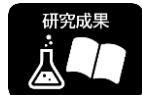


2026年2月5日



分野：社会科学系/工学系

キーワード：大規模言語モデル（LLM）、単語の心理特性、言語発達、AIの心理学、SDGs

AIが「ことばの印象」をどこまで再現できるかを検証 —単語に対するイメージは人間とLLMでどのくらい似ているの?—

【研究成果のポイント】

- ◆ 英語圏で発達初期に習得される約700単語を対象に、「具体性」「感情価」「身体性」など21の心理的特徴を大規模言語モデル(LLM)に評定させ、人間の評定値と比較
- ◆ LLMによる評定は、「具体性」(単語がどのくらい具体的だと感じるか)など一部の指標において人間と高い一致率を示したが、「類像性」(単語の音と意味がどのくらい近いと感じるか)などの指標では人間と乖離を示すことを発見
- ◆ 人間の評定値を収集するには多大なコストがかかるが、LLMを活用することで、単語の心理的特徴を網羅的に評定できる可能性が示され、心理学・言語学研究の効率化や、言語発達のメカニズムの解明、保育・教育的支援への応用に繋がると期待

❖ 概要

大阪大学大学院人間科学研究科の萩原広道講師、同大学院基礎工学研究科の宮澤和貴助教の研究グループは、ChatGPTなどに代表される大規模言語モデル(LLM)^{※1}が、「単語に対する人間の感覚」をどれほど再現できるのかを、21の指標にわたって網羅的に検証し、現代のLLMがもつ可能性と限界を明らかにしました。

単語の心理的特徴は、幅広い研究分野で用いられていますが、指標によって人間による評定がある単語とない単語とが混在しているため、網羅性に欠けていました。また、既存の指標は特定の言語に偏っているため、さまざまな言語を比較しながら研究を進めるための土壌は未整備でした。

今回、研究グループは、人間の代わりにLLMに単語の評定値を推定させることで、これらの課題を克服できるのではないかと着想し、その最初の一步として本研究を行いました。

対象としたのは、英語圏の乳幼児が早期におぼえる約700の単語です。これらの各語について、「具体性」「社会性」「感情価」「視覚性」など既存の21の指標にわたる、単語の心理的特徴^{※2}をLLMに評定させ、既存研究で報告されている人間の評定データと比較しました。

その結果、単語がどのくらい具体的かを示す「具体性」や、単語の具体的なイメージをどのくらい簡単に

以下の単語はどのくらい「具体的」？

抽象的 ← → 具体的
apple 1---2---3---4---⑤

because ①---2---3---4---5

以下の単語はどのくらい「音と意味が似ている」(類像性)？

似ていない ← → 似ている

kick 1---2---3---④---5

better 1---②---3---4---5

早期習得 約700語(英語)
心理的特徴21項目



図1 本研究の概略図

乳幼児の早期習得約700語(英語)を対象に、21項目の心理的特徴をLLMに評定させ、人間の評定値と比較することで、LLMが「ことばの印象」をどこまで再現できるかを検証した。

思い描けるかを示す「想像性」などいくつかの指標では、LLM による評価は全体的に、人間による評価とよく一致することが明らかになりました。一方で、単語の音と意味がどのくらい似ていると思うかを示す「類像性」や、単語がどのくらい強い感情を喚起するかを示す「覚醒度」などの指標では、LLM は人間と大きく異なる評価をすることを発見しました。さらに、全体としては高い一致率を示す指標でも、単語のカテゴリーによっては、LLM と人間の評価との間に乖離がみられることを特定しました。

本研究は、ことばの意味や感覚に関する人間の認知が、身体をもたない LLM にどの程度内在しているのかを検証するとともに、心理学や言語学などの研究において、LLM が人間に代わって一部のデータを補完する新しい方法論的な可能性を示したものです。今後、心理指標が未整備な他言語への応用や、新しい心理的特徴の開発時に LLM を補助的に活用する際に本研究の知見が役立つと期待されます。

本研究成果は、心理学・認知科学分野の国際誌「Behavior Research Methods」に、2月2日(月)に公開されました。

【萩原講師のコメント】

言語発達研究では、「子どもがおぼえやすい単語ってどういうもの？」という疑問が広く共有されています。こうした疑問の解決を図る場合に、本研究で得られた成果が役に立てば嬉しく思います。

【宮澤助教のコメント】

「AIやロボットがどのように言葉を学習し、その意味を理解していくのか？」という問いに関心があります。本研究の成果が、人間とは異なる機械の言語獲得についても探求を進める一助になれば幸いです。

❖ 研究の背景

近年、ChatGPTなどのLLMは、文章生成だけでなく、論理的推論や感情分析など、さまざまな認知タスクにも応用されています。一方で、「単語に対する人間の感覚や印象」といった、心理的・主観的な情報をどこまで再現できるのかについては、まだ十分に検証されていませんでした。

「具体性」「類像性」といった単語の心理的特徴は、心理言語学、発達心理学、認知科学、情報学など幅広い研究分野で用いられており、発達心理学においては、「具体性の高い単語は早期に習得されやすい」など、単語のどのような特徴が子どもにとっての覚えやすさに関連するのかを明らかにする研究が精力的に進められてきました。しかし、人間による評価値がある単語とない単語とが混在しているために網羅性に欠けていました。

また、さまざまな単語の評価値を得るためにデータを収集するには大きなコストがかかるため、既存の心理的特徴の網羅性を高めたり、新しい指標を開発したりすることは極めて困難でした。さらに、公開されている単語の心理的特徴のほとんどは英語圏のものであり、さまざまな言語でデータを収集することには大きなハードルがあるという現状がありました。

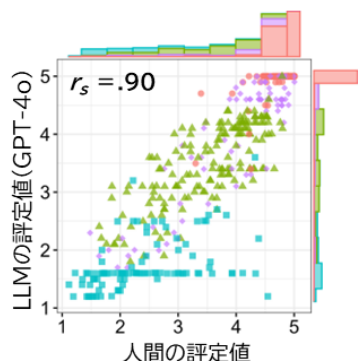
❖ 研究の内容

今回、研究グループは、人間による評価値の既存研究が豊富な英語を対象に、乳幼児が早期に習得する約700単語について、多岐にわたる既存の心理的特徴をLLMに評価させることで、単語の心理的特徴を人間の代わりにLLMが評価・再現できる可能性を探りました。研究には、GPT-4o(OpenAI社)やLlama-3.1(Meta社)を含む4種類のLLMが用いられました。その結果、**「具体性」などのいくつかの指標で、LLMは人間の評価を高い精度で再現できる一方、「類像性」などの指標では、LLMは人間と大きく**

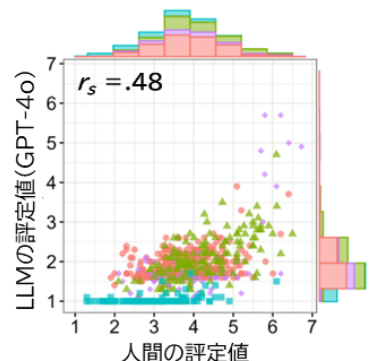
ことなる評定をすることを発見しました。また、全体としては高い一致率を示す指標でも、「名詞」「動詞・形容詞」などのカテゴリーごとに分析すると、特に前置詞や接続詞などの「機能語※³」において、LLM と人間の評定との間に乖離がみられることを明らかにしました。

例えば、人間は“in”や“more”などの単語は、“of”や“because”などの単語よりも具体的だと判断する傾向がありますが、LLM はこれらの単語を総じて「具体的でない」と判断しやすいことがわかりました。さらに、一部の LLM では、特定の心理的特徴がもつ「単語の習得時期」に対する予測力を、過剰に大きく推定してしまう場合があることを特定しました。

具体性 Concreteness (意味がどのくらい具体的か)



類像性 Iconicity (音と意味がどのくらい似ているか)



● 名詞 ▲ 述語(動詞・形容詞) ■ 機能語(前置詞・接続詞など) ◆ その他(挨拶など)

図2 人間と LLM の評定値の比較(例)

「具体性」は人間と LLM とで高い一致率を示すが、「類像性」は人間と LLM とで評定の傾向が大きく異なる。また、全体的に高い一致率を示す「具体性」でも、前置詞や接続詞などの「機能語」に注目すると、人間の評定では値が高いものと低いものがあるが、LLM では一貫して低い評定値となった。

「機能語」に対する予測力を、過剰に大きく推定してしまう場合があることを特定しました。

❖ 本研究成果が社会に与える影響(本研究成果の意義)

本研究は、AI が人間の「ことばに対する感じ方」をどれだけ再現できるのかを網羅的かつ精密に評価した初の試みの一つです。本研究成果を活用することで、心理学や言語学の研究をより効率化したり、心理指標が未整備な他言語への展開も視野に入れたりすることができると期待されます。

実際に、心理指標によっては、人間データでは欠損が多い一方で、LLM では対象となった単語のほぼすべてに対して、欠損なく即座に評定可能であることが示され、LLM の方法論的利点も確認されました。加えて、本研究成果は、子どもの言語発達に影響する単語ごとの要因を明らかにする基礎研究として、言語発達のメカニズムの解明や保育・教育的支援への応用も期待されます。

近年、AI が人間のように振る舞えるという知見が増えていますが、本研究成果から、実際には「得意なこと」と「苦手なこと」があるということも明らかになりました。本研究は、AI を心理学や認知科学における「研究参加者」として扱い、物理世界や人間の認知をAIがどのくらい反映しているかを明らかにするという「マシン・サイコロジー」(機械の心理学)という新しい研究潮流にも貢献する知見だといえます。

❖ 特記事項

本研究成果は、2026 年 2 月 2 日(月)に国際科学誌「Behavior Research Methods」(オンライン)に掲載されました。

タイトル: “How well do large language models mirror human cognition of word concepts?: A comparison of psychological ratings for early-acquired English words”

著者名: Hiromichi Hagihara, and Kazuki Miyazawa

DOI: <https://doi.org/10.3758/s13428-025-02938-2>

なお、本研究は、JSPS 科研費(JP23K11786)からの助成を受けて実施されました。

❖ 用語説明

※1 大規模言語モデル(LLM)

人間が書いた膨大な文章などをもとに、言語パターンを学習した AI(人工知能)モデルの一種。ChatGPT(OpenAI 社)や Llama(Meta 社)などがあり、文章の生成や要約、翻訳などを行うことができる。

※2 単語の心理的特徴

人間が、ある単語に対して感じる印象や性質のこと。たとえば「リンゴ」は具体的で、五感に訴える特徴(赤い、甘いなど)があるのに対して、「平和」は抽象的で、特定の五感との関連性は低く、社会的な性質が強いなどが挙げられる。

※3 機能語

名詞や動詞などの「内容語」と対比して、単語それ自体には明確な意味が乏しく、文法的な関係などを示すようなものをいう。前置詞や接続詞などが該当する。日本語の場合、「が」「は」「へ」などの助詞が当てはまる。

❖ SDGs目標**❖ 参考 URL**

萩原広道講師 研究者総覧 <https://rd.iai.osaka-u.ac.jp/ja/f2ec3ddc37fb11c3.html>

Researchmap <https://researchmap.jp/hagiii?lang=ja>

宮澤和貴助教 研究者総覧 <https://rd.iai.osaka-u.ac.jp/ja/9756c3fd91835615.html>

Researchmap <https://researchmap.jp/kazukimiyazawa>

❖ 本件に関する問い合わせ先

大阪大学 大学院人間科学研究科 講師 萩原広道(はぎはらひろみち)

TEL: 06-6879-8045

E-mail: hiromichi.h.hus@osaka-u.ac.jp

❖ 発信先 報道機関

大阪大学から 大阪科学・大学記者クラブ、文部科学記者会、科学記者会