bodo	Zero-Shot	0.2	0.0060	0.0117	0.7	113.5
	Few-Shot	0.4	0.0098	**0.0201**	0.9	130.5
	Proposed	**4.38**	**0.1318**	0.0090	**35.50**	**92.56**
en-trp	Zero-Shot	0.0	0.0011	0.0023	10.2	155.1
	Few-Shot	0.1	**0.0090**	0.0087	**18.0**	158.9
	Proposed	**0.18**	0.0063	**0.0153**	5.60	**105.49**

Table 2: Comparison of Zero-Shot, Few-Shot, and Proposed Architecture Results. Bold indicates best performance per metric.

The English → Kokborok and English → Bodo experiments were trained for 10 epochs with small batch sizes, a fine-tuning learning rate of 1.5e-5, and regularization through weight decay and warmup. A cosine restart scheduler and gradient clipping ensured stable optimization, while bf16 precision improved speed and memory efficiency, making the setup well-suited for low-resource translation tasks.

6 Results and Analysis

We compare our proposed architecture against zero-shot and few-shot prompting baselines, where the few-shot setup uses the limited bilingual examples from the WMT 2025 shared task as in-context demonstrations. Table 2 reports results across BLEU, METEOR, ROUGE-L, chrF, and TER for English→Bodo (en-bodo) and English→Kokborok (en-trp).

For en-bodo, the proposed model achieves a BLEU score of 4.38, over 10× higher than both zero-shot (0.2) and few-shot (0.4). METEOR similarly improves to 0.1318, compared to 0.0060 and 0.0098. While ROUGE-L (0.0090) trails the few-shot baseline (0.0201), chrF rises sharply to 35.50, far above prompting (0.7/0.9), and TER drops to 92.56, indicating far fewer edits than either baseline. These gains show that our architecture captures structure and meaning beyond what prompting alone enables.

For en-trp, improvements are more modest but still consistent. The model reaches a BLEU of 0.18 (vs. 0.0/0.1) and ROUGE-L of 0.0153 (vs. 0.0087). TER improves markedly to 105.49, compared to 155.1 and 158.9. METEOR (0.0063) is slightly below few-shot (0.0090), and chrF (5.60) remains low, reflecting severe data sparsity and script mismatches.

Overall, our architecture substantially outperforms zero-shot and few-shot prompting across most metrics. Gains are especially pronounced for Bodo, while Kokborok results highlight the extreme difficulty of this language pair but still show measurable progress over prompting-only baselines.

7 Conclusion

Existing literature supports the efficacy of curriculum based training, monolingual continued pre-training, and Parameter Efficient Fine-Tuning (PEFT) techniques for neural machine translation in low-resource settings. Furthermore, preceding investigations have illustrated the advantages of instruction-following fine-tuning for enhancing cross-lingual generalization. Nonetheless, certain gaps persist. Specifically, there has been no research concentrating on the translation of Kokborok and Bodo through the application of instruction-following LLMs, such as LLaMA3-Instruct. Additionally, there has been no exploration of monolingual continued pre-training followed by Alpaca-style instruction-tuning facilitated by Low-Rank Adaptation (LoRA).

Our research endeavors to bridge these gaps through the following approaches: Leveraging monolingual corpora in Kokborok and English to pre-train LLaMA3-8B prior to instruction-tuning. Implementing LoRA-based adapters to tailor the model for translation from English to Kokborok and English to Bodo. Conducting evaluations using WMT25 parallel datasets to establish foundational metrics, such as a BLEU score of 0.17 for Kokborok. Moreover, we underscore the existing methods' limitations and accentuate the necessity for further research in data augmentation and domain specific or architectural adaptation.

Limitations

The proposed (WMT25-INDIC-MT$_{\text{proposed}}$) model demonstrates notable improvements for English →

Bodo translation, particularly with a BLEU score of 4.38 and a significant reduction in TER, its performance on English → Kokborok remains limited, with BLEU scores reaching only 0.18. This highlights the strong dependency of model performance on the size and quality of available corpora, as Kokborok suffers from a much more severe data sparsity compared to Bodo. Furthermore, although gains are observed across most metrics, certain inconsistencies persist; for example, the METEOR score for English → Kokborok in the proposed setup is lower than in the few-shot setting. These results underline that while instruction-tuned LLMs with LoRA-based fine-tuning are promising, their effectiveness remains constrained by low-resource conditions, limited evaluation coverage, and reliance on automatic metrics, which may not fully capture translation quality in morphologically rich languages.

Acknowledgement

The authors gratefully acknowledge the COIL-D Project under Bhashini, funded by MeitY, for providing support and resources that enabled the successful conduct of this research.

References

Parvez Aziz Boruah, Kuwali Talukdar, Mazida Akhtara Ahmed, and Kishore Kashyap. 2023. Neural machine translation for a low resource language pair: English-Bodo. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 295–300, Goa University, Goa, India. NLP Association of India (NLPAI).

Changjiang Gao, Hongda Hu, Peng Hu, Jiajun Chen, Jixing Li, and Shujian Huang. 2024. Multilingual pretraining and instruction tuning improve cross-lingual knowledge alignment, but only shallowly. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6101–6117, Mexico City, Mexico. Association for Computational Linguistics.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

Nabam Kakum, Sahinur Rahman Laskar, Koj Sambyo, and Dr. Partha Pakray. 2023. Neural machine translation for limited resources english-nyishi pair. *Sādhanā*, 48.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Hele-Andra Kuulmets, Taido Purason, Agnes Luhtaru, and Mark Fishel. 2024. Teaching llama a new language through cross-lingual knowledge transfer. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3309–3325, Mexico City, Mexico. Association for Computational Linguistics.

Partha Pakray, Reddi Krishna, Santanu Pal, Advaitha Vetagiri, Sandeep Dash, Arnab Kumar Maji, Saralin A Lyngdoh, Lenin Laitonjam, Anupam Jamatia, Koj Sambyo, Ajit Das, and Riyanka Manna. 2025. Findings of 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. Held at EMNLP 2025.

Partha Pakray, Santanu Pal, Advaitha 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.

Rahul Raja and Arpita Vats. 2025. Parallel corpora for machine translation in low-resource Indic languages: A comprehensive review. In *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*, pages 129–143, Albuquerque, New Mexico, U.S.A. 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.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Loitongbam Sanayai Meetei, Ningthoujam Justwant Singh, Thoudam Doren Singh, and Sivaji Bandyopadhyay. 2023a. 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, Goa University, Goa, India. NLP Association of India (NLPAI).

Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Loitongbam Sanayai Meetei, Sivaji Bandyopadhyay, and Thoudam Doren Singh. 2023b. NITS-CNLP low-resource neural machine translation systems of English-Manipuri language pair. In *Proceedings of the Eighth Conference on Machine Translation*, pages 967–971, Singapore. Association for Computational Linguistics.

Ningthoujam Justwant Singh, Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Sanjita Phijam, and Thoudam Doren Singh. 2024. WMT24 system description for the MultiIndic22MT shared task on Manipuri language. In *Proceedings of the Ninth Conference on Machine Translation*, pages 797–803, Miami, Florida, USA. 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.