

Challenges in Finding Metaphorical Connections

Katy Ilonka Gero

Columbia University

katy@cs.columbia.edu

Lydia Chilton

Columbia University

chilton@cs.columbia.edu

Abstract

Poetry is known for its novel expression using figurative language. We introduce a writing task that contains the essential challenges of generating meaningful figurative language and can be evaluated. We investigate how to find metaphorical connections between abstract themes and concrete domains by asking people to write four-line poems on a given metaphor, such as “death is a rose” or “anger is wood”. We find that only 24% of poems successfully make a metaphorical connection. We present five alternate ways people respond to the prompt and release our dataset of 186 categorized poems. We suggest opportunities for computational approaches.

1 Introduction

Poetry expresses the feelings or emotions of an experience, often relying on figurative language to communicate an otherwise elusive idea. This makes poetry an exciting genre for those interested in generating figurative language.

Recently, researchers have made progress in computationally generating poetry (Ghazvininejad et al., 2016; Veale, 2013). However, in a survey of computer generated poetry, Oliveira (2017) notes that while poetic text must convey a conceptual message, this requirement is “often only softly satisfied”.

We focus on creating intentionally meaningful lines of poetry. Poems generated from a single theme such as “love” can rely on language related to the theme, but are often ambiguous and have no clear meaning. Although ambiguity can be a desirable property in poetry, it makes it difficult to evaluate whether the meaning is intentional, or being attributed by the reader. We propose generating poetry from a metaphor such as “love is a rock”. These poems can still have some ambiguity, but we can evaluate whether readers can detect

Surrender is a book
it's pages contain paragraphs of regret
chapters of inaction
an epilogue of defeat

Figure 1: Example poem for “surrender is a book”.

their metaphorical meaning or not.

In this paper, we introduce a short poetry writing task that contains the essential challenges of generating meaningful figurative language. We establish a baseline for how well amateur writers perform and show that evaluators achieve high agreement.

The task is to write a four-line poem containing a given metaphor such as “love is a rock” or “death is a stream.” Although these poems leave room for interpretation and novelty, we can evaluate whether or not they successfully express the given metaphor. An example poem from our dataset is shown in Figure 1.

Our study generates a dataset that includes successful poems, which generative computers systems may model or use as inspiration, as well as unsuccessful ones, which let us better understand the task and discover common failure points.

This paper makes the following contributions:

- Introducing a writing task that is short and contains the essential challenges of meaningful figurative language.
- A dataset of 186 poems, and their associated meta-data, annotated with their coherence to the prompt metaphor.¹
- A categorization of common failure cases in how a poem relates to its prompt.

¹<http://github.com/kgero/metaphorical-connections>

2 Related Work

Procedural poetry, in which poets use algorithmic processes to create their work, has a long history preceding the invention of modern computers and continues strong today (Parrish, 2018; Montfort, 2017). In computer science, the generation of poetry represents a challenge to generate emotional, creative, and meaningful text.

Some work analyzes the stylistic features of contemporary poetry (Kao and Jurafsky, 2012; Kaplan and Blei, 2007) and others build generative systems that output poems (Netzer et al., 2009; Colton et al., 2012; Manurung et al., 2000). A recent neural-network based system, Hafez (Ghazvininejad et al., 2016), produces rich sounding sonnets. This is a promising computational approach to achieve the stylistic aspects of poetry. However, it is an open problem whether computational approaches can produce the structural or meaningful aspects of poetry.

Generating metaphors is a challenge in artificial intelligence (Veale et al., 2016). Gagliano et al. (2016) use word embeddings to find connector words between two conceptual domains to aid in making metaphorical connections. Veale and Hao (2007) mine metaphorical relations using Google search results for adjectives that describe both terms. Later work (Veale, 2013) generates one line expressions from conceptual metaphors. It remains a challenge to expand a metaphor into a poem that expresses the feelings or emotions of an experience.

3 Experiment and Methodology

In this experiment, we ask 200 amateur writers to write four-line poems that use a given metaphor. Each writer is given one metaphorical prompt.

We base this poetry-writing task on expressing a metaphor because metaphors are a common but challenging aspect of poetry, and we can evaluate whether the poem expresses the given metaphor.

The metaphorical prompts are created by randomly combining one concrete noun and one poetic theme, a technique introduced by Gagliano et al. (2016). We use their lists of concrete nouns and poetic themes, a subset of which are shown in Table 1. Because the concrete and poetic words are paired randomly, we expect this task to be difficult—people may struggle to find a metaphorical connection between the words.

concrete nouns	poetic themes
bed	loss
horse	confusion
bell	faith
book	freedom
ship	grace
wing	hate
wood	jealousy
room	love

Table 1: Example words in the concrete noun and poetic theme lists, from (Gagliano et al., 2016). An example prompt metaphor, created by randomly drawing one word from each list, could be “faith is a horse”.

We recruit 200 people from Amazon Mechanical Turk. Each writer is given one of the following 10 randomly generated metaphorical prompts:

- “Anger is wood”
- “Compassion is blood”
- “Death is a rose”
- “God is a breath”
- “Grace is a garden”
- “Hate is a mist”
- “Hope is a ship”
- “Immortality is a room”
- “Peace is a rock”
- “Surrender is a book”

We ask them each to write a four-line poem coherent with the prompt metaphor. They are told to not use the exact words of the metaphor as given but rather express the idea the metaphor represents. They are also told to use stylistic elements of poetry such as rhyme, alliteration, and line breaks. We collect 20 poems on each of the 10 metaphors. Workers are only allowed to write one poem and are paid \$1 for the task.

The authors of the paper independently evaluate the poems. We analyze the success of the poems by indicating whether or not a poem contained its given metaphor. For poems that did not contain the given metaphor, we used grounded theory (Strauss and Corbin, 1990) to develop categories of how they failed. These categories include: not related at all, containing only one of the concepts, and three non-metaphorical connections. Example poems for each category are found in Figure 2.

<p>A. NO CONNECTION: JUST ABOUT ANGER My anger is solid and vast The ghosts of my present, the ghosts of my past I can't break through the tunnel of time My anger is vicious, my anger is mine</p>
<p>B. ATTRIBUTIONAL CONNECTION: <u>WOOD</u> IS ANGRY The bench of pine wood planks Held up peoples glutes Till the people went home for the night leaving the <u>bench</u>, angry, in abandonment</p>
<p>C. OFFSET CONNECTION: ANGER IS A <u>FIRE</u> <u>Wood</u> is a place where a <u>fire of anger</u> can be lit like a sparrow. Both anger and <u>fire</u> work to light up a room. Wood is the conductor of rage, the spite that turns heads. Like ants marching through its hollow shell, <u>wood</u> is a source of fury.</p>
<p>D. INCOHERENT CONNECTION Thoughts of breeze, my anger is teaming, of rapture I freeze, my boiling pot is steaming, tired of despair, my wood is drying, once so full of anger & rage, now I feel like I am dying.</p>
<p>E. METAPHORICAL CONNECTION the anger grew, like a tree this large, immovable object had taken root casting shade on even the happiest parts of my life I could let it consume me, or <u>cut it down</u></p>

Figure 2: Example poems for the given metaphor “anger is wood”. We show one example for each of the four failure cases and one for a successful metaphorical connection.

4 Results

On average people take 13.6 minutes on this writing task. 14 poems were plagiarized and removed from consideration, leaving 186 poems for the resulting analysis. The two evaluators had 97% observed agreement on whether the poem successfully made the given metaphorical connection. 24% of poems, or 45 poems, were found to be successful by at least one of the evaluators. 7% of poems were off-topic. Similarly the evaluators had 97% observed agreement on whether the poems were off-topic or not.

In the remaining poems, the poem used the words in the metaphor but did not make a metaphorical connection between the words. Our grounded theory found four alternate ways of relating the given concepts in the poem: *no connection*, *attributional connection*, *offset connection*, and *incoherent connection*.

Raters had a 69% agreement on these categories, indicating that it is sometimes ambiguous

which error is made. Sometimes this is due to different interpretations of the poem and sometimes this is due to evaluators determining that a given poem didn't cleanly sit into a single category. For the remaining analysis, if evaluators disagreed on which category to place a poem in, a poem is considered to be in both categories.

The fraction of poems in each category is reported in Table 2. By looking at the other ways poems relate to the prompt, we learn the tactics people use when attempting to complete this task.

4.1 Categorization of Poems

We categorize six distinct ways poems relate to the prompt. We define and discuss the categories below. Figure 2 provides example poems for each category, while Table 2 reports the fraction of poems in each category.

4.1.1 Off-Topic

A poem is *off-topic* if it fails to include aspects of either word in the metaphor. For the prompt “surrender is a book” a poem might be about the loss of a lover, which has no relation to “surrender” or “book”. 7% of poems are off topic. Although people write a poem, this is a case when the worker does not truly attempt to do the task.

4.1.2 No Connection

A poem has *no connection* if it explores the conceptual domain of only one word in the metaphor or does not relate the two conceptual domains. In Figure 1A, the poem talks only about feeling angry, “My anger is vicious”, with no reference or connection to wood. There is only a vague attempt to connect anger with wood in the line “my anger is solid”; although wood is solid, many things are solid and this is not enough to establish a metaphorical connection.

This is the most common failure case for poems, with 41% of all poems placed in this category. Possibly these poems intended to express a connection, but the result was too vague and evaluators couldn't detect one. Alternatively, the writer couldn't find a metaphorical connection and simply wrote what they could about one of the words.

4.1.3 Attributional Connection

A poem has an *attributional connection* if it attributes the abstract concept directly to the concrete noun. In Figure 1B, the poem says the “[wooden] bench, [is] angry”. Although this

prompt metaphor	<i>off-topic</i>	<i>no connection</i>	<i>attributional</i>	<i>offset</i>	<i>incoherent</i>	<i>metaphorical</i>
surrender is a book	0.11	0.53	0.47	0.11	0.05	0.05
death is a rose	0.00	0.30	0.55	0.40	0.20	0.10
god is a breath	0.00	0.26	0.21	0.11	0.47	0.11
grace is a garden	0.00	0.67	0.11	0.22	0.39	0.17
immortality is a room	0.05	0.58	0.05	0.11	0.47	0.21
compassion is blood	0.11	0.37	0.05	0.05	0.21	0.26
peace is a rock	0.10	0.45	0.10	0.10	0.05	0.35
hope is a ship	0.07	0.36	0.14	0.14	0.36	0.36
anger is wood	0.05	0.37	0.00	0.32	0.21	0.37
hate is a mist	0.11	0.26	0.05	0.00	0.26	0.47
all	0.07	0.41	0.17	0.15	0.26	0.24

Table 2: Success rates of the 10 metaphorical prompts. The fraction of successful poems is highlighted in blue. The bold number represents the most common connection for each prompt. Because poems can be placed in two categories if evaluators disagree, numbers do not add to 1 horizontally.

poem uses figurative language by personifying the bench, it is not coherent with the given metaphor.

This category is an especially common error for poems about the prompts “death is a rose” (55%) and “surrender is a book” (53%). Many poems said “the rose died” or “I surrender to the book”. We posit that these connections are easier than metaphorical connections because they do not require a shared third aspect which writers have to generate themselves.

4.1.4 Offset Connection

A poem has an *offset connection* if it expresses a shared feature between one word in the metaphor and another word very related to the other word in the metaphor. In Figure 1C, the poem talks about the “fire of anger” for which “wood is a source of fury”; the poem is about the offset metaphor “anger is fire”. “Death is a rose” had 40% of poems categorized as an offset connection; most commonly these poems talk about “life is a rose” and note that life, like roses, must end in death.

We suggest that writers make this error because they are looking for any connection they can find, even if the connections are not directly linked to the given metaphor. An offset connection increases the search space by allowing for connections within a broader set of domains.

4.1.5 Incoherent Connection

A poem has an *incoherent connection* if it relates the two words in the metaphor but in an unclear way. In Figure 1D, the poem says “anger is teaming, ... my wood is drying” with no supporting text to explain how these two concepts are related.

In this case writers acknowledge both words in the prompts but either do not attempt to connect them or connect them in an incoherent way.

4.1.6 Metaphorical Connection

A poem has a successful *metaphorical connection* if it relates the two words metaphorically in the way provided by the given metaphor and understood by the evaluators. In Figure 1E, the poem says that “anger grew, like a tree ... it had taken root”. This poem takes several aspects of wood and coherently applies them to anger. Although this poem talks primarily about a tree, we do not consider this an offset connection because trees are the only source of wood.

Each of the given metaphors had at least one successful poem. All of our successful poems made creative connections, like “Immortality lies just down the hall / The path to it is not easy to find”. Failed poems tended to repeat the same connections, like “I am surrounded by four walls indefinitely”.

5 Discussion

The rate of success between different prompts varies greatly, from 5% for “surrender is a book” to 47% for “hate is a mist”. Some prompts are more likely to result in different kinds of connections, like *offset connections*, than others. What explains these varying success rates?

We first explore whether word similarity between the two words in the prompt could account for this variability. In Figure 3, we plot WORD2VEC² word similarity against success rate for our 10 prompts. Based on these 10 data points, it seems that word similarity is not a strong predictor of users making a metaphorical connection. This suggests that people are not picking up on existing connections but finding new, creative ways to relate the words.

²<https://code.google.com/archive/p/word2vec/>

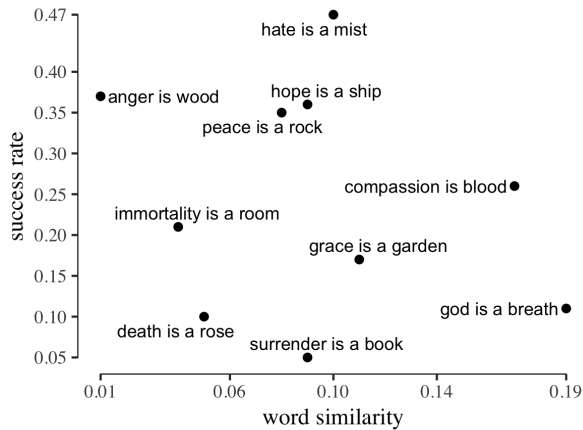


Figure 3: Fraction of successful poems as a function of the similarity between the words in the prompt.

Although we see no correlation between word similarity and success rate, it could be that WORD2VEC is not accurately modeling previous associations people may bring to the task. Other models of semantic relatedness may be able to better predict the success of people in the task.

Looking at the least successful prompts, we note that they use sensible but not metaphorical connections. The prompt “death is a rose” has many *attributional connections* saying “the rose died”. Though sensible, it is not a metaphor. Similarly, the prompt “surrender is a book” often resulted in poems saying “I surrendered to the book” which is a connection, but does not express the target metaphor. In contrast, “anger is wood” had a high success rate. These words could also be connected by saying “the wood is angry” but this rarely happened, possibly because this phrase is not as sensible as “the rose died.”

We hypothesize that if two words can be sensibly connected, people are likely to write a poem with this connection without checking whether the connection meets the target metaphor. If this does explain the varying success rates, it is likely that computational systems will have similar problems.

6 Future Work

We believe this task is a good candidate to test the ability of computers to automatically generate coherent poetry or to see how computational techniques could help novices better complete the task.

Further work could explore how computational techniques can aid in the evaluation of this task. This feedback could help people write successful poems, particularly if told which error they

are making. Can metaphor detection techniques, such as those based on conceptual metaphor theory (Shutova and Sun, 2013), evaluate whether a poem expresses its given metaphor? Can we detect what connections are being made?

Computer evaluation would also help further computer generation. Can the work of Veale (2013), which generates poetic metaphorical expressions, be extended to produce poems similar to the successful ones found in the paper? If we could express the target metaphor as a constraint, can computational techniques like those used in Hafez (2016) write poems based on metaphors, not just themes?

There is high potential for computational tools to aid people in this task. Given that only 24% of writers successfully wrote poems to a metaphorical prompt, there is an open problem of how to improve on this baseline. Future work could design computational aids, like those in (Gagliano et al., 2016), to suggest possible metaphorical connections that writers could accept or reject, similar to other creative writing aids (Clark et al., 2018).

Beyond poetry, helping people find connections between two domains has far-reaching applications from science education (Glynn, 1991) to product design (Hope et al., 2017). This is a hallmark of human intelligence that can be computationally supported.

7 Conclusion

In this paper we introduce a short poetry writing task that gets at the heart of meaningful figurative language. We collect 186 amateur examples and find that only 24% of poems successfully make the metaphorical connection, indicating that this task is hard but possible. The most common failure case is when poems make no connection between the words (41%). Other poems may fail by making a non-metaphorical connection or a connection with the wrong word.

We see potential in this task as a demonstration of computational creativity and figurative language generation. By analyzing the common errors we show ways in which improvements can be made. We believe that computational systems can improve upon this baseline.

Acknowledgements

This work is supported by NSF Graduate Research Fellowship Grant No. DGE 16-44869.

References

- Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories.
- Simon Colton, Jacob Goodwin, and Tony Veale. 2012. Full-face poetry generation. In *ICCC*, pages 95–102.
- Andrea Gagliano, Emily Paul, Kyle Booten, and Marti A Hearst. 2016. Intersecting word vectors to take figurative language to new heights. In *Proceedings of the Fifth Workshop on Computational Linguistics for Literature*, pages 20–31.
- Marjan Ghazvininejad, Xing Shi, Yejin Choi, and Kevin Knight. 2016. Generating topical poetry. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1191.
- Shawn M Glynn. 1991. Explaining science concepts: A teaching-with-analogies model. *The psychology of learning science*, pages 219–240.
- Tom Hope, Joel Chan, Aniket Kittur, and Dafna Shahaf. 2017. [Accelerating innovation through analogy mining](#). *CoRR*, abs/1706.05585.
- Justine Kao and Dan Jurafsky. 2012. A computational analysis of style, affect, and imagery in contemporary poetry. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 8–17.
- David M Kaplan and David M Blei. 2007. A computational approach to style in american poetry. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 553–558. IEEE.
- Hisar Manurung, Graeme Ritchie, and Henry Thompson. 2000. Towards a computational model of poetry generation. Technical report, The University of Edinburgh.
- Nick Montfort. 2017. *The Truelist*. Counterpath.
- Yael Netzer, David Gabay, Yoav Goldberg, and Michael Elhadad. 2009. Gaiku: Generating haiku with word associations norms. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pages 32–39. Association for Computational Linguistics.
- Hugo Gonalo Oliveira. 2017. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 11–20.
- Allison Parrish. 2018. *Articulations*. Counterpath.
- Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 978–988.
- Anselm Strauss and Juliet M Corbin. 1990. *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.
- Tony Veale. 2013. Less rhyme, more reason: Knowledge-based poetry generation with feeling, insight and wit. In *ICCC*, pages 152–159.
- Tony Veale and Yanfen Hao. 2007. Comprehending and generating apt metaphors: a web-driven, case-based approach to figurative language. In *AAAI*, volume 2007, pages 1471–1476.
- Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov. 2016. Metaphor: A computational perspective. *Synthesis Lectures on Human Language Technologies*, 9(1):33–54.