個別選好の異質性を考慮した大喜利ユーモア選好要因の分析

村上 聡一朗 [1] 上垣外 英剛 [1,2] 高村 大也 [3] 奥村 学 [3]
[1] 株式会社サイバーエージェント [2] 奈良先端科学技術大学院大学 [3] 東京科学大学
murakami_soichiro@cyberagent.co.jp
kamigaito.h@is.naist.jp {takamura,oku}@pi.titech.ac.jp

概要

ユーモアの好みは個人差・文化差が大きく，LLM を用いたユーモア評価を難しくする．本研究では，日本語の大喜利を対象に，投票履歴に基づいてユーザをクラスタリングし，各クラスタに対して解釈可能な選好要因上の重みを Bradley–Terry–Luce モデルで推定することで，選好の異質性をモデル化する．さらに，LLM に「最も面白い回答の選択」を行わせて選好データを収集し，同一の要因集合で BTL 推定した重みをユーザクラスタと比較した．その結果，ユーザクラスタ間で異なる選好パターンが観察され，LLM の選好が特定クラスタに近い場合があることを示す．最後に，ペルソナプロンプトにより LLM の選好を特定クラスタへ誘導できることを示す．

1 はじめに

大規模言語モデル（LLM）は，人間に近い創造的推論を可能にするものとして注目を集めている．文脈理解や微妙なニュアンスの把握を要するユーモア理解・生成は，その能力を評価する有用な題材である．一方で，ユーモアは主観的で文化依存性も高く，「面白さ」の定量化や再現は自然言語処理における難題である [1]．本研究では，大喜利に着目する．大喜利はお題に対して機知に富んだ回答をする質問応答型のユーモア形式である．

LLM のユーモア理解・生成を改善するには，人間がどのようなユーモアを好むかを明確化し，LLM と人間のギャップを特定することが重要である．しかし，既存研究には 2 つの課題が残る．第一に，人間のユーモア選好分析では個人差が十分に扱われてこなかった．従来は複数評定者の評価を平均して「全体スコア」を作り分析することが多いが [2, 3]，主観評価では一致率が低いことが多い [4]．したがって，ユーザ間で選好が異なるとみなすのが自然である [5, 6]．第二に，LLM と個々の人間選好のギャップは明確化されていない．Sakabe ら [3] は，人間は *empathy* を，LLM は *novelty* を重視すると報告したが，議論は集約された選好に基づくもので，個人（あるいはユーザ群）との整合性は未検証である．

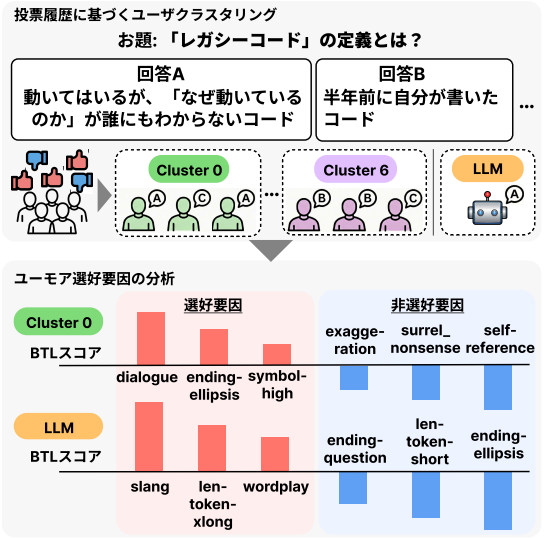

図 1 ユーザクラスタと LLM を横断したユーモア選好要因分析の概要．各クラスタ，LLM の選好要因を推定する．

そこで本研究では，投票データに付随するユーザ ID を用い，投票履歴に基づいてユーザをクラスタリングし，クラスタ単位でユーモア選好要因を分析する（図 1）．各クラスタについて，言語的特徴量（例：長さ）およびユーモア戦略ラベル（例：ブラックジョーク）からなる解釈可能な要因集合を定義し，Bradley–Terry–Luce（BTL）モデル [7] により要因重みを推定する．また，LLM に各お題の回答集合から最も面白いものを 1 つ選ばせることで LLM 選好データを構築し，同一の要因集合で BTL 推定した重みをユーザクラスタと比較する．さらに，LLM 選好を特定クラスタへ誘導する方法を検討する．

本研究は次の研究課題に答える．（RQ1）ユーザクラスタのユーモア選好要因はどう異なるか．（RQ2）LLM の選好は特定のユーザクラスタと整合するか．（RQ3）LLM 選好を特定クラスタへ誘導できるか．

表 1　BTL 分析で用いたユーモア選好要因の概要

要因グループ (数)	概要
言語的特徴量 (45)	
表層 (11)	表層統計（例：文字数，文字種比率）
形態素 (10)	形態素解析による特徴（例：品詞比率）
特殊記号 (5)	特殊記号等の使用（例：引用符）
文末 (9)	文末パターン（例：疑問符で終わるか）
文体 (4)	日本語特有の文体（例：丁寧体，誇張）
関係性 (6)	お題と回答の関係（例：長さ比）
ユーモア戦略 (11)	ユーモア理論に基づくラベル（例：`incongruity`, `black_joke_satire`）

本研究の主な貢献は 3 つある．(i) ユーザクラスタごとの要因重み推定により，クラスタ間で異なる選好パターンが存在することを定量化した．(ii) 集約選好ではなくクラスタ単位で比較することで，LLM の選好は特定クラスタに近いことを示した．(iii) ペルソナプロンプトが LLM 選好を特定クラスタへ誘導することを示した．これらの結果は，ユーモア理解・生成の個別最適化の必要性と可能性を示す．

2　関連研究

ユーモア理解・生成研究は活発化している [8]．ユーモアには駄洒落など様々な形式があり，データセットや手法が提案されてきた [9, 10, 11]．近年，大喜利については大喜利プラットフォーム（例：Bokete）が普及し，Zhong ら [12] は Oogiri-GO データセットを構築し，研究を加速させた [2, 3]．

ユーモア選好要因を定量化することは，ユーモア理解研究の前進に有用である [2]．Sakabe ら [3] は人間と LLM の差異を分析したが，集約選好に基づくため個人差や群差の観点が残る．Chakrabarty ら [5] は投票データからユーザクラスタを構築したが，クラスタ別の選好パターンは分析されていない．本研究の新規性は，ユーザクラスタ単位で選好要因を分析し，LLM と人間の関係を詳細分析する点にある．

3　分析データセットの構築

既存の大喜利データセットはお題・回答ペアと得票数から構成されるが，誰がどの回答に投票したかという情報を欠くため，個人選好の追跡ができない [3]．本研究では，Oogiri-Corpus [2] を拡張し，ユーザ単位の投票履歴を含むデータセットを構築した．

データソース　Oogiri-Corpus は，大喜利総合サイト [13] から収集されたお題・回答データであり，908 件のお題と 82,536 件の回答からなる．

構築手順および統計量　全回答について，元サイトからユーザ ID を含む投票データを収集し，分析の信頼性のためフィルタリングした．具体的には，100 票以上の投票実績を持つアクティブユーザを抽出し，その投票のみを残して得票数を再計算したうえで，得票数 3 未満の回答を除外した．その結果，908 件のお題，14,389 件の回答，57,751 票，276 人のユーザからなるデータセットを得た．これにより各投票にユーザ ID が付随し，投票履歴を追跡できる．

4　分析手法

図 1 に分析手順を示す．選好の異質性を考慮するために投票履歴に基づくユーザクラスタリングを行い（§4.1），予め定義した解釈可能な選好要因（§4.2）について BTL モデルにより重み推定する（§4.3）．

4.1　ユーザ表現とクラスタリング

各ユーザ u を投票履歴ベクトル $\mathbf{x}_u \in \mathbb{R}^N$ で表す．N は回答数であり，$\mathbf{x}_u[i]$ は回答 i への投票数（未投票は 0）とする．多くのユーザに選ばれる回答の影響を抑えるため TF-IDF で再重み付けし $\tilde{\mathbf{x}}_u$ を得る．次に，Truncated SVD により 100 次元へ次元削減して $\mathbf{z}_u = \mathrm{SVD}_{100}(\tilde{\mathbf{x}}_u)$ を得て，$\mathbf{y}_u = \mathbf{z}_u / \|\mathbf{z}_u\|_2$ で正規化する．最後に K-means で $\{\mathbf{y}_u\}$ をクラスタリングする．

4.2　ユーモア選好要因

ユーモア選好に影響する要因として，言語的特徴量とユーモア戦略ラベルの 2 種類を定義した（表 1）．言語的特徴量はお題・回答から直接抽出できる 45 個の特徴である．ユーモア戦略ラベルは，言語的特徴だけでは捉えにくいユーモアの型を表す 11 個のラベルある．要因設計には先行研究 [2] とユーモア理論 [14] を参照した．詳細定義は付録 A に示す．

言語的特徴量　言語的特徴量は，長さや品詞比率などの基本的な特徴から，お題と回答の関係性に基づく特徴の 45 種類を含む．これらは表層・形態素・特殊記号・文末・文体・関係性の 6 群に分類される．

ユーモア戦略ラベル　言語的特徴量だけでは表現しにくいユーモアの型を捉えるため，ユーモア理論に基づくユーモア戦略ラベルを付与する．例えば `incongruity` は不一致理論 [15] に対応する意外性を表し，`black_joke_satire` は無害な逸脱理論 [16] に関連するブラックユーモアや社会風刺を表す．全回答に対するアノテーションは，GPT-5.1 を用いた．各ラベルについて定義と例を含むプロンプトを設計

し，自己一貫性プロトコル [17] に従い各回答について 3 回実行し，多数決で最終ラベルを決定した．

4.3 選好モデル化

要因の重み推定には DecipherPref [18] を基礎として BTL モデル [7] を用いる．DecipherPref は，解釈可能な要因集合に基づき，選好（比較）データから要因ごとの重みを推定して選好を要因レベルで説明する枠組みである．データ $\mathcal{D} = \{(p_i, \mathbf{r}_i, \mathbf{v}_i)\}_{i=1}^{M}$ において，p_i はお題，$\mathbf{r}_i = (r_i^1, \ldots, r_i^{n_i})$ は回答列，$\mathbf{v}_i = (v_i^1, \ldots, v_i^{n_i})$ は各回答の得票数とする．解釈可能な要因集合 $\mathbb{F}$ を定義し，各回答 r に含まれる要因集合を $f(r) \subseteq \mathbb{F}$ とする．なお，連続値の特徴量は，先行研究 [18] に従い四分位により離散化し扱う[1]．

得票数からペア比較を導出する．各お題 p_i について回答ペア (r_i^j, r_i^k) を比較し，$v_i^j > v_i^k$ なら r_i^j を勝者，r_i^k を敗者とし，同点は除外する．勝者固有の要因 $F_i^+ = f(r_i^j) \setminus f(r_i^k)$，敗者固有の要因 $F_i^- = f(r_i^k) \setminus f(r_i^j)$ を定義し，共通要因は情報を持たないため除外する．各 $f \in F_i^+$ が各 $g \in F_i^-$ に「勝った」とみなし，要因レベルの勝敗を得る．

BTL モデルでは各要因 $k \in \mathbb{F}$ にパラメータ $\theta_k \in \mathbb{R}$ を割り当て，k が ℓ に勝つ確率を

$$P(k \succ \ell) = \frac{\exp(\theta_k)}{\exp(\theta_k) + \exp(\theta_\ell)}$$

で表す．推定には Luce spectral ranking（LSR）[19] を用いた．推定された $\hat{\theta}_k$ が高い要因ほど選好に寄与し，低い要因ほど非選好と関連する．

5 LLM 選好データの収集

LLM の選好分析のために選好データを収集する．

タスク定義 各お題 p に対し，回答集合 $\mathcal{R}(p) = \{r_{p,1}, \ldots, r_{p,K_p}\}$ を LLM に提示し，最も面白い回答を 1 つ選択させる．選択結果を $y_p \in \{1, \ldots, K_p\}$ として記録し，$\hat{r}_p = r_{p,y_p}$ を得る．これにより，LLM 選好データ $\{(p, \mathcal{R}(p), y_p)\}_{p \in \mathcal{P}}$ を構築する．

評価データ 分析データセットから，回答数が 5 未満のお題を除外し，897 件のお題と 14,352 件の回答（1 お題あたり平均約 16 回答）を評価対象とした．

モデル Gemini 3 Pro，GPT-5.1，Claude Sonnet 4.5 の 3 モデルを対象とした．

ペルソナプロンプト プロンプトの違いが LLM の応答特性に影響を与えることが知られている [20]．本研究ではペルソナプロンプト [21] を用い，LLM の選好が変化し得るか，また特定クラスタの選好へ近づき得るかを検証する．具体的には，`{male, female}_20`，`{male, female}_45`，`{male, female}_65`，およびペルソナを与えない `no_persona` を定義し，プロンプトに各ペルソナ固有の属性（性別・年齢層）や背景情報（趣味や肩書）を与えた．例えば，`male_20` 設定では「あなたは 20 歳男子大学生です。SNS をよく見ています。...」と指示した．

1） 例えば，連続値特徴量の一つである文字数は四分位に基づき 4 つの要因 `len-char-{short, medium, long, xlong}`に変換し，それぞれの重みを推定する．（付録 A を参照）

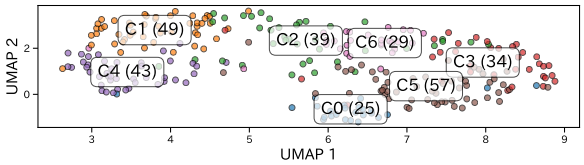

図 2 投票履歴に基づくユーザクラスタの UMAP 可視化．C0〜C6 は各クラスタであり，括弧内に人数を示す．

6 ユーモア選好要因の分析

BTL モデルにより，各ユーザクラスタおよび LLM の選好要因を推定した．投票履歴に基づくクラスタリングでは，クラスタ数 K は 7 とした[2]．図 2 は UMAP[24] により 2 次元可視化したものであり，低いシルエットスコア（$s = 0.025$）に対応して境界は明瞭ではないが，クラスタの偏りが観察できる．図 3 に各クラスタおよび LLM の BTL スコアを示す．以降，各研究課題に対する分析結果を報告する．

6.1 RQ1：各クラスタの選好要因の異質性

選好の多様性と共通点 各クラスタは異なる要因を好む．例えば C0 は会話文を含む回答（`dialogue`）や語数が多い回答（`len-token-xlong`）といった長めの回答を好む一方で，C5 は長い回答（`len-char-xlong`）を嫌う傾向がある．`slang` や自虐（`self_reference`）等はクラスタ間で好みが大きく異なる．一方，各クラスタの共通点も存在する．言葉遊び（`wordplay`）や適切な長さ（`len-char-medium`）など，多くのクラスタで正のスコアを示す．誇張表現を含む回答（`exaggeration-rule`）のように，多くのクラスタで負のスコアを示す要因も観察された．

クラスタ間の相関分析 クラスタ間の差異を定量化するため，各クラスタの BTL スコア間のピアソン相関を計算した．特定クラスタ対の間に弱〜中程度の正・負の相関が観察された（例：C0 と C3 は 0.41，C0 と C6 は-0.39）．これらの結果はクラスタ間で選

2） K はエルボー法 [22] とシルエットスコア [23] により決定．

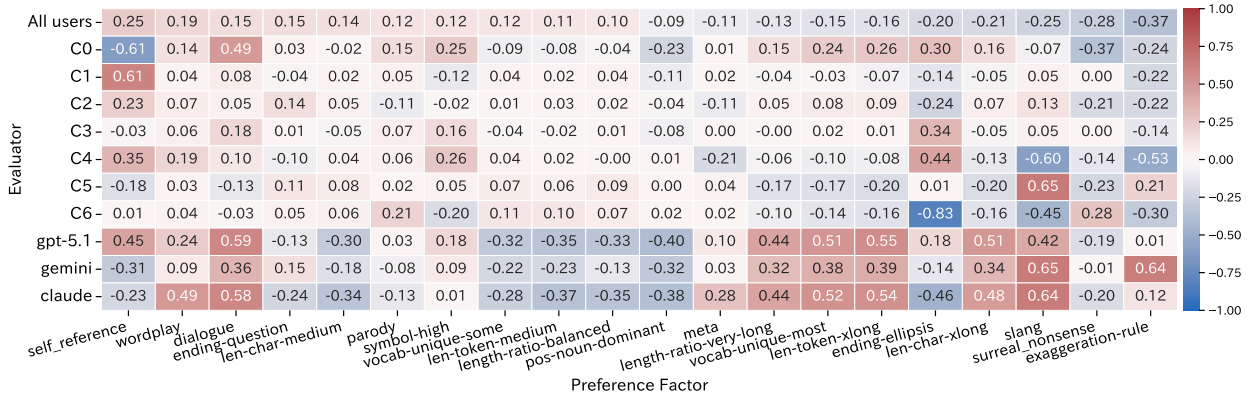

図 3　各ユーザクラスタおよび LLM のユーモア選好要因の BTL スコア．C0〜C6 は各ユーザクラスタを表す．All users はクラスタリングを行わず全ユーザの投票データで推定したスコアである．各 LLM は no_persona 設定で収集した選好データから推定したスコアである．紙面の都合上，“All users” の BTL スコアに基づく上位・下位 10 要因のみを提示する．

好が異なることを支持し，ユーモア選好の異質性を示すものである．相関行列の詳細は付録 B に示す．

6.2　RQ2：LLM とユーザクラスタの整合

ユーザクラスタとの差異・共通点　LLM はクラスタに比べ，過度に長い回答（len-char-xlong），語彙多様性が高い回答（vocab-unique-most），スラング（slang）に高いスコアを示す傾向がある．一方で，wordplay や dialogue は LLM・ユーザの双方で好まれやすく，高い名詞比率（pos-noun-dominant）やナンセンス・不条理なユーモア（surreal_nonsense）は双方で好まれにくいなど，共通点もある．また，LLM の選好が特定クラスタに近い場合も観察された（例：C5 はスラングに高いスコア，C0 は len-token-xlong に相対的に高いスコア）．

ユーザクラスタとの相関分析　LLM と各クラスタの BTL スコア間の相関を計算すると，特定クラスタ（例：C0）と LLM の間に中程度の正の相関（0.52 から 0.57）が得られた．一方，クラスタリング無しで推定した All users と LLM の間には弱い負の相関（-0.36 から-0.22）が観察された．この結果は，LLM が「ユーザ全体平均」とは選好が異なるという先行研究 [3] を支持するとともに，特定クラスタに近い選好を持つ可能性を示す新たな知見を提供する．

6.3　RQ3：特定クラスタへの選好誘導

各ペルソナ設定 LLM とクラスタの BTL スコアの相関を計算した（図 4）．その結果，ペルソナにより特定クラスタとの相関が増大する場合があることが分かった（例: female_45 を与えた場合，C0 との相関は 0.39 から 0.63 に増大)[3]．この結果は，ペルソナプロンプトが LLM 選好を特定クラスタへ誘導することを示すものであり，LLM によるユーモア理解・生成の個別最適化への可能性を提示する．

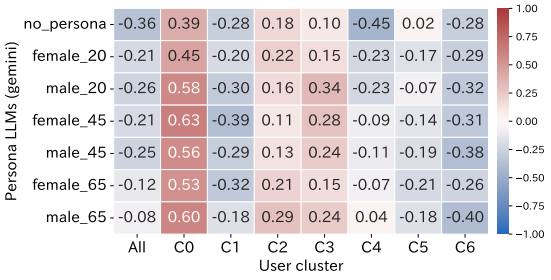

図 4　Gemini 3 Pro とクラスタ間のピアソン相関行列．

7　おわりに

日本語大喜利を対象に，ユーザクラスタと LLM のユーモア選好を，要因の BTL スコアを通して分析した．その結果，各クラスタの選好は異質であること，LLM 選好はユーザ全体平均とは異なるが特定クラスタと類似すること，ペルソナプロンプトにより特定クラスタへ選好を誘導できることを示した．

本研究の枠組みは，ユーモアをはじめとする主観評価タスクにおける一致率の低さを「ノイズ」ではなく「異質性」として扱い，解釈可能な要因を通じてユーザ群と LLM の関係を定量化する新たな視点を提供する．これは，LLM 評価の公平性・妥当性（例: LLM は誰の好みを代弁しているか）に関する議論を発展させ，ユーザごとに異なる価値観を前提とした生成・推薦・評価の設計へ繋がる基盤となる．

3)　なお，クラスタの真の属性や背景情報は不明であり，本結果はペルソナがそれを再現したと主張するものではない．ここでの主張は，ペルソナにより LLM 選好を特定クラスタへ誘導し得る点にある．

参考文献

[1] Tyler Loakman, William Thorne, and Chenghua Lin. Who's laughing now? an overview of computational humour generation and explanation, 2025. Preprint, arXiv:2509.21175.

[2] Soichiro Murakami, Hidetaka Kamigaito, Hiroya Takamura, and Manabu Okumura. Oogiri-master: Benchmarking humor understanding via oogiri, 2025. Preprint, arXiv:2512.21494.

[3] Ritsu Sakabe, Hwichan Kim, Tosho Hirasawa, and Mamoru Komachi. Assessing the capabilities of llms in humor:a multi-dimensional analysis of oogiri generation and evaluation, 2025. Preprint, arXiv:2511.09133.

[4] Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. Evaluation of text generation: A survey, 2021. Preprint, arXiv:2006.14799.

[5] Navoneel Chakrabarty, Srinibas Rana, Siddhartha Chowdhury, and Ronit Maitra. Rbm based joke recommendation system and joke reader segmentation. In Bhabesh Deka, Pradipta Maji, Sushmita Mitra, Dhruba Kumar Bhattacharyya, Prabin Kumar Bora, and Sankar Kumar Pal, editors, **Pattern Recognition and Machine Intelligence**, pp. 229–239. Springer International Publishing, 2019.

[6] Jifan Zhang, Lalit Jain, Yang Guo, Jiayi Chen, Kuan Lok Zhou, Siddharth Suresh, Andrew Wagenmaker, Scott Sievert, Timothy Rogers, Kevin Jamieson, Robert Mankoff, and Robert Nowak. Humor in ai: Massive scale crowdsourced preferences and benchmarks for cartoon captioning. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, **Advances in Neural Information Processing Systems**, Vol. 37, pp. 125264–125286. Curran Associates, Inc., 2024.

[7] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. **Biometrika**, Vol. 39, No. 3/4, pp. 324–345, 1952.

[8] Miriam Amin and Manuel Burghardt. A survey on approaches to computational humor generation. In Stefania DeGaetano, Anna Kazantseva, Nils Reiter, and Stan Szpakowicz, editors, **Proceedings of the 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature**, pp. 29–41. International Committee on Computational Linguistics, December 2020.

[9] Graeme Ritchie. Computational mechanisms for pun generation. In Graham Wilcock, Kristiina Jokinen, Chris Mellish, and Ehud Reiter, editors, **Proceedings of the Tenth European Workshop on Natural Language Generation (ENLG-05)**. Association for Computational Linguistics, August 2005.

[10] Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque. MemoSen: A multimodal dataset for sentiment analysis of memes. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 1542–1554. European Language Resources Association, June 2022.

[11] Jack Hessel, Ana Marasovic, Jena D. Hwang, Lillian Lee, Jeff Da, Rowan Zellers, Robert Mankoff, and Yejin Choi. Do androids laugh at electric sheep? humor "understanding" benchmarks from the new yorker caption contest. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 688–714. Association for Computational Linguistics, July 2023.

[12] Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, and Pan Zhou. Let's think outside the box: Exploring leap-of-thought in large language models with creative humor generation. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 13246–13257, June 2024.

[13] 大喜利総合サイト. ちんすこう大喜利. https://chinsukoustudy.com/, 2026. アクセス日： 2025 年 12 月 2 日.

[14] John Morreall. Philosophy of Humor. In Edward N. Zalta and Uri Nodelman, editors, **The Stanford Encyclopedia of Philosophy**. Metaphysics Research Lab, Stanford University, Fall 2024 edition, 2024.

[15] P. McDonald. **The Philosophy of Humour**. Philosophy Insights. HEB Humanities E-Books, 2013.

[16] A Peter McGraw and Caleb Warren. Benign violations: making immoral behavior funny: Making immoral behavior funny. **Psychol. Sci.**, Vol. 21, No. 8, pp. 1141–1149, August 2010.

[17] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In **The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023**. OpenReview.net, 2023.

[18] Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Hassan Foroosh, and Fei Liu. DecipherPref: Analyzing influential factors in human preference judgments via GPT-4. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 8344–8357. Association for Computational Linguistics, December 2023.

[19] Lucas Maystre and Matthias Grossglauser. Fast and accurate inference of plackett-luce models. In **Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1**, NIPS'15, pp. 172–180. MIT Press, 2015.

[20] Jia He, Mukund Rungta, David Koleczek, Arshdeep Sekhon, Franklin X Wang, and Sadid Hasan. Does prompt formatting have any impact on llm performance?, 2024. Preprint, arXiv:2411.10541.

[21] Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. Two tales of persona in LLMs: A survey of role-playing and personalization. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Findings of the Association for Computational Linguistics: EMNLP 2024**, pp. 16612–16631. Association for Computational Linguistics, November 2024.

[22] Robert L Thorndike. Who belongs in the family? **Psychometrika**, Vol. 18, No. 4, pp. 267–276, 1953.

[23] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, Vol. 20, pp. 53–65, 1987.

[24] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018. Preprint, arXiv:1802.03426.

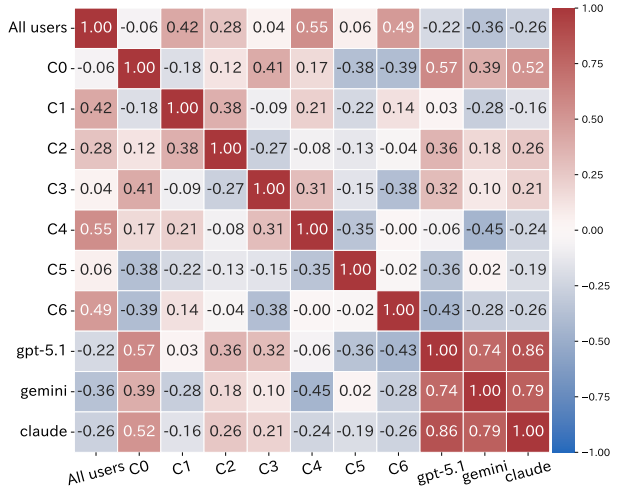

図 5　ユーザクラスタと LLM 間の BTL スコアのピアソン相関行列．行・列の C0〜C6 は各ユーザクラスタを表す．

A　ユーモア選好要因

表 2 および表 3 に言語的特徴量とユーモア戦略ラベルの定義を示す．言語的特徴量は，表層・形態素・文末・特殊記号・文体・関係性の 6 群に分類される．前者 5 群は回答から抽出する特徴量であり，関係性群はお題と回答の関係に基づく特徴量である．

BTL モデルで要因の重みを推定するため，本研究では全要因を離散値として扱う．そこで DecipherPref [18] に従い，連続値の特徴量はデータ全体の分布に基づき離散化した．基本的には四分位により 4 水準（例：`len-char-{short, medium, long, xlong}`）へ分割し，重複が多く四分位分割が不安定な特徴量は中央値により 2 水準（例：`pos-verb-{minimal, high}`）へ分割した．

B　BTL スコアの相関行列

図 5 に，各ユーザクラスタ間および LLM との BTL スコアのピアソン相関行列を示す．各 LLM はペルソナ情報を与えない `no_persona` 設定である．クラスタ間には正・負の相関が混在し，選好が異なることが示唆される（例：C0 と C3 は 0.41，C0 と C6 は-0.39）．LLM は特定クラスタ（例：C0）と中程度の正の相関（0.52 から 0.57）を示す一方で，All users とは弱い負の相関（-0.36 から-0.22）を示す．

表 2　言語的特徴量の定義．

特徴名	定義
表層特徴（11 種）	
len-char-{short, medium, long, xlong}	回答の文字数
{hiragana, kanji}-{minimal, low, medium, high}	ひらがな, 漢字の比率
{katakana, alphabet, digit, punct, space, symbol}-{minimal, high}	カタカナ, アルファベット, 数字, 句読点, 空白, 記号の比率
punct-count-{few, most}	句読点の数
sentences-{one, many}	文数
形態素解析特徴（10 種）	
pos-noun-{low, medium, high, dominant}	名詞比率
pos-particle-{minimal, low, medium, high}	助詞比率
pos-{verb, adj, adverb, auxiliary}-{minimal, high}	動詞, 形容詞, 副詞, 助動詞比率
len-token-{short, medium, long, xlong}	語数
vocab-unique-{few, some, many, most}	ユニーク語数
vocab-diversity-{repetitive, very-diverse}	語彙多様性（ユニーク語率）
proper-noun	固有名詞を含む
特殊記号（5 種）	
dialogue, parentheses, tilde, number, slang	「」, 括弧, ~, 数字, スラング（ww）を含む
文末パターン（9 種）	
ending-{period, question, exclamation, ellipsis}	特定の記号で終わる (句点, 疑問符, 感嘆符, 省略記号)
ending-{noun, verb, adjective, particle ,auxiliary}	特定の品詞で終わる（名詞, 動詞, 形容詞, 助詞, 助動詞）
文体（4 種）	
polite-style	丁寧体（です・ます調）
casual-style	常体（だ調）
exaggeration-rule, negation-rule	誇張表現, 否定表現を含む
お題・回答関係性特徴（6 種）	
prompt-overlap-{minimal, low, medium, high}	お題と回答の文字重複率
prompt-kanji-share-{minimal, high}	お題と回答の漢字共有率
prompt-{noun, verb, proper-noun}	お題に含まれる名詞, 動詞, 固有名詞を含む
length-ratio-{short, balanced, long, very-long}	回答長/お題長の比

表 3　ユーモア戦略ラベルの定義

Label	Description
`incongruity`	想定される期待を裏切る予想外のオチ
`wordplay`	言葉遊び（駄洒落，音・綴りのひねり等）
`exaggeration`	量・感情・規模などを過剰または過小に誇張
`black_joke_satire`	ブラックジョーク／風刺
`parody`	既存作品の設定や物語構造の利用
`shared_experience`	多くの人が共有する日常経験に基づく共感
`surreal_nonsense`	論理の整合性を崩すナンセンス／不条理
`personification`	非人間的対象の擬人化を活用したユーモア
`meta`	メタ的ユーモア（大喜利の枠組みへの言及）
`mini_story`	短い物語として構成されたユーモア
`self_reference`	自分自身に言及する自虐的なユーモア