

認知言語学的イメージスキーマの生成と解釈における大規模言語モデルと画像生成モデルの評価

本田恭平¹ 松崎孝介² 吉田遥音² 坂口慶祐^{2,3}

¹ 株式会社ブリヂストン ² 東北大学 ³ 理化学研究所

kyohei.honda@bridgestone.com matsuzaki.kosuke.r7@dc.tohoku.ac.jp

haruto.yoshida@dc.tohoku.ac.jp keisuke.sakaguchi@tohoku.ac.jp

概要

イメージスキーマとは認知言語学において、認知プロセスに繰り返し現れるパターンを表現する図である。しかしこのパターンは人間が同定する必要があるため、全てを網羅的に発見することは困難であった。本研究では英語の動詞を対象に、大規模言語モデル (LLM) と画像生成モデルを用いたイメージスキーマの生成を行った。評価の結果、画像生成モデルでは生成が困難であったが、LLM は LaTeX (TikZ) のコードとして、人間と LLM 双方にとって解釈性の高いイメージスキーマを生成できることが示された。これは LLM を用いたイメージスキーマの網羅的发现の可能性を示唆するものである。

1 はじめに

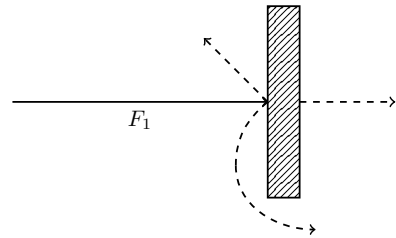
イメージスキーマとは認知言語学において、運動や知覚に繰り返し現れる共通したパターンや規則性などをダイアグラムとして表現したものである [1,2]。代表的な例として、日常的な空間的移動全般に共通するパターンを表現した「道」(図 1a) や、環境中の対象や人々と力を伴う相互作用を行おうと試みたときに会う、自分の力を妨害する障害物を表現した「妨害」(図 1b) などが挙げられる [1]。

イメージスキーマについて、詩や色彩表現に関する研究 [3,4] や異言語間での比較研究 [5,6] が行われている。また近年では文章に対応するイメージスキーマを選定するモデルの研究といった、LLM を活用した研究も行われるようになってきている [7]。しかし、イメージスキーマの発見は人間が経験や感覚、言語表現などに共通するパターンを逐一特定する必要があるため、網羅的な発見が困難であるという課題があった。

この課題を解決するため LLM および画像生成モ

A $\xrightarrow{\hspace{2cm}}$ B

(a) 道のイメージスキーマ。「起点、ないし出発点」と「目的地、あるいは到達点」と「起点と目的地を結びつける一連の隣接する場所」によって構成される



(b) 妨害のイメージスキーマ。障害にぶつかった後に任意の可能な方向を取る力のベクトルによって妨害を表現している

図 1: 「道」と「妨害」のイメージスキーマ

デルによって様々な概念や経験、感覚を表すイメージスキーマを生成し、それらに共通する部分を調査することによって網羅的にイメージスキーマを発見することを考えた。そこで本研究では、LLM と画像生成モデルを用いたイメージスキーマの生成と評価を行った。その結果、画像生成モデルでは装飾的で複雑な図になってしまう一方、LLM では LaTeX (TikZ) のコードの形式で単純な図形で構成されたシンプルなダイアグラムとしてのイメージスキーマを生成できることが分かった。また、このイメージスキーマは人間と LLM 双方にとって解釈性の高いことを示唆する結果が得られた。

2 分析データの作成

2.1 データセットの作成

本研究では、英語の動詞を対象にイメージスキーマの生成と評価を行う。対象とする動詞は、既存研

究 [8] で取り扱われている 3084 個の英語動詞のうち、動詞としての使用頻度の高い 146 個の単語とした。使用頻度の計算にあたっては、2024/10/20 時点の Wikipedia のダンプデータ [9] を用いた。

2.2 イメージスキーマの生成

本研究では英語の動詞に対して、その動詞が示す行動をシンプルかつ抽象的に表現したダイアグラムを画像生成モデルと LLM に生成させた。画像生成モデルは DALL-E 3 [10] を使用し、LLM は GPT-4o [11]、Claude (claude-3-5-sonnet-20240620) [12] を使用した。なお、プロンプトはすべて英語で記述した。実際の文面は付録 B を参照されたい。

画像としての生成 0, 3 ショットの 2 つの条件で、DALL-E 3 を用いて画像としてのイメージスキーマの生成を実施した。0 ショットでの生成は API 経由で行った。3 ショットでの生成は API 経由で参照例を入力することができないため、WebUI 上で行った。また、3 ショットでの生成時には、既存研究 [1] で定義されている「強制」、「道」、「循環」のイメージスキーマを参照例として使用した。

LaTeX (TikZ) のコードとしての生成 0, 3, 5 ショットの 3 つの条件で、GPT-4o 並びに Claude を用いて LaTeX (TikZ) のコードとしてのイメージスキーマの生成を行った。3 ショットでの生成時には、画像としての生成時と同じイメージスキーマを使用した。5 ショットではこの 3 つに加えて、同じく既存研究 [1] で定義されている「バランス」と「妨害」を用いた。

3 実験方法

3.1 人間による定性評価

生成されたイメージスキーマに対して、以下の 5 つの項目で 5 点満点の定性評価を行った。

- 変換の適切さ (コードとしての生成時のみ)
生成された LaTeX のコードと、LLM の説明が合致しているか。すなわち、LLM は表現しようとしたものをコードとして正しく表現できているか。
- 抽象度の高さ
生成されたイメージスキーマが特定の具体的な場面を描写したものではなく、抽象的な概念を表現したものになっているか。
- 構成の簡潔さ

生成されたイメージスキーマは、その動詞を表現するにあたり余分な構成要素がないか。

- 構成の充足度
生成されたイメージスキーマは、その動詞を表現するにあたり不足している要素がないか。
- 人間の解釈性
生成されたイメージスキーマを人間が見て、元になった動詞を表現したものであると理解 (解釈) できるか。

3.2 生成されたイメージスキーマの解釈

画像生成モデル並びに LLM は解釈可能な形でイメージスキーマを生成できているのか、または LLM は生成されたイメージスキーマを解釈できるのかを評価するために、LLM による解釈の実験も行った。具体的には、生成されたイメージスキーマを LLM に入力し「何の動詞を表現したダイアグラムであるか」という質問に回答させ、その結果を定量的に分析した。画像生成モデルは生成されたイメージスキーマに対する解釈をテキストで出力できないため、GPT-4o 並びに Claude の 2 つのモデルで解釈を行った。生成時と同様に、プロンプトはすべて英語で記述した。実際の文面は付録 C を参照されたい。

3.3 解釈結果の評価

LLM によるイメージスキーマの解釈結果を、以下の手順で定量的に評価した。

1. 解釈結果の類似度算出
イメージスキーマの元になった単語と解釈結果の単語の組み合わせに対して、埋め込みベクトルのコサイン類似度を計算した。
2. 解釈結果の類似度の評価
元単語に対する解釈結果の類似度がベンチマークの類似度の分布に対して有意に高いか、または低いかを統計的に検定した。検定にはウィルコクソンの符号順位検定 [13] を用いた。
なお、ベンチマークは以下の手順で作成した。
 1. 評価用データセットの作成
既存研究 [8] から抽出した 3084 単語それぞれに対して、自身以外の全ての単語とのコサイン類似度の平均を算出。この平均類似度の上位 500 単語を比較用データセットとした。
 2. ベンチマークの算出
イメージスキーマの元になった単語と比較用

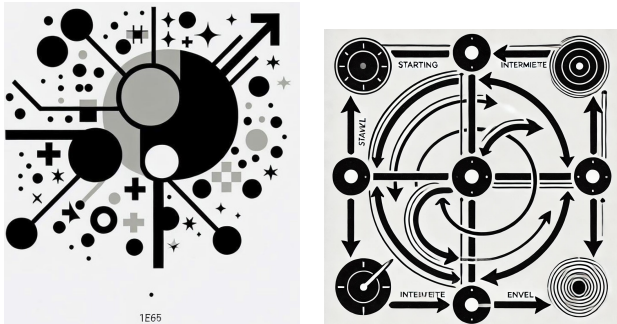
データセット内の全単語とのコサイン類似度を計算することで類似度の分布を求め、これをベンチマークとした。

なお、単語の埋め込みには gensim の訓練済みモデル (glove-wiki-gigaword-300 [14]) を使用した。

4 実験結果および考察

4.1 イメージスキーマ生成

画像としての生成 図2に、DALL-E 3に0ショットで生成させた“leave”のイメージスキーマと、3ショットで生成させた“travel”のイメージスキーマを示す。画像として生成すると、図2のように装飾的で複雑な構造を持ち、図1に示したような認知言語学的なイメージスキーマ [1,2] とは異なる例が多かったため、これ以上の分析は行わなかった。



(a) “leave”, 0 ショット (b) “travel”, 3 ショット

図2: DALL-E 3 が生成したイメージスキーマ

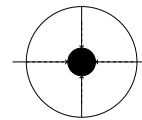
LaTeX (TikZ) のコードとしての生成 0, 3, 5 ショットの条件で Claude と GPT-4o が生成した “appear” のイメージスキーマを図3に示す。画像として生成させた場合と異なり、単純な図形のみで構成されたシンプルなダイアグラムが得られた。なお、GPT-4o が0ショット時に “appear” のイメージスキーマとして生成したコードの一部を以下に掲載する。応答の全文は付録Dを参照されたい。

```
\begin{tikzpicture}
  \useasboundingbox (0,0) rectangle (10,5);

  % Initial state: faded circle
  \draw[fill=black!10] (1,2.5) circle (0.5);

  % Arrow indicating the process of appearing
  \draw[->, thick] (2,2.5) -- (4,2.5);

  .....
\end{tikzpicture}
```



(a) Claude, 0 ショット



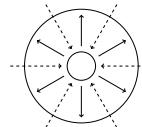
(b) GPT-4o, 0 ショット



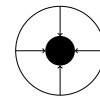
(c) Claude, 3 ショット



(d) GPT-4o, 3 ショット



(e) Claude, 5 ショット



(f) GPT-4o, 5 ショット

図3: Claude および GPT-4o による “appear” のイメージスキーマ

4.2 人間による定性評価

コードとして生成させたイメージスキーマの一部に対して、定性評価を行った。評価対象は動詞としての使用頻度の上位5単語である、bring, include, have, help, travel とした。各単語のモデル並びにショット数毎の評価結果の平均値を図4に示す。なお、この評価は4人の評価者で実施した。

図4から、両モデルとも変換の適切さの点数が高いことがわかる。これは、LaTeX (TikZ) の形式で生成されたイメージスキーマが LLM の意図を忠実に反映していることを示しており、本研究の正当性を支えるものである。

抽象度や簡潔さといった生成物自体の評価項目に注目すると、GPT-4o の方が Claude よりも抽象的で簡潔な (余分な要素のない) イメージスキーマを生成していることがわかる。一方で Claude は抽象度が低く構成の充足度が高い (不足している要素が少ない) ため、より具体的に動詞の内容を描写していると言える。なお、両モデル間で人間の解釈性やショット数による大きな差は見られなかった。

4.3 生成されたイメージスキーマの解釈と類似度の評価

コードとして生成されたイメージスキーマについて、元単語と解釈結果の類似度の検定結果を表1に示す。また、GPT-4o および Claude が解釈した動詞

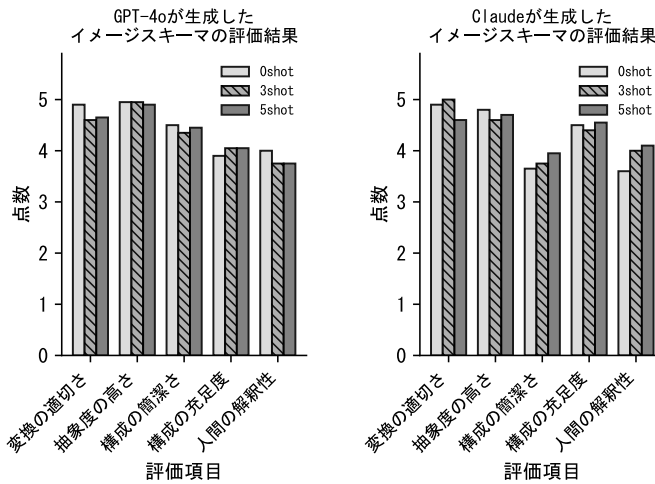


図 4: 人間による定性評価の結果

表 1: 元単語と解釈結果の類似度評価結果

モデル	参照数	一致	高い ($p < 0.01$)	低い ($p < 0.01$)	有意差なし ($p \geq 0.01$)
GPT-4o	0	4	138	5	3
	3	6	127	9	10
	5	4	116	18	12
Claude	0	7	107	26	13
	3	5	116	20	10
	5	3	116	22	8

の分布を図 5 に示す。これらの図には解釈結果として現れた個数の上位 6 単語のみを掲載している。

表 1 から、GPT-4o は Claude に比べて、与えられたイメージスキーマを元単語と近い（コサイン類似度が有意に高い）語に解釈することが多いことがわかる。また図 5 から、GPT-4o は解釈結果が特定の単語に偏っていることがわかる。これは、GPT-4o が複数の動詞に共通する抽象的なダイアグラムを生成しており、そのダイアグラムが元単語を適切に表現できていることを示唆している。一方で Claude は、解釈結果が元単語から乖離している場合が多く、再解釈された単語も多様であった。これは、Claude がイメージスキーマを生成する際に十分な抽象化が行えていないためと考えられる。これらの結果は、「GPT-4o の方が Claude に比べて、より抽象的で構成要素の少ないイメージスキーマを生成している」という人間による定性評価の結果とも一致する。

0, 3, 5 ショットの条件で生成した 3 つのイメージスキーマの解釈にあたって、解釈結果と元単語のコサイン類似度が 2 回以上 $p < 0.01$ (低い) となった単語を表 2 に示す。表 2 より、静的な動作 (have, keep)、内面的な動作 (think)、動作そのものではな

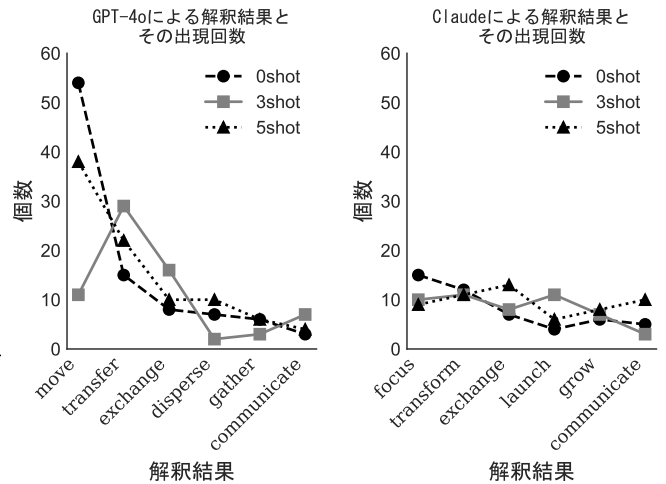


図 5: 解釈結果として得られた単語とその回数

表 2: 解釈結果との類似度が有意に低かった単語

モデル	単語群
GPT-4o	derive, indicate, keep, lie, play, tell
Claude	have, keep, make, marry, need, perform, play, remove, serve, think

く「どう実行したか」というプロセスに着目した動詞 (perform)、「不足」や「除去」の表現が必要な動詞 (need や remove) といった、ダイアグラムでの表現が難しい単語において、イメージスキーマの解釈結果が元単語と乖離していることがわかる。

5 終わりに

本研究では LLM と画像生成モデルを用いたイメージスキーマの生成と評価を行った。特に LLM による LaTeX (TikZ) のコード形式での生成では、シンプルかつ人間と LLM 双方にとって解釈性の高いダイアグラムが得られることを示した。一方で画像生成モデルでは装飾的で複雑な図が得られ、イメージスキーマの生成は困難であることが分かった。今後の研究課題としては、動詞以外の単語や他言語への適用、他の LLM モデルでの実験が挙げられる。また、生成されたイメージスキーマに共通するパターンを抽出することで、未発見のイメージスキーマを網羅的に明らかにする可能性が期待される。さらに、本研究では動詞の多義性を考慮していないため、特定の意味に限定した実験や、各意味に対応するイメージスキーマの生成方法の検討が求められる。

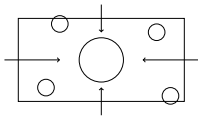
謝辞

本研究は JSPS 科研費 JP21K21343 の助成を受けたものです。本研究を進めるにあたり多大なご助言、ご協力を賜りました Tohoku NLP グループの皆様に感謝いたします。

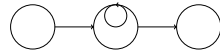
参考文献

- [1] Mark Johnson. **The body in the Mind. The Bodily Basis of Meaning, Imagination, and Reasoning.** The University of Chicago Press, 1987. 邦題：心のなかの身体 想像力へのパラダイム 転換、訳：菅野盾樹、中村雅之、出版：株式会社紀伊國屋書店、出版年：2001.
- [2] George Lackoff. **Women, fire, and dangerous things. what categories reveal about the mind.** The University of Chacago Press, 1987.
- [3] Margaret H Freeman. Momentary stays, exploding forces: A cognitive linguistic approach to the poetics of emily dickinson and robert frost. **Journal of English Linguistics**, Vol. 30, No. 1, pp. 73–90, 2002.
- [4] 三田寛真. 色彩表現の意味解釈と視覚的变化のイメージスキーマ「スタジアム」が「真っ赤」になるとき—. **言語情報科学**, Vol. 20, pp. 55–69, 2022.
- [5] Ellen Dodge and George Lakoff. Image schemas: From linguistic analysis to neural grounding. In Beate Hampe, editor, **From Perception to Meaning**, pp. 57–91. De Gruyter Mouton, Berlin, New York, 2005.
- [6] Wenlian HUANG. 日中同形多義動詞「上がる (agaru)」「上 (shàng)」の認知対照研究: イメージスキーマ・ネットワークの分析. 2024. <https://dspace.jaist.ac.jp/dspace/handle/10119/18983>.
- [7] Lennart Wachowiak and Dagmar Gromann. Systematic analysis of image schemas in natural language through explainable multilingual neural language processing. In **Proceedings of the 29th International Conference on Computational Linguistics**, pp. 5571–5581, 2022.
- [8] Beth Levin. **English Verb Classes and Alternations: A Preliminary Investigation.** University of Chicago Press, 1993.
- [9] 英語版 Wikipedia ダンプデータ, (2024-11 月閲覧). <https://dumps.wikimedia.org/enwiki/20241020/>.
- [10] OpenAI. DALL·E 3: OpenAI’s Image Generation Model, 2024. <https://openai.com/dall-e>.
- [11] OpenAI. GPT-4: OpenAI’s Large Language Model, 2024. <https://openai.com/gpt-4>.
- [12] Anthropic. Claude 3.5: Large Language Model by Anthropic, 2024. Model version: claude-3-5-sonnet-20240620. <https://www.anthropic.com/claude>.
- [13] Frank Wilcoxon. Probability tables for individual comparisons by ranking methods. **Biometrics**, Vol. 3, No. 3, pp. 119–122, 1947.
- [14] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In **Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks**, pp. 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.

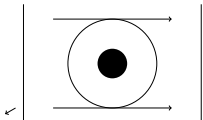
A LaTeX(TikZ) のコードとして生成されたイメージスキーマの例



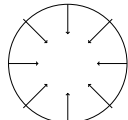
(a) Claude, 0 ショット



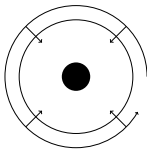
(b) GPT-4o, 0 ショット



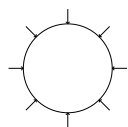
(c) Claude, 3 ショット



(d) GPT-4o, 3 ショット



(e) Claude, 5 ショット



(f) GPT-4o, 5 ショット

図 6: Claude および GPT-4o による “keep” のイメージスキーマ

B イメージスキーマ生成に用いたプロンプト

Create a diagram representing the verb to {verb} using Latex code. However, avoid using objects such as human limbs or clocks to depict specific situations. Express the essence of the action indicated by this verb simply and abstractly using figures such as circles and arrows. Please note the following points when creating it: Express using only white and black, without using colors such as red or blue. Do not include specific labels, such as 'water' or 'finger,' within the diagram. Use `useasboundingbox (0,0) rectangle (10,5);` to create a diagram with a size of 5cm in height and 10cm in width.

画像生成モデルを用いて画像生成をした際は、第一文の “using Latex code” と最後の文を削除した。

C イメージスキーマ解釈に用いたプロンプト

The following latex code represents diagram about an abstract representation that conveys the essential image of a certain verb. Please tell me which verb you think it represents, along with your reasoning for this judgment.

Please follow the format below for your answer: 'This diagram represents the verb xxxx. The reason is xxxx.' Feel free to fill in the xxxx parts with your own words.

D 0 ショットでの GPT-4o による “appear” のイメージスキーマ生成結果

To create a simple and abstract representation of the verb "to appear" using LaTeX, we can utilize circles and arrows to convey the concept of something coming into view or emerging. Below is an example of how you might achieve this using the `tikz` package in LaTeX. This diagram will use circles to represent objects and arrows to indicate the process of appearing.

```

\documentclass{standalone}
\usepackage{tikz}

\begin{document}

\begin{tikzpicture}
  \useasboundingbox (0,0) rectangle (10,5);

  % Initial state: faded circle
  \draw[fill=black!10] (1,2.5) circle (0.5);

  % Arrow indicating the process of appearing
  \draw[->, thick] (2,2.5) -- (4,2.5);

  % Final state: solid circle
  \draw[fill=black] (5,2.5) circle (0.5);
\end{tikzpicture}

\end{document}

```

Explanation:

- **Bounding Box**: The `\useasboundingbox (0,0) rectangle (10,5);` command sets the diagram size to 10cm width and 5cm height.
- **Faded Circle**: A circle at position `(1,2.5)` with a light fill (using `black!10`) represents the initial, less visible state.
- **Arrow**: A thick arrow from `(2,2.5)` to `(4,2.5)` indicates the transition or process of appearing.
- **Solid Circle**: A circle at position `(5,2.5)` with a solid black fill represents the final, fully visible state.

This abstract representation uses simple geometric shapes and transitions to convey the concept of "appearing" without relying on specific objects or labels.