通じることと分かり合うことは同じか –記号的一致と表象的一致からみた共通基盤形成–

Are “understanding” and “communication” the same? –Sharing a common ground from the perspective of symbolic and representational agreement–

山本 大暉 [†1], 森田 純哉 [†1], 東中 竜一郎 [†2], 竹内 勇剛 [†1]
Taiki Yamamoto[†1], Junya Morita[†1], Ryuichiro Higashinaka[†2], Yugo Takeuchi[†1]
[†1] 静岡大学, [†2] NTT 株式会社
[†1] Shizuoka University, [†2] NTT, Inc.
yamamoto.taiki.20@shizuoka.ac.jp

概要

本研究は，コミュニケーション成功率の上昇が内的表象の整合を保証するかを検証した．知覚側を更新する条件と言語生成側を更新する条件を設定し，さらに候補間の違いを明確化する学習の有無を操作した．評価には, 記号的一致と表象的一致を用いた．その結果, 言語生成側を更新し弁別を導入した条件では記号的一致が大きく向上する一方，表象的一致は低下した．対して知覚側を更新した条件では表象的一致が維持されやすかった．以上より，共通基盤の評価には単一指標ではなく，多層的指標の併用が必要である．

キーワード：共通基盤，非還元性，記号的一致，表象的一致，創発コミュニケーション

1. はじめに

今日の情報社会では，文化的背景，専門性，世代，社会的文脈の異なる主体どうしが協働する機会が増えている．その結果，身近な語が相手には別の意味で解釈されたり，同じ状況を異なる参照枠で理解したりするずれが生じやすい．

こういった問題と関連して，反復参照ゲームなどを用いて，当初はコミュニケーションが成立しない状態から意思疎通に至る条件が検討されてきた (Hawkins et al., 2020; Lazaridou et al., 2017)．こういった課題では，試行の進行に伴い，受け手が送り手の意図する対象を正しく同定する割合（記号的一致）が向上する．しかし，記号的一致が高い場合でも，送り手と受け手の内部表象が整合しているとは限らない．すなわち，「課題は解けるが，理解の中身は揃っていない」可能性が残る．

送り手と受け手の内部表象の不整合は，コミュニケーションにおける共通基盤の成立を何で評価するかという問題に直結する (Clark, 1996)．本研究は，共通基盤形成における一致の成立を，記号的一致（指示語と指示対象の対応関係における一致）と表象的一致（指示語と内部表象の対応関係における一致）に区別し，その関係を検討することを目的とする．その際，森田他 (2024) が提示した，外界の知覚・内部表象の形成・言語表現を段階的に接続するモデルを継承する．具体的には，モデル内のどのモジュールを主に更新するかという学習の経路（チャネル）と指示対象の記号的な弁別（弁別学習の有無）を操作要因とし，両要因が 2 種類の一致指標に与える影響を検証する．これにより，従来の研究では把握されてこなかった共通基盤形成の多面性を明らかにする．

2. 関連研究と理論的ギャップ

反復参照ゲーム研究は，反復的な相互作用を通じて指示表現が短縮・慣習化され, 送り手と受け手の間で参照が一致するようになる過程を示してきた (Hawkins et al., 2020)．創発コミュニケーション研究でも，エージェント間で効率的な記号体系が形成される条件が議論されており，課題達成率などを指標とした検討が行われてきた (Lazaridou et al., 2017; Bouchacourt & Baroni, 2018)．こういった研究に対し，単純な参照の成功率のみでは不十分であり，コミュニケーションに関わる複数の指標を併用する必要が指摘されている (Lowe et al., 2019).

この系譜を踏まえ，森田他 (2024) は，知覚・内部表象・言語表現を段階的に接続した計算モデルにより，入力–出力の成否だけでは捉えにくい内部過程を分析可能とするモデルを提案した．さらに，Baba et al. (2026) は，表層（言語）チャネルと深層（内部表象）チャネルを区別した学習シミュレーションを行い，表層チャネルの更新が記号的一致の改善に寄与することを示した．さらに，山本他 (2025) は，この枠組みに弁別学習（記号推論）を導入した．ここでいう弁別学習とは，言語的指示に対して複数の候補刺激が存在す

る状況において，各候補との対応関係を比較し，他の候補との差異に基づいて適切な指示対象を選択する学習過程を指す．この考え方は，刺激に対する反応の弁別を扱う古典的な EPAM (Elementary Perceiver and Memorizer) の系譜と整合する (Feigenbaum & Simon, 1962).

一方，これらの研究では，メッセージの受け手が送り手の意図と整合する指示対象を選択できたことをコミュニケーションの成功とみなしており，受け手と送り手で同一の内部表象が成立したかを検討していない．共通基盤の理論的観点では，合意の成立は表層的な記号一致だけでなく，相互行為のなかで更新される理解状態の調整として捉えられる (Clark, 1996).

このギャップを踏まえ，本研究では次の研究課題を検討する．

RQ 学習チャネル（知覚学習条件／言語学習条件）と弁別学習（記号推論の有無）は，記号的一致および表象的一致にどのような影響を与えるか．

3. 方法

3.1 モデル構成と学習設計

本研究では Baba et al. (2026) をベースとしつつ，山本他 (2025) の弁別学習を取り入れたモデルを利用する．図 1 に示すように，本モデルは送り手と受け手が，複数の深層学習モジュールを段階的に実行することでコミュニケーションを行う．この構成は，人どうしの対象指示コミュニケーション，すなわち一方が対象を言葉で記述し，他方が候補から対象を選ぶやり取りを模したものであり，送り手は記述する話者（director），受け手は同定する聞き手（matcher），ラベル・キャプションは参照表現，生成画像は各自が思い浮かべる心的イメージに対応する．

送り手の処理では，知覚モジュール（ViT: Vision Transformer）を用いて入力タングラム画像からラベルを推定する．次に，そのラベルに基づいて画像生成（Stable Diffusion, img2img）を実行し，対象の内部表象を生成する．最後に，言語モジュール（Vision Encoder–Decoder[*1]）は，画像生成モジュールで生成された表象をエンコーダ入力としてキャプションを生成する．この際，知覚モジュールが出力したラベルをデコーダ初期入力として与えることで，生成過程を条件付ける．

受け手の処理では，受信したキャプションを画像生成モジュール（Stable Diffusion, txt2img）へ入力して内部表象を再構成する．続いて，生成イメージと 6 種類の候補タングラム画像を同一の視覚空間上（ViT）に入力し，出力層のコサイン類似度を計算することで，最も指示されている可能性の高いタングラムを同定する．

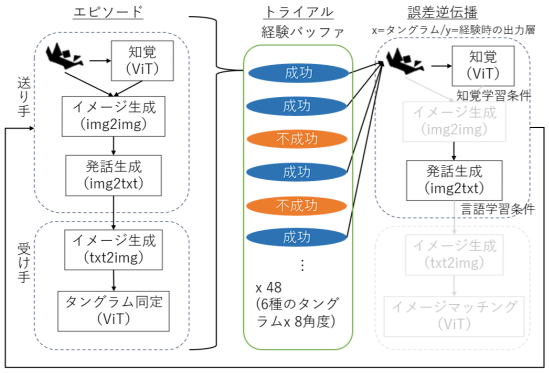

図 1　モデル構成と学習設計の統合図．送り手/受け手の段階的処理，1 試行の構成，および条件別の更新対象

ここで，知覚モジュール（ViT）の出力をラベル，言語モジュールの出力をキャプションと定義する．

学習は 6 種類のタングラムと 8 方向回転を組み合わせた試行（Trial）を基本単位とする．試行で得られた 48 の受け手の応答をもとに成功事例を正解とした誤差逆伝播により，ネットワークパラメータを更新した．損失は当該モジュールの出力誤差に基づいて定義し，勾配は更新対象モジュールのみに適用した．更新対象モジュールは Baba et al. (2026) と同様，知覚モジュールと言語モジュールとした．これらのモジュールのパラメータを独立に操作（他のモジュールを固定）することで，学習チャネル（知覚学習条件／言語学習条件）を操作要因とした．

さらに，本研究では山本他 (2025) によって導入された弁別学習を操作要因とした．弁別学習がある条件では，試行内の成功事例の出力（知覚学習条件ではラベル，言語学習条件ではキャプション）をタングラムごとにカウントし，各出力（ラベル／キャプション）に対して最も多く対応付けられたタングラムとの組を形成する．後の試行において，送り手は，その組を利用することでタングラムを名付ける．また，画像生成モジュールでは，他のタングラムと組まれたラベルを避けるようネガティブプロンプトを与え，それらと類似しない表象を生成させる．

この学習プロセスを 10 試行繰り返すシミュレーションを独立に 20 回実行した．Stable Diffusion の seed 値は画像生成結果に影響するため，事前生成した 20 組の seed 値ペア（送り手・受け手）を各実行に割り当てた．

[*1] https://huggingface.co/nlpconnect/vit-gpt2-image-captioning

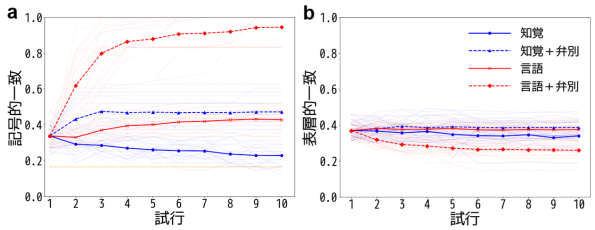

図 2　4 条件の主要結果の比較図．a. 記号的一致推移，b. 表象的一致推移．薄線は seed ごとの推移，太線は平均．言語 + 弁別条件では記号的一致が最大化する一方で表象的一致は継続的に低下し，記号的一致と表象的一致の乖離が最も明瞭に現れる．

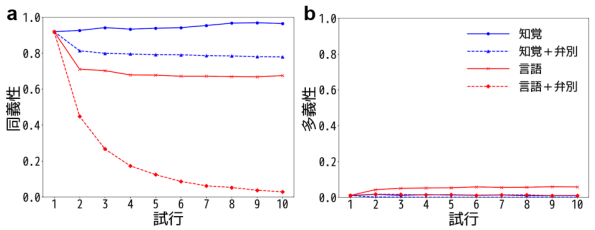

図 3　4 条件の語彙構造指標の比較図．a. 同義性推移，b. 多義性推移．薄線は seed ごとの推移，太線は平均．とくに言語 + 弁別条件で語彙構造変化が大きい傾向を示す．

3.2　評価指標

本研究では，記号的一致と表象的一致を評価し，語彙構造指標を補助的に用いる．

- 記号的一致: ゲームにおける同定正解率（受け手が送り手の意図したタングラムを正しく選択できた割合）．
- 表象的一致: 送り手が生成した画像と受け手が生成した画像を同一の視覚空間モデル（本システムの Vision Transformer）に入力，分類ヘッド出力（1000 次元）を特徴ベクトルとしたコサイン類似度．
- 同義性: 回転に対する頑健性を示す指標．1 つのタングラム（8 角度）に対し，何種類のユニークな語彙が用いられたかの平均値（低いほど良い）．知覚学習条件ではラベル，言語学習条件ではキャプションを対象とする．
- 多義性: 言葉の意味の明確さを示す指標．1 つの語彙が，いくつの異なるタングラムを指しているかの平均値（低いほど良い）．知覚学習条件ではラベル，言語学習条件ではキャプションを対象とする．

4.　結果

4.1　記号的一致と表象的一致へ及ぼす影響

図 2 は，学習チャネルと弁別学習を操作した 4 条件における記号的一致と表象的一致の推移を示す．なお，受け手は 6 種類の候補タングラムから 1 つを選択するため，記号的一致のチャンスレートは 1/6 (=0.167) である．

記号的一致に関する結果は概ね先行研究で得られた知見を再現している．すなわち記号的一致には，知覚チャネルの更新ではなく言語チャネルの更新が寄与し (Baba et al., 2026)，さらに弁別学習を含めることで記号的一致が向上する (山本他, 2025)．加えて，本研究ではこれら 2 要因を組み合わせた条件（言語学習＋弁別学習）において，送り手と受け手の言語利用が完全一致に近づくことが示された．

一方で，表象的一致に関しては異なる傾向が観測された．いずれの条件においても表象的一致の顕著な向上は認められず，コミュニケーションの反復のみでは内部表象の整合が容易には達成されないことが示唆される．

さらに重要なのは，両指標の関係である．特に言語チャネルかつ弁別学習あり条件では，記号的一致がほぼ完全一致に近づく一方で，表象的一致は顕著に低下した．すなわち，記号的一致と表象的一致は同方向に改善するとは限らず，条件によっては両者が乖離することが確認された．この結果は，課題成功のために言語的表現の一致を高めることが，必ずしも内部表象の整合を伴わず，むしろ表象の乖離を生じさせる可能性を示唆する．

4.2　言語的表現の遷移

図 3 は各条件において生成された言語キャプションの分析結果を示す．まず，多義性に関わる図より，本研究で設定されたいずれの条件も語彙の多義性が抑えられたことがわかる．一方で同義性に関しては本研究で設定した両要因の効果が生じた．すなわち言語チャネルの更新および弁別学習の導入により送り手が生成するキャプションは一意なものに近づくことが示される．この結果は図 2 における記号的一致の遷移と整合するものである．

4.3　内部表象生成の事例

図 4 は，本研究で得られた学習過程の具体例である．ここでは高い記号的一致を示した言語 + 弁別条件のなかで相対的に表象的一致の高かった事例と低かった事例を比較している．

図 4 より，相対的に表象的一致の高かった事例では，送り手のキャプションから再構成された受け手画

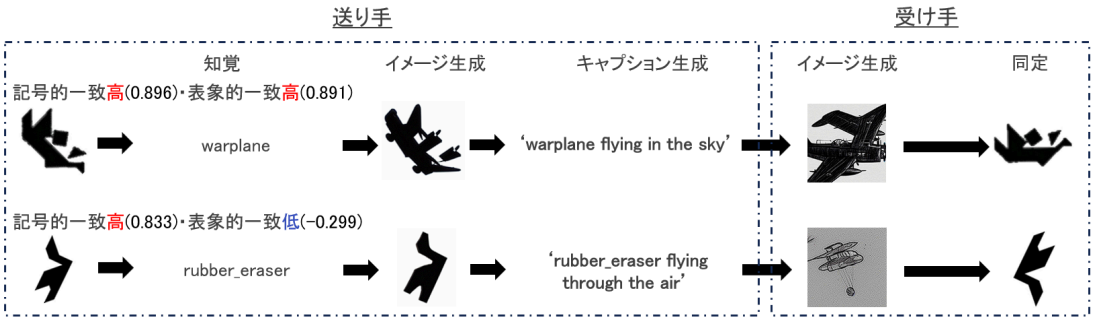

図 4　具体例の比較図．言語 + 弁別条件における記号高条件（trial 記号的一致が高い集合）から，表象的一致の相対上位例と相対下位例を示す．括弧内は各指標（記号的一致・表象的一致）の値を示す．

像が対象タングラムの輪郭（warplane の形状）を比較的保持していることが読み取れる．これに対し表象的一致が相対的に低い事例では，結果的には記号的に一致したタングラムが選択されたものの，送り手と受け手のイメージでは幾何的特徴が大きく異なっている．これは図 2 で示した言語 + 弁別条件における記号的一致の高止まりと表象的一致の低下，および図 3 で示した語彙構造の収束傾向と整合する．

5.　考察とまとめ

本研究における重要な結果は，言語 + 弁別条件で記号的一致が最大化する一方，表象的一致が継続的に低下した点である．これは，先行研究を踏まえつつ，本研究で設定した両要因（学習チャネル，弁別学習）が 2 指標を一様に改善するのではなく，条件に応じて寄与先が分かれることを示す．従来の共通基盤形成や記号創発に関する研究では記号的一致が主要な成功指標とされてきたが (Hawkins et al., 2020; Lazaridou et al., 2017)，本結果はその有効性を否定するのではなく，記号的一致の改善だけでは表象的一致が同時に保証されない条件があることを示した．

この結果は，共通基盤を単一指標へ還元する立場の限界を示しており，少なくとも記号的一致と表象的一致の二水準で評価する必要がある．とくに言語＋弁別条件では，同定に成功した表現を強化する形で学習が進むため，言語表現は相手に通じる方向へ収束する一方，各エージェントの内部表象を互いに近づける学習は働かず，記号的一致だけが先行して高まりやすい．これに対し知覚チャネルでは，表象空間の制約により両者の対応が保持されやすい．こうした乖離は，人どうしでいえば，専門家と非専門家が同じ用語で別の内容を想定したまま協働が進むようなすれ違いに対応し，図 4 の事例はその計算論的な例示である．

ただし，本解釈には限界がある．第一に，表象的一致は単一指標に依存しており，異なる表象指標では乖離の大きさが変化しうる．第二に，本課題は送り手と受け手の役割を固定した一方向的コミュニケーション設計であり，双方向の相互調整過程を直接には扱っていない．第三に，学習更新は送信側の更新対象に限定され，受け手の学習を明示的に含まないため，受信側適応を伴う共通基盤形成へは直ちに一般化できない．

したがって，本研究の理論的含意は，共通基盤を単一一致ではなく，条件依存の多層整合として捉える点にある．記号的一致のみを最適化した系は，単純課題では成功しても，文脈が変化すると内部表象の不整合が露呈しうる．とくに，文化的背景や専門性の異なる主体が協働する状況では，記号的一致だけでは長期的・複雑な問題への対応が不十分になりうる．今後は，代替指標と双方向学習への拡張を通じて，表象的一致の達成可能性と汎化への寄与を検証する．

文　献

Baba, R., Morita, J., Higashinaka, R., & Takeuchi, Y. (2026). Modeling the Depth of Grounding through a Generative Cognitive Model in Referential Communication. *Computers in Human Behavior: Artificial Humans*, 100262.

Bouchacourt, D., & Baroni, M. (2018). How Agents See Things: On Visual Representations in an Emergent Language Game. *arXiv preprint arXiv:1808.10192*.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Feigenbaum, E. A., & Simon, H. A. (1962). A Theory of the Serial Position Effect. *British Journal of Psychology*, *53*, 307–320.

Hawkins, R. D., Franke, M., & Goodman, N. D. (2020). Characterizing the Dynamics of Learning in Repeated Reference Games. *Cognitive Science*, *44* (12), e12947.

Lazaridou, A., Peysakhovich, A., & Baroni, M. (2017). Multi-Agent Cooperation and the Emergence of (Natural) Language. *arXiv preprint arXiv:1612.07182*.

Lowe, R., Foerster, J., Boureau, Y.-L., Pineau, J., & Dauphin, Y. (2019). On the Pitfalls of Measuring Emergent Communication. *arXiv preprint arXiv:1903.05168*.

森田 純哉・東中 竜一郎・竹内 勇剛 (2024). 認知アーキテクチャと生成 AI が織りなす人と機械の共通基盤　人工知能, *39* (2). 特集「生成 AI 時代における認知のモデリング」.

山本 大暉・馬場 龍之介・森田 純哉・東中 竜一郎・竹内 勇剛 (2025). 共通基盤形成における知覚学習の停滞メカニズム解明と弁別学習による解決 . 第 23 回情報学ワークショップ (WiNF2025) 抄録集.