

Web アンケートの形式と設問文の口調が自由記述設問の回答の質に及ぼす影響：画像説明タスクを用いた検証

藤代 航希¹ 加藤 純世¹ 三山 貴也¹ 中村 聡史¹

概要：Web アンケートは短期間かつ低コストで大量のデータを収集できる一方、低品質な回答が混入しやすく、特に自由記述設問では設問で求められる内容を伴わない回答が生じやすい。我々はこれまで、アンケートの形式と設問文の口調に着目し、チャット形式のアンケートやカジュアルな口調の設問文が自由記述設問の回答に及ぼす影響を調査してきた。しかし、これまでの調査で用いたアンケートでは、回答内容が回答者の知識や興味などの個人差に左右されていた問題があり、アンケートの形式や設問文の口調による影響を十分に検証できていなかった。そこで本研究では、回答者の知識や興味による影響を抑えるため、回答者全員が同じ画像を見てその内容を説明する画像説明タスクを用いたアンケートを実施した。これにより、回答者間の個人差による影響を軽減したうえで、Web アンケートの形式および設問文の口調が自由記述回答の質に及ぼす影響を再検証した。その結果、アンケート全体では自由記述の回答トピック数に有意な差はみられなかったが、作成した画像刺激の内容の違いによっては回答の質に有意な傾向が見られた。

1. はじめに

Web アンケートは、紙媒体のアンケートと比較して短期間かつ手軽に多くのデータを集めることができるというメリットがある。また、紙媒体のアンケートよりもコストを削減でき [1]、手軽に依頼が可能である点でも有効である。さらに Denscombe ら [2] は、自由記述設問を含むアンケートを Web と紙媒体それぞれで実施したところ、Web アンケートでより記述量の多い回答を得られたと報告している。こうした点を踏まえて、Web アンケートは、大規模な基礎データの収集として学術研究や市場調査といった幅広い分野で活用されている。

しかし Web アンケートにおける自由記述設問では、回答が短く、含まれるトピックの数や詳細な記述が少ないなど、回答内容が不十分な問題がある [3]。また、回答者自身にとって回答可能な設問内容にもかかわらず、自由記述欄に「特になし」と記入したり、意味のない適当な文字列を入力したりするなど、不適切な回答も生じやすい [4]。さらに Reja ら [5] は、Web アンケートの選択式設問と自由記述設問を比較したところ、自由記述設問のほうが選択設問より欠損データが多く存在すると報告している。以上のことから、特に自由記述設問を含む Web アンケートでは、設問で求められる内容に対して十分な回答を得るための設

計が重要となる。

こうした背景から、Web アンケートの自由記述設問において回答に影響を及ぼす要因に関する研究が行われている。Yamazaki ら [6] は、自由記述設問の提示位置、テキストボックスのサイズ、PC やスマートフォンといったデバイスの違いが不真面目回答率やアンケートの途中離脱に影響を及ぼすことを報告している。また Kim ら [7] は、Web アンケートの形式（フォーム形式・チャット形式）と設問文の口調（カジュアル・フォーマル）に着目し、選択式設問ではチャット形式とカジュアルな口調の設問文を組み合わせることで回答の質やアンケートの体験が向上することを報告している。我々もこれまでの研究 [8] で Kim らと同様にアンケートの形式と設問文の口調に着目し、それらが自由記述設問の回答に及ぼす影響について調査を行ったが、回答の質に差は見られなかった。しかし、これまでの調査で用いたアンケートでは、回答内容が回答者の知識や興味などの個人差に左右されていた可能性があり、アンケートの形式や設問文の口調による影響を十分に検証できていなかった。

そこで本研究では、回答者間の個人差による影響を軽減したうえで、Web アンケートの形式および設問文の口調が自由記述回答の質に及ぼす影響を再検証する。具体的には、畑中ら [9] が提案した、回答者全員が同じ画像を見てその内容を説明する画像説明タスクを用いたアンケートを実施する。画像説明タスクは、回答者が画像から読み取っ

¹ 明治大学
Meiji University

た情報をもとに説明を行うため、回答が特定のテーマに関する知識や興味の有無に左右されにくい。そのため全ての回答者が必ず何らかの説明を行うことが可能であり、またどの範囲に注目するかやどれだけ詳しく説明するかといった点で自然な個人差が現れる。このような特徴から、画像説明タスクは、アンケートの形式および設問文の口調が自由記述回答の質に及ぼす影響を検証するうえで適していると考えられる。以上を踏まえ、両要因の影響をこれまでの研究よりも適切に評価することを目指す。

2. 関連研究

2.1 アンケート形式に関する研究

アンケート調査における形式に関して、これまで様々な研究が行われている。Celino ら [10] は、従来の Web フォーム形式とチャット形式のアンケートを比較し、チャット形式は従来形式と同等以上の回答品質を維持しながら、利用者の満足度や受容性を向上させる可能性を明らかにした。さらに、Xiao ら [11] は、従来型 Web アンケートと AI チャットボットを用いた対話型アンケートを比較した結果、AI チャットボットを用いたアンケートでは、自由記述回答の情報量や具体性、明確さが向上し、回答者の積極性も高まることが明らかになっている。

これらの研究は、チャット形式というアンケート形式が回答品質や利用者の回答体験を向上させる可能性を示している。本研究はこうしたアンケート形式が自由記述に与える影響を、個人差による影響を軽減したうえで検証するものである。

2.2 口調に関する研究

アンケート調査における口調に関しても、これまで様々な研究が行われている。Okoso ら [12] は、AI が提示する文章において、AI の役割や年齢や性別といったユーザの属性によって異なる形で、人間の意思決定に有意な影響を及ぼすことを明らかにした。また、調査協力の場面でカジュアルな言葉遣いを用いると、回答者の心理的負担を軽減する効果があることも報告されている。Kim ら [7] は、フォーム形式のアンケートおよびチャット形式のアンケートにおいて、システムが提示する文章の口調をフォーマルとカジュアルで比較した。カジュアル口調には日常会話で用いられる親しみやすい表現を採用し、その効果を検証した。その結果、カジュアル口調単独では大きな効果は見られなかったものの、チャット形式と組み合わせることで回答者の満足度の抑制や回答体験の向上につながる可能性を明らかにした。

これらの研究から、システムが提示する文章の口調はユーザの行動や回答に影響を与える可能性が示されている。本研究はこうした口調の効果がアンケート調査の自由記述

回答にどのように影響するかを検証し、さらにアンケート形式との交互作用を検証するものである。

3. 実験

3.1 実験概要

Web アンケートの形式および設問文の口調が自由記述設問の回答の質に及ぼす影響を明らかにすることを目的として、画像説明タスクを用いたアンケート調査を実施する。実験は、参加者間要因計画による 2 要因 2 水準のデザインで実施する。要因にはアンケートの形式と設問文の口調を設定し、水準としてアンケート形式にはチャット形式とフォーム形式、設問文の口調にはフォーマルとカジュアルを設定する。参加者には、これらを組み合わせた 4 条件のうちいずれか 1 条件を無作為に割り当て、各条件に対応するアンケートに回答してもらう。

本実験では、自由記述回答の質を評価する指標として、回答者が言及した意味的に独立した情報の数（回答トピック数）を用い、以下の 3 つの仮説を設定する。これは、Kim ら [7] の研究において、チャット形式の主効果とカジュアル口調との交互作用における回答品質の向上が明らかとなったためである。また、口調の主効果についても、Kim ら [7] の研究においては結果は得られなかったが、本研究においては回答者が任意の文章を回答できる自由記述設問を対象としているため、口調による影響がより大きく出る可能性がある。

- **H1**: チャット形式のアンケートの場合、フォーム形式の場合と比較して回答トピック数が有意に向上する
- **H2**: 設問文の口調がカジュアルな場合、フォーマルな場合と比較して回答トピック数が有意に向上する
- **H3**: チャット形式かつカジュアルな口調の条件では、他の条件と比較して回答トピック数が有意に向上する

3.2 実験設計

本実験で用いるアンケートは、実験用アンケートと事後アンケートの 2 種類で構成される。実験用アンケートでは 3 枚の画像を提示し、各画像に対して選択式設問 2 問と自由記述設問 1 問（画像説明タスク）の計 3 問ずつ、合計 9 問を実施した（表 1）。事後アンケートでは、参加者属性およびアンケートに対する印象を取得し、表 2 に示す選択式設問で構成した。なお、事後アンケートの形式と設問文の口調は、実験条件によらずフォーム形式かつフォーマルな口調に統一している。

3.2.1 画像説明タスクに用いる刺激の設計

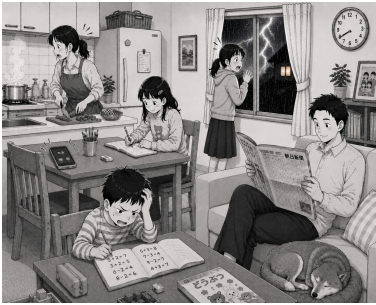
本実験では、回答者の知識や興味による影響を抑えた上で自由記述回答の質を比較するため、先述のとおり画像説明タスクを採用した。実験で用いる画像は、畑中ら [9] の画像を参考に、複数の人物や出来事が含まれ、回答者が複

表 1 実験用アンケートの設問一覧

設問番号	回答形式	フォーマルの設問文	カジュアルの設問文
1	選択式	この画像の雰囲気として最も近いものを選択してください	この画像の雰囲気にどれが一番近い気がする？
2	選択式	この画像は全体としてまとまりがあると思いますか	この画像は全体としてまとまっていると思う？
3	自由記述	この画像の中で起こっている出来事について、できるだけ多く記述してください	この画像の中にある出来事をできるだけたくさん書いてー

表 2 事後アンケートの設問一覧

設問番号	回答形式	設問文	選択肢
1	選択式	あなたの年齢が含まれるものを選択してください	10-19 / 20-29 / 30-39 / 40-49 / 50-59 / 60-69 / 70 以上
2	選択式	あなたの性別を選択してください	男性/女性/その他/回答しない
3	選択式	普段どれだけチャットのやり取りを行っていますか	ほぼ毎日やり取りしている/週に数回程度やり取りしている/月に数回程度やり取りしている/ほとんどやり取りしない
4	選択式	アンケートの文章はどのくらい親しみやすいものでしたか	とても親しみやすかった/やや親しみやすかった/どちらともいえない/あまり親しみやすくなかった/全く親しみやすくなかった
5	選択式	アンケートはどのくらい回答しやすいものでしたか	とても回答しやすかった/やや回答しやすかった/どちらともいえない/あまり回答しやすくなかった/全く回答しやすくなかった



(a) リビングの場面



(b) 公園の場面



(c) 学校の場面

図 1 生成 AI を用いて作成した画像刺激

数の要素について記述できる日常的な場面を描いたものとした。また、特定の専門知識や経験を必要とせず、画像を見るだけで誰でも回答できる内容となるよう設計した。本実験で用いた画像刺激を図 1 に示す。これらの画像は生成 AI を用いて作成した。画像間で情報量に大きな差が生じないように、各画像には 5 人の人物が登場するよう設定した。また、回答者が画像内の人物や物、出来事について幅広く記述できるよう、各場面にはその状況に応じた出来事や変化を含めるよう生成条件を設定した。さらに、提示順による影響を抑えるため、画像刺激の提示順は参加者ごとにランダム化した。

3.2.2 アンケート UI の設計

本実験では、アンケート形式を要因として扱うため、フォーム形式とチャット形式の 2 種類の UI を実装した。

フォーム形式の UI は図 2 に示すように、一般的な Web アンケートに近い構成とした。選択式設問では、設問文、画像、選択肢を上から順に表示し、選択肢をタップすると次の設問へ遷移するようにした (図 2a)。自由記述設問で

は、設問文、画像、テキストボックス、送信ボタンを上から順に表示し、回答入力後に送信ボタンを押すことで次の設問へ遷移するようにした (図 2b)。

チャット形式の UI は図 3 に示すように、LINE などのメッセージアプリに類似した構成とした。選択式設問では、画像の直下に設問文を表示し、その下に選択肢ボタンを配置することで、チャットの流れに沿って回答できるようにした (図 3a)。自由記述設問では、設問文と画像を画面左側の吹き出しとして表示し、参加者が画面下部の入力欄から回答を送信すると、その回答が画面右側の吹き出しとして追加されるようにした。また、回答は複数回に分けて送信できるようにし、十分に回答した後に「次の質問に進む」ボタンを押すことで、次の設問へ遷移するようにした (図 3b)。

なお、これらの UI は基本的に加藤ら [8] の研究を参考に設計した。ただし、加藤らの研究では、フォーム形式とチャット形式とで自由記述設問のテキストボックスのサイズが異なっていた。テキストボックスのサイズの違いが自

(a) 選択式設問 (b) 自由記述設問

図 2 フォーム形式のアンケート画面

(a) 選択式設問 (b) 自由記述設問

図 3 チャット形式のアンケート画面

由記述回答の量や内容などの回答品質に影響を与える可能性を考慮し、本実験では両形式におけるテキストボックスのサイズを統一した。

3.2.3 設問文の口調の設計

本実験では、設問文の口調を要因として扱うため、フォーマル条件とカジュアル条件の2種類の設問文を作成した。加藤ら[8]の研究では、カジュアル条件の設問文を著者が独自に作成していたが、本研究では作成者の主観による影響を抑え、より客観的にカジュアルな表現を選定するため、複数人による書き換えと投票を行った。具体的には、大学生、大学院生22名にフォーマル条件の設問文をカジュアル表現へ書き換えてもらい、著者が各設問につき5候補を選定した後、別の大学生、大学院生22名による投票で最多票を得たものを採用した。

3.3 実験参加者

実験参加者は、アカウントの信頼性と完了率が高いとされる[13]Yahoo!クラウドソーシング^{*1}(YCS)を通じて募集した。また時給1250円を基準として、実験を正規に完了した参加者に想定所要時間に相当する146円を報酬として支払った。また、スマートフォンやPCといった回答デバイスの違いによっても回答行動に差がみられる[6]ことが報告されているため、それを考慮して回答デバイスはスマートフォンに限定した。本実験では、男性100名、女性100名の計200名を募集した。

3.4 実験手順

実験参加者は、スマートフォン上のGoogle Chromeを用いて、YCSのページからアンケートシステムにアクセスする。なお、参加者がアンケート開始画面に到達した時

点で、実験条件の4条件のうち1条件がランダムに割り当てられる。開始画面ではアンケートの概要が提示され、参加者は「アンケートを開始する」ボタンを押すことで回答画面に遷移する。遷移後は、割り当てられた条件に応じて表示されるUIに従ってアンケートに回答し、9問の実験用アンケートすべてに回答することが求められる。実験用アンケートが終了すると、事後アンケートへ進み、参加者がすべての設問の回答を終えると、報酬を受け取るためのコードとユーザ固有のIDが表示され、そのコードをYCSのタスク提出欄に入力することで報酬を受け取ることができる仕組みとした。

3.5 自由記述回答の評価方法

自由記述回答に含まれる情報量を適切に評価するため、本研究では回答文字数ではなく回答トピック数に基づいて評価を行う。回答トピック数とは、画像中に含まれる出来事や状況に関して回答者が言及した意味的に独立した情報の数を表す指標である。回答トピック数を効率的かつ正確に計測するために、畑中ら[9]の研究を参考に専用のアノテーションシステムを構築した。このシステムでは、各画像刺激に対して作成した「出来事(イベント)」のリストを提示し、評価者は該当する記述がある場合にチェックボックスを選択してもらった。この数を回答トピック数として集計した。

アノテーション作業は2名の評価者によって、回答データがどの実験条件のものか分からないよう盲検化した上で独立に実施し、一致しなかった回答については、著者が最終的な回答トピック数を判断することとした。

^{*1} <https://crowdsourcing.yahoo.co.jp>

表 3 画像ごとの回答トピック数に対する評価者間一致度

画像	n	平均差	級内相関係数	完全一致率
公園	194	-0.22	0.941	70.6%
学校	194	-0.17	0.911	62.9%
リビング	193	-1.75	0.627	15.5%

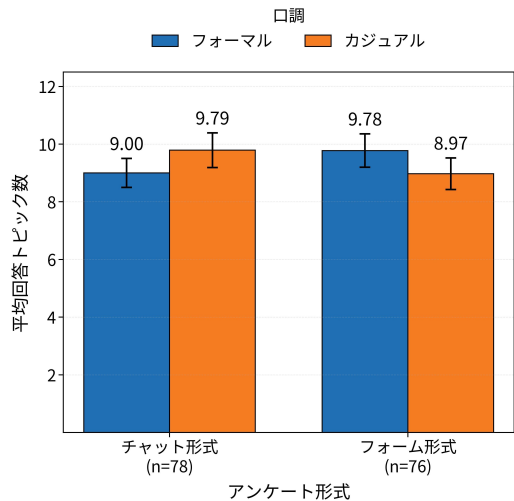


図 4 条件ごとの回答トピック数の平均合計値

4. 結果

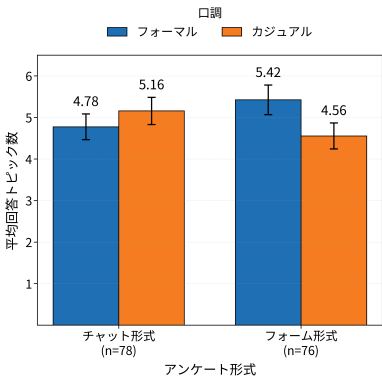
4.1 除外データ

実験に参加した男性 100 名、女性 100 名の計 200 名のうち、タスク完了時に入力された ID に不備があり、その結果システムログとの照合が不可能であった 6 名をアノテーション実施前に除外した。

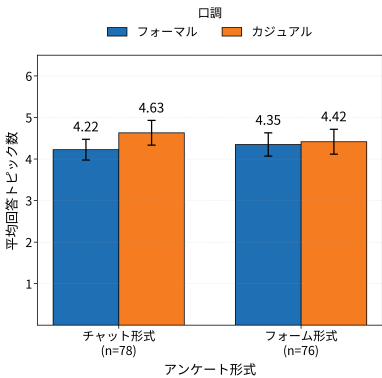
表 3 に、画像ごとの回答トピック数に対する評価者間の完全一致度と級内相関係数を示す。評価者間の完全一致度は、公園画像で 70.6 %、学校画像で 62.9 %、リビング画像で 15.5 %であり、画像によって評価の安定性に差が認められた。そこで、評価者間にブレが大きいリビング画像に関しては、回答トピック数に対する分析対象から除外した。

さらにアノテーションを実施した 194 名のうち、設問に対して不適切な回答を行った参加者、および自由記述設問 2 問のうちいずれかで回答トピック数が 0 であった参加者を分析対象から除外した。なお、不適切な回答については、設問の意図から明らかに逸脱している回答を著者らが目視で確認した。また、回答トピック数が 0 であった参加者を除外したのは、内容を含まない回答の影響を避け、2 問すべてに回答内容が含まれている参加者について、回答トピック数の合計を比較するためである。

その結果、不適切な回答を行った参加者 21 名、およびいずれかの自由記述設問で回答トピック数が 0 であった参加者 19 名を除外し、最終的な分析対象は 154 名となった。



(a) 公園の画像



(b) 学校の画像

図 5 各画像刺激における回答トピック数の平均合計値

4.2 仮説に対する結果

図 4 に、各条件における自由記述回答の合計トピック数を示す。二元配置分散分析の結果、アンケート形式の主効果および設問文の口調の主効果、アンケート形式と設問文の口調の交互作用はいずれも有意ではなかった。また、各条件の平均値を比較すると、回答トピック数は「チャット形式×カジュアル」条件で最も高く、次いで「フォーム形式×フォーマル」条件で高かった。

以上の結果から、チャット形式がフォーム形式より回答品質を向上させるという H1、カジュアル口調がフォーマル口調より回答品質を向上させるという H2、チャット形式かつカジュアルな口調が他の条件と比較して回答品質が向上するという H3 は支持されなかった。

4.3 画像ごとの分析結果

前節では、アンケート形式および設問文の口調の主効果はいずれも認められなかった。そこで、画像によって回答トピック数の傾向が異なるかを確認するため、画像ごとの分析を行った。

図 5 に、各画像刺激における回答トピック数の平均値を示す。画像ごとに回答トピック数を分析した結果、リビング、学校のいずれの画像においても、アンケート形式の主効果、設問文の口調の主効果、および両者の交互作用は有

意ではなかった。一方、公園画像では形式と口調の交互作用が有意傾向を示した。

4.4 設問順の分析結果

前節では、画像ごとの回答トピック数について分析した。一方、本実験では画像の提示順を回答者ごとにランダム化しているため、各設問番号で提示される画像は参加者間で異なる。そこで本節では、自由記述設問における設問順の影響について分析した。

図6に設問番号ごとの回答トピック数を示す。回答トピック数について分析した結果、設問順の主効果は有意ではなく、設問が進んでも回答トピック数に有意な変化は認められなかった。

そこで、画像ごとの設問順の影響を確認するため、探索的分析を行った。その結果、公園画像においてのみ設問順の主効果が有意であった。多重比較の結果、2問目と3問目の間に有意差が認められた。一方、学校画像では、設問順による有意な変化は認められなかった。

回答時間について分析した結果、設問順の主効果が有意であり、設問が1つ進むごとに回答時間は平均約28.5秒短縮した。一方、回答トピック数については設問順による有意な変化は認められなかった。

そこで、回答時間の推移がアンケート形式および設問文の口調によってどのように異なるかを分析した。図7に、各条件における自由記述設問の回答時間の推移を示す。いずれの条件においても、設問が進むにつれて回答時間は短くなる傾向がみられた。また、「チャット×フォーマル」条件および「フォーム×カジュアル」条件では、他の2条件と比較して設問1から設問3にかけての回答時間の減少量が大きい傾向がみられた。

回答時間の推移について分析した結果、設問順とアンケート形式の交互作用が有意であり ($p = .032$)、回答時間の短縮量はアンケート形式によって異なることが示された。さらに、設問順、アンケート形式、設問文の口調の三者交互作用も有意であった ($p = .049$) ことから、設問順に伴う回答時間の変化は、アンケート形式と設問文の口調の組み合わせによって異なることが示された。

5. 考察

5.1 回答トピック数について

本研究では、アンケート形式および設問文の口調が自由記述回答の品質に及ぼす影響を明らかにすることを目的とし、H1「チャット形式のアンケートの場合、フォーム形式アンケートと比較して回答トピック数が有意に向上する」、H2「設問文の口調がカジュアルな場合、フォーマルな場合と比較して回答トピック数が有意に向上する」、H3「チャット形式かつカジュアルな口調の条件では、他の条件と比較

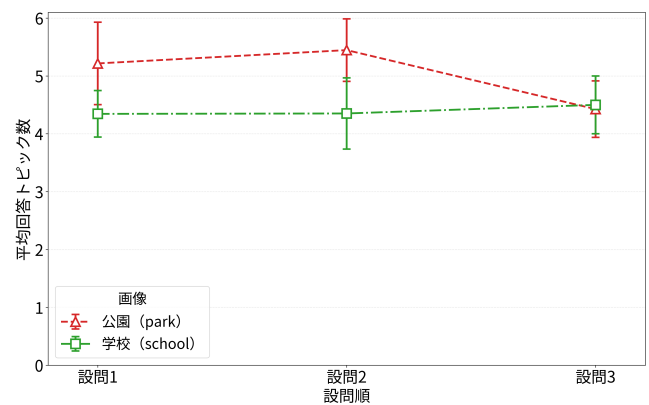


図6 設問ごとのトピック数推移

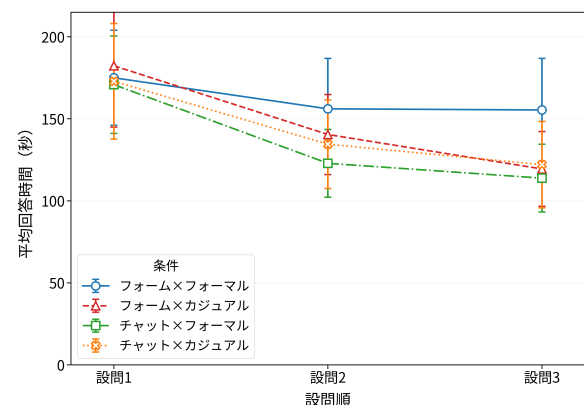


図7 条件ごと回答時間の推移

して回答トピック数が有意に向上する」という3つの仮説を検証した。しかし、リビング画像を除いた2つの画像の回答トピック数を対象とした分析では、アンケート形式および設問文の口調による有意な影響は確認されなかった。

一方、画像ごとの分析では、公園画像においてのみアンケート形式と設問文の口調の交互作用が有意水準に近い値を示した。さらに、画像ごとの設問順の分析では、公園画像においてのみ設問順の主効果が認められ、2問目と3問目の間で回答トピック数が有意に減少した。これらの結果から、公園画像は他の画像と比較して回答者間の違いが現れやすい画像であった可能性が考えられる。

公園画像では、犬、ブランコ、水をこぼす場面、セミ、鳥など、複数の出来事や対象が明確に区別できる構成となっていたため、回答内容とアノテーション項目との対応が比較的明確であった可能性がある。その結果、公園画像では回答トピック数の評価者間一致度が高く、回答内容を他の画像よりも安定して回答トピック数として評価できていた可能性がある。

また、公園画像は平均回答文字数が84.6文字、標準偏差91.6であり、リビング画像が平均77.6文字、標準偏差70.9および学校が画像が平均78.8文字、標準偏差57.7よりも平均回答文字数および標準偏差がともに大きかった。このことから、公園画像では回答者による記述量のばらつきが

大きく、回答者ごとに着目する対象や記述内容が異なっていた可能性がある。以上より、公園画像における結果には、回答者ごとの記述内容の違いが現れやすい画像刺激であったことに加え、回答トピック数を比較的安定して評価できていたことの双方が関係していた可能性がある。

5.2 実験設計について

表3に示すように、画像間における評価者の完全一致度には大きな差があり、リビング画像において評価者間の一致度が著しく低下する結果が得られた。リビング画像では、複数の人物が同一の行動をしている場面について、「男の子と女の子が勉強している」を2トピックとみなすか、「子どもが勉強している」を1トピックとみなすかなどで判断が分かれやすかった。また、リビング画像では、事前に設定したアノテーション項目に含まれない記述や、複数人物の類似した行動をまとめて表現する回答もみられた。そのため、回答内容を既定のトピックに対応付ける際の判断が曖昧になったり、評価指標が十分に適合しなかったりしたため、回答トピック数を安定して評価できなかった可能性がある。

以上を踏まえると、回答トピック数を安定して評価するためには、画像刺激及びアノテーションシステムの双方を見直す必要がある。まず、画像刺激を作成する際には、事前に設定する各言及箇所が意味的に独立した内容になるようにし、それ以外の内容が回答対象になり得ないように余分な要素を画像に含めないことが重要である。しかしその一方で、想定していない要素への言及が生じる可能性を完全に排除することは難しい。そのため、事前に設定した項目に該当しない記述についても、評価者が独立したトピックとして適切であると判断した場合には新たにカウントできるように、アノテーションシステムを改善する必要がある。

5.3 アンケートシステムの影響について

前節では、画像説明課題の性質や画像刺激の特徴が回答トピック数に及ぼす影響について考察した。一方、本研究ではアンケート形式としてチャット形式とフォーム形式を比較したが、チャット形式の主効果は認められなかった。そこで本節では、本研究で構築したアンケートシステムの設計という観点から、その要因について考察する。

本研究で構築したチャット形式アンケートは、外観上は一般的なチャットアプリケーションに近い構成であったものの、回答に対する「いいですね」や「ありがとうございます」といったフィードバックや、回答内容に応じた追加質問などの対話的な機能は実装していなかった。そのため、チャット形式によって期待される回答意欲や記述内容を促進する効果が十分に発揮されなかった可能性がある。

近年のチャットボットを用いたアンケートでは、応答に

対するフィードバックや追加質問など、回答者との対話を継続する機能が回答意欲や回答内容に影響を及ぼすことが報告されている[14]。本研究で用いたチャット形式は、主に見た目や入力方法をチャット風にしたものであり、対話性そのものは限定的であった。このことが、フォーム形式との差が小さくなった一因である可能性が考えられる。

5.4 アノテーション手法と回答トピックの限界について

本研究で用いたアノテーションシステムでは、評価可能な回答トピック数が最大9個に制限されていたため、多くの情報を含む回答であっても、9個を超える言及は回答トピック数に反映されないという問題があった。そのため、今後はトピックについて追加を可能にするとともに、事前に複数人でどのようなトピックがあるかを精査し、アノテーション評価を行っていく予定である。

また、今回用いた回答トピック数では、同一のトピックについてどの程度詳しく記述されているかを評価することはできないという問題があった。そのため、本研究では回答トピック数に加えて、1トピック当たりの記述量についても評価を行った。その結果、1トピックあたりの記述量については、2画像の合計値および各画像のいずれにおいても、フォーム形式がチャット形式より有意に高かった。この要因の一つとして、チャット形式におけるメッセージ単位の入力に影響した可能性がある。チャット形式では回答を複数のメッセージに分けて送信できるため、回答者が一つのメッセージを短い発話としてまとめ、補足的な説明を省略した可能性が考えられる。一方、フォーム形式では一つの入力欄に回答全体をまとめて記述するため、複数の情報を接続した文章や、人物・出来事に関する補足説明が記述されやすく、1トピック当たりの記述量が多くなった可能性がある。こうした指標は、今後の研究において導入および検討を行っていく予定である。

5.5 回答時間について

自由記述設問の回答時間について分析した結果、設問が進むにつれて回答時間は短縮していた。このことから、回答者は実験の進行に伴って画像説明課題やアンケートの操作に慣れ、より効率的に回答できるようになった可能性がある。

また、回答時間の変化はアンケート形式および設問文の口調の組み合わせによって異なり、「チャット×フォーマル」および「フォーム×カジュアル」では、他の条件と比較して回答時間の減少量が大きい傾向がみられた。回答時間の推移に形式と口調の組み合わせによる違いが認められたことから、回答者がアンケート形式や設問文の特徴に適応する過程は、条件によって異なっていた可能性が考えられる。特に、「チャット×フォーマル」および「フォーム×

カジュアル」は、アンケート形式と設問文の口調が一致しない条件であり、実験開始時には形式と口調の組み合わせに対して違和感や回答方法の迷いが生じ、回答時間が長くなっていた可能性がある。

一方、設問を重ねるにつれて回答者がその組み合わせに慣れたことで、回答時間がより大きく短縮したと考えられる。これに対し、「フォーム×フォーマル」のような日常的に見慣れた組み合わせでは、実験開始時から比較的自然的に回答できたため、回答時間の短縮幅が小さかった可能性がある。

以上より、本研究で観察された回答時間の短縮は、回答者がアンケート形式や設問文の特徴に徐々に適応していったことを反映している可能性が示唆された。

6. おわりに

本研究では、Web アンケートにおけるアンケート形式および設問文の口調が自由記述回答に及ぼす影響を調査した。その結果、全体の回答トピック数では、アンケート形式および設問文の口調による有意な影響は認められず、チャット形式やカジュアル口調が回答品質を向上させるという仮説は支持されなかった。一方、画像ごとの分析では、公園画像において形式と口調の交互作用が有意傾向を示した。探索的分析の結果、画像刺激の特徴や回答トピック数の評価の安定性が、この結果に影響した可能性が示唆された。また、画像刺激によって回答内容や評価のしやすさが異なることから、画像刺激の構成や情報量がアンケート形式および設問文の口調の効果に影響を及ぼす可能性が示された。さらに、回答時間の分析では、不一致条件において設問が進むにつれて回答時間がより大きく短縮する傾向が認められ、回答者がアンケート形式や設問文の組み合わせに徐々に適応していく可能性が示唆された。

今後は、チャット形式アンケートに相づちや追加質問などの対話的要素を導入し、より実際のチャットに近い環境において回答品質への影響を検証する必要がある。また、本研究では回答品質を回答トピック数、探索的分析として詳細度を用いて評価したが、画像刺激の選定やアノテーション方法、評価指標についても改善の余地がある。特に、画像刺激の内容や構成によって回答品質の評価結果が変化する可能性が示されたことから、今後は画像刺激の難易度や情報量を考慮した実験設計についても検討する必要がある。さらに、回答者の適応過程や回答行動の変化に加え、より汎用的かつ客観的な回答品質の評価手法についても検討していく予定である。

謝辞 本研究の一部は JSPS 科研費 JP26K02962 の助成を受けたものです。

参考文献

- [1] Vergnaud, A. C., Touvier, M., Méjean, C., Kesse-Guyot, E., Pollet, C., Malon, A., Castetbon, K. and Hercberg, S.: Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Santé study, *International Journal of Public Health*, Vol. 56, No. 4, pp. 407–417 (2011).
- [2] Denscombe, M.: The Length of Responses to Open-Ended Questions, *A Comparison of Online and Paper Questionnaires in Terms of a Mode Effect*, *Social Science Computer Review*, Vol. 26, No. 3, pp. 359–368 (2008).
- [3] Smyth, J. D., Dillman, D. A., Christian, L. M. and McBride, M.: Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality?, *Public Opinion Quarterly*, Vol. 73, No. 2, pp. 325–337 (2009).
- [4] Holland, J. L. and Christian, L. M.: The influence of topic interest and interactive probing on responses to open-ended questions in web surveys, *Social Science Computer Review*, Vol. 27, No. 2, pp. 196–212 (2009).
- [5] Reja, U., Manfreda, K. L., Hlebec, V. and Vehovar, V.: Open-ended vs. Close-ended Questions in Web Questionnaires, *Advances in Methodology and Statistics*, Vol. 19, No. 1, pp. 159–177 (2003).
- [6] Yamazaki, I., Hatanaka, K., Nakamura, S. and Komatsu, T.: The Effects of Order and Text Box Size of Open-ended Questions on Withdrawal Rate and the Length of Response, *Proceedings of the 35th Australian Conference on Human-Computer Interaction*, pp. 446–453 (2023).
- [7] Kim, S., Lee, J. and Gweon, G.: Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1–12 (2019).
- [8] 加藤純世, 三山貴也, 畑中健彦, 中村聡史: Web アンケートの形式と設問文の口調が自由記述設問の回答に及ぼす影響, 情報処理学会 研究報告コラボレーションとネットワークサービス (CN), Vol. 2026-CN-127, No. 6, pp. 1–8 (2026).
- [9] 畑中健彦, 中村聡史: 自由記述回答の定量評価に向けた画像説明タスクの提案: 進捗バー速度と回答欄サイズの影響比較, 情報処理学会 研究報告コラボレーションとネットワークサービス (CN), Vol. 2026-CN-127, No. 10, pp. 1–8 (2026).
- [10] Celino, I. and Re Calegari, G.: Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness, *International Journal of Human-Computer Studies*, Vol. 139, p. 102410 (2020).
- [11] Xiao, Z., Zhou, M. X., Liao, Q. V. and Mark, G.: Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions, *ACM Transactions on Computer-Human Interaction*, Vol. 27, No. 3, pp. 1–37 (2020).
- [12] Okoso, A., Yang, M. and Baba, Y.: Do Expressions Change Decisions? Exploring the Impact of AI's Explanation Tone on Decision-Making, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 1–18 (2025).
- [13] Seaborn, K. and Nakamura, S.: Quality and

- representativeness of research online with Yahoo! Crowdsourcing, *Frontiers in Psychology*, Vol. 16 (2025).
- [14] Zhao, Y., Gao, X. S. and Paas, L.: Chatbot-Administered Surveys in Marketing: A Literature Review, *International Journal of Market Research*, Vol. 68, No. 4, pp. 472–493 (2026).