A Comparative Analysis of LLMs in interpreting Idioms and Proverbs in Konkani

Anonymous ACL submission

Abstract

It is difficult for machines to interpret idioms and proverbs, and more so in lesser known languages such as Konkani. This paper categorizes the ability of the LLMs to comprehend the actual meaning of 50 Konkani idioms and 50 Konkani proverbs. We have compared three language models i.e., GPT-4, Gemini 2.5 Flash, and LLaMA 3 in our work and examined how well each model was able to comprehend the meaning of every idiom and proverb not only literally, but whether it was able to catch the figurative or cultural connotation behind it. Gemini 2.5 Flash provided the best results, whereas GPT-4 performed better in comprehending meanings. Although as observed throughout the research, Gemini 2.5 Flash initially struggled but appeared to get better gradually as if it learned from previous trends. GPT-4 did very well right from the start, although its scores were not always the highest. LLaMA 3 performed comparatively low. Our results indicate the issues which still exist when dealing with regional languages and figurative language.

1 Introduction

India being a large country has so many languages spoken across different places. Language is not only meant for communication but also serves as a means of identity, culture, history and common understanding. One interesting aspect of a language is that it uses figurative speech in the form of idioms and proverbs. But interpreting them involves more than wordy comprehension. These phrases capture the native wisdom, cultural know-how and sense of humor that cannot be literally deciphered and its true meaning may be entirely different. Although much work is already undertaken in Natural Language Processing by targeting international languages like English, low-resource languages lie untouched particularly when it comes to interpreting the figurative sense. In this case study paper, we're highlighting Konkani, a language with more than 2.5 million speakers, mainly across India's western coast in areas such as Goa, coastal Karnataka, Kerala and parts of Maharashtra (census, 2011). Konkani is rich in culture and tradition and boasts idiomatic phrases and proverbs distinctive to the language. But the majority of Large Language Models (LLMs) have been trained and tested heavily on English as well as a few other popular languages, and regional languages such as Konkani are badly represented.

It is necessary to analyze how these models respond to such phrases particularly in low resource languages. This prompted us to try out the models such as GPT-4, Gemini 2.5 Flash and LLaMA 3 on how these models will understand the figurative meaning of the idioms and proverbs expressions in Konkani. We took 50 proverbs and 50 idioms as our dataset and provided them as inputs to three various LLMs models (GPT-4, Gemini 2.5 Flash and LLaMA 3). Then we performed analysis over three parameters: The expression length, the interpretation accuracy (scale 0-3) and faithfulness towards literal and figurative meaning (0/1). We want to know which model is the most accurate in interpretation, which model does not grasp the figurative faithfulness and whether the expression length impacts the model. The paper is divided into 6 sections starting from the abstract. Then comes the introduction of this paper. Next, the related work in other languages. This is followed by the proposed methodology section. The subsequent sections detail the implementation which covers the dataset and the experimental setup. Then we talked about the analysis of experiments,followed by this experiment's limitations and conclusion.

2 Related Work

A great deal of work has been done on English idioms, particularly with state-of-the-art models such as BERT, BiLSTM, and other neural networks. But for Konkani, there is not a great deal of work yet. One recent paper by (Shaikh and Pawar, 2024)employed a neural network to try to determine if idioms in Konkani sentences were being used literally or figuratively and it worked reasonably well. Still, none of the studies so far have looked into how these models understand or explain the meaning behind idioms and proverbs. That's the main focus of our research. The establishment of the Konidioms Corpus by (Shaikh et al., 2024) is a significant advancement, consolidating an extensive collection of konkani idioms that will facilitate future research. Also prior work by (Shaikh, 2020) demonstrated the use of grammatical rules and classification strategies to identify expression in a statement. Shaikh has also attempted to comprehend and translate proverbs in Konkani using context and grammar, and also a trie-based method(Shaikh, 2022). This provides a new perspective to how figurative language in Konkani can be dealt with. Recently, a paper by (Shaikh et al., 2025)on efficient idioms and metaphor classification in Konkani using mBERT with pruning technique.

There are also useful online resources of idioms and proverbs in Konkani, such as Dhyas Konkani (dhy, n.d.a,n), where many of the expressions are listed, and the popular Konkani literature collection (Hegde, n.d.) of Konkani proverbs, which is quoted frequently in language research. Even with all this work, most efforts so far have been about spotting idioms, not understanding what they mean or how models make sense of them. That's the gap we're trying to address by exploring how neural models interpret and explain idioms and proverbs in Konkani.

3 Proposed Methodology

We wanted to evaluate how well language models could understand idioms and proverbs in Konkani. So, we gave each expression to three different LLMs and asked them to interpret the meaning. We focused on three main things: how well the models got the actual meaning right, whether they picked up on the figurative or non-literal sense, and if the length of the expression had any effect on how well the models performed. To find out how accurately LLMs understand, we utilized three models: GPT-4, Gemini 2.5 Flash, and LLaMA 3. We manually input a prompt of each idiom and proverb into each of the three models individually. Each model's output was then copied into an Excel sheet where we manually tagged the response against the actual meaning of proverbs and idioms. We ensured that we utilized the same kind of prompt and input format for each model to make things as uniform as possible. This enabled us to measure their raw understanding impartially without any external influence.

For example in case of Idioms: Prompt "Identify and Interpret the meaning of this sentence: उमेदीर विरजण पडप" Response from ChatGPT-4.0: "To face disappointment after expecting something."

For example in case of Proverbs: Prompt "Identify and Interpret the meaning of this sentence: कोणाच्यो म्हशी, कोण काडटा उटाबशी" Response from Gemini 2.5 Flash : "Someone is taking on a responsibility that isn't theirs, especially when the rightful owner or person responsible is absent or not taking action."

After copying the responses from LLMs into Excel sheet , we then evaluated each LLMs based on how accurately they interpreted the meaning of expression on the scale of 0-3 , 0 point for completely wrong interpretation ,1 point for partially correct interpretation, 2 point for mostly correct but out of context and 3 point for Fully correct interpretation. We then evaluated on the basis of faithfulness to literal vs figurative meaning on a scale of 0/1 . 0 point means the model gave literal or incorrect interpretation and 1 point means figurative meaning. After the rating was complete , we then calculated the average accuracy and also the average figurative understanding of each of the LLMs.

Metric	Average Accuracy	Avg. Figurative Faithfulness	Avg. Length
`ChatGPT-4.0`	1.56	0.58	2.54
`Gemini 2.5 Flash`	2.02	0.74	2.54
`LLaMA 3`	1.40	0.52	2.54

Table 1: Idioms Evaluation

Metric	Average Accuracy	Avg. Figurative Faithfulness	Avg. Length
`ChatGPT-4.0`	1.94	0.68	4.78
`Gemini 2.5 Flash`	2.38	0.70	4.72
`LLaMA 3`	1.90	0.62	4.74

Table 2: Proverbs Evaluation

3.1 Dataset Used

Our dataset included 50 idioms and 50 proverbs in Konkani. These were collected from a mixture of sources like blogs (Dhyas Konkani), books, and local sayings from local resources. After gathering the expressions, we translated them into English and wrote down their actual intended meanings. We tracked and organized all of this in a simple Excel sheet to keep things manageable.

3.2 Experimental Setup

We used a zero-shot setup, meaning the models were not trained or shown examples from our dataset beforehand. To test the models, we gave each expression to all three language models and collected their explanations. Once we had collected the outputs from all three models, our team of five members carefully reviewed each response. We tested them based on two main criteria. First, we assigned a score from 0 to 3 to check how accurately it explained the true meaning of the expression. Then, we assessed whether the model successfully captured the figurative or metaphorical meaning of the expression, giving it a score of either 0 (no) or 1 (yes). Once we finished these evaluations, we averaged the scores for all the idioms and proverbs to see how each model performed overall.

4 Results & Analysis

When we looked at how the models tackled Konkani proverbs, Gemini 2.5 Flash ended up giving the most spot-on responses, with an average score of 2.38. GPT-4 wasn't far behind, scoring 1.94, while LLaMA 3 came in last with 1.90 .As for really getting the deeper meaning of the proverbs, Gemini 2.5 Flash still led the pack with a score of 0.70. ChatGPT was close at 0.68, and LLaMA 3 scored 0.62 (see table 2). The same kind of trend popped up when we tested the models on idioms, too. Gemini 2.5 Flash again had the highest accuracy score at 2.02, followed by GPT-4 with 1.56, and LLaMA 3 at 1.40. Figurative understanding for idioms showed Gemini 2.5 Flash ahead again (0.74), GPT-4 with a moderate score 0.58, and LLaMA 3 lagging 0.52(see table 1). Regarding length, we found that for proverbs there was almost no link between length and performance and all idioms were of similar length, so no clear conclusion could be made there.

So, here's the interesting thing we noticed: Gemini 2.5 Flash started off a bit offbeat. At first, it would miss the point a lot, kind of like it was stumbling over itself. But the more we used it, the better it got, almost like it was learning from its earlier mistakes. By the end, it was actually giving some of the most accurate answers. GPT-4, was pretty consistent. It didn't really change or improve much, but from the start, it was solid ,just dependable answers, especially when it came to understanding things like figurative language and the subtleties of expression. LLaMA 3 was the wild card. It was all over the place, sometimes getting it right, sometimes completely missing the mark. It had a habit of giving literal translations, which didn't always make sense when trying to capture the deeper meaning or cultural context. One thing all of them struggled with, though, was idioms. They're tricky since they're more abstract, and there's less context to lean on. Proverbs, on the other hand, were easier for them to handle probably because they tend to have more of a clear context and meaning.

5 Limitations

The primary limitations observed in this study is the lack of training that is given to all three LLMs in terms of Konkani data. Due to these the LLMs faces difficulty in generating the accurate interpretation.Idioms with dual meanings often posed challenges for the models.Many idioms and proverbs are rooted in regional customs and traditions which leads to misinterpretation. Few models also gave literal translation that were grammatically correct but missed the figurative meaning.

6 Conclusion

This study evaluated how GPT-4, Gemini 2.5 Flash and LLaMA 3 can understand the Konkani idioms and proverbs on three parameters: length, accuracy and figurative meaning. Results showed that Gemini 2.5 Flash gave better results, followed by GPT-4 whereas LLaMA 3 struggled the most with understanding. The length has a very low impact on accuracy of interpretation. This depicts that understanding of figurative expressions depends more on figurative knowledge of models. Our paper can serve as a reference for future research on LLM performance. These models need more training on regional languages and cultural expressions.

References

n.d.a. Konkani idioms and phrases. https://dhyaskonkani.com/konkani-grammar-/konkani-idioms-phrases/.

n.d.b. Konkani proverbs - dhyas konkani. https://dhyaskonkani.com/konkani-proverbs/.

Amey Hegde. n.d. Konkani proverbs compilation. http://www.ameyhegde.in/uploads/8/3/9/2/8392019/amey_hegde_konkani_proverbs.pdf.

Naziya Shaikh. 2020. Determination of idiomatic sentences in paragraphs using statement classification and generalization of grammar rules. In *Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation*, pages 45–50, Marseille, France. European Language Resources Association (ELRA).

Naziya Shaikh. 2022. Grammar and context based approach for identification and translation of proverbs using trie-based ontology. https://rclibrary.rosarycollege.org/wp-content/uploads/2022/02/2-2.pdf.

Naziya Shaikh and Jyoti D. Pawar. 2024. Identification of idiomatic expressions in konkani language using neural networks. In *Proceedings of ICON 2024*.

Naziya Shaikh and 1 others. 2025. Pruning for performance: Efficient idiom and metaphor classification in low-resource konkani using mbert. https://arxiv.org/abs/2506.02005. Preprint.

Naziya Mahamdul Shaikh, Jyoti D. Pawar, and Mubarak Banu Sayed. 2024. Konidioms corpus: A dataset of idioms in konkani language. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9932–9940, Torino, Italia. ELRA and ICCL.