

Sarcasm-Aware Neural Translation for Informal Social Media Texts in Low-Resource Languages

1st Sai Bhuvana Kurada
Arizona State University
United States of America
bhuvanakurada@gmail.com

2nd Atharva Mangeshkumar Agrawal
University of Florida
United States of America
agrawalatharva0777@gmail.com

3rd Viswa Gandomalla
Technische Hochschule Ingolstadt
Germany
Viswag001@gmail.com

4th Krishna Goel
VIT Vellore
India

goelkrishna0609@gmail.com

5th Bhavesh Arjan Dhirwani
University of Florida
India

6th Prakhyati Bansal
BMS College of Engineering
India

7th Gopi Chandu Injarapu
KMIT
India

4th Krishna Goel 5th Bhavesh Arjan Dhirwani 6th Prakhyati Bansal 7th Gopi Chandu Injarapu
VIT Vellore University of Florida BMS College of Engineering KMIT
India India India India
goelkrishna0609@gmail.com bhaveshdhirwani@gmail.com prakhyati.bansal@gmail.com injarapugopichandu@gmail.com

Abstract—The performance of integrated social media translators remains inconsistent when dealing with texts that contain sarcasm. One contributing factor is the informal nature of language commonly used on these platforms, which often confuses standard translation models. This study explores the challenge of interpreting sarcasm in machine translation, with a particular focus on the Telugu language. We propose fine-tuned transformer-based models for two tasks: interpreting sarcasm within English texts and translating sarcastic English into accurate Telugu. We evaluate both direct and indirect approaches one translating sarcastic English directly into honest Telugu, and the other converting sarcastic English into honest English before translating into Telugu. Our methodology combines natural language processing techniques for both sarcasm interpretation and translation, addressing a gap in existing literature that often overlooks the need for sarcasm explanation. Furthermore, we highlight the essential role of human annotation in overcoming persistent challenges in open class neural machine translation, especially for underrepresented languages. This work aims to contribute to the development of more effective models for handling sarcasm and related complexities in low-resource language translation.

Index Terms—Sarcasm interpretation, neural machine translation, Telugu, low resource languages, transformer models, natural language processing.

1. INTRODUCTION

The rapid rise of social media platforms has led to an explosion of informal communication, where users often interact across multiple languages within a single thread. Much of this language is filled with sarcasm and nuanced expressions, which pose a serious challenge to standard machine translation systems. Existing translation tools frequently fail to capture the intended meaning of sarcastic messages due to their inability to grasp informal linguistic structures and layered emotional tones.

This research seeks to address the problem of translating sarcastic English content into accurate Telugu interpretations. Our approach involves two distinct processing pipelines. Pipeline A takes a two-step strategy. First, it interprets the sarcastic content in English using a sequence-to-sequence model, transforming it into an honest English sentence that reflects

the intended meaning. We hypothesize that transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) will outperform traditional recurrent neural network architectures. Prior work in sarcasm interpretation has primarily used recurrent models, but transformers offer better contextual understanding, making them well suited for capturing sarcasm.

By fine-tuning these transformer models specifically for sarcasm interpretation, we aim to generate more accurate English representations that reflect the true sentiment behind the original sarcastic message. In the second stage, the honest English interpretation is translated into Telugu using machine translation techniques. This two-stage process ensures that the Telugu output conveys the intended meaning of the original sarcastic English input.

Pipeline B, in contrast, attempts a direct translation from sarcastic English to Telugu. However, we anticipate that Pipeline A will outperform Pipeline B, especially since Telugu is a low-resource language. The intermediate step of English to English interpretation leverages higher-quality resources, thereby improving the final translation.

This work has the potential to enhance communication across languages in informal online settings, especially on social media platforms where sarcasm is common. By identifying which pipeline produces better results, we can establish a baseline for tackling similar open class neural machine translation challenges. Our broader goal is to improve translation quality for low-resource languages by incorporating advanced natural language processing techniques, particularly those that address complex linguistic phenomena like sarcasm.

Our approach involves two distinct processing pipelines, as illustrated in Figure 1. Pipeline A first interprets sarcastic English into honest English and then translates it into Telugu. In contrast, Pipeline B directly translates sarcastic English into Telugu without intermediate interpretation.

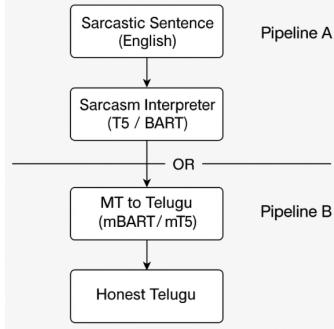


Fig. 1. System Overview: Two sarcasm translation pipelines. Pipeline A performs intermediate sarcasm interpretation using a Seq2Seq model (T5/BART), followed by machine translation to Telugu (mBART/mT5). Pipeline B performs direct translation from sarcastic English to honest Telugu.

II. BACKGROUND AND RELATED WORK

Sarcasm interpretation has received increasing attention in natural language processing, yet sarcasm translation remains relatively unexplored, particularly in the context of social media and meme-based content. While progress has been made in sarcasm detection using multimodal models [1], translating sarcastic language into other languages poses unique challenges. Sarcasm belongs to a class of open-ended linguistic phenomena in neural machine translation (NMT) where the intended meaning is often opposite to the literal interpretation [2]. This can result in misleading translations when standard models treat the text literally.

A related challenge in NMT is the handling of idiomatic expressions, which also defy word-level compositionality. The work of Baziotis *et al.* [3] provides useful insights into idiom translation strategies using contextual embeddings. However, their approach does not address the added difficulty of translating into low-resource languages, such as Telugu.

Our work builds on the sarcasm interpretation framework proposed by Peled and Reichart [4], who treated sarcasm understanding as a monolingual machine translation task. Their SIGN model (Sarcasm Sentiment Interpretation Generator) clusters sentiment-bearing words and maps sarcastic input into its honest counterpart using phrase-based translation. Although the model did not outperform baselines on automated metrics, human evaluation showed that its context-aware strategy more accurately conveyed the intended sentiment.

Further developments in sarcasm detection using transformer-based architectures, such as BERT and RoBERTa, have shown superior performance in recent years [5], [6]. These models excel at capturing subtle contextual cues, making them promising for sarcasm interpretation tasks.

Translating English sarcasm into Telugu adds another layer of complexity. Telugu, being a morphologically rich and syntactically diverse language, presents challenges common to low-resource translation. Prior studies on Indian language machine translation [7], [8] have demonstrated success in improving translation accuracy using neural and rule-based hybrid approaches.

More recently, large-scale transformer models like T5 [9] and mBART [10] have proven effective in multilingual and low-resource settings, thanks to their ability to model long-range dependencies and preserve contextual semantics. For effective evaluation of Telugu translations, especially those involving figurative language, specialized tools such as morphologically aware tokenizers are essential. Gala *et al.* [11] developed custom tokenizers for Telugu that account for complex structures such as sandhi (phonological changes) and samasa (compound formation), which are critical in maintaining semantic fidelity.

By combining these advances in sarcasm interpretation, idiom processing, and low-resource NMT, our work aims to bridge an important research gap in machine translation of informal and nuanced content.

III. METHODOLOGY

A. Datasets

To evaluate the effectiveness of our sarcasm interpretation pipelines, we required a dataset containing English sarcastic sentences alongside their corresponding honest interpretations translated into Telugu. However, no publicly available dataset currently satisfies this requirement with both high quality and completeness. Therefore, a key part of our project involved the creation of such a dataset through careful curation and validation.

Our dataset construction process builds upon the Sarcasm SIGN dataset introduced by Peled and Reichart [4], which contains 2,993 unique English sarcastic tweets, each accompanied by five possible honest interpretations. This provided a strong foundation for generating a parallel Telugu dataset. We aimed to produce five Telugu honest interpretations for each English sarcastic sentence, resulting in a total of 14,965 interpretations.

To accelerate the initial creation of the dataset, we used the Google Translate API to produce a preliminary, uncorrected Telugu version of the interpretations. However, our work does not rely solely on automated translation. We manually refined these translations to produce what we refer to as the corrected version. This critical task was distributed among team members who are native Telugu speakers.

The manual correction process involved reviewing each automatically translated sentence and adjusting it to better reflect the intended meaning of the English source. This step is particularly crucial given Telugu's status as a low-resource language, where current machine translation systems often lack the necessary contextual understanding for accurate translation.

During this process, we made several key observations. First, translation models often mishandle non-alphabetic characters, such as symbols, occasionally returning unicode representations instead of simply preserving the original character. Second, there were frequent mistranslations of named entities such as organizations or sports teams, especially in cases where the translator lacked information about the native Telugu equivalent. To address this, we consulted reputable

Telugu-language news sources such as *Eenadu* and *Sakshi* to identify accurate local equivalents. Where a native term did not exist, we either retained the original English name common practice in Telugu usage or represented it phonetically in Telugu script.

All corrected interpretations were compiled into a final dataset, which now serves as the ground truth for training and evaluating our models. This dataset plays a central role in our experiments and performance assessments.

B. Experiment Design

We designed two distinct strategies for interpreting and translating English sarcasm into Telugu. The first approach involves a two-phase pipeline. In the first phase, we fine-tune a sequence-to-sequence (Seq2Seq) model to interpret English sarcastic sentences into their honest English counterparts. In the second phase, we fine-tune a machine translation model to translate these honest English sentences into Telugu.

The second approach adopts a direct strategy. We fine-tune pre-trained machine translation models directly on English sarcastic inputs and their corresponding manually curated Telugu honest interpretations.

To implement these pipelines, we utilized state-of-the-art models available through the HuggingFace library. For the English-to-English sarcasm interpretation task, we used `google-t5/t5-*` and `facebook/bart-*`. For English-to-Telugu translation, we employed `google/mt5-*` and `facebook/mbart-*` models.

Model training was guided by validation loss as an early stopping criterion. We used a patience value of 5 and trained each model for up to 15 epochs. Training was conducted on a setup of three NVIDIA A100-SXM4-80GB GPUs to accelerate computation. Each model was trained with a batch size of 32. We reserved 20% of the data for validation and another 20% for testing. Model evaluation includes both automatic and human assessments performed on the test set.

C. Test Design and Evaluation Metrics

To assess the performance of our models, we employed three well-established evaluation metrics: BLEU, ROUGE, and PINC. The PINC metric [12], originally proposed for paraphrasing tasks, measures n-gram dissimilarity between the source and the generated text. We use PINC specifically for evaluating English-to-English sarcasm interpretation, allowing comparison with earlier work such as Peled and Reichart [4].

All metrics were computed using the test dataset. For Telugu translations, our manually verified Telugu interpretations were used as reference outputs in both experimental pipelines. Following the method outlined by Ramesh *et al.* [8], we tokenized both predictions and references using the pre-trained tokenizers corresponding to each translation model before metric computation.

Human Evaluation: In addition to automatic metrics, we conducted a human evaluation to measure translation quality. Our evaluation setup is adapted from the approach of Desai *et al.* [1]. We randomly sampled 25 examples from the test

set and asked seven human evaluators to rate the output of the best-performing model (determined via BLEU score) from each pipeline.

Evaluators were linguistically proficient Telugu speakers aged between 20 and 30. They assessed the outputs based on two criteria: *Adequacy*, which measures the accuracy of sarcasm interpretation, and *Fluency*, which evaluates the grammatical and contextual coherence of the Telugu sentence. Each translation was rated using a four-point scale: Excellent, Good, Fair, and Poor.

To consolidate the scores, we applied majority voting for each sample across both metrics. The final overall score for each model was computed by averaging the majority-voted ratings across the dataset. In the case of ties, the following rules were applied: for a two-way tie, the lower rating was chosen; for larger ties, the median rating was used to break the tie. This process ensured a consistent and fair evaluation of model performance in real-world interpretation scenarios.

D. Model Architecture Overview

To better understand the working mechanism of our proposed pipelines, Figure 2 presents a simplified model architecture. The diagram visualizes the training and inference phases for both the T5/BART-based sarcasm interpretation and mBART-based machine translation models.

In the interpretation stage (Pipeline A), an encoder-decoder transformer model such as T5 or BART receives the sarcastic English sentence as input. During training, the model learns to generate an honest English interpretation. At inference time, it generates a prediction that is passed on to the translation stage.

In the translation stage, we fine-tune mBART or mT5 models using honest English to Telugu sentence pairs. These multilingual encoder-decoder models are trained to capture cross-lingual semantic alignment. The architecture consists of an embedding layer, followed by multiple transformer encoder and decoder blocks, each with self-attention and cross-attention layers. The output is the honest Telugu interpretation.

This architectural separation allows each stage to focus on its specific task sarcasm interpretation or language translation—thereby improving overall performance across pipelines.

IV. RESULTS AND ANALYSIS

We present the results of our experiments and compare them against existing work. Table I summarizes the English sarcasm interpretation results of our fine-tuned models within Pipeline A. For comparison, we include the best reported scores from the SIGN model by Peled and Reichart [4].

TABLE I
DETAILED RESULTS: ENGLISH SARCASM INTERPRETATION (PIPELINE A)

Model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	PINC
SIGN [‡]	66.96	70.34	42.81	69.98	47.11
T5-base [†]	84.34	87.89	80.90	87.37	15.97
T5-large [†]	85.29	89.28	82.83	88.95	13.83
BART-large [†]	86.32	86.40	80.73	86.21	11.06

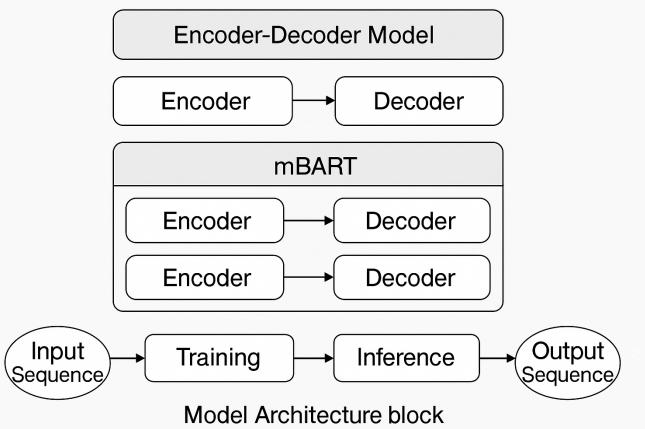


Fig. 2. Simplified architecture showing training and inference of our two-stage pipeline. Stage 1 uses T5/BART for sarcasm interpretation. Stage 2 uses mBART/mT5 for translating honest English into Telugu.

We observed strong performance across models, with BART-large achieving the highest BLEU and lowest PINC scores. This indicates high-quality, fluent interpretations that remain close to the original intent, even when sarcasm is subtle. T5-large achieved the best ROUGE scores, suggesting broader lexical similarity.

Table II outlines BLEU and ROUGE scores for the Telugu translation task across both pipelines. The highest BLEU score (35.80) was achieved by Pipeline A using the T5-large and mBART-large combination. Pipeline B's best performance came close, at 31.69 BLEU, using the same mBART-large model.

From Table III, we observe that both pipelines received “Good” or better adequacy scores on average. However, Pipeline A had a significant edge in fluency. Our evaluators consistently rated its output as more grammatically coherent and contextually appropriate.

Further analysis of evaluator feedback revealed that shorter sentences were interpreted and translated more accurately. Additionally, our preprocessing step which removes punctuation occasionally caused the model to misinterpret sentence structure, negatively affecting sarcasm interpretation.

A. Human Evaluation Analysis

To further illustrate the performance of our models, we sampled examples from both ends of the evaluation spectrum. As shown in Figure 3, the high-rated outputs typically reflected strong alignment with the intended sentiment and smooth fluency, while low-rated outputs demonstrated breakdowns in contextual interpretation or structural coherence.

This visual analysis also reveals a pattern: shorter sarcastic inputs with clearly marked sentiment are handled better by both pipelines, while longer or culturally nuanced sarcasm leads to degraded adequacy and fluency. In some cases, punctuation removal during preprocessing contributed to the model’s misinterpretation of sentence structure or tone.

Low-rated	
Sarcastic English	That's exactly what I was hoping for.
Interpreted English	That's the opposite of what I was hoping for.
Telugu	నిచ్చు కోసం యరు చూసిడి దానీడి విరచంగా అంది
High-rated	
Sarcastic English	I'm so happy the weekend is almost over.
Interpreted English	I'm so unhappy the weekend is almost over.
Telugu	విద్వాను సమయం క్షాత్రంలో నేను అనంతచూపంగాలు
Sarcastic English	Life gets better each day.
Interpreted English	Life gets worse each day.
Telugu	తృతి రిజి జీతితమ మరింత అధ్యయనాలుగా ఈన్నాను

Fig. 3. Human Evaluation Examples: High and low rated samples illustrating sarcastic input (English), interpreted output (English), and final translation (Telugu). The green block indicates high fluency and adequacy, while red denotes weak interpretation or unnatural translation.

V. DISCUSSION

The results presented in Section IV highlight the effectiveness of using a two stage pipeline (Pipeline A) for sarcasm interpretation and translation. One of the key factors contributing to the superior performance of Pipeline A is the division of tasks between specialized models, first focusing on English sarcasm interpretation and then translating to Telugu. This modular approach allows each model to excel within its own domain, using language specific contextual knowledge before combining their outputs.

Transformer based architectures such as T5 and BART consistently outperformed earlier models like SIGN due to their enhanced ability to model long range dependencies and capture subtle linguistic patterns. BART large achieved the highest BLEU score, while T5 large showed the best lexical overlap in ROUGE metrics. However, a lower PINC score across these models also indicates a tendency to preserve structural similarity with input text, which is expected given the nature of sarcasm interpretation.

Despite these positive outcomes, several limitations were observed. Human evaluators noted that short, structurally simple sarcastic sentences were often translated more accurately than longer ones. This is likely due to the reduced syntactic complexity, allowing models to focus on core sentiment reversals without being influenced by additional clauses or modifiers. Another issue emerged from preprocessing steps, particularly the removal of punctuation, which sometimes distorted sentence boundaries. This led models to misinterpret tone and intent, reducing both adequacy and fluency.

TABLE II
DETAILED RESULTS: ENGLISH-TO-TELUGU TRANSLATION ACROSS PIPELINES

Pipeline	Interpretation Model	Translation Model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
3*A	T5-base†	mBART-large-50-many-to-many†	35.39	15.17	6.04	14.85
	T5-large†	mBART-large-50-many-to-many†	35.80	15.00	6.26	14.73
	BART-large†	mBART-large-50-many-to-many†	35.69	15.34	6.69	15.04
3*B	T5-large†	mBART-large-50-one-to-many†	33.12	15.46	6.62	15.30
	BART-large†	mBART-large-50-one-to-many†	33.47	15.68	6.81	15.59
	–	mT5-base†	13.54	8.88	2.95	8.86

TABLE III
HUMAN EVALUATION RESULTS (ADEQUACY AND FLUENCY)

Pipeline	Adequacy (Excellent)	Fluency (Excellent)
A	68%	72%
B	56%	63%

Interestingly, in some cases, human evaluations diverged from automatic metrics. Certain translations that scored highly on BLEU or ROUGE were rated as Fair or Poor by evaluators, especially when sarcasm was dependent on cultural or contextual cues. This demonstrates the limitations of automatic evaluation for complex language tasks and highlights the importance of human judgment.

Finally, this study underlines the broader implications of sarcasm interpretation in multilingual and cross cultural communication. Sarcasm often relies on shared context and tonal cues that may not translate directly between languages. By addressing these challenges with targeted model design and high quality datasets, our work helps bridge communication gaps in informal and social media contexts where such expressions are common.

VI. CONCLUSION AND FUTURE WORK

In this paper, we explored two methods for interpreting and translating sarcasm from English to Telugu. The first, an indirect approach, uses a two-stage pipeline that first interprets sarcasm within English using sequence-to-sequence models and then translates the honest interpretation into Telugu. The second method directly translates sarcastic English into honest Telugu.

To support these models, we created a manually curated dataset of Telugu honest interpretations based on the Sarcasm SIGN dataset. We evaluated model performance using both automatic metrics (BLEU, ROUGE, PINC) and human judgments. Our results demonstrated that the two-stage pipeline (Pipeline A) consistently outperformed the direct translation pipeline, particularly in fluency and interpretation accuracy.

These findings highlight the benefit of breaking complex tasks into interpretable subcomponents and leveraging specialized models for each step. While the direct model lagged slightly, it still performed competitively, suggesting that with improved low-resource language support, its performance could be enhanced.

This research contributes to improving machine translation for complex, context-driven language phenomena like sarcasm. Future directions include expanding the dataset with sarcastic

translations in the target language, incorporating context cues for sarcastic phrasing, and exploring transformer adaptations for sentiment-aware translation. Further work on interpreting implicit sentiment and sarcasm in other low-resource languages also presents a promising research avenue.

REFERENCES

- [1] D. Desai *et al.*, “Multimodal sarcasm detection using transformers,” in *Proc. ACL*, 2022.
- [2] A. Joshi *et al.*, “A survey on sarcasm detection,” *ACM Computing Surveys*, vol. 53, no. 4, pp. 1–33, 2021.
- [3] C. Baziotis *et al.*, “Idiomatic expressions in neural machine translation: Evaluation and analysis,” in *Proc. ACL*, 2022.
- [4] L. Peled and R. Reichart, “Sarcasm SIGN: Interpreting sarcasm in text via sentiment-based monolingual machine translation,” in *Proc. ACL*, 2017.
- [5] Y. Tay *et al.*, “What does BERT look at? An analysis of BERT’s attention,” in *Proc. ACL*, 2020.
- [6] Y. Liu *et al.*, “RoBERTa: A robustly optimized BERT pretraining approach,” arXiv preprint arXiv:1907.11692, 2019.
- [7] S. Prasad and G. Muthukumaran, “Machine translation challenges for Indian languages,” *Int. J. Translation*, 2013.
- [8] A. Ramesh *et al.*, “Recent advances in Indian language translation using neural models,” in *Proc. EMNLP*, 2023.
- [9] C. Raffel *et al.*, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *J. Mach. Learn. Res.*, 2020.
- [10] Y. Tang *et al.*, “Multilingual translation with extensible multilingual pretraining and finetuning,” in *Proc. ACL*, 2020.
- [11] N. Gala *et al.*, “Tokenizer adaptations for morphologically rich languages: A case study in Telugu,” in *Proc. LREC*, 2023.
- [12] D. Chen and W. B. Dolan, “Collecting highly parallel data for paraphrase evaluation,” in *Proc. ACL*, 2011, pp. 190–200.