

多様な客観的解釈を反映した主体性コーパス構築と予備的分析

林純子¹ 伊藤和浩¹ 永井宥之¹ 矢田竣太郎^{1,2} 若宮翔子¹ 荒牧英治¹
¹ 奈良先端科学技術大学院大学 ² 筑波大学
 {hayashi.junko.hh5,ito.kazuhiro.ih4,wakamiya,aramaki}@is.naist.jp
 hiro.nagai@naist.ac.jp yada@slis.tsukuba.ac.jp

概要

主体性とは「行為を引き起こしているのは自分だ」という行為主体としての自己の感覚を指す。主体性は well-being やうつ病などと密接に関連しており、文章を通じて主体性をモニタリングすることは非常に重要である。しかし、ある行為の記述に対してどの程度の主体性を認めるかという判定は人によって異なる。本研究ではアノテーションのばらつきを考慮したコーパスを構築し、主体性の多寡を推定するタスクと解釈のばらつきを推定するタスクを実施し予備的分析を行なった。前者のタスクでは高い精度を示したものの、後者のタスクでは課題が残った。主体性という心理学的概念に対して自然言語処理によるアプローチの可能性が示唆された。

1 はじめに

主体性とは、「行為を引き起こしているのは自分だ」という行為主体としての主観的な自己の感覚 [1] を指す。例えば、「私は挑戦した」という文章からは、書き手がその行為を自ら進んで行おうとしている感覚が伝わる。一方で「私は自転車でこけた」という文章は、書き手がその行為は偶発的なものであり、自ら起こしたわけではないと考えている可能性が高い。主体性はさまざまな心理的状态に作用することが知られている。例えば、主体性が高いほど well-being が高く [2]、人生に対する満足度が高く [3] SNS の使用においても幸福度が高い [4] ことが明らかにされてきた。そのような背景の中、大規模データを対象とした分析やモニタリングの必要性が高まっている。実際に、SNS のデータ [5, 6] や映画脚本 [7] を対象に主体性の研究が行われてきた。

文章における主体性は二つに分けられる。書き手が出来事に対して感じている主体性、すなわち主観的な主体性（以下、**主観主体性**）は書き手本人にしかわからない情報である。そのため、テキストに書

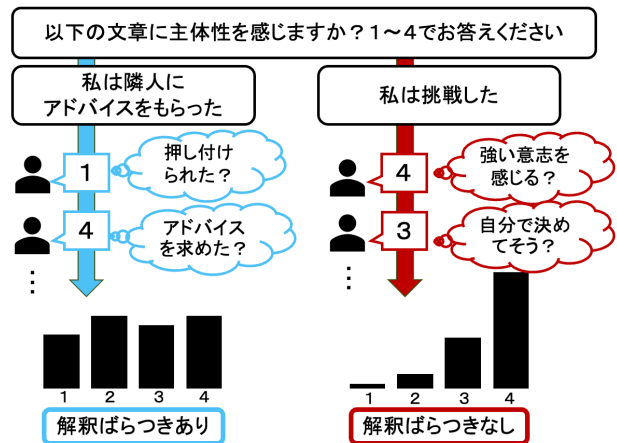


図1 本研究で作成するコーパスの例。横軸は読み手が想像する書き手の主体性（客観主体性）を定量化する「主体性スコア（1～4）」を表している。各グラフの高さは頻度を表しており、左図はスコアがばらついている反面、右図は一定の傾向が見られる。

かれた主体性を推定するためには、読み手が書き手の主体性を想像するという、客観的な主体性（以下、**客観主体性**）が用いられる。先行研究 [6, 7] においても、客観主体性をアノテーションしている。

しかし、主体性の判断は、人の経験や文化差により大きく左右される [8] にも関わらず、先行研究においては客観主体性の多様性については触れられてこなかった。このことは、客観主体性に内在する多様な解釈の一面のみを過剰に汎化してしまう可能性を意味する。例えば、「隣人にアドバイスをもらおう」という行為に関して、ある読み手が「隣人にアドバイスを押し付けられた」と考えて主体性の低い行為と判断する一方で、また別の読み手が「隣人にアドバイスを求めた」と考えて、主体性が高い行為と解釈する可能性は十分にある。これまで自然言語処理においてアノテータ間の不一致（ばらつき）は、コーパスの質を下げるものという扱いであった [9, 10] が、本来ばらつきがある客観主体性に対して上述のような方法を取ることは適切ではない。むしろ

る、多くの人に客観主体性を問い、ばらつきをありのまま取り入れたコーパスを構築する必要がある。多様な解釈を包含したコーパスを構築することにより、一見して捉え難い主体性という概念に自然言語処理の手法を用いてアプローチすることが可能になると考えた。

本研究では、これらを踏まえ、以下の二つの視点を重視しコーパスを構築した。

- 主観主体性と客観主体性の違い
- 客観主体性の解釈のばらつき

図 1 にコーパスの概要を示す。コーパス構築のため、主体性の強度を主体性スコア (1~4) として定量化した。コーパスの評価として言語モデルを用いて学習を行い、タスクがどの程度解けるのかを調査した。具体的には、客観主体性に関する以下のタスクを実施した。

1. 客観主体性のスコアの平均値を推定可能か？
2. 客観主体性スコアのばらつきを判定可能か？

結果として、スコアの平均値は高精度に推定できた一方、スコアのばらつきの判定では中程度の性能にとどまった。

2 関連研究

文法的特徴に着目して主体性を測定した研究 [5] では受動態が多いほど主体性が少ないとみなして、うつ病患者の書く文章は主体性が低くなる傾向があることを示した。受動態という文法的特徴は、英語であれば利用可能であるが、日本語にはそのまま用いることが難しい¹⁾。

次に客観主体性をラベル付けした研究を二つ挙げる。一つ目は、「走る」「食べる」のような動詞そのものに対して客観主体性を高、中、低の3段階でラベル付けした研究が存在する [7]。この研究においては、アノテータ間一致率の低さについても言及しており、客観主体性の一致が困難であることを示唆している。二つ目に、主体性の所在をラベル付けした研究 [6] がある。主体性の所在とは個人が出来事の結果を自己の努力や能力 (内的統制) に帰属させるか、環境や運といった外部要因 (外的統制) に帰属させるかの傾向を示すものである。[6] は SNS

1) 日本語で受動文に相当する形式は助動詞「(ら)れる」であるが、「(ら)れる」は受け身以外に自発・可能・尊敬の解釈をもつ。例えば、単純に「(ら)れる」を数え上げるだけでは受け身の「私が部長にほめられた」と同じく自発の「昔が思い出される」も受け身とされてしまう。

上の各文章に対して、内部統制もしくは外部統制どちらかのラベル付けをしている。客観主体性が人によって違うということを考慮してはいない。

アノテーションの解釈の多様性についての研究はいくつか存在する [11, 12, 13]。これらの研究は自然言語理解タスクコーパスを対象にして人々の解釈や意見の多様性を調査することに加え、解釈のばらつきの重要性を指摘しており、本研究の着眼点の重要性を補足するものとなっている。

3 データセット構築

データ収集は次の手順で実施した。まず行動について書かれた文章を収集し、その文章に対する客観主体性をラベル付けした。幅広い年代や性別の人々からデータを収集するため、Yahoo!クラウドソーシングを用いて参加者を募った。

初めに、ラベル付けに用いるための行動文章 (「私は~した」) を 500 名から収集した。男女比は 67:33 であり、年齢は 40 代~50 代が半数以上を占めた。質問は以下の通りとした。

- 主体性を持っておこなった行動について「私」を主語にして3つ記述してください。
- やや主体性を持っておこなった行動について「私」を主語にして3つ記述してください。
- やや主体性を持たずにおこなった行動について「私」を主語にして3つ記述してください。
- 主体性を持たずにおこなった行動について「私」を主語にして3つ記述してください。

得られる行動のバリエーションを増やすため、一人あたり 12 文ずつ、全 6000 文を収集した。

次に、新たに 500 名の参加者を募集し、各文章に対し主体性のスコアをラベル付けした。具体的には、「私は 0 時前に寝る」といった文章に対して、1 点 (主体性がない) から 4 点 (主体性がある) でスコアをつけてもらった。500 名のうち、有効な回答者数は 485 名であった。「以下の文章にどれくらい『主体性』が感じられるかを 1~4 のスケールで評価してください。数字が小さいほど文章から主体性が感じられず、数字が大きいほど文章から主体性が強く感じられることを意味します。」という指示を与え、スコアをつけてもらった。本研究においては、おおよその結果を調査するため 6000 文中 500 文をサンプルした。そして、後述するタスクのため、それぞれの文章に付随した 485 名分の客観主体性スコ

表1 主体性データセットの例. 平均値, 分散値, ばらつきの有無は全て主体性スコアについてである.

文章	客観主体性平均値	客観主体性分散値	ばらつきの有無
私はお友達とおしゃべりした	2.92	0.84	1
私はたまにはと思いゴミ捨て場の掃除をしてみた	3.21	0.63	0
私はついでに夕食の買い物もしてきた	3.13	0.61	0

アの平均値と、分散値をそれぞれ算出した。表 1 に収集したテキストの例および客観主体性のスコアを示す。今後データは拡張し公開予定である。

なお、本実験は、奈良先端科学技術大学院大学の倫理審査委員会にて承認を受けたものである(承認番号: 2024-I-36)。

4 実験

コーパス評価のため、以下のタスクを実施した。

- 主体性スコアの平均値推定
- 主体性スコアの分散判定

推定モデルのベースには、Bidirectional Encoder Representations from Transformers (BERT) [14] を採用した。日本語 Wikipedia 等で事前学習されたモデル²⁾を用いる。モデルは、文章を入力とし、数値を出力する。なお、学習率は 0.000002, エポック数は 15 とし、オプティマイザーには AdamW[15] を用いた。実験においては 5 分割交差検証を行い、平均値をモデルの性能として評価に使用する。

主体性スコアの平均値推定

- 入力：文章 (私は～した)
- 出力：主体性スコア推定値 (1~4 の連続値)

本タスクでは言語モデルは主体性の多寡を人間と同じ感覚で文章から判断できるのかを調査した。評価指標としてピアソンの相関係数, MSE を用いた。

主体性スコアのばらつき判定

- 入力：文章 (私は～した)
- 出力：主体性スコアのばらつきの有無 (0 か 1)

本タスクにおいて、言語モデルは人々の客観主体性の解釈のばらつきを文章から判定できるのか調査する。構築したコーパスに対して新たに分散値を基準に二値分類のためにラベルをつけた。分散値が平均より上だった場合、つまり多くの人がつけた客観主体性のスコアが分散する場合に 1 を付与した。そして分散値が平均より下、つまり多くの人がつけた客観主体性のスコアが一致する傾向がある場合に 0

を付与した (表 1 「ばらつきの有無」を参照)。評価指標として、正解率と F1 スコアを用いた。

5 結果と考察

主体性スコアの平均値推定

BERT を用いて平均値を推定した結果、ピアソン相関係数は 0.88 であり、有意な相関が見られた。また、MSE は 0.074 であった。主体性は文章内に表現されており、一貫した特徴から BERT が学習可能であったことが示されている。よって大規模に機械学習を用いて主体性の多寡を調査できることを示唆している。一方で、今回構築したコーパスはあくまでクラウドソーシングに答えた人々による客観主体性の評価であることを考慮する必要がある。1 章で言及した通り、客観主体性の判断は読み手の経験などに左右されるものである。よって、推定モデルとしての利用を考える際には学習データの特徴を考慮する必要がある。

エラー分析

エラー分析の結果、特定の単語や行為に対する一般的な認識がモデルの推定値に偏りを生じさせる要因である可能性が示唆された。特に、図 2 の文章においてモデルの推定値が実際の値から大きく乖離していた。「私のマンションの掃除が行われた」という文章では書き手ではなく第三者によって行われたと解釈されるべきである。しかしモデルが高い主体性を推定した要因としてデータ収集時に「掃除」に関連する文が多く含まれており、それらが主体性の高い行為としてラベル付けされていた可能性がある。そのためにモデルが「掃除」という単語そのものを主体性の高い行為と誤認識したのかもしれない。

二つ目の文章である「私は福引でポケットティッシュが当たった」に関してコーパス内に類似した内容の文章が存在しないことが確認された。このため、本研究で使用したモデルが該当文章の主体性スコアを適切に推定できなかった可能性がある。

三つ目の文章である「私が知らないうちに食事が

2) <https://github.com/cl-tohoku/bert-japanese>

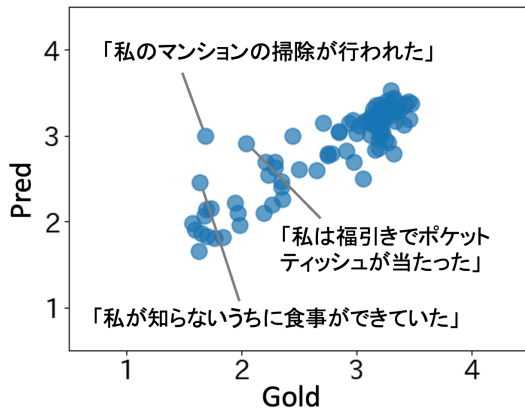


図2 客観主体性スコアの平均値を予測した散布図. Goldは各文章に付与された客観主体性平均値を示しており, PredはBERTを用いて推定した客観主体性平均値である. 注釈は, エラー分析の結果最も点数を外していた三つの文章.

できていた」も, モデルは「食事」という行為を主体性の高いものと評価する傾向を示した. この現象は, 一般に「食事」が主体性が関与する行為として認識されやすいことに起因し, 文脈を無視した形でモデルの推定値に影響を及ぼした可能性が高い.

主体性スコアのばらつき判定

BERTを用いて分散を判定した結果, 正解率は0.63であり, F1は0.63, Recallは0.65, Precisionは0.64であった. この結果の原因の一つは, 客観主体性という概念自体によるものであり, もう一つは収集したデータに起因する. まず, 2節で述べたように, 客観主体性の判断は文脈や個人の経験を反映するが, モデルはそうした情報を持たないことから, 人手による判断とは異なる. さらに, 今回収集したデータでは分散値が正規分布に近い分布を示しており, 平均値付近に値が集中していた. 本タスクは平均値に対する上下の二値分類であったが, 比較的判定が難しい平均値付近のデータを判定することが多かった. 加えて, 客観主体性の解釈のばらつきは, データを収集した母集団, すなわちアノテータによって変動する要素であり一貫していない. こうした背景から, 客観主体性スコアのばらつき判定が難しかったと考えられる.

エラー分析

表2にエラー分析の結果を示す. まず, False Positive (実際はばらつきがないのに, ばらつきがあると判定されているもの)を考察する. (1)は, 「健康維持」が書き手の意志によるものか, 医師などの

表2 主体性スコアのばらつき判定のエラー分析において False Positive(FP), False Negative (FN)と判定された例.

	判定
(1) 私は健康維持のため運動した。	FP
(2) 私が知らないうちに食事ができていた。	FP
(3) 私は家事をした。	FN
(4) 私は朝にランニングをする。	FN

指示によるものかで解釈が分かれるため, モデルの判断を完全に誤りと断定することは難しい. この文章が学習データ作成時に分散の平均値のボーダーライン上にあった可能性が考えられる. こうした曖昧な事例に対処するため, ボーダー上の文章について意見を再収集するプロセスの導入が有効である可能性がある. 一方で(2)は, 「知らないうちに」から主体性が低いと予想されるが, 動詞「できる」³⁾の解釈が曖昧であることが影響したと考えられる.

次に, False Negative (実際はばらつきがあるのに, ばらつきがないと判定されているもの)について考察する. (3), (4)は文脈や背景情報によって主体性の解釈が変動し得る文である. しかし本コーパスには類似した表現の多様な状況の文章が含まれており, それがモデルのばらつき判定に影響を与えた可能性がある⁴⁾. このような背景から, 本モデルが該当文に対して適切な主体性判定を行うことは困難であった可能性がある.

6 おわりに

本研究では, 客観主体性の解釈のばらつきに焦点を当てコーパスを構築し, 主体性の平均スコア推定と解釈のばらつき判定という2つのタスクを設定し評価を行った. 結果, 言語モデルが主体性を捉えられる可能性を示すとともに主体性評価における課題を提示した. 今後は客観主体性の解釈の分布を推定するという発展も考えられる. 本研究は主体性に対する幅広い解釈を定量化する基盤としての可能性を示しており, 心理学的な意思決定や認知の研究, さらには社会的な意見形成プロセスの分析にも寄与することが期待される.

3) 「食事ができる」は「食事が作られる」という意味と「食事することができる」という意味で曖昧である.

4) 例えば(3)については「私は妻の機嫌を伺いながら家事の手伝いをした(ばらつき無し)」「私は家族のために家事をした(ばらつきあり)」といった文章が含まれていた.(4)については「私は以前からやろうと思っていた早朝ジョギングを家族の手助けを得て始めた(ばらつき無し)」「私は毎日, 朝に2kmのウォーキングをしている(ばらつきあり)」といった文章が含まれていた.

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 の支援を受けたものである。

参考文献

- [1] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. **Trends in cognitive sciences**, Vol. 4, No. 1, pp. 14–21, 2000.
- [2] James W Moore and Paul C Fletcher. Sense of agency in health and disease: a review of cue integration approaches. **Consciousness and cognition**, Vol. 21, No. 1, pp. 59–68, 2012.
- [3] Albert Bandura. Social cognitive theory: An agentic perspective. **Annual review of psychology**, Vol. 52, No. 1, pp. 1–26, 2001.
- [4] Angela Y. Lee, Nicole B. Ellison, and Jeffrey T. Hancock. To use or be used? the role of agency in social media use and well-being. **Frontiers in Computer Science**, Vol. 5, , April 2023.
- [5] Almog Simchon, Britt Hadar, and Michael Gilead. A computational text analysis investigation of the relation between personal and linguistic agency. **Commun Psychol**, Vol. 1, No. 1, p. 23, September 2023.
- [6] Masoud Rouhizadeh, Kokil Jaidka, Laura Smith, H. Andrew Schwartz, Anneke Buffone, and Lyle Ungar. Identifying locus of control in social media language. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1146–1152, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [7] Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, and Yejin Choi. Connotation frames of power and agency in modern films. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2329–2334, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [8] Shaun Gallagher. **Ambiguity in the Sense of Agency**, p. 118–135. Oxford University Press, March 2013.
- [9] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. **Computational Linguistics**, Vol. 34, No. 4, pp. 555–596, 12 2008.
- [10] Milagro Teruel, Cristian Cardellino, Fernando Cardellino, Laura Alonso Alemany, and Serena Villata. Increasing argument annotation reproducibility by using inter-annotator agreement to improve guidelines. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).
- [11] Yixin Nie, Xiang Zhou, and Mohit Bansal. What can we learn from collective human opinions on natural language inference data? In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 9131–9143, Online, November 2020. Association for Computational Linguistics.
- [12] Ellie Pavlick and Tom Kwiatkowski. Inherent disagreements in human textual inferences. **Transactions of the Association for Computational Linguistics**, Vol. 7, pp. 677–694, 2019.
- [13] Barbara Plank. The “problem” of human label variation: On ground truth in data, modeling and evaluation. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 10671–10682, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Tamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [15] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **International Conference on Learning Representations**, 2019.

A 特徴的な文章の例

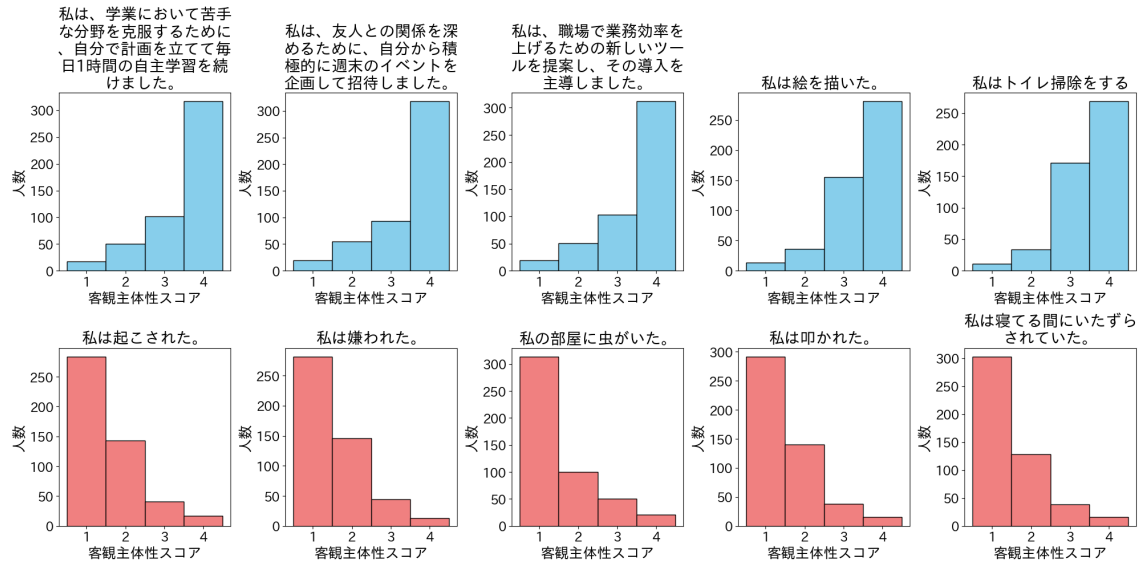


図3 客観主体性の平均値が大きい例（水色）と平均が小さい例（オレンジ）。

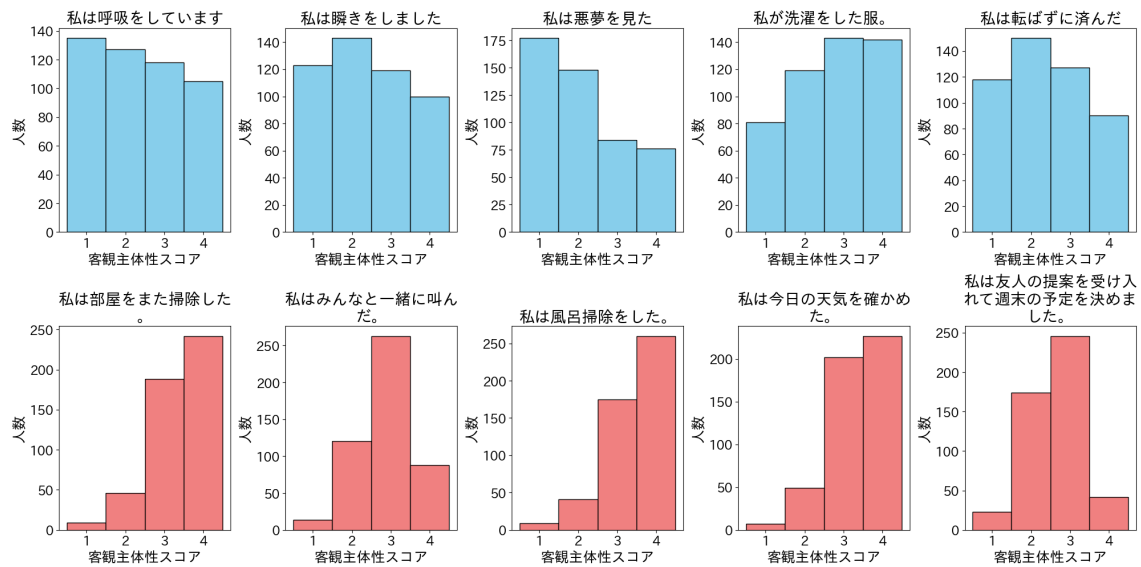


図4 客観主体性の分散値が大きい例（水色）と分散が小さい例（オレンジ）。