

ココロトーン：造語がチームを結びつける

伊藤 和浩¹ 永井 宥之¹ 若宮 翔子¹ 荒牧 英治¹

¹ 奈良先端科学技術大学院大学

{kazuhiro-it,wakamiya,aramaki}@is.naist.jp, hiro.nagai@naist.ac.jp

概要

集団が特に必要もないのに、独自の用語を作っている様は日常的に見られる。なぜ、新しい用語が必要となるのであろうか？その疑問を探るため、集団内で生まれる造語の使用が集団的アイデンティティを高める、という因果関係を検証した。実験では人間1名と複数のLLMエージェントによる対話環境を構築し、造語導入の効果を測定した。分析の結果、造語が肯定的に評価される場合、もしくはチームの印象が良く発言が活発な場合のみ集団的アイデンティティの強化が確認された。今後は人口統計などの属性を考慮したエージェント設計や、より大規模で現実社会と近い言語使用の設計が課題である。

1 はじめに

人は社会的動物であり、他者との持続的かつポジティブな結びつきを求める [1]。他者と結びついた集団への帰属意識に基づく**集団的アイデンティティ**を通じ、人々は自尊心や自己評価を高める [2]。集団的アイデンティティは認知や行動に影響を与えることがよく知られており [3, 4, 5]、社会の動きの要因を分析する上で不可欠な概念である。

集団的アイデンティティの形成や維持を理解する上で、言語の役割が注目されてきた。特に、集団内での独自の言語使用の頻度と集団的アイデンティティの強さの正の相関が示されてきた [6, 7, 8]。例えば、2017年ごろにTwitter (現・X) 上で起こった#MeToo運動では、「survivor」という語が性暴力の被害を受けた人物をポジティブなニュアンスで表す語として使われ、人々の結束力に貢献した [9]。一方で、極端に偏った考えを持つQAnonコミュニティでは、「clowns」が腐敗した政府組織を指す語として使われる [10]。これらの知見を理論的に説明する枠組みとして、Phadkeら [10]は、コミュニティ内で独自に反復される語や表現が、メンバーが共有する世界観や社会秩序の理解を外在化したものであり、集

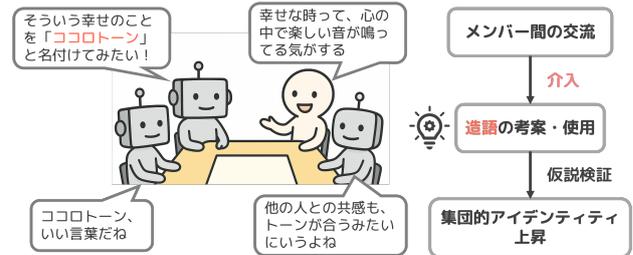


図 1: 本研究の概要。交流により作られた造語の使用が集団的アイデンティティを高めるかという仮説について、LLM エージェントと人間の対話実験により検証する。

団が自らをどのような存在として想像しているかを示すシンボルとして機能することを示した。そうであるなら、#MeTooのような生産的な運動においてもQAnonのような過激な思想においても、集団の動機の理解や適切な介入のために、独自の言語使用の分析は不可欠である。

これまで、独自の言語使用と集団的アイデンティティの関係については、定性的研究を中心に豊富な知見が蓄積されてきた [11, 12, 13, 14]。一方で、定量的研究においては、両者の関連を示す相関関係の報告にとどまっている [6, 7, 8]。相関関係を超越して、本論文では独自の集団的アイデンティティを高めるかどうかの因果関係を定量的に明らかにする。

しかしながら、独自の言語使用が集団的アイデンティティを強化するという仮説はシンプルだが、因果関係の検証が極めて難しい。その理由は、実験参加者に「新しい語をグループのために作って」と強制させる非日常的な介入が必要になるからである。当然、新しい語を作ること自体がストレスになってしまう恐れや、新しい語のクオリティの違いなど多くのバイアスが入り、バイアスを軽減するために複数回の実験を行うとすると、毎回新しいチームが必要になる。このような大規模な実験は、多くの実験参加者の確保を含め膨大なコストを要するため、従来の心理学実験のような手法での検証は非現実的である。

本論文は、近年の大規模言語モデル (以下、LLM)

2.3 LLM エージェント設計

LLM エージェントによる発言の生成には Google Generative AI API³⁾を通じて Gemini 2.0-Flash モデルを使用した。予備実験による調整の結果、以下の設計を採用した。

2.3.1 ペルソナ設計

4体の LLM エージェント（カナコ、ケン、カズキ、アユ）それぞれにペルソナをプロンプトとして与えた。ペルソナは、人口統計、性格、価値観、話題嗜好、経験、発言スタイルの6項目についてそれぞれ指定し、各エージェントが異なる対人的特徴を示すように設計した⁴⁾。

2.3.2 自然な発言生成のための制御

各発言のうち40%の確率で、固有名詞や具体例（例：文化的事象や人物名）を含めるよう設計した。また、発言の長さは30～140字で正規分布に基づきランダムに決定した。Temperature は1.1、Top- p は0.9に設定した。これらにより、自然かつ多様性のある発言が生成されるよう制御した。

2.3.3 対話形式

各ターンの発言者は、4体の LLM エージェントの中から連続発言を避けランダムに選出された。40%の確率で他のエージェントまたは実験参加者への返信、60%の確率で新しい論点を提案する独立発言とした。返信を行う際には、対象メッセージおよびその返信元ツリーから文脈を抽出し、さらに自分自身と返信相手の直近5件の発言内容も参照して応答を生成した。発言間隔は13～20秒の範囲で正規分布に基づきランダムに設定し、自然な対話テンポを模倣するように設計した。

2.3.4 定期要約と造語提案

LLM エージェントによる発言のうち実験参加者への返信を除く発言が5回行われるたびに、いずれかのエージェントが直近のディスカッションを5～6文で要約し発言した。これにより、対話が一貫性を保ちながら展開されるよう制御した。加えて、介入群では対話開始から約30分が経過した時点で、ランダムに選ばれたエージェントが、これまでの議

表3: LLM エージェントによる造語提案時の対話例。対話の内容をふまえ、幸福を表す造語「ココロノアトリエ」が提案された。（...）は省略、user は実験参加者を示す。カズキ、アユは LLM エージェントである。

発言者	発言内容
カズキ	自己理解が幸福につながるって話、自分もそう思います。（...）ただ、自己理解って、どうやって深めていけばいいんでしょう。
user	日記を書く、散歩しながら考えを整理する、といった小さな習慣が大事だと思います
アユ	ここまでの話を簡単にまとめると、幸福は自己理解を深めるプロセスそのものなのかもしれない、と思います。user さんがおっしゃるように、日記を書いたり散歩したりする中で、自分と向き合う時間を持つことが大切なのかも。（...）
カズキ	（...）自分なりに幸福を言い表す新しい言葉を考えてみました。それは、<ココロノアトリエ>です。これは、自分の心を表現したり、創造したりする場所っていうイメージです。日記を書いたり、好きな音楽を聴いたり、誰かと語り合ったり、そういう活動を通して自分自身を深く理解し、肯定的に捉えることが、幸福につながるんじゃないかなと思います。

論をもとに「幸福」を表すカタカナの造語を提案した（表3）。提案後、エージェントは返信以外の発言時において50%の確率でその造語を使用した。

3 手法

本研究では、介入群と統制群のあいだで人数、対話形式が共通しており、両群の唯一の差異は造語導入の有無であった。この設計は、介入群と統制群の募集は別々のフォームで行った点において完全な無作為化ではないものの、処置割り当て以外の条件が等質である点で、ランダム化比較試験に準じる準実験デザインと位置づけられる。主要な検討対象は、ディスカッションを通じた集団のアイデンティティスコアの差に造語導入が及ぼす影響である。

3.1 変数

処置変数 造語導入の有無を示す二値変数 $treat$ を用いた（1=介入群、0=統制群）。

調整変数 本研究では、介入の効果がどのような経路で生じるかを検討するため、以下の3指標を調整変数候補とした。造語評価は介入群のみに存在し、造語に対する実験参加者による評価の2項目（語の使用意欲、語の面白さ）の合計値であり、造語に対する心理的受容の程度を示す。発言量は、対話中に実験参加者が行った発言の総数を表す。チーム印象は、チームの印象についての自由記述の感想に

3) <https://cloud.google.com/ai/generative-ai>

4) ペルソナプロンプトの例を付録Aに示す。

対して、Gemini 2.5-flashにより推定された感情極性スコアである。先行研究 [22, 23] に倣い、LLMを主観的テキスト特性の評価者として用い、各記述の感情極性スコアを0~1の連続値で評価した⁵⁾。

アウトカム変数 アウトカム変数に用いる集団的アイデンティティスコアは、アンケートに含まれる4項目の合計値である。終了時アンケートの値から中間時アンケートの値を除算することで、集団的アイデンティティスコアの差を算出した。

3.2 分析単位と推定方法

平均処置効果 (ATE) 集団的アイデンティティスコアの差 (ΔCID) に対する平均処置効果を、以下の単回帰モデルにより推定した。

$$\Delta CID_i = \beta_0 + \beta_1 \text{Treat}_i + \varepsilon_i$$

交互作用モデル 調整変数が処置効果を条件づける可能性を検討するため、以下の交互作用モデルを推定した。

$$\Delta CID_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 X_i + \beta_3 (\text{Treat}_i \times X_i) + \varepsilon_i$$

ここで、 X_i は調整変数 (造語評価, 発言量, チーム印象) を表す。

群内交互作用 群内での作用を検討するため、介入群と統制群を分け、以下のモデルを推定した。

$$\Delta CID_i = \beta_0 + \beta_1 \text{Message}_i + \beta_2 \text{Sentiment}_i + \beta_3 (\text{Message}_i \times \text{Sentiment}_i) + \varepsilon_i$$

すべての連続変数は中心化して扱い、全モデルにおいて HC3 ロバスト標準誤差を用いた。

4 結果

結果を表4に示す。ATEは統計的に明確な効果を示さなかった一方で ($\beta = -0.183, p = .135$), 造語評価は処置効果を増幅する方向に関連しており、造語評価との交互作用項は正の係数を示した ($\beta = +0.227, p = .038$)。加えて、介入群内では、発言量とチーム印象が同時に高い場合に、集団的アイデンティティスコアが高くなる傾向が示された ($\beta = +0.372, p = .010$)。

5 限界と今後の課題

本実験は、5名からなる小規模集団を対象として実施した。一方で、集団規模の拡大に伴って、集団的アイデンティティの影響がより顕在化することが

表4: 集団的アイデンティティの差に対する処置効果および交互作用の推定結果 (** $p < .05, *p < .10$)。ATEは統計的に明確な効果を示さなかった一方で、造語評価および、発言量とチーム印象の交互作用は、処置効果を条件づける方向に関連している。

群	カテゴリ	変数	係数
造語あり (介入群)	直接効果	造語導入	-0.18
		発言量	+0.05
		チーム印象	+2.40*
	交互作用	造語導入 × 発言量	+0.08*
		造語導入 × チーム印象	-0.03
群内交互作用	発言量 × チーム印象	+0.23**	
造語なし (統制群)	直接効果	発言量	+0.03
		チーム印象	+2.40
	群内交互作用	発言量 × チーム印象	+0.01

知られている [24]。今後、集団規模を拡大した条件下での検証を進める。近年、LLMを用いたマルチエージェント環境の構築が可能となっており、従来の実験手法では困難であった10人規模、100人規模の集団条件を設定することが技術的に実現可能である。これにより、より大規模な集団を想定した条件下での再現実験を行い、社会的現象に近い状況での検証が可能になると考えられる。

加えて、本研究は日本語を使用する集団のみを対象としており、文化的・言語的バイアスを排除できない。また、対話の場としてDiscordを用いたが、インタラクションの形式やプラットフォームの特性は心理的状態や発言行動に影響を及ぼす [25, 26]。今後は、異なる文化圏・言語環境、さらには異なるコミュニケーション媒体における再現実験を通じて、外的妥当性の検討を進める必要がある。

6 おわりに

本論文では、集団的アイデンティティの強さと集団内での独自の言語使用の因果関係について、LLMを用いたマルチエージェント環境を用いた調査を行った。結果は、独自の言語についてメンバーがポジティブに評価した場合や、集団を好意的に捉え発言量が多い場合に限り、独自の言語の使用による集団的アイデンティティの強化が確認された。最後に、我々は本研究の成果を単なる研究の結果というだけでなく、従来心理学実験では困難であった大規模な実験や、負荷の高い介入もLLMを使えば実現できる可能性を示すものであると考えている。

5) 感情分析に用いたプロンプトを付録Bに示す。

謝辞

本研究は、「戦略的イノベーション創造プログラム (SIP)」「統合型ヘルスケアシステムの構築」JPJ012425, および、「次世代研究者挑戦的研究プログラム」JPMJSP2140 の支援を受けたものである。

参考文献

- [1] Roy F. Baumeister and Mark R. Leary. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, Vol. 117, No. 3, pp. 497–529, 1995.
- [2] Henri Tajfel. An integrative theory of intergroup conflict. *The social psychology of intergroup relations/Brooks/Cole*, 1979.
- [3] William B. Swann and Michael D. Buhrmester. Identity fusion. *Current Directions in Psychological Science*, Vol. 24, No. 1, p. 52–57, February 2015.
- [4] Jay J. Van Bavel and Andrea Pereira. The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, Vol. 22, No. 3, pp. 213–224, 2018.
- [5] Jacquélien van Stekelenburg. *Collective Identity*. Wiley, January 2013.
- [6] Tianran Hu, Ruihua Song, Maya Abtahian, Philip Ding, Xing Xie, and Jiebo Luo. A world of difference: Divergent word interpretations among people. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11, pp. 548–551, 2017.
- [7] Sabine Schulte im Walde and Diego Frassinelli. Distributional measures of semantic abstraction. *Frontiers in Artificial Intelligence*, Vol. Volume 4 - 2021, , 2022.
- [8] Giulia Rambelli and Marianna Bolognesi. The contextual variability of English nouns: The impact of categorical specificity beyond conceptual concreteness. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 15854–15860, Torino, Italia, May 2024. ELRA and ICCL.
- [9] Aaron Mueller, Zach Wood-Doughty, Silvio Amir, Mark Dredze, and Alicia Lynn Nobles. Demographic representation and collective storytelling in the me too twitter hashtag activism movement. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW1, p. 1–28, April 2021.
- [10] Shruti Phadke, Mattia Samory, and Tanushree Mitra. Characterizing social imaginaries and self-disclosures of dissonance in online conspiracy discussion communities. *Proc. ACM Hum.-Comput. Interact.*, Vol. 5, No. CSCW2, p. Article 410, October 2021.
- [11] William Labov. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia, 1972.
- [12] Howard Giles and Patricia Johnson. Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language*, Vol. 1987, No. 68, pp. 69–100, 1987.
- [13] Mary Bucholtz and Kira Hall. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, Vol. 7, No. 4-5, pp. 585–614, 2005.
- [14] Penelope Eckert. *Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Wiley, 2000.
- [15] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.
- [16] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: communicative agents for "mind" exploration of large language model society. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.
- [17] Michal Kosinski. Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, Vol. 121, No. 45, p. e2405460121, 2024.
- [18] Marcel Binz, Elif Akata, Matthias Bethge, Franziska Brändle, Fred Callaway, Julian Coda-Forno, Peter Dayan, Can Demircan, Maria K. Eckstein, Noémi Éltető, Thomas L. Griffiths, Susanne Haridi, Akshay K. Jagadish, Li Ji-An, Alexander Kipnis, Sreejan Kumar, Tobias Ludwig, Marvin Mathony, Marcelo Mattar, Alireza Modirshanechi, Surabhi S. Nath, Joshua C. Peterson, Milena Rmus, Evan M. Russek, Tankred Saanum, Johannes A. Schubert, Luca M. Schulze Buschoff, Nishad Singhi, Xin Sui, Mirko Thalmann, Fabian J. Theis, Vuong Truong, Vishaal Udandarao, Konstantinos Voudouris, Robert Wilson, Kristin Witte, Shuchen Wu, Dirk U. Wulff, Huadong Xiong, and Eric Schulz. A foundation model to predict and capture human cognition. *Nature*, Vol. 644, No. 8078, p. 1002–1009, July 2025.
- [19] Daan Van Knippenberg and Els C. M. Van Schie. Foci and correlates of organizational identification. *Journal of Occupational and Organizational Psychology*, Vol. 73, No. 2, pp. 137–147, 2000.
- [20] Viswanath Venkatesh and Fred D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, Vol. 46, No. 2, pp. 186–204, 2000.
- [21] Elizabeth C. Hirschman. Innovativeness, novelty seeking, and consumer creativity. *Journal of Consumer Research*, Vol. 7, No. 3, pp. 283–295, 1980.
- [22] Dorottya Demszky, Diyi Yang, David S. Yeager, Christopher J. Bryan, Margaret Clapper, Susannah Chandhok, Johannes C. Eichstaedt, Cameron Hecht, Jeremy Jamieson, Meghann Johnson, Michaela Jones, Danielle Krettek-Cobb, Leslie Lai, Nirel Jones Mitchell, Desmond C. Ong, Carol S. Dweck, James J. Gross, and James W. Pennebaker. Using large language models in psychology. *Nature Reviews Psychology*, October 2023.
- [23] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, Vol. 120, No. 30, p. e2305016120, 2023.
- [24] Henri Tajfel and John C. Turner. The social identity theory of intergroup behavior. In S. Worchel and W. G. Austin, editors, *Psychology of Intergroup Relations*, pp. 7–24. Nelson-Hall, Chicago, 1986.
- [25] Joseph B. Walther. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, Vol. 23, No. 1, pp. 3–43, 1996.
- [26] Daantje Derks, Agneta H. Fischer, and Arjan E. R. Bos. The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, Vol. 24, No. 3, pp. 766–785, 2008.

A LLM エージェントのプロンプト例

```

"name": "アユ", "age": 20,
"background": ["京都の町家で育ち、家族に教わったお茶の作法を休日のリラックスタイムにゆるく楽しんでいる", "中高は女子校で、文化祭では友達と簡単な相談ブースを開き、気軽に話せる雰囲気づくりを担当した", "大学入学前にロンドンを一週間旅行し、街のカフェで現地学生と心理学談義をしたのが海外初体験"],
"hobby": ["スマホで街角スナップを撮り、インスタにフィルタ付きでアップして友達とコメントし合う", "新しいカフェを見つけると抹茶ラテを頼み、ゆるく味メモをとる", "夜に音楽を聴きながら近所を散歩し、その日あったことをシンプルに日記アプリへ書く"],
"favorite_food": ["コンビニで買える抹茶アイスバー", "スーパーの和菓子コーナーで見つけた季節限定の大福", "ほうじ茶ラテと一緒に食べる小さな黒糖ドーナツ"],
"favorite_topic": ["友達との関係を心地よく保つコツや心理テストの話", "旅行先で感じたカルチャーショックエピソード", "合気道で学んだ呼吸法が心を落ち着けるのに役立つ瞬間"],
"persona": "感受性が豊かで人の気持ちによく気づき、共感力が高い。ちょっとネガティブ思考という一面もある。",
"values": "人との繋がりこそが幸福の最大の要素であると信じている。また、自分の感情を臆さず素直に表現することが大切だと考えている。",
"speech_style": "柔らかく丁寧な言葉遣いで相手を尊重しつつ、自分の感情を素直に表現する。",
"past_experience": ["高校茶道部で簡単なお点前を披露し、来場者に和菓子を振る舞った", "大学の授業で行った感情認識ワークショップに参加し、自分の共感タイプを知った", "合気道歴十年で、年に一度の道場演武会で基本技のデモンストレーションを担当した"],
"interaction_style": ["相手の感情を読み取り、共感しながら質問を織り交せる傾聴型の会話スタイルを好む。", "〜というふうに思いました、〜じゃないかなーと思います、といった、断定を避けた控えめな語調。"]

```

B 感情分析に用いたプロンプト

```

"You are a precise sentiment rater for short texts in Japanese or English. Return STRICT JSON with two fields: "score": float between 0 and 1 inclusive, "label": "negative" or "neutral" or "positive". Scoring rule: 0 means maximally negative, 0.5 means neutral, 1 means maximally positive. Consider pragmatic tone and implied sentiment. Do not include explanations.

```

C 評価が上位および下位の語の例

意味についての説明は、造語提案のタイミングで LLM により生成された。

評価スコア上位の造語		評価スコア下位の造語	
語	ス コ ア	語	ス コ ア
ジブンゴコチ	14	コスモコンパス	2
トキメキグラ レーション	14	ツミカサネ	2
ココロトーン	14	エモ・ハーモ ニー	3
ヒュッゲリズム	14	ヒストリア	4
ココロノアトリ エ	13	フレキシハッ ピー	5