Improving LLM Abilities at Idiomatic Translation

Anonymous ACL submission

Abstract

For large language models (LLMs) like NLLB and GPT, translating idioms remains a challenge as the non-computational nature of idioms may cause traditional Transformer-based systems to translate idioms literally, failing to convey the proper meaning. Previous work has utilized knowledge bases like IdiomKB by providing the LLM with the meaning of an idiom to use in translation. Although this method yielded better results than a direct translation, it is still limited in its ability to preserve idiomatic writing style across languages. Our goal is to enhance translation fidelity by improving LLM processing of idiomatic language while preserving the original linguistic style, ensuring translated texts retain their cultural nuances and emotional resonance. In this research, we expand upon the knowledge base to find corresponding idioms in the target language. We benchmark two methods: The first method employs the SentenceTransformers model to semantically generate cosine similarity scores between the meanings of the original and target language idioms, selecting the best idiom (Semantic Idiom Alignment method, or SIA). The second method uses an LLM to find a corresponding idiom in the target language for use in the translation (LLM-based Idiom Alignment method, or LIA). As a baseline, we performed a direct translation without providing additional information. Human evaluations on the English -> Chinese, Chinese -> English, and Hindi -> English show the SIA method outperformed others in all GPT4o translations. To further build upon IdiomKB, we developed a low-resource Urdu dataset and Hindi dataset containing idioms and their translations. Despite dataset limitations, the SIA method and LLM-based Idiom Alignment method shows promise, potentially overcoming language barriers and enabling the exploration of diverse literary works in Chinese, Urdu, and Hindi.

1 Introduction

The primary challenge faced was enabling large language models (LLMs) to capture the cultural and emotional essence of the original author's words—frequently lost in direct translations (Levin et al.,2014). Idioms particularly highlight this difficulty; they differ significantly across languages and are deeply embedded in cultural contexts, requiring additional cultural knowledge for accurate translation (Fadaee et al., 2018). Previous work has made efforts to enhance LLMs like NLLB and GPT for idiomatic translation and has primarily relied on augmenting these models with knowledge bases such as IdiomKB (Li et al., 2023). These knowledge bases provide meanings to assist in translating idioms. However, current methods still face challenges in preserving the idiomatic style and cultural nuances of the original text (Levin et al., 2014). Despite advancements, existing methods often struggle to maintain the idiomatic writing style in translated texts. The difficulty lies in accurately capturing the cultural and emotional essence embedded in idiomatic expressions, which are highly context-dependent and vary across languages (Shao et al., 2017). This research addresses these challenges by expanding upon existing knowledge bases to include idiomatic expressions from both source and target languages. Specifically, we introduce a novel method termed SIA that utilizes a refined dataset of the chosen language of translation with corresponding idioms that are inserted according to the direct translation from the previous language. This is optimized for SentenceTransformer embeddings (Li et al., 2023). We introduced and benchmarked methods to ensure fidelity in translating idiomatic sentences across languages, validated through human evaluation metrics, alongside compiling a benchmark dataset of Urdu idioms indexed by their English meanings.

2 Related Works

2.1 Limitations in Translation Technology

From a literary standpoint, idioms are figurative, institutionalized expressions that enrich speech and writing, demonstrating mastery of a language. Language models must understand and interpret idioms, especially when translating from one language to another. Recent work has used IdiomKB as a knowledge base for translating idioms, achieving some success with language models (Li, Shuang, et al. “Translate Meanings, Not Just Words: IdiomKB’s Role in Optimizing Idiomatic Translation with Language Models.” ArXiv abs/2308.13961 (2023): n. pag.). This knowledge base pairs idioms in a language with their meanings in English, Chinese, and Japanese. In their method, they use this to provide the translation model with the figurative meaning of the idiom in the sentence. However, this technique fell short of consistently super accurate results. The knowledge base is also relatively small, limited to only three languages, and it does not include any low-resource languages. Building on these techniques for idiomatic translation is the use of retrieval-augmented models (KNN-MT) and the upweighting of training loss on potentially idiomatic sentences (Liu, Emmy, et al. “Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting.” Conference on Empirical Methods in Natural Language Processing (2023)). This showed improvements in translations for idiomatic sentences along with slight improvements in non-idiomatic sentences as well. However, limitations include the use of synthetic data, limited languages, and the heavy reliance on high-quality training data. Past research has focused on translating an idiom in the original language to the figurative meaning in the target language. Although this may convey the message, it fails to be a true translation because the idiomatic sentence style is lost.

2.2 Next Steps to Build On IdiomKB

As evidenced by Li and Chen, the use of specialized knowledge bases such as IdiomKB has proven beneficial. However, the limited scope of these resources, covering only a few languages, constrains their utility in broader linguistic contexts (Li, Shuang, et al. “Translate Meanings, Not Just Words: IdiomKB’s Role in Optimizing Idiomatic Translation with Language Models.” ArXiv abs/2308.13961 (2023): n. pag.). This highlights the need to expand these databases to encompass a wider array of languages and idiomatic expressions. We also hope to build on the use of a knowledge base in idiomatic translation by using it to translate an idiom in one language to an idiom in another language. This would better capture cultural nuances and help maintain the style of the idiomatic sentence across languages.. The inherent complexity of idioms is underscored by research from Dankers and Lucas, who analyze the compositional challenges faced by Transformer models in handling idiomatic expressions. Their findings reveal that while these models adeptly process standard grammatical constructions, they frequently misinterpret the non-compositional nature of idioms, leading to incomplete or incorrect translations ("Can Transformers be Too Compositional? Analyzing Flexibility in Multi-Word Expression Translation," Semantic Scholar (2023)). This suggests that current models need enhancements in semantic flexibility to better accommodate the abnormalities of idiomatic language. Further highlighting the translation challenges, Shao and Sennrich’s evaluation of machine translation performance on idiomatic texts points out that even advanced models struggle to maintain the expressive depth and cultural nuances of idioms, often resulting in translations that are either too literal or misleading ("Evaluating Machine Translation Performance on Text with Idiomatic Expressions," Semantic Scholar (2023)). The necessity for more refined training datasets specifically tailored to improve the handling of idiomatic expressions within translation systems becomes an emphasized need after understanding the limitations of such technology.

2.3 Newer Idiom Knowledge Resources

In response to these challenges, new resources such as the EPIE dataset introduced by Saxena and Paul are emerging. This dataset aims to enhance the identification and translation of idiomatic expressions by providing context-rich examples of their usage across various languages ("EPIE Dataset: A Corpus For Possible Idiomatic Expression Identification," Semantic Scholar (2023)). Such resources are invaluable for developing more sophisticated models capable of recognizing and translating idioms accurately. The work of Liu et al. offers a promising direction through the application of retrieval-augmented models and idiomatic sentence-focused training techniques. Their approach shows improvements in translating idiomatic sentences and enhances the overall fluency of translated texts, suggesting a viable pathway to overcome some inherent limitations of current translation models ("Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting," Semantic Scholar (2023)).

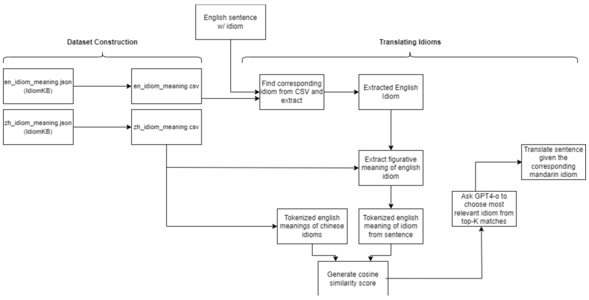
3 Method

3.1 Dataset construction

For the English-to-Chinese translation, we used the “MWE-PIE” (Zhjjn, 2021) dataset that had 1,197 English idioms with around 5 sentences per idiom for a total of 5,170 sentences. For the Chinese-to-English translation, we used the CCT "cheng yu" dataset (Jiang et al., 2018) which had 108,987 Chinese sentences that contained 7,397 unique idioms. For future use with the SIA method, we re-formatted the datasets so the English meaning was the key with the meanings and idioms from other languages as the values. We are indexing on the English meanings so that semantically comparing the English meanings of idioms is made easier(Li et al., 2023). For the Urdu dataset construction, we found a dataset with 2,111 Urdu idioms(with repeats) (ul Hassan, 2024) and their English meanings/idioms. We then found matching English idioms when they existed from our English idiom dataset and, using GPT4o, generated English sentences for those that we did not already have sentences that we flagged. For the Hindi dataset construction, we manually compiled 990 Hindi idioms, Hindi meanings, and Hindi sentences from reputable websites, ensuring there are no duplicates. We generated the English meanings for these idioms from the Hindi meanings using GPT-4o.

3.2 Translating Idioms

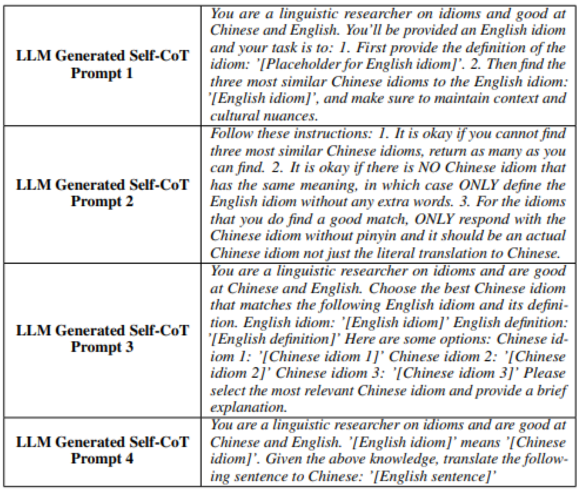
We tested three translation methods: (1) SIA, (2) LLM-Generated Idioms, and (3) Direct Translation. For the EN -> ZH, ZH -> EN, and HI -> EN we evaluated a random subset of 500 sentences and for the EN -> UR we evaluated on 216 sentences. The Urdu idiom dataset was limited because we only translated the idiomatic sentences that had corresponding English and Urdu idioms. All methods were translated with GPT-3.5-Turbo and GPT-4o. For all translations, we set the temperature to 0.7.


English sentence w/ idiom
Dataset Construction
Translating Idioms
en_idiom_meaning.json (IdiomKB)
en_idiom_meaning.csv
zh_idiom_meaning.json (IdiomKB)
zh_idiom_meaning.csv
Find corresponding idiom from CSV and extract
Extracted English Idiom
Extract figurative meaning of english idiom
Translate sentence given the corresponding mandarin idiom
Ask GPT4-o to choose most relevant idiom from top-K matches
Tokenized english meanings of chinese idioms
Tokenized english meaning of idiom from sentence
Generate cosine similarity score

SIA Method In the SIA method, we extracted idioms from sentences and searched for their meanings in the data. Using SentenceTransformers paraphrase-MiniLM-L6-v2, we generated embeddings for English meanings and compared them with target language idioms using cosine similarity with a threshold of 0.7 to find the best match. If no match was found, we used the English meaning for translation. For the idioms that did find a match, we prompted GPT4o to choose/confirm an idiom if the lookup method found corresponding idioms in the dataset. We then translate the sentence while providing the matching target language idiom.

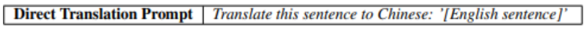
Cosine Similarity Lookup Prompt 1	*You are a linguistic researcher on idioms and are good at Chinese and English. Choose the best Chinese idiom that matches the following English idiom and its definition. English idiom: '[English idiom]' English definition: '[English definition]' Here are some options: '[Chinese idioms]'*
Cosine Similarity Lookup Prompt 2	*'[Chinese idiom 1]' (0.78), '[Chinese idiom 2]' (0.72), '[Chinese idiom 3]' (0.70), '[Chinese idiom 4]' (0.72) Please select the most relevant Chinese idiom and provide a brief explanation.*
Cosine Similarity Lookup Prompt 3	*'[English idiom]' means '[Chinese idiom]'. Given the above knowledge, translate this sentence to Chinese: '[English sentence]'.*

LLM-Generated Idioms Method For the LLM-Generated Idioms method we first use GPT 4o to generate corresponding idioms in the target language that match the idiom in the original language. We give an option for the model to find up to 3 matches, specifically clarifying that it is acceptable to not find any match at all to minimize hallucinations. Then we prompt the model again to choose the best match from the top 3. We do this in order to stay consistent with the GPT confirmation performed in the SIA method. Lastly, we prompt the model to use the top LLM-generated idiom when translating the sentence.



LLM Generated Self-CoT Prompt 1	*You are a linguistic researcher on idioms and good at Chinese and English. You'll be provided an English idiom and your task is to: 1. First provide the definition of the idiom: '[Placeholder for English idiom]'. 2. Then find the three most similar Chinese idioms to the English idiom: '[English idiom]', and make sure to maintain context and cultural nuances.*
LLM Generated Self-CoT Prompt 2	*Follow these instructions: 1. It is okay if you cannot find three most similar Chinese idioms, return as many as you can find. 2. It is okay if there is NO Chinese idiom that has the same meaning, in which case ONLY define the English idiom without any extra words. 3. For the idioms that you do find a good match, ONLY respond with the Chinese idiom without pinyin and it should be an actual Chinese idiom not just the literal translation to Chinese.*
LLM Generated Self-CoT Prompt 3	*You are a linguistic researcher on idioms and are good at Chinese and English. Choose the best Chinese idiom that matches the following English idiom and its definition. English idiom: '[English idiom]' English definition: '[English definition]' Here are some options: Chinese idiom 1: '[Chinese idiom 1]' Chinese idiom 2: '[Chinese idiom 2]' Chinese idiom 3: '[Chinese idiom 3]' Please select the most relevant Chinese idiom and provide a brief explanation.*
LLM Generated Self-CoT Prompt 4	*You are a linguistic researcher on idioms and are good at Chinese and English. '[English idiom]' means '[Chinese idiom]'. Given the above knowledge, translate the following sentence to Chinese: '[English sentence]'*

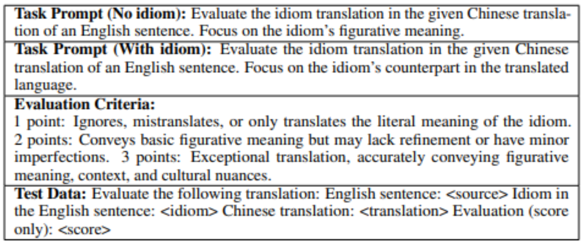
Direct Translation Method The direct translation method simply prompts the model to translate the sentence without providing additional information about the idiom. This method is the baseline that we compare the performance of the other two methods.



Direct Translation Prompt	*Translate this sentence to Chinese: '[English sentence]'*

3.3 Evaluation method

To evaluate the translations, we compared the original sentence and the translated sentence. We used both GPT4 and GPT4o as well as human evaluations. The focus of the evaluation depended on whether the model was instructed to use a specific idiom in the translation. If there was an idiom in the translated sentence we instructed the model to focus on the idiom counterpart, but if there was not an idiom in the translated sentence we instructed the model to focus if the figurative meaning of the idiom was maintained. We did this to ensure that the evaluation prompt was fairly tailored for each translation. We also set the temperature to 0.1 for the evaluations so there is less randomness. Every translation received a score from 1-3 based on the scale outlined in the table below:



Task Prompt (No idiom): Evaluate the idiom translation in the given Chinese translation of an English sentence. Focus on the idiom's figurative meaning.
Task Prompt (With idiom): Evaluate the idiom translation in the given Chinese translation of an English sentence. Focus on the idiom's counterpart in the translated language.
Evaluation Criteria: 1 point: Ignores, mistranslates, or only translates the literal meaning of the idiom. 2 points: Conveys basic figurative meaning but may lack refinement or have minor imperfections. 3 points: Exceptional translation, accurately conveying figurative meaning, context, and cultural nuances.
Test Data: Evaluate the following translation: English sentence: <source> Idiom in the English sentence: <idiom> Chinese translation: <translation> Evaluation (score only): <score>

4 Results

The evaluations from our run presented below reveal the performance of different models for translating idiomatic expressions from English to Chinese, Chinese to English, and English to Urdu. The GPT-4o translations, expectantly, outperformed the GPT3.5-Turbo translations. Regarding the translation model, the GPT-4o evaluations consistently score the translations lower than the GPT4 evaluations(Page 6); the evaluation done by GPT-4o matched more closely with the human evaluations. Using a binary correlation we found that the GPT4o score matched the human evaluation score 65% of the time while the GPT4 score only matched 53% of the time. The superior GPT4o model was more critical of the idiom translations than GPT4, making it a more human-like evaluation. Although the LLM evaluations typically did not score the SIA method the highest, the GPT-4o SIA method scored the highest on the human evaluations(which were evaluated using the same criteria as the LLM), making it a promising and viable method. For the EN->ZH translation, 238 idioms did not find a match, and 262 did. For ZH->EN, 386 idioms did not find a match and 114 did. Despite the dataset not being designed for idiom-to-idiom correlation, the method still found success in translation. The translations that did not find an idiom scored better than the translations that did find an idiom in the LLM evaluations. However, the human evaluations show that the translations that did find an idiom were mostly better translations. This suggests that the LLM is not adequately equipped to assess the accuracy of translations that contain idioms as it prefers the usage of the figurative meaning in the translation over a corresponding idiom. This is likely why the LLM evaluations also favored direct translation as it was better able to assess the accuracy of an idiom -> meaning translation rather than an idiom -> idiom translation. Occasionally the SIA method fell short when the meanings were semantically similar but not the same. For example, "having extremely poor or no vision" ("blind as a bat") was paired with "having small and narrow vision; lacking in foresight ("目光如豆"). These two idioms being considered semantically similar is reasonable but the differences in the meaning account for the poor idiomatic translation. The majority of SIA method usages are successful such as pairing "to remain silent or keep a secret" ("zip one’s lips") with "keep one’s lips sealed, remain silent" ("缄口不言"). The LLM-Generated Idiom method scored lower likely due to the model not producing good idiom translations in the first place compared to the SIA method. The outputted idioms were very sensitive to the prompt as slight variations in the prompt led to varying idioms which could be a reason for the method’s worse performance. The direct translation performed surprisingly well because for simple idioms such as "quality time" it was able to successfully translate it without additional information. For the EN -> UR sentences, 48 sentences were found in the English sentences dataset while 168 were generated by GPT4o. The low resource language results showed the SIA underperforming. We attribute this to the LLM evaluations previously favoring the usage of the figurative meaning in the translation rather than a corresponding idiom, which is especially true here because, for the Urdu idioms dataset, we had a 1:1 correspondence for idioms. Following the trend of the previous translations we hypothesize that human evaluations would show positive results for the SIA method. Similarly, for the HI -> EN translation, the LLM-generated idiom method and direct translation were favored by the LLM evaluations. The human evaluations for the HI -> EN translations show the LLM-generated idiom method performing the best for the GPT3.5-turbo translations and the direct translation performing the best for GPT-4o translations, with the SIA method only scoring slightly worse. Our SIA method and LLM-Generated idiom method prove to be viable, promising methods by being on par and even at times exceeding the direct translation. GPT-4o’s direct translations were successful because they provided simple translations that captured the meaning of the original sentence, even though they lost the idiomatic essence, whereas our methods preserved that idiomatic essence. Overall, both the SIA method and LLM-Generated idiom method had the most complete translations when the corresponding idiom that was chosen was high quality, but direct translation still proved to be adequate at times.

Table 1: Cosine similarity look-up evaluations(En->Zh)

Translation Model	Evaluation Model	Cosine Evaluations	Non-Cosine Evaluations
GPT 3.5	GPT 4.0	2.6527	2.8109
GPT 3.5	GPT-4o	2.3092	2.3193
GPT-4o	GPT 4.0	2.7290	2.9286
GPT-4o	GPT-4o	2.3779	2.5588

Table 2:Cosine similarity look-up evaluations(Zh->En)

Translation Model	Evaluation Model	Cosine Evaluations	Non-cosine Evaluations
GPT 3.5	GPT 4.0	2.4561	2.7798
GPT 3.5	GPT-4o	1.7719	1.8964
GPT-4o	GPT 4.0	2.5439	2.8938
GPT-4o	GPT-4o	2.0526	2.2668

Table 3: LLM-generated idioms evaluations(En->Zh)

Translation Model	Evaluation Model	Idiom:No Idiom Ratio	No Idiom Evaluations	Idiom Evaluations	Total Average Score
GPT 3.5	GPT 4.0	486:14	2.8571	2.7840	2.786
GPT 3.5	GPT-4o	486:14	2.4286	2.3786	2.380
GPT-4o	GPT 4.0	486:14	2.8571	2.7901	2.792
GPT-4o	GPT-4o	486:14	2.6429	2.4403	2.446

Table 4: LLM-generated idioms evaluations(Zh->En)

Translation Model	Evaluation Model	Idiom:No Idiom Ratio	No Idiom Evaluations	Idiom Evaluations	Total Average Score
GPT 3.5	GPT 4.0	494:6	2.8333	2.6356	2.638
GPT 3.5	GPT-4o	494:6	2.0000	1.9291	1.930
GPT-4o	GPT 4.0	494:6	2.8333	2.8036	2.804
GPT-4o	GPT-4o	494:6	2.3333	2.3016	2.302

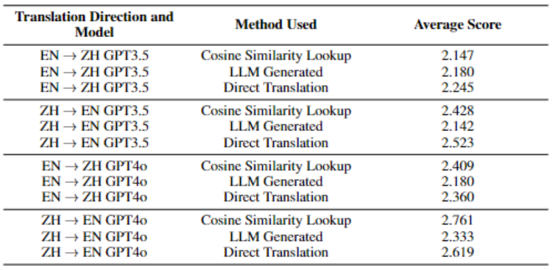
Table 5:Direct translation evaluations(En->Zh)

Translation Model	Evaluation Model	Average Score
GPT 3.5	GPT 4.0	2.776
GPT 3.5	GPT-4o	2.322
GPT-4o	GPT 4.0	2.898
GPT-4o	GPT-4o	2.638

Table 6:Direct translation evaluations(Zh->En)

Translation Model	Evaluation Model	Average Score
GPT 3.5	GPT 4.0	2.754
GPT 3.5	GPT-4o	2.014
GPT-4o	GPT 4.0	2.922
GPT-4o	GPT-4o	2.452

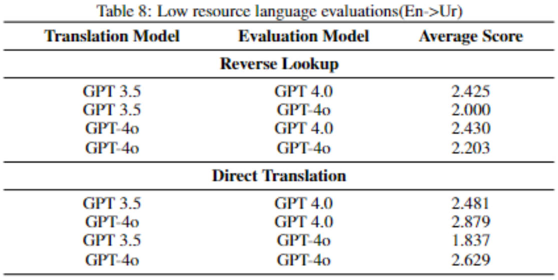
Table 7: Human evaluations



Translation Direction and Model	Method Used	Average Score
EN → ZH GPT3.5	Cosine Similarity Lookup	2.147
EN → ZH GPT3.5	LLM Generated	2.180
EN → ZH GPT3.5	Direct Translation	2.245
ZH → EN GPT3.5	Cosine Similarity Lookup	2.428
ZH → EN GPT3.5	LLM Generated	2.142
ZH → EN GPT3.5	Direct Translation	2.523
EN → ZH GPT4o	Cosine Similarity Lookup	2.409
EN → ZH GPT4o	LLM Generated	2.180
EN → ZH GPT4o	Direct Translation	2.360
ZH → EN GPT4o	Cosine Similarity Lookup	2.761
ZH → EN GPT4o	LLM Generated	2.333
ZH → EN GPT4o	Direct Translation	2.619

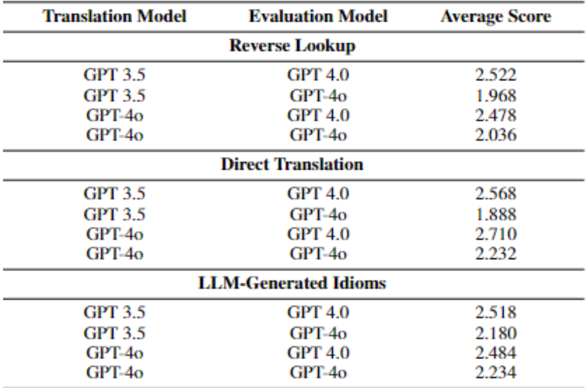
The human evaluator was a part-time Mandarin teacher, who evaluated the idioms using the same evaluation prompt as the LLMs. She received the sentences for evaluations anonymously and wasn't aware of which method was used to create each sentence.

Table 8: Low resource language evaluations(En->Ur)

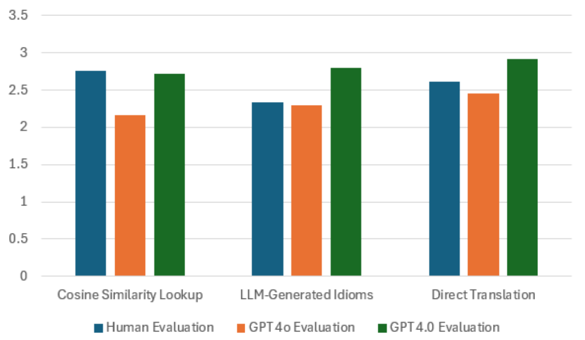


Translation Model	Evaluation Model	Average Score
	Reverse Lookup	
GPT 3.5	GPT 4.0	2.425
GPT 3.5	GPT-4o	2.000
GPT-4o	GPT 4.0	2.430
GPT-4o	GPT-4o	2.203
	Direct Translation	
GPT 3.5	GPT 4.0	2.481
GPT-4o	GPT 4.0	2.879
GPT 3.5	GPT-4o	1.837
GPT-4o	GPT-4o	2.629

Table 9: Low resource language evaluations (Hi->En)

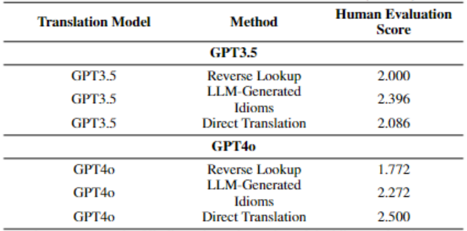


Translation Model	Evaluation Model	Average Score
	Reverse Lookup	
GPT 3.5	GPT 4.0	2.522
GPT 3.5	GPT-4o	1.968
GPT-4o	GPT 4.0	2.478
GPT-4o	GPT-4o	2.036
	Direct Translation	
GPT 3.5	GPT 4.0	2.568
GPT 3.5	GPT-4o	1.888
GPT-4o	GPT 4.0	2.710
GPT-4o	GPT-4o	2.232
	LLM-Generated Idioms	
GPT 3.5	GPT 4.0	2.518
GPT 3.5	GPT-4o	2.180
GPT-4o	GPT 4.0	2.484
GPT-4o	GPT-4o	2.234



ZH->EN Evaluations using GPT-4o for translation

Table 10: Human Evaluations of Translation Methods (Hi->En)



Translation Model	Method	Human Evaluation Score
	GPT3.5	
GPT3.5	Reverse Lookup	2.000
GPT3.5	LLM-Generated Idioms	2.396
GPT3.5	Direct Translation	2.086
	GPT4o	
GPT4o	Reverse Lookup	1.772
GPT4o	LLM-Generated Idioms	2.272
GPT4o	Direct Translation	2.500

The human evaluator was a native Hindi speaker, who evaluated the idioms using the same evaluation prompts as the LLMs. He received the sentences for evaluations anonymously and wasn't aware of which method was used to create each sentence.

5 Limitations

Although the results of the SIA method have been promising thus far, there have been limitations in our work that prevented the method from being an even bigger success.

Finite amount of idioms As stated earlier in the LLM-generated idioms method, we could generate a corresponding idiom in the target language for nearly every original idiom. This yielded a much higher percentage of idioms that found a match, even if they were not all perfect matches. However the IdiomKB datasets, which were used in the SIA method, were composed of English and Chinese idioms without a 1:1 correspondence. There were 8,643 Chinese idioms and 3,990 English idioms. As a result, only about 1/2 of the idioms had a match in the SIA method. Had there been a comprehensive dataset that had both the English idiom and its corresponding Chinese idiom, the method would have been much more effective, which we leave to future work. Further, we leave the expansion of the knowledge base to more low-resource languages as well as exploration of more sophisticated ways to measure semantic similarity that cosine similarity for future work.

Inferior GPT evaluation GPT evaluation does not always strongly mimic human evaluation, especially for Urdu translation, where we lacked access to an Urdu human evaluator.

6 Potential Risks

Although relatively risk-free, some risks associated with translation can come to fruition if left overlooked. Data bias and representation issues within the knowledge base could lead to culturally insensitive or offensive translations. Along the same line of reasoning, language is always evolving, which is why it is important that the knowledge base remains up-to-date, and as comprehensive as possible. If it fails to fit such criteria, misunderstandings could arise, which in important contexts, such as legal, medical, or diplomatic communications could create dire situations.

7 Conclusion

In this paper, we presented advancements in translating idiomatic expressions using LLMs. We evaluated two methods, Semantic Idiom Alignment, and LLM-based Idiom Alignment, using Direct Translation as a baseline. Our findings indicate that the SIA method is particularly effective in preserving idiomatic integrity and achieving higher translation fidelity. Despite sometimes yielding worse results than other methods, the SIA method proved to be an effective and viable option. LIA performed well but fell short compared to the SIA, while Direct Translation often missed idiomatic nuances. Human evaluations confirmed the effectiveness of the Cosine Similarity Look-up method, emphasizing the need for context-aware translations. The impact of this technology can be proven significant when used to enhance communication through more accurate and culturally resonant translations of literary and educational materials. By making literary works more accessible, this research can help bridge cultural gaps and promote cross-cultural literacy and education globally. It profoundly impacts literary and educational communities by preserving the original tone and style of literary works, allowing readers worldwide to experience texts as intended. By enhancing LLMs to maintain the style and tone of messages across languages, we acknowledge the crucial role idioms play in communication and how they can express authors' intent in their work, something that is often lost with direct translation from two languages.

References

Dankers, V., Lucas, C., Titov, I. (2022). Can Transformer be too compositional? Analysing idiom processing in neural machine translation. ArXiv, abs/2205.15301.

Fadaee, M., Bisazza, A., Monz, C. (2018). Examining the tip of the iceberg: A data set for idiom translation. ArXiv, abs/1802.04681.

Haviv, A., Cohen, I., Gidron, J., Schuster, R., Goldberg, Y., Geva, M. (2022). Understanding Transformer memorization recall through idioms. Conference of the European Chapter of the Association for Computational Linguistics.

Levin, L., Mitamura, T., Fromm, D., MacWhinney, B., Carbonell, J., Feely, W., Frederking, R., Gershman, A., Ramirez, C. (2014). Resources for the detection of conventionalized metaphors in four languages. In Proceedings of the 17th International Conference on Computational Linguistics.

Li, S., Chen, J., Yuan, S., Wu, X., Yang, H., Tao, S., Xiao, Y. (2023). Translate meanings, not just words: IdiomKB's role in optimizing idiomatic translation with language models. ArXiv, abs/2308.13961.

Liu, E., Chaudhary, A., Neubig, G. (2023). Crossing the threshold: Idiomatic machine translation through retrieval augmentation and loss weighting. Conference on Empirical Methods in Natural Language Processing

Ram. (2023, November 7). Hindi muhavare(idioms). NCERT Books. https://www.ncertbooks.guru/hindi-muhavare/

Salton, G., Ross, R., Kelleher, J. (2014). Evaluation of a substitution method for idiom transformation in statistical machine translation. In V. Kordoni, M. Egg, A. Savary, E. Wehrli, S. Evert (Eds.), Proceedings of the 10th Workshop on Multiword Expressions (MWE) (pp. 38-42). Association for Computational Linguistics.

Saxena, P., Paul, S. (2020). EPIE Dataset: A corpus for possible idiomatic expressions. ArXiv, abs/2006.09479.

Shao, Y., Sennrich, R., Webber, B.L., Fancellu, F. (2017). Evaluating machine translation performance on Chinese idioms with a blacklist method. ArXiv, abs/1711.07646.

Tang, K. (2022). PETCI: A parallel English translation dataset of Chinese idioms. ArXiv, abs/2202.09509.

Tanwar, R. S. (2023, July 30). The SimpleHelp. https://thesimplehelp.com/hindi-idioms-with-meanings-and-sentences/google$_{v}ignette$

Wehrli, E. (1998). Translating idioms. In COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics.