Making the Familiar Unfamiliar: AI-Driven Synectics Metaphor Generation for Computational Creativity

Georgios Palamas
Department of Technology and Society
Malmö University
georgios.palamas@mau.se

Abstract

Metaphors are central to human creativity, enabling abstract reasoning and novel conceptual connections. This paper explores how AI can generate and evaluate metaphors using Synectics, a structured creativity method that makes the familiar unfamiliar. We introduce a GPT-4-based pipeline to generate metaphors across four analogy types: Direct, Personal, Symbolic, and Fantasy. A comparison with 1,000 human-created metaphors is conducted using semantic coherence (BERT), novelty (TF-IDF), and GPT-4-assessed creativity. Results show that AI-generated metaphors tend to be more novel but less coherent than human ones, with Fantasy analogies rated highest in creativity but lowest in structure. Abstract themes like Dreams and Knowledge lead to more imaginative outputs. While the system produces stylistically rich metaphors, it also reflects biases and lacks functional evaluation. This work contributes to computational creativity by demonstrating how structured prompting supports controllable metaphor generation. We suggest future directions including hybrid evaluation frameworks that combine computational and human judgment to better assess metaphor effectiveness in context.

Introduction

Creativity often emerges from the ability to perceive the familiar in a new light or to find coherence in seemingly unrelated ideas. Metaphors serve as a fundamental mechanism for human cognition and creativity, enabling abstract reasoning, novel conceptual connections, and artistic expression (Lakoff and Johnson 2008). Mapping familiar ideas onto unfamiliar domains is central to innovation and problem-solving.

A striking example of this process appears in surrealist art, where everyday objects are reconfigured into unexpected, thought-provoking compositions. Dora Maar's *Untitled (Hand-Shell)* (Figure 1) exemplifies this, merging a human hand with a seashell to create a novel hybrid form. This perceptual shift mirrors how metaphors bridge disparate concepts to create new meanings. We explore whether AI can perform similar conceptual transformations through metaphor generation, systematically making the familiar unfamiliar.

Metaphor generation has traditionally been a human task, shaped by cultural, emotional, and experiential factors. With the advent of large language models (LLMs), there is growing interest in whether AI can produce metaphors that are both creative and meaningful. However, existing AI systems often lack conceptual depth, coherence, and contextual awareness.

To address this, we incorporate Synectics, a structured creativity framework encouraging conceptual shifts through analogy-based reasoning (Gordon 1961). Unlike conventional approaches based on direct analogies or statistical associations, Synectics guides metaphor generation using four distinct analogy types: Direct, Personal, Symbolic, and Fantasy. This structure enables AI to move beyond surface-level mappings toward more imaginative transformations.

We systematically evaluate AI-generated metaphors through three key metrics:

1. **Semantic Coherence:** Logical consistency via contextualized sentence embeddings.

2. **Novelty Score:** Conceptual distinctiveness measured against 1,000 human-created metaphors using TF-IDF similarity.

3. **Creativity Score:** GPT-4 self-assessment of metaphor originality and impact.

Our human reference set spans 1,000 literary and culturally significant metaphors categorized into themes such as Time, Love, Memory, and Fear. By applying computational methods, we offer insights into how AI can function as a creative partner in linguistic abstraction and symbolic reasoning.

Results indicate that AI-generated metaphors exhibit higher novelty but lower coherence compared to human-created ones. Fantasy analogies demonstrate the most significant conceptual shifts, aligning with Synectics' emphasis on making the familiar unfamiliar, but they also suffer from reduced structural coherence.

This research contributes a scalable approach to AI-driven metaphor modeling, evaluation, and application. Practical implications arise for creative writing, AI-assisted storytelling, and design ideation. The remainder of the paper is structured as follows: Section 2 reviews related work on metaphor generation, Synectics, and computational evaluation methods. Section 3 details our methodology. Section

4 presents comparative results, followed by a discussion in Section 5. Section 6 concludes with future research directions.

Figure 1: Dora Maar's surrealist photograph *Untitled (Hand-Shell)*. The image illustrates the transformation of familiar objects, a hand and a shell, into an unfamiliar context, analogous to AI metaphor generation.

Related Work

Metaphor Generation. Early approaches to metaphor generation used structured knowledge bases such as ConceptNet and WordNet (Miller 1995), producing template-based metaphors with limited flexibility (Veale, Shutova, and Klebanov 2022). Later, distributional semantics enabled the identification of latent conceptual mappings (Shutova, Sun, and Korhonen 2013). Transformer-based models such as GPT-3 and GPT-4 (Brown et al. 2020), (OpenAI 2023) now generate metaphors through large-scale language modeling, but their outputs often lack grounding or coherent structure (Bondla 2024), (de Lima et al. 2025). Recent surveys highlight the evolution of figurative language generation from rule-based systems to large language models (Lai and Nissim 2024), as well as advances in computational metaphor processing including identification, interpretation, and generation techniques (M. Ge and Cambria 2023). These developments underscore the growing capabilities of LLMs while emphasizing the need for structured creative frameworks to guide metaphor production. Recent tools like Metaphorian (Kim et al. 2023) and MetaPro (Mao et al. 2023) show progress in metaphor assistance and processing, but do not apply structured creative frameworks during generation.

Synectics as Structure. Recent advances in computational creativity and natural language processing have explored metaphor generation through various approaches. Chakrabarty et al. (Chakrabarty et al. 2021) introduced MERMAID, a model that fine-tunes BART using a parallel corpus of metaphorical and literal sentences, employing a generator-discriminator framework to produce expressive metaphors. Stowe et al. (Stowe et al. 2021) extended this work by incorporating conceptual mappings to guide metaphor generation using FrameNet-based embeddings and a controlled sequence-to-sequence model. These methods emphasize large-scale model-based generation with varying degrees of symbolic structure.

Within the computational creativity community, several contributions have addressed creative language modeling and metaphor evaluation. Siekiera et al. (Siekiera et al. 2020) proposed ranking methods for creative language in small data scenarios, including metaphor and humor. Elzohbi and Zhao (Elzohbi and Zhao 2023) provided a comprehensive review of creative data generation, with emphasis on poetic and metaphorical outputs. Muller et al. (Muller, Candello, and Weisz 2023) explored analogy-based design in human-AI co-creative settings, highlighting how conceptual reframing supports ideation and interpretability.

Our work complements these efforts by introducing a structured prompting framework inspired by Synectics, emphasizing analogy type (Direct, Personal, Symbolic, Fantasy) as a controllable axis for metaphor variation. This allows for a more interpretable and targeted exploration of metaphor generation, bridging generative power with stylistic and thematic guidance.

Synectics promotes creativity by encouraging analogical thinking at increasing levels of abstraction (Gordon 1961). While analogy-based reasoning has been explored in computational creativity (Goel, Rugaber, and Vattam 2009), and structured ideation has been supported through tools such as AI_deation (Palamas, Guerra, and Møsbæk 2022), Synectics remains underutilized in metaphor generation. Our work addresses this by applying its four analogy types: Direct, Personal, Symbolic, and Fantasy, as an explicit prompting framework. Unlike chain-of-thought or few-shot methods, Synectics explicitly guides conceptual divergence across multiple axes.

Conceptual Blending and Creative Reasoning. Conceptual Blending Theory provides a complementary perspective on how creativity arises from the merging of disparate mental spaces (Fauconnier and Turner 2002). Unlike Synectics, which offers operational categories for analogy-making, conceptual blending emphasizes the emergent structure formed by integrating partial elements from multiple sources. Prior computational models of blending have demonstrated success in analogy-making and story generation (Veale 2012), but typically lack the structured prompting mechanisms employed in this work. Integrating Synectics into AI systems enables more targeted exploration across creative axes while maintaining interpretability.

Evaluating Metaphor Quality. Evaluating metaphors is challenging due to their subjective and cultural complexity.

Human evaluation remains the gold standard (Dankers et al. 2019) but is resource-intensive. Scalable metrics include semantic coherence via contextual embeddings (e.g., BERT) (Jacob Devlin and Toutanova 2019) and novelty through TF-IDF similarity to human corpora (Shutova 2015). However, TF-IDF lacks semantic nuance. Alternatives such as Word Mover's Distance (WMD) (Kusner et al. 2015) provide deeper semantic comparison but are computationally costly. GPT-4 self-evaluation, while scalable, may introduce bias; we use it as a proxy, not a substitute, and recommend future work incorporating human ratings.

Methodology

This study explores structured metaphor generation through GPT-4 using the Synectics creativity framework, evaluating results via scalable computational metrics. We compare AI-generated metaphors to a curated set of human-created examples across thematic domains and analogy types. To provide a clear overview of the process, Figure 2 summarizes the overall workflow from thematic input to evaluation.

Synectics-Based Prompting

Each prompt requested four analogies for a given theme, corresponding to the four Synectics types:

- **Direct:** Concrete mappings (e.g., "Fear is a shadow.")

- **Personal:** Personification-based (e.g., "If I were fear, I'd be a whisper in the dark.")

- **Symbolic:** Abstract associations (e.g., "Fear is a cracked mirror.")

- **Fantasy:** Surreal, imaginative mappings (e.g., "Fear is a beast that feeds on silence.")

Across both human and AI-generated sets, we used a consistent set of ten thematic domains to prompt metaphor creation: *Love, Time, Memory, Fear, Hope, Dreams, Knowledge, Power, Loss*, and *Freedom*. Each theme served as a seed for generating four distinct analogies based on Synectics' analogy types: Direct, Personal, Symbolic, and Fantasy. This structure enabled comparative analysis across both thematic and stylistic dimensions.

This controlled variation promotes conceptual diversity and allows us to systematically assess the effects of analogy type on metaphor quality. Unlike general prompt-based generation, Synectics ensures a structured exploration across creative axes and facilitates more interpretable comparative analysis across styles of metaphorical abstraction.
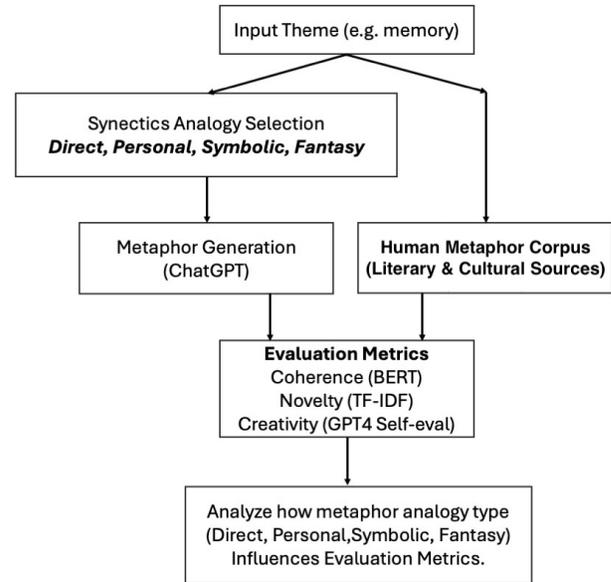

Figure 2: Overview of the metaphor generation and evaluation pipeline.

Human Reference Dataset

We constructed a comparison set of 1,000 human metaphors across ten themes (e.g., Love, Time, Dreams), drawn from literary, idiomatic, and cultural sources. Each metaphor was manually categorized to ensure thematic consistency. This dataset serves as a benchmark for evaluating coherence and novelty. By selecting metaphors from diverse sources, we sought to capture a wide range of figurative expressions and ensure a robust comparison with AI-generated outputs.

Evaluation Metrics

We evaluate metaphor quality across three dimensions:

- **Semantic Coherence**: Assesses the plausibility and interpretability of the metaphor's mapping. Measured via cosine similarity between BERT sentence embeddings of source and target, with higher scores indicating stronger alignment.

- **Novelty**: Captures deviation from familiar expressions, estimated via inverse TF-IDF similarity to a 1,000-metaphor human corpus. Lower similarity implies higher novelty. While interpretable, TF-IDF has lexical limitations; embedding-based alternatives (e.g., Word Mover's Distance) are proposed for future work.

- **Creativity**: Reflects originality, emotional resonance, and conceptual surprise. Each AI-generated metaphor was scored by GPT-4 on a 1–5 scale using the following prompt:

  Evaluate the following metaphor according to creativity on a 1–5 scale. How original, evocative, or surprising is the metaphor? Consider whether it introduces a novel association or emotional depth beyond conventional phrasing. Provide one or two short sentences justifying your score.

GPT-based scoring offers scalable approximation but may inherit stylistic bias—favoring poetic or surreal constructions. Prior work on computational aesthetics and affect (Lunterova, Spetko, and Palamas 2022) stresses the need to consider audience perception. Future iterations will integrate human-in-the-loop evaluations to validate and contextualize results.

Experimental Setup

GPT-4 generated 1,000 metaphors (100 per theme), evenly distributed across analogy types. Each generation batch was conducted under temperature sampling parameters encouraging diversity (temperature = 0.9). We analyzed performance not only by analogy type but also by theme, enabling fine-grained insight into how topic abstraction impacts metaphor quality.

Preliminary observations suggest that abstract themes such as Dreams and Knowledge yielded higher creativity scores, while culturally saturated themes like Love and Power resulted in more conventional, less novel metaphors. This indicates that the degree of thematic abstraction modulates the creative potential of AI-driven metaphor generation, aligning with theories of conceptual space expansion in creativity research.

Results and Discussion

We evaluated 1,000 AI-generated metaphors and compared them to 1,000 human metaphors using semantic coherence (BERT), novelty (TF-IDF), and GPT-4 self-evaluated creativity. These comparisons reveal how structured prompting influences metaphor quality and expose the interplay between analogy type, thematic abstraction, and metaphor effectiveness.

AI vs. Human Comparison

Figure 2 summarizes the performance across all metrics. AI metaphors were more novel but generally less coherent than human metaphors. Creativity scores were comparable, although slightly favoring human references.

Analogy Type Effects

Fantasy analogies scored highest in creativity (4.8) but lowest in coherence (3.8), reflecting Synectics' principle that greater conceptual distance fosters imaginative output at the cost of structure. Direct analogies, in contrast, were highly coherent (4.5) but less novel (3.9). This confirms that analogy type strongly influences output balance and suggests that different analogy styles may be preferentially suited to different application contexts.

Thematic Influence

Themes such as *Dreams*, *Hope*, and *Knowledge* yielded more creative and novel outputs, while themes like *Love* and *Power* resulted in more conventional expressions. This suggests that abstract or less culturally saturated domains offer AI greater conceptual flexibility.

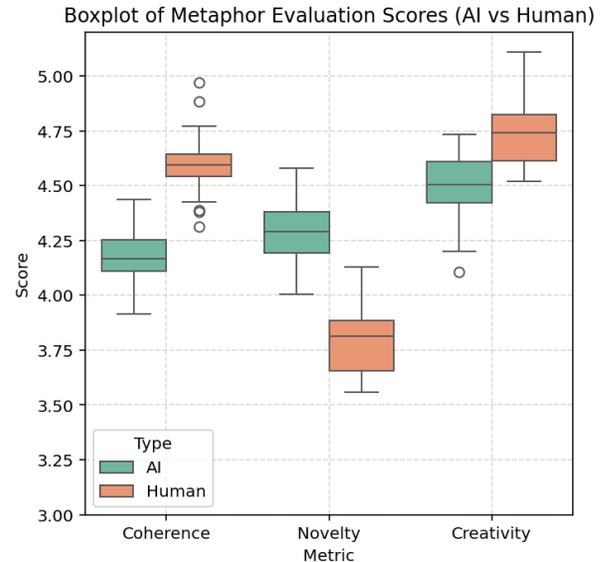Table 1 highlights the top and bottom performing themes by average creativity score.

Figure 2: Boxplot comparing AI and human metaphor scores across coherence, novelty, and creativity.

Theme	Creativity Score (AI)
Dreams	4.8
Knowledge	4.7
Hope	4.7
Power	4.3
Love	4.2

Table 1: Creativity scores across selected themes (AI-generated metaphors).

These results suggest that thematic abstraction modulates AI creative performance, with abstract domains offering a broader space for novel metaphorical constructions, while highly familiar domains may anchor outputs in conventional associations.

Interpretation

Our findings suggest four key observations grounded in the data:

1. **Analogy type affects tradeoffs between creativity and coherence.** Fantasy analogies consistently yielded higher creativity scores but lower semantic coherence, while Direct analogies exhibited the opposite trend. This aligns with Synectics' principle that greater conceptual distance enables more imaginative outputs but increases the risk of disjointed mappings.

2. **Thematic abstraction modulates metaphorical diversity.** Themes classified as more abstract (e.g., *Dreams, Knowledge*) enabled GPT-4 to explore less conventional associations, often resulting in higher novelty and creativity scores. Conversely, culturally saturated themes (e.g., *Love, Power*) tended to anchor metaphor generation in fa-

miliar idioms.

3. **Prompt structure shapes stylistic tone.** The poetic or surreal tone of many AI metaphors, especially in the Fantasy category, can be attributed in part to the Synectics style prompting, which implicitly encouraged narrative or image rich outputs. However, this style may not always align with functional communicative needs.

4. **Utility of metaphors remains unexamined.** While this study focused on stylistic and structural characteristics (creativity, novelty, coherence), we did not evaluate metaphor *function*, i.e., whether a metaphor effectively communicates, persuades, or aids understanding. Future work should assess metaphor performance in goal-directed contexts to evaluate usefulness beyond stylistic metrics.

Example Comparison

To illustrate stylistic differences, Table 2 compares human and AI metaphors across themes using Fantasy analogies.

Theme	Human Metaphor	AI (Fantasy Analogy)
Time	Time is a thief.	Time is a god juggling destinies.
Love	Love is a battlefield.	Love is a star slowly forgetting how to shine.
Memory	Memory is a filing cabinet.	Memory is a time traveler lost in old dreams.
Fear	Fear is a shadow.	Fear is a beast that feeds on silence.

Table 2: Illustrative human vs. AI metaphors by theme and analogy type.

Qualitative Observations

We note an important stylistic asymmetry between the two sets: the human metaphors were typically short, often using familiar "A is B" constructions drawn from idiomatic or literary sources. In contrast, the AI-generated metaphors, particularly in the Fantasy category, were prompted to produce more expressive, poetic outputs.

This discrepancy stems from the lack of explicit formatting constraints placed on the human corpus, which was compiled from pre-existing metaphors rather than instructing humans to follow the same prompting protocol used with GPT-4. As a result, AI metaphors often exhibit more figurative richness but may also reflect prompting artifacts rather than inherent superiority.

Future work will address this imbalance by collecting human-authored metaphors using matched Synectics-style prompts and constraints, enabling a more direct and fair comparison of metaphorical capabilities across agents. Human metaphors tend to draw from idioms and shared cultural knowledge, resulting in accessible but sometimes predictable constructions.

Potential Applications

This structured metaphor generation pipeline has immediate applications across creative writing, education, and ideation. Writers may use Synectics-guided AI outputs as prompts or conceptual seeds, stimulating new narrative directions or thematic elaborations. In education, structured metaphor generation can help students reframe abstract concepts, making difficult ideas more approachable through imaginative mapping.

For example, an instructor teaching abstract scientific concepts such as "entropy" might use Fantasy analogies like "Entropy is a tired orchestra forgetting its symphony" to provoke discussion and deepen understanding. Similarly, design teams could leverage metaphorical prompts during ideation phases to reimagine user experiences or brand narratives in more emotionally resonant terms.

Conclusion and Future Directions

This study introduced a structured prompting strategy for metaphor generation using GPT-4, grounded in the Synectics framework. By varying analogy types and thematic domains, we explored stylistic and semantic differences in AI-generated metaphors and compared them to a human-authored baseline. Our results suggest that certain combinations, such as abstract themes with Fantasy analogies, tend to produce more novel and creative outputs, sometimes at the expense of coherence.

However, our evaluation focused on stylistic aspects such as novelty, creativity, and coherence, without examining the functional utility of metaphors. In real-world contexts such as education or design, the value of a metaphor lies in its clarity and relevance to a given purpose. Relying solely on stylistic metrics may favor poetic abstraction over communicative effectiveness. Future work should include task-based or audience-specific evaluations to assess metaphors in context.

The use of GPT-4 as an evaluator also introduces potential bias, as it may prefer poetic or surreal forms. This raises ethical questions about how language models reflect and reinforce aesthetic norms. To address this, we propose incorporating human evaluation, broader prompt strategies, and assessments that account for both form and function.

Our approach contributes to computational creativity by framing metaphor generation as a process that combines structured prompting with generative power. It supports both theoretical exploration and applied use in fields where metaphor helps transform familiar ideas into unfamiliar perspectives, and vice versa. This aligns with computational creativity's broader aim: balancing structure, expression, and intent.

References

Bondla, S. 2024. Peeping into ai: A literary deconstruction of an essay by ai gpt3 and an interview with ai lamda. *International Review of Literary Studies* 6(1):13–29.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell,

A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33:1877–1901.

Chakrabarty, T.; Zhang, X.; Muresan, S.; and Peng, N. 2021. MERMAID: Metaphor generation with symbolism and discriminative decoding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4250–4261. Online: Association for Computational Linguistics.

Dankers, V.; Rei, M.; Lewis, M.; and Shutova, E. 2019. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2218–2229.

de Lima, E. S.; Neggers, M. M.; Feijó, B.; Casanova, M. A.; and Furtado, A. L. 2025. An ai-powered approach to the semiotic reconstruction of narratives. *Entertainment Computing* 52:100810.

Elzohbi, M., and Zhao, R. 2023. Creative data generation: A review focusing on text and poetry. *arXiv preprint arXiv:2305.08493*.

Fauconnier, G., and Turner, M. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Goel, A. K.; Rugaber, S.; and Vattam, S. 2009. Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *Ai Edam* 23(1):23–35.

Gordon, W. J. 1961. Synectics: The development of creative capacity.

Jacob Devlin, Ming-Wei Chang, K. L., and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Kim, J.; Suh, S.; Chilton, L. B.; and Xia, H. 2023. Metaphorian: Leveraging large language models to support extended metaphor creation for science writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 115–135.

Kusner, M.; Sun, Y.; Kolkin, N.; and Weinberger, K. 2015. From word embeddings to document distances. In *International conference on machine learning*, 957–966. PMLR.

Lai, H., and Nissim, M. 2024. A survey on automatic generation of figurative language: From rule-based systems to large language models. *ACM Computing Surveys* 56(10):1–34.

Lakoff, G., and Johnson, M. 2008. *Metaphors we live by*. University of Chicago press.

Lunterova, A.; Spetko, O.; and Palamas, G. 2022. Computational aesthetics of the collective affective dynamics of imdb movie reviews. In *Porto, Lisbon: 10th Conference on Computation, Communication, Aesthetics & X: 10th Conference on Computation, Communication*, 26.

M. Ge, R. M., and Cambria, E. 2023. A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*.

Mao, R.; Li, X.; Kai, H.; Ge, M.; and Cambria, E. 2023. Metapro online:: A computational metaphor processing online system. In *Proceedings of the 61st annual meeting of the association for computational linguistics*. Association for Computational Linguistics (ACL).

Miller, G. A. 1995. Wordnet: a lexical database for english. *Communications of the ACM* 38(11):39–41.

Muller, M.; Candello, H.; and Weisz, J. 2023. Interactional co-creativity of human and ai in analogy-based design. In *International Conference on Computational Creativity*.

OpenAI. 2023. Gpt-4 technical report. *arXiv preprint* arXiv:2303.08774.

Palamas, G.; Guerra, A. M.; and Møsbæk, L.-D. 2022. Ai_deation: A creative knowledge mining method for design exploration. *EAI Endorsed Transactions on Creative Technologies* 9.

Shutova, E.; Sun, L.; and Korhonen, A. 2013. Metaphor identification using verb and noun clustering. *Proceedings of ACL* 178–187.

Shutova, E. 2015. Design and evaluation of metaphor processing systems. *Computational Linguistics* 41(4):579–623.

Siekiera, J.; Köppel, M.; Simpson, E.; Stowe, K.; Gurevych, I.; and Kramer, S. 2020. Ranking creative language characteristics in small data scenarios. *arXiv preprint arXiv:2010.12613*.

Stowe, K.; Chakrabarty, T.; Peng, N.; Muresan, S.; and Gurevych, I. 2021. Metaphor generation with conceptual mappings. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 6724–6736. Online: Association for Computational Linguistics.

Veale, T.; Shutova, E.; and Klebanov, B. B. 2022. *Metaphor: A computational perspective*. Springer Nature.

Veale, T. 2012. *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity*. Bloomsbury Academic.