

共通基盤の構築に及ぼすイメージ生成の個体差に関するシミュレーション

由井 達也¹ 森田 純哉¹ 天谷 武琉¹ 東中 竜一郎² 竹内 勇剛¹

¹ 静岡大学 ² 日本電信電話株式会社

{yui.tatsuya.20, amaya.takeru.19}@shizuoka.ac.jp, {j-morita,takeuchi}@inf.shizuoka.ac.jp
ryuichiro.higashinaka@ntt.com

概要

コミュニケーションの送り手が発した記号は、受け手の有する認知的フレームを介してイメージとして解釈される。つまりコミュニケーションを効率化するためには、送り手と受け手の間で認知的フレームをすり合わせる必要がある。本稿では、そのような送り手と受け手で共有されたフレームを共通基盤と呼び、タングラム命名課題を対象とした共通基盤形成のモデルを提示する。特に、本稿ではイメージ生成における個体差を、モデルに与えるノイズパターン（シード値）によって操作した2つのシミュレーションを示す。

1 はじめに

コミュニケーションとは、送り手が有する意図を、メディアを介し、受け手に伝達するプロセスである [1]。人間のコミュニケーションにおける主要なメディアは（自然）言語である。世界には様々な言語が存在する。いずれも膨大な語彙を持ち、それらを定められた文法構造によって関係づけることで、複雑な意図を表現する。しかし、いずれの言語にも曖昧性があり、送り手の意図は完全には記述されない。また、言語を解釈するフレームの背景にある抽象性の高い知識・記憶の層は、そもそも他者からは直接アクセスできないという問題もある [2]。よって、人間のコミュニケーションを成立させるためには、言語表現を工夫するだけでなく、送り手と受け手の間で、意図と言語を対応づける認知的フレームを調整する必要がある。

本研究では、コミュニケーションを成立させる個人（送り手・受け手）が有する認知的フレームを共通基盤 (common ground) と呼び、その構築プロセスに焦点を当てる。コミュニケーションの共通基盤

は、過去、多くの認知科学的な研究によって検討されてきた [4, 5, 6, 7]。しかし、従来の共通基盤の構築に関する認知科学的なモデル [6, 7] は、コミュニケーションにおけるイメージ化のプロセスを扱っていない。実世界において、送り手の伝達意図は、物理的世界と接地するアナログなイメージを含み、離散的な記号のみでは表現しきれない。この観点に立てば、コミュニケーションの共通基盤は、送り手内部によるイメージから言語への変換プロセスと、受け手による言語からイメージへの変換プロセスの関係において検討されるべきである。

著者らは、上記の観点のもとで、Convolutional Neural Network (CNN)、画像生成モデル (Stable Diffusion) [8]、大規模言語モデル (GPT) [9] を組み合わせたモデルを検討してきた [3, 10]。このモデルは、タングラム命名課題 [11] と呼ばれる抽象図形を題材とし、2者間で共通の対象に同一の言語ラベルを割り当てる過程を表現する。特に、著者らによるはじめの研究 [10] は、送り手と受け手が固定された状況での一方向的なコミュニケーションが表現された。すなわち、送り手がタングラムからイメージを生成し、そのイメージを表現する言語を受け手に送信する。受け手は言語からイメージを生成し、自身が見ているタングラムのいずれに当てはまるかを答える。ここで送り手の意図したタングラムと受け手の解釈したタングラムが一致した場合に、そのコミュニケーションは成功したものとみなす。

直近の報告 [3] では、著者らのモデルが、タングラム命名課題におけるチャンスレベルを有意に上回る正解率を示すこと、またイメージ生成に関わるモデルのパラメータを、コミュニケーションの成功事例をもとに調整することで、その成績が向上しうることを示した。

なお、この検討において、イメージの生成に利用

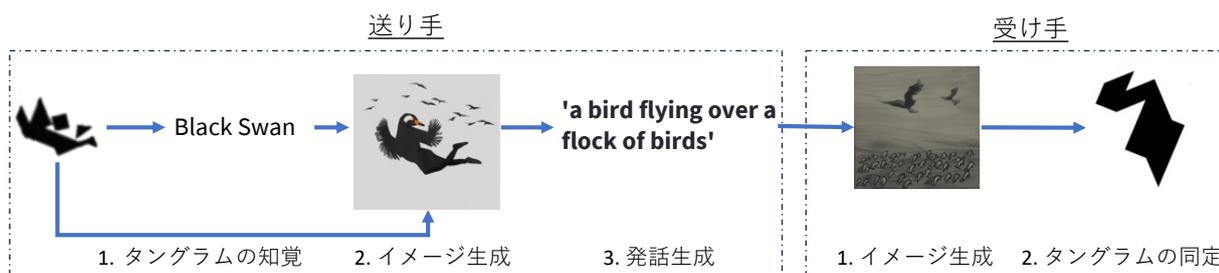


図 1 全体的発話の内部表現. [3] をの図 5 を改変.

されるシード値は単一の値に固定されていた. イメージ生成モデルにおけるシード値は, 生成されるイメージの基となるノイズのパターンを決定する. モデルは, 与えられたプロンプトに即して, ノイズパターンを変形し, 人間が意味を感じるイメージを生成する. この値をモデルごとに固定化することで, シード値は異なるプロンプト間で共通する, そのモデルによるイメージ生成のスタイルが表現される.

イメージ生成のスタイルは, 人間同士のコミュニケーションにおいても重要な要因である. コミュニケーションの文脈によっては, イメージを簡略化するスタイルが好まれることもあれば, その逆も有り得る [5]. あるいは, イメージ生成のスタイルを文化 (画派やジャンルに相当) と関連付けて考えることもできる. 異文化コミュニケーションで生じる問題などを考えれば, 送り手と受け手のスタイルを一致させることは, コミュニケーションの成功率に寄与すると考えられる. 本研究では, これらイメージ生成の個人差に関する仮説を著者らの先行研究のモデルを用いることで検討する (シミュレーション 1). そのうえで, シード値による初期の成績の違いが, 共通基盤の構築に及ぼす影響を検討する (シミュレーション 2).

2 タングラム命名モデル

先行研究 [3] において提示したモデルの概略を述べる. はじめに, このモデルにおける送り手と受け手のコミュニケーションのフローを示し, その後にフローの出力を学習により変化させる方法を示す.

2.1 コミュニケーションフロー

コミュニケーションのフローは図 1 に示される. 図の各ステップは以下のように説明される.

送り手の処理

1. タングラムの知覚: 各タングラムの形状に対

して第一印象を形成する. 本研究では一般物体認識のために構築された Convolutional Neural Network (CNN) を利用する.

2. イメージ生成: タングラムの第一印象を詳細化する. Stable diffusion の `img2img` を用いて, タングラムの画像と物体認識のラベルを組み合わせたイメージを生成する.
3. 発話生成: イメージ生成によって得られた画像にイメージキャプションを行う. 言語生成には GPT-2 を利用する.

受け手の処理

1. イメージ生成: 送り手が生成した言語をもとにイメージを生成する. Stable diffusion の `txt2img` を使用する.
2. タングラムの同定: 生成されたイメージと手元にあるタングラムの類似度を計算し, 送り手の意図を同定する. 類似度の計算には, 送り手のはじめのステップと同じ構造を持つ CNN の出力ベクトルを用いる.

2.2 学習

図 2 は, 上記のフローを繰り返すことによる学習プロセスを表現している. 図の一番左 (エピソード) は, 図 1 のフローを縦方向に変形したものである. 各エピソードは, N のタングラムと K の角度の組み合わせに適用される. これら $N \times K$ のエピソードを 1 つの試行として, 成功事例を抽出する. この教師データに対して, 各ニューラルネットのパラメータを較正する. 特に先行研究 [3] においては最上部の CNN (図のハイライト部分 (赤矢印)) に対する学習が行われた. この学習は, 入力されたタングラム画像と成功時における CNN の出力ベクトルのペアを教師データとした誤差逆伝播法によるものである [12].

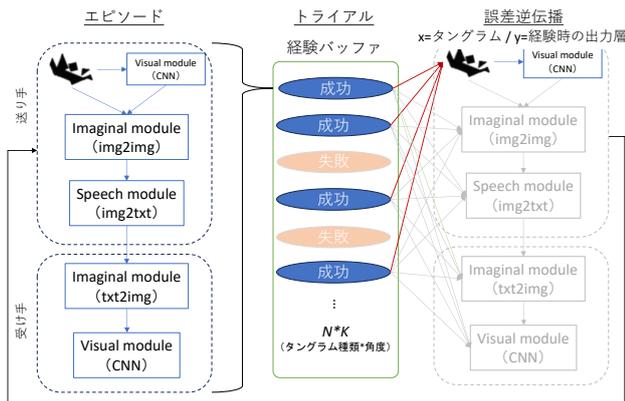


図2 学習プロセスの図解. 各ステップは図2のモジュールによって遂行される. [3]の図4を改変.

3 シミュレーション1

3.1 目的と設定

1節の最後に述べた仮説を検討するシミュレーションを行った. すなわち, イメージ生成におけるシード値を変化させることで, コミュニケーションの成功率がどのように変化するか. また, 送り手と受け手のイメージ生成のスタイル (シード値) が一致することで, コミュニケーションの成功率がどのように変化するか. これらを検討するために, 以下の2つの条件を設定した.

1. 独立条件: 送り手と受け手のそれぞれのイメージ生成に利用するシード値を独立にサンプリングする.
2. 共変条件: 送り手と受け手のイメージ生成に共通して利用するシード値をサンプリングする.

両条件とも, サンプリングの後に図2におけるトライアル (6種のタングラムと8の角度からなる48エピソード) を1回のみ実行し (誤差伝播はおこなわず), それを独立に100回繰り返した ($n = 100$).

3.2 結果

得られたエピソードの正解率を計算し, 条件ごとに平均, 最大値, 最小値をまとめた (表1). 表より, 両条件ともに最小値と最大値に大きな違いがあることがわかる. つまり, イメージ生成のシード値を変化させることで, モデルによるコミュニケーションの成否に幅を持たせることができた. 一方, 両条件では平均値に大きな差が見られず ($t = 0.45, p = 0.65$), モデル間でのシード値の共通性 (イメージ生成スタイルの共通性) がコミュニ

表1 シード値を操作したシミュレーション結果

	平均	最大	最小
独立	0.187	0.354	0.062
共変	0.184	0.375	0.083

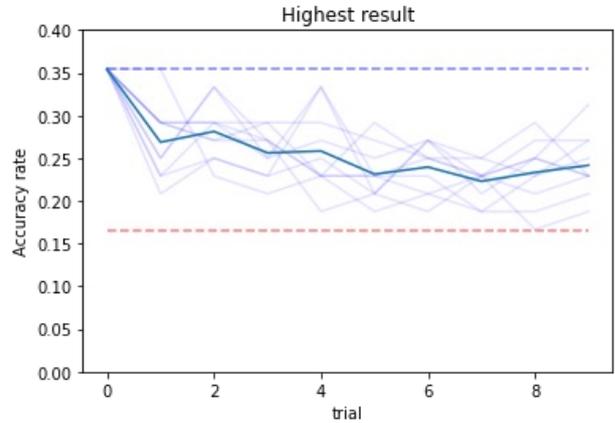


図3 独立条件の最大値と対応付けられるシード値ペアを用いた学習.

ケーションの成功に寄与するとは言えない結果となった.

4 シミュレーション2

4.1 目的と設定

このシミュレーションでは, イメージ生成のスタイルによる初期のコミュニケーション成績の違いが, 後続する共通基盤の構築にどのように影響するのかを検討する. 具体的には, 表1における独立条件の最大値と最小値と対応するシード値のペアを抽出し, それぞれに対して図2のトライアルを10回繰り返す学習を10回独立に実行した.

4.2 結果

図3と図4は, それぞれ独立条件の最大値と最小値と対応付けられるシード値ペアによる学習の結果である. 結果は対比的である. 正解率の初期値が高い図3は, トライアルが進行するに従って正解率が減少した. それに対して, 正解率の初期値が低い図4は, トライアルの進行に伴い正解率が上昇した.

5 考察

本研究では, 画像生成モデルと大規模言語モデルを組み合わせた共通基盤の構築モデルにおいて, イメージ生成におけるスタイル (シード値) の効果を2つのシミュレーションによって検討した. 結果は

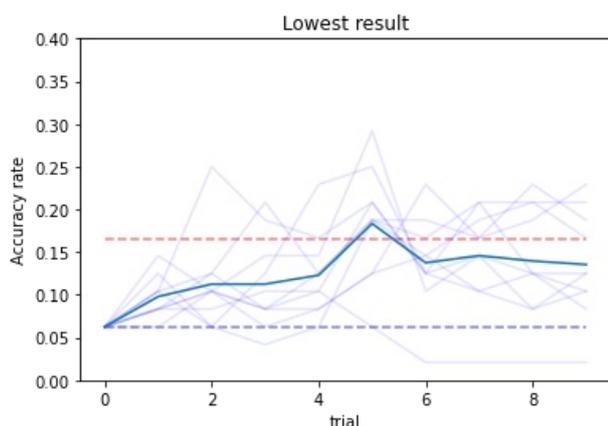


図 4 独立条件の最小値と対応付けられるシード値ペアを用いた学習。

以下の 4 点にまとめられる。

1. イメージ生成のスタイルにより、コミュニケーションの成績が変化する。
2. 送り手と受け手のイメージ生成のスタイルの一致は、コミュニケーションの成績に影響しない。
3. 初期に高いコミュニケーションの成績を示すイメージ生成スタイルの組み合わせは、学習によりコミュニケーションの成績を低下させる。
4. 初期に低いコミュニケーションの成績を示すイメージ生成スタイルの組み合わせは、学習によりコミュニケーションの成績を向上させる。

上記の結果のうち、1 は 1 節において事前に想定され、2 は想定されなかった。また、4 は設計された学習メカニズムに即したものであり、4 は図 2 における学習メカニズムの有効性に疑問を投げかけるものといえる。

事前に想定された結果である 1 や学習メカニズムの有効性を示す 4 からは、人間のコミュニケーションのモデルとしての本研究のアプローチの有効性が示唆される。世の中には、多様なイメージ生成のスタイルを有する個人が存在し、その個人間でのコミュニケーションがなされている。よって、イメージ生成のスタイルを考慮するコミュニケーション支援は、現代社会における複雑な問題の解決に貢献する可能性がある。

その一方で 2 は、送り手と受け手の認知的フレームが一致することがコミュニケーションに有効であるという素朴な仮定に疑問を投げかけるものといえる。なお、過去の認知科学的なモデル [7] においても、意図と記号を結びつけるルールが異なるモデル

により、コミュニケーション課題の達成が可能であることが示されている。そのような先行研究を考慮すれば、本研究のような送り手と受け手が明確に分離された状況で、送り手と受け手にとって有効なイメージ生成のスタイルが異なったという結果は不思議なものではないといえる。

トライアルの進行によりコミュニケーションの成績が低下した 3 については、図 2 の学習の欠点を示すものとみなせる。モデルが受け入力はいずれも似通った抽象図形である。また、学習においては成功事例のみが利用されるため、その事例における行動が極度に強化されやすい。実際、このモデルにおいて、トライアルの進行により、学習対象となった CNN の出力の画一化（同一のラベルの出力）が進行することが確かめられている。著者らの別の研究は、このような過学習を、人間が、命名対象となるタングラム間を比較する分析的な処理によって抑制している可能性を示している [13]。今後、そのような分析処理をモデルに含めることで、共通基盤の構築に関する全体的なプロセスを明らかにしていく必要がある。

6 まとめ

本研究では、コミュニケーションにおけるイメージ生成のスタイルを、画像生成モデルにおけるシード値と捉えた検討を行った。2 つのシミュレーションによって、このアプローチの可能性と今後の課題が示された。

言語とコミュニケーションの創発に関する研究における本研究のアプローチの特徴は、コミュニケーションのプロセスにアナログなイメージ生成のプロセスを含めたことにある。従来、言語進化におけるダイナミクスは、言語の有する曖昧性に起因するという議論がなされてきた [14, 15]。本研究で取り扱ったイメージ生成の個人差は、そのような曖昧性を自然かつ操作可能な形で表現する有効な手段である。このアプローチに即した研究を進めることで、人間レベルのコミュニケーションの創発に迫る計算論的な理解が導かれると考えている。

参考文献

- [1] Claude Elwood Shannon. A mathematical theory of communication. **The Bell system technical journal**, Vol. 27, No. 3, pp. 379–423, 1948.
- [2] 名塩征史. 「解釈の枠組み」の形成と共有: 「何を表現しているか」を考え話し合う活動の事例研

- 究. メディア・コミュニケーション研究, Vol. 64, pp. 87–106, 2013.
- [3] Junya Morita, Tatsuya Yui, Takeru Amaya, Ryuichiro Hishinaka, and Yugo Takeuchi. Cognitive architecture toward common ground sharing among humans and generative ais: Trial modeling on model-model interaction in tangram naming task. In **Proceedings of the 2023 AAAI Fall Symposium on Integrating Cognitive Architectures and Generative Models**. AAAI Press, 2023.
- [4] Herbert H Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. **Cognition**, Vol. 22, No. 1, pp. 1–39, 1986.
- [5] Simon Garrod, Nicolas Fay, John Lee, Jon Oberlander, and Tracy MacLeod. Foundations of representation: where might graphical symbol systems come from? **Cognitive science**, Vol. 31, No. 6, pp. 961–987, 2007.
- [6] D. Reitter and C. Lebiere. A cognitive model of spatial path-planning. **Computational and Mathematical Organization Theory**, Vol. 16, No. 3, pp. 220–245, 2010.
- [7] 森田純哉, 金野武司, 奥田次郎, 鮫島和行, 李冠宏, 藤原正幸, 橋本敬. 協調的コミュニケーションを成立させる認知的要因. ヒューマンインタフェース学会論文誌, Vol. 20, No. 4, pp. 435–446, 2018.
- [8] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In **2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, 2022.
- [9] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.
- [10] 森田純哉, 由井達也, 天谷武琉, 光田航, 東中竜一郎, 竹内勇剛. 抽象図形への命名を介したコモングラウンド形成モデルの構想. 言語処理学会第 29 回年次大会.
- [11] 須藤早喜, 浅野恭四郎, 光田航, 東中竜一郎, 竹内勇剛. 推測的かつ暫定的な対話による共通基盤の形成過程. 電子情報通信学会和文論文誌 (D), Vol. 106, .
- [12] Maxim Lapan. **Deep Reinforcement Learning Hands-On: Apply modern RL methods, with deep Q-networks, value iteration, policy gradients, TRPO, AlphaGo Zero and more**. Packt Publishing Ltd, 2018.
- [13] 天谷武琉, 由井達也, 森田純哉, 東中竜一郎, 竹内勇剛. 効率的な共通基盤形成を行うための投機的な対話計画の提案. HCG シンポジウム 2023.
- [14] Steven Pinker, Martin A. Nowak, and James J. Lee. The logic of indirect speech. **Proceedings of the National Academy of Sciences**, Vol. 105, No. 3, pp. 833–838, 2008.
- [15] Naoki Inoue and Junya Morita. A behavioral task for exploring dynamics of communication system in dilemma situations. **Artificial Life and Robotics**, Vol. 26, No. 3, p. 329–337, aug 2021.