Literature Review

Musab Iskandar 444003841 **Ahmad Arif** 444002984 **Yousef Koshak** 444000774

1 FLUTE: Figurative Language Understanding through Textual Explanations

The paper introduces FLUTE, a novel dataset designed to advance the understanding of figurative language. By leveraging a model-in-the-loop approach with GPT-3, crowd workers, and expert annotators, the researchers created a comprehensive dataset of 9,000 instances spanning sarcasm, similes, metaphors, and idioms. Each instance includes a literal premise, a figurative hypothesis, entailment/contradiction labels, and crucially, a textual explanation that justifies the labeling. This methodology allows for a more nuanced evaluation of how language models comprehend figurative expressions, moving beyond simple label prediction to understanding the reasoning behind the interpretation.

The key findings demonstrate the dataset's effectiveness in revealing the limitations of current language models. When a T5 model was fine-tuned on FLUTE, it produced significantly higher-quality explanations compared to a model trained on the existing e-SNLI dataset. In human evaluations, the FLUTE-trained model generated explanations that were more contextually relevant, logically consistent, and truly explanatory. Notably, crowd workers found the explanations from the FLUTE-trained model to be much more satisfactory, with a 43.4% increase in "Yes" responses about explanation justification. The research underscores the importance of not just achieving high accuracy, but developing models that can genuinely explain their understanding of complex linguistic phenomena like figurative language.

2 I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors

This paper introduces an innovative approach to generating visual metaphors by collaborating Large Language Models (LLMs), diffusion-based image generation models, and human experts. Using a three-step process with Chain-of-Thought prompting, the researchers developed HAIVMet, a high-quality dataset of 6,476 visual metaphors, demonstrating that DALL-E 2 outperforms Stable Diffusion in creating metaphorical images.

The study highlights the power of human-AI collaboration, with expert evaluations showing that the HAIVMet dataset was preferred 45% of the time over unfiltered outputs and achieved a significant 23-point improvement in visual entailment tasks. By carefully integrating linguistic interpretation, image generation, and human verification, the research establishes a new benchmark in transforming abstract metaphorical language into concrete visual representations.

3 FigCLIP: A Generative Multimodal Model with Bidirectional Cross-attention for Understanding Figurative Language via Visual Entailment

The paper introduces FigCLIP, an innovative multimodal model designed to understand figurative language through visual entailment. By incorporating a bidirectional cross-attention mechanism between text and visual modalities, FigCLIP creates a bridge between figurative expressions and their visual representations. The model architecture extends the CLIP framework

by adding a generative component that can both interpret figurative language and create corresponding visual representations.

The key innovation lies in how FigCLIP handles the alignment between figurative expressions and visual content. The bidirectional cross-attention mechanism allows the model to capture both text-to-image and image-to-text relationships, enabling a more nuanced understanding of figurative language. In experimental evaluations, FigCLIP demonstrated superior performance on visual entailment tasks involving figurative language, achieving a 15% improvement over baseline models. The research also showed that the generative capabilities of FigCLIP could produce contextually appropriate visual representations of figurative expressions, making it particularly valuable for multimodal tasks involving idiomatic and metaphorical language.

4 Enhancing Idiomatic Representation in Multiple Languages via an Adaptive Contrastive Triplet Loss

The paper introduces a novel approach to improve how language models represent idiomatic expressions. The authors propose using adaptive contrastive learning with triplet loss to build an "idiomatic-aware" language model. Their method involves fine-tuning pre-trained models using in-batch positive-anchor-negative triplets, where sentences with idiomatic expressions and their synonyms serve as positive and anchor pairs, while other sentences act as negatives.

The approach achieves state-of-the-art results on the SemEval-2022 Task 2 dataset, demonstrating significant improvements in overall score: 0.548 (compared to previous best of 0.428). A key innovation is their use of a specialized training process that employs a triplet loss function and mining technique to generate high-quality training samples without requiring similarity scores. Their best model shows particularly strong performance on multilingual idiom detection, achieving substantial gains over baselines while using relatively modest computational resources. The paper demonstrates that their method effectively captures the nuanced differences between literal and figurative meanings of expressions across multiple languages.

Our Baseline

For our baseline implementation, we selected "Enhancing Idiomatic Representation in Multiple Languages via an Adaptive Contrastive Triplet Loss," which achieved state-of-the-art results (0.548 overall score) on SemEval-2022 Task 2. We will adapt their triplet loss architecture, which uses sentences with idiomatic expressions and their synonyms as positive-anchor pairs, modifying it to focus specifically on English text processing for our SemEval-2025 task. The baseline's proven effectiveness in distinguishing between literal and figurative meanings, combined with its modest computational requirements, makes it an ideal foundation for our system.

References

Tuhin Chakrabarty, Arkadiy Saakyan, Debanjan Ghosh, and Smaranda Muresan. 2022. FLUTE: Figurative Language Understanding through Textual Explanations. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7139–7159, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tuhin Chakrabarty, Arkadiy Saakyan, Olivia Winn, Artemis Panagopoulou, Yue Yang, Marianna Apidianaki, and Smaranda Muresan. 2023. I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages

7370–7388, Toronto, Canada. Association for Computational Linguistics.

Qihao Yang and Xuelin Wang. 2024. FigCLIP: A Generative Multimodal Model with Bidirectional Cross-attention for Understanding Figurative Language via Visual Entailment. In *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 92–98, Mexico City, Mexico (Hybrid). Association for Computational Linguistics.

Wei He, Marco Idiart, Carolina Scarton, and Aline Villavicencio. 2024. Enhancing Idiomatic Representation in Multiple Languages via an Adaptive Contrastive Triplet Loss. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 12473–12485, Bangkok, Thailand. Association for Computational Linguistics.