

ロボット間の対話における指示語・省略が第三者に与える印象の検討

井上 南^{1,a)} 前川 知行¹ 松森 匠哉¹ 柴田 遼一¹ 吉野 哲平¹ 今井 倫太¹

概要：コミュニケーションロボットがユーザと親密な関係を築くことは重要である。人同士の対話において親密さを表出する方法として、共通に理解されている事物に対する代名詞の使用や単語の省略が挙げられる。しかし、ロボットの発話生成においてユーザとの親密さを表出するために適切に代名詞の使用や単語の省略を行う体系的な手法は確立されていない。本論文では大規模言語モデルである BERT を用いて対話文中の単語の省略および代名詞化を行う手法を提案する。さらに、提案手法で生成した発話をロボットの音声対話に適用することが観察者に与える印象を調査する。実験では、2体の人型ロボットが対話する動画について、対話文中の単語の省略および代名詞化の度合いが異なる複数の動画を用意し、観察者が主観的に感じるロボット間の関係性を測定した。結果として、省略・代名詞化によって対話内容の理解が困難になっても、観察者が知覚するロボット間の親しみは有意に下がらないことが確認された。

キーワード：省略、発話生成、親密さ、自然言語処理、印象評価、ロボット

Impressions given to Third Parties by Demonstrative Pronouns and Ellipsis in Interrobot Conversations

1. はじめに

人間と関係性を構築するコミュニケーションロボットの実現には、相手との共有体験を加味して個人適応した発話生成を行うことが重要である [1]。本研究では対話における指示語・省略に着目して適応的な発話を行う手法の検討を行う。

人間同士の対話では、当事者間で共通に理解されている情報を省略したり、「あれ」等の代名詞に置き換えたりするなど非明示的な形で伝達することがある [2], [3]。このような省略や代名詞化は、当事者間の関係性を表した重要な情報であるとして着目されてきた。一方で、省略・代名詞化を行うことで関係性が近くなるという効果も考えられる。対話における代名詞化の効果に関して、山崎 [4] は、代名詞の使用は「話手と聞き手との間にしばしば親近感をもたらす」と指摘している。

このような共通認識による省略や代名詞化をロボットの発話に取り入れることは、人間とロボットの関係性の構築及び持続を促進する可能性がある。

エージェントの発話文における省略や代名詞化に関連する先行研究として、ロボットの言葉足らずな発話が人間の関与にどのような変化をもたらすかを調べた研究 [5]、スマートスピーカーがユーザの予定を伝える際に代名詞化・省略された発話を行った場合の印象を調査する研究 [6] などがあある。先行研究では、共通に理解されている情報が何であるかを既知として、構文解析によって省略あるいは代名詞化した発話を生成している。

しかし、共通に理解されている情報が何であるかを対話の文脈から推測し、その上で任意の構造の発話文に対して省略・代名詞化した発話を生成する手法は検討されていない。

そこで本論文ではロボットが人間との関係性や共有体験に応じて親密感を表出するために、対話の文脈を用いて省略・代名詞化した発話を生成する手法を提案する。本研究の概略を図 1 に示す。具体的には、ロボットの発話における省略・代名詞化の度合いを調節するパラメータを定義し、

¹ 慶應義塾大学理工学部 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1
Faculty of Science and Technology, Keio University, 3-14-1
Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa 223-8522, Japan

^{a)} inoue@ailab.ics.keio.ac.jp

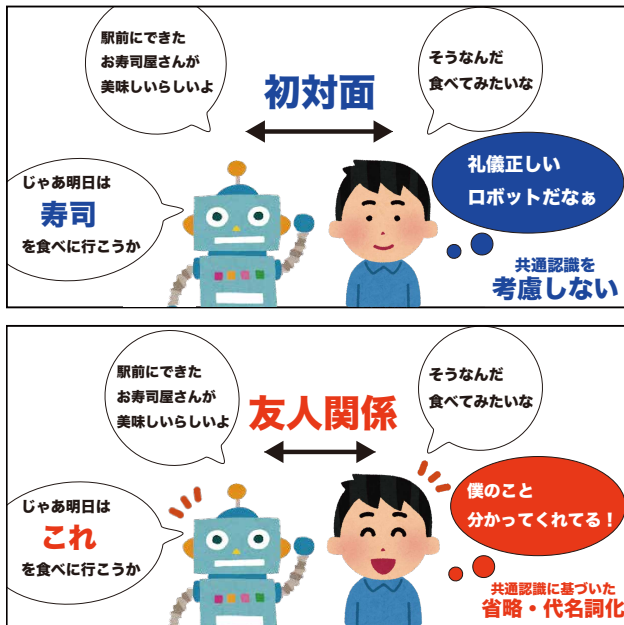


図 1 本研究の概略。初対面の関係(上)では既知の「寿司」という単語を再度ロボットが使用しているのに対し、友人関係(下)では「寿司」の代わりに「これ」という指示代名詞を用いている。

大規模言語モデル BERT [7] のマスク予測手法を用いて省略・代名詞化の対象となる単語を決定する。また、提案手法で生成した発話をロボット同士の音声対話に適用し、観察者に与える印象を調査する。

本論文の新規性は以下の 2 点である。第 1 に、BERT のマスク予測手法をロボットの発話における親密感の調節に初めて適用した点である。対話において共通に理解されている単語を省略することは親しい人間同士の間で見られる現象なので、ロボットがユーザーに親密感を与えるのにも有効であると考えた。対話における共通理解は、それまでの対話文の文脈に基づいて形成される。したがって、言語モデルのマスク予測手法を用いて文脈から予測できる単語を省略あるいは代名詞化することで発話の親密感を調節できると考えられる。第 2 に、発話における省略・代名詞化の度合いを調節するパラメータを定義した点である。マスク予測の確信度はある文脈における単語の出現頻度に基づいており、確信度が高いほど文脈から単語を予想することが可能であると考えられる。したがって、マスク予測の確信度が一定の閾値以上のときに省略・代名詞化を行う。対話相手との関係性に応じて閾値を調節することで適切な親密感を与えられると予想する。

2. 関連研究

2.1 日本語の対話における単語の省略および代名詞の使用

省略や代名詞化はコミュニケーションを円滑に進めたり、相手との関係性を親密にしたりすることに寄与すると考えられる。佐藤 [8] は話し手と聞き手が共通で理解している

単語を省略することで「経済的なコミュニケーション」が実現すると指摘している。また、Grice[3] が提唱する「協調の原理」の 1 つである「量の準準」では、必要な情報を全て提供し、過剰に多くの情報を提供しないことがコミュニケーションにおいて大切であるとしている。適切な省略や代名詞化は過剰な情報提供を防ぐ役割があると考えられる。指示代名詞が当事者間の関係に与える影響について、山崎 [4] は代名詞の使用が「話手と聞き手が同じ体験をもち合っているという印象を喚び起す」ことで親近感をもたらすと指摘している。また、省略及び代名詞の使用は対話の参加者同士が共通認識としている物事に対して行われることが多い [4], [8]。

2.2 ロボットの発話における省略および代名詞の使用

ロボットやエージェントの発話傾向と話者の関係性を明らかにする研究の多くは非言語的な側面に着目することが多かった [9]。一方、ロボットやエージェントの発話における省略や代名詞化が人間に与える印象を調査した研究は少ない。その中で西脇ら [5] はロボットが発達途中の幼児の発話のように一部の文節が欠落した発話をするすることで、ユーザの能動的な関与を引き出し、ユーザとの協調的な対話を促進することを示した。また、板敷ら [6] はエージェントが省略および指示代名詞の使用をすることで、人間がエージェントに対して親和性や友好性を感じることを示した。

しかし、西脇らの研究における省略及び指示代名詞の使用はユーザと共通の認識を前提としたものではない。また、板敷らの研究ではどの情報が共通基盤であるかを既知として省略や代名詞化を行っている。これに対して、ロボットが対話内容から共通基盤に相当する情報を推測した上で省略や代名詞化を行う研究は我々の調べる限りでは存在しなかった。

2.3 自然言語処理による省略の解決と照応生成

自然言語処理の分野において代名詞が指す対象や省略された単語を解決する照応解決は多く研究されている [10], [11]。これに対して、照応を用いた文章生成は照応解決に比べあまり研究されていない。例を挙げると、新聞記事をコーパスとして人手で作成した決定木をもとに照応表現を選択するモデルの研究 [12] や、文章中の共参照連鎖を利用することで参照表現を省略するか否かの 2 値問題を解く研究 [13] などがある。大規模言語モデルを用いて単語を省略した日本語文を生成する研究は我々の調べる限りでは存在しなかった。

2.4 本研究の問題設定

本研究では、省略のない対話文を入力として、省略や代名詞化が施された対話文を出力するという問題設定を取り

扱う。その際、ユーザとの関係性や対話履歴が与えられていることを前提条件とする。リアルタイムの対話における逐次的な発話処理については本研究では取り扱わない。

また、観察者が対話を聞いて主観的に感じる話者間の親密度を提案手法の評価とする。本研究の最終的な目標は省略及び代名詞を用いた発話により話者間の親密度を制御することにある。提案手法を用いて省略や代名詞化が施された対話文に対して観察者の感じる話者間の親密度が高くなることが望ましい。

3. 提案システム

本論文で提案するシステムは一連の対話文を入力として、省略及び代名詞化された新たな対話文を出力する(図2)。入力された対話文から省略・代名詞化する箇所の候補を選択するには大規模言語モデル BERT[7] のマスク予測手法を用いる。さらに、省略及び代名詞化の度合いを表すパラメータを定義し、これを閾値としてマスク予測手法の確信度を大小比較することで省略・代名詞化の判断を行う。代名詞化する場合は該当箇所を適切な代名詞または形式名詞に置き換える。また、パラメータとして省略限界 α を定義し、省略及び代名詞化の度合いを調節する。

提案システムにおいて、省略及び代名詞化は大きく以下の3つの手順で処理される。

- (1) STEP1: 日本語係り受け解析器 CaboCha[14] を用いた省略・代名詞化する候補の抽出
- (2) STEP2: BERT によるマスク予測
- (3) STEP3: 省略・代名詞化の判断と置き換え

3.1 STEP1: 日本語係り受け解析器 CaboCha を用いた省略・代名詞化する候補の抽出

入力された対話文の中から省略・代名詞化する単語の候補を抽出するために、入力文に事前処理を行い、省略・代名詞化候補となるトークン群を作成する。 N_S 個の発話文からなる対話文 S の i 番目の発話文を S_i とし、 N_S 個のインデックスを表す集合を $\mathcal{I}(N_S)$ とすると、入力対話文は $S = \{S_i | i \in \mathcal{I}(N_S)\}$ と表すことができる。

まず、 S を日本語係り受け解析器 CaboCha に入力して取得した解析情報から、各発話の分かち書きと文節区切りの情報、各単語の品詞細分類の情報を抽出する。分かち書き後の文に含まれる個々の単語をトークンとすると、各発話文 S_i は、 N_v 個のトークン v の集合として $S_i = \{v_{ij} | j \in \mathcal{I}(N_v)\}$ と表すことができる。

次に、省略・代名詞化候補となるトークンのリストを作成する。本研究では共通基盤となっている事柄を暗黙的に伝達することを目的として省略・代名詞化を行うので、対話文に含まれる単語の中でも共通基盤として成立しやすい

物事や事柄などのみを省略・代名詞化する単語の候補とする。具体的には、条件は以下の3つであり、形態素解析の結果を参照することで判断する。(1) 文節の最初のトークンである、(2) 品詞が名詞である、(3) 細分類が非自立、代名詞、数のいずれでもない。以上の条件を満たすトークンのインデックス (i, j) の集合を、次式で表す省略・代名詞化候補 I_C とする。

$$I_C = \{(i, j) | C(i, j), i \in \mathcal{I}(N_S), j \in \mathcal{I}(N_v)\} \quad (1)$$

ここで、 $C(i, j)$ はトークン v_{ij} が条件を満たすか否かを判断する命題である。

3.2 STEP2: BERT によるマスク予測

省略・代名詞化候補に含まれるトークン $\{v_{ij} | (i, j) \in I_C\}$ に対して、BERT でマスク予測を行い、各トークンが省略可能かどうかを判断する。

BERT は Transformer に基づいた双方向言語モデルである。BERT のマスク予測とは、文中で隠されているトークン(これをマスクと呼ぶ)に入りうる単語とその確信度を得る手法である。省略・代名詞化候補の位置にもともと入っていた単語が BERT のマスク予測によって高い確信度で予測できるならば、その単語は文脈上の共通理解になっていると考えられるので省略可能であると判断される。本研究では東北大学が作成した BERT の日本語事前学習済みモデル^{*1}を使用した。

BERT への入力文は、省略・代名詞化候補のトークンごとにそれぞれ作成する。この際、対話の文脈に基づいた予測を行うために、入力文には省略・代名詞化候補のトークンを含む発話文だけでなく、それ以前の発話文もコンテキストとして含めることとする。 i 番目の発話文における j 番目のトークン v_{ij} が省略・代名詞化候補であるとき、BERT に対する入力文は $\bar{S}_{ij} = [S_{i-1-k:i-1}, S'_{ij}]$ と表すことができる。ここで S'_{ij} は発話文 S_i について v_{ij} を [MASK] に置き換えた文であり、 k は考慮するコンテキストの長さである。

次に、入力文 \bar{S}_{ij} を BERT に入力し、マスクされた箇所に入りうる単語とその確信度の組み合わせ E_{ij} を取得する。

$$E_{ij} = \mathbf{F}_{\text{BERT}}(\bar{S}_{ij}) \quad (2)$$

ここで E_{ij} は、 $r(w)$ をトークン w の確信度の降順の順位、 $p(w)$ をトークン w の確信度、BERT の全トークンの集合を W_{vocab} とすると、 $E_{ij} = \{(r(w), p(w)) | w \in W_{\text{vocab}}\}$ と表すことができる。

3.3 STEP3: 省略・代名詞化の判断と置き換え

BERT から取得した各トークンの確信度、順位の集合

^{*1} <https://github.com/cl-tohoku/bert-japanese/tree/v1.0>

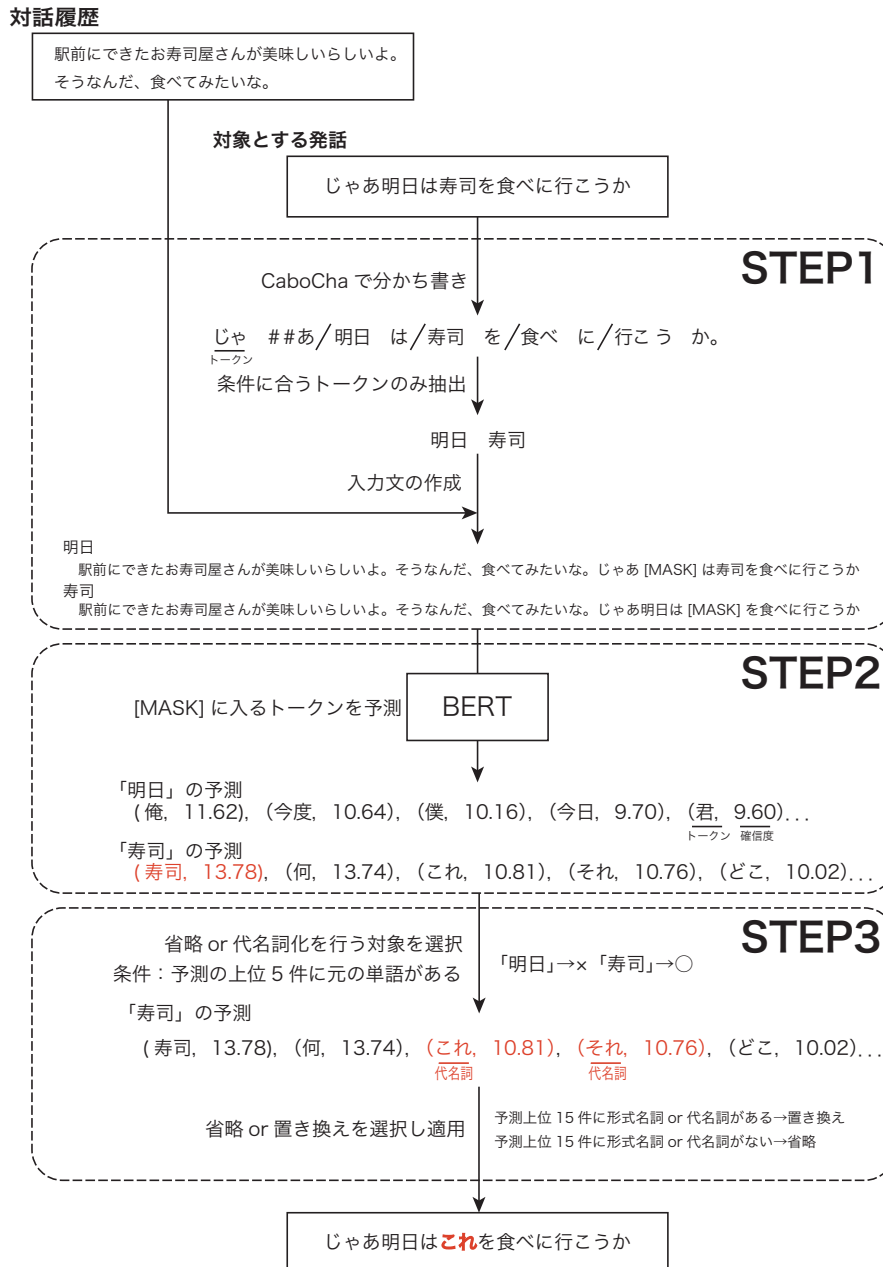


図 2 提案システムによる対話文の処理の流れ。

E_{ij} を用いて、 v_{ij} に対して省略・代名詞化を適用するか、適用する場合はどちらの手法を用いるかを決定する。

システムが発話文に対して省略・代名詞化を行う度合いをここでは省略限界と定義しパラメータ α で表す。省略限界はロボットとユーザとの関係性に基いて決定されるものとし、省略限界が高いほど発話文中のより多くの単語を省略・代名詞化する。

省略・代名詞化するかどうかは、省略・代名詞化候補のトークンごとにマスク予測の確信度を省略限界と比較することで判断する。則ち、 E_{ij} が次の式 (3) で表される条件を満たすときに限り v_{ij} を省略・代名詞化するという判断

を行う。

$$r(v_{ij}) \leq 5 \wedge \alpha \leq p(v_{ij}) \quad (3)$$

上記の式 (3) で表される条件を満たすトークン v_{ij} に対して省略と代名詞化のどちらを行うかを、確信度の高い形式名詞または代名詞の有無によって次のように判断する。

$$v_{ij} \leftarrow \begin{cases} w_m & (\text{if } r(w_m) \leq l) \\ "" & (\text{otherwise}) \end{cases} \quad (4)$$

ここで、 w_m は確信度の最も高い形式名詞または代名詞を表し、"" は空白を意味する。上式 (4) に示すとおり、確信

度上位 l 件のトークンに w_m が含まれる場合は、代名詞化が適当であると判断し該当のトークンを置き換える。一方、代名詞化の条件を満たさない場合は適当な置き換え候補が存在しないと判断し該当のトークンを単に省略する。

なお、通常 v_{ij} のみを省略するだけでは助詞が残って不自然な文になってしまう。そのため、省略を行う際は v_{ij} を含む文節ごと削除するものとする。

4. 実験

4.1 実験概要

本研究で実施した実験の目的は、提案手法により生成した対話が観察者に与える印象を、異なる省略限界の間で比較することである。実験は参加者間計画で行われた。実験参加者は2体の人型ロボットが対話する動画を視聴した後、対話の雰囲気やロボット間の関係性に関する質問と、話題の内容に関する質問のいずれかに回答した。

4.2 実験条件

本実験では段階的な省略・代名詞化の影響を調べるため、同一の対話例について省略限界が異なる3種類の対話を比較条件とした。具体的にはパラメータ α が以下のように異なる対話をそれぞれ作成した。また、全ての条件において $k = 10, l = 15$ とした。

- 省略なし: $\alpha = \infty$,
- 省略少: $\alpha = 13$,
- 省略多: $\alpha = 0$.

4.3 実験刺激

実験刺激となる動画の作成に使用した対話文は大阪大学マルチモーダルコーパス Hazumi [15] に収録されている2者間の対話を元に生成された。元となった対話の内容は対話者の学生時代についての雑談であった。

本実験では口調や表情などの発話内容以外の要因で印象が変化してしまうのを避けるため、刺激動画に人間を登場させず、代わりに2体の人型ロボットに合成音声が発話させてその様子を録画した。人型ロボットとして2体の Pepper を用いた。

合成音声は CeVIO Creative Studio^{*2} という音声創作ソフトウェアを用いて作成した。2者の発話はそれぞれ CeVIO Creative Studio で用意されている「さとうさらら」と「すずきつづみ」というキャラクターの声をを用いて合成した。

刺激動画は図3のように2体の Pepper 同士を向かい合わせた状態で撮影した。また、観察者にとってどちらの Pepper が発話しているかを容易に判別できるようにする

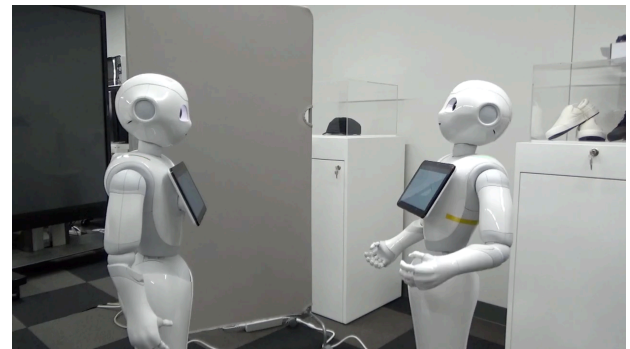


図3 実験に使用した Pepper の対話動画。

こと、自然な対話に近づけることの2つを目的として、どちらの Pepper も発話開始時に手を広げたり頷いたりする単純な動作を行うようにした。発話開始時の Pepper の動作及び発話タイミングは全ての条件で統一した。

4.4 実験手続き

クラウドソーシングを通して18~71歳の男女600人の実験参加者を集めた。実験参加者は男性327名、女性273名である。参加者の年齢平均は42歳であり、年齢の標準偏差は10.4である。実験参加者は省略なし、省略少、省略多の3条件にそれぞれ200人ずつランダムに割り当てられた。さらに、各条件の参加者は「対話内容に関するアンケートに回答する群」と「対話の雰囲気に関するアンケートに回答する群」の2群にそれぞれ100人ずつランダムに割り当てられた。

実験はインターネット上のアンケートフォームを用いて行った。参加者がフォームにアクセスすると、まず該当する条件の刺激動画を視聴するよう求められた。その後、以下の表1に示すアンケートの質問項目について、どのくらい自分が同意するかを7段階のリッカート尺度で回答してもらった。これに加えて、すべての群で対話内容について気になる点を自由記述で回答してもらった。参加者は実験終了後50円の報酬を得た。

集めたデータのうち回答に不備のある参加者のデータを取り除いた結果、有効なデータ数は合計で593件となった。

4.5 実験結果の予測

「対話内容に関する質問項目」では、「意思疎通」を除き省略が多くなるほど評価が低くなると予測する。これは省略箇所が多いほど対話の内容は理解しにくく、元のコーパスより不自然な対話になるためである。一方省略を行うことで観察者と話者の間に理解度の差が生じるため、話者間に共通認識があると観察者が判断する可能性がある。そのため、省略箇所が多いほど「意思疎通」の評価が高くなると予測する。

「対話の雰囲気に関する質問項目」では、「初対面」、「介

^{*2} <https://cevio.jp/product/ccs>

表 1 質問項目. 各質問項目における矢印は、省略限界が高くなった場合に評価が変化すると予測される方向を示している.

尺度名	質問内容
対話内容に関する質問項目	
内容理解	Q1. 対話の内容が理解できた↓
自然さ	Q2. 対話は自然だった↓
意思疎通	Q3. ロボット同士の間でお互いの言いたいことが伝わっている感じがした↑
知性	Q4. 2体のロボットに知性があると感じた↓
対話の雰囲気に関する質問項目	
ロボット間の親しみ	Q1. ロボット同士は互いに親しみを感じている↑
盛り上がり	Q2. 対話は盛り上がっている↑
初対面	Q3. 二体のロボットは初対面だと思った↓
有意思性	Q4. ロボットは自分の意思で話していると感じた↑
活気	Q5. 2体のロボットは生き生きと話しているように感じた↑
介入可能性	Q6. ロボット同士の会話に途中参加できそうと感じた↓
人間らしさ	Q7. 2体のロボットは人間のように話していると感じた↑
親しみやすさ	Q8. 2体のロボットに親しみやすさを感じた↑
楽しさ	Q9. 2体のロボットは会話を楽しんでいるように感じた↑

介入可能性」を除き省略箇所が多くなるほど評価が高くなると予測する。これは省略箇所が多いほど話者間に共通認識があるという感覚を受け、話者が仲良く話していると感じるのではないかと考えたからである。一方で話者同士でのみ共通認識が形成されていると観察者が対話に途中参加しにくくなるため「介入可能性」の評価が低くなると予測した。また、「初対面」の評価については、共通認識のある話者同士は関係性が近く、初対面ではないと判断されやすいことから評価が低くなると考える。

4.6 実験結果

まず、対話内容に関する質問項目に対する回答者の評価の平均を図 4 に示す。Tukey-Kramer 法に基づいた多重比較 ($p < 0.05$) によると、全ての項目において省略多条件と省略なし条件の間に有意差が認められた。また、「内容理解」、「自然さ」、「意思疎通」において省略少条件と省略なし条件の間で有意差が認められ、「内容理解」のみにおいて省略少条件と省略多条件の間で有意差が認められた。

次に、対話の雰囲気に関する質問項目に対する回答者の評価の平均を図 5 に示す。Tukey-Kramer 法に基づいた多重比較 ($p < 0.05$) によると、「人間らしさ」において省略なし条件と省略少条件の間でのみ有意差が認められた。その他の項目においては全ての条件の間で有意差が認められなかった。

5. 考察

5.1 質問項目の結果

この項ではアンケートの各質問項目の回答結果を、結果の予測と照らし合わせて考察する。

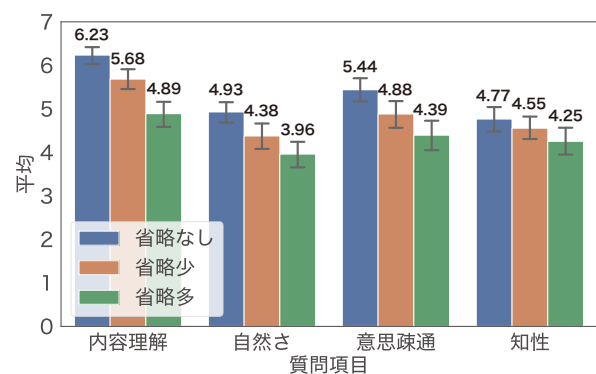


図 4 対話内容に関する質問への回答の平均値. エラーバーは 95% の信頼区間を表す.

「対話内容に関する質問項目」では、全ての項目で省略箇所が多くなるほど評価が低くなっていくことが分かった。これは、「内容理解」、「自然さ」、「知性」の項目に関しては予想通りであるが、「意思疎通」の項目に関しては予想に反していた。ロボットの動作を条件間で統一しているにもかかわらず「意思疎通」の項目の評価が省略するほど低くなることは、実験参加者が自分の理解できない内容に対してロボット同士も理解できていないと判断したことを示している。ロボット同士が対話内容を理解していないという印象を与える理由として、省略によって発話文の整合性が失われていることが考えられる。提案システムでは省略によって係り受け関係が崩れる可能性があるため、代名詞化に比べて省略が多く行われた場合は整合性のない発話文が生成されやすい。実際、省略少条件では代名詞化された単語は 1 個もなく、省略多条件では代名詞化された単語が 2 個だったのに対して省略された単語は 17 個と圧倒的に多かった。

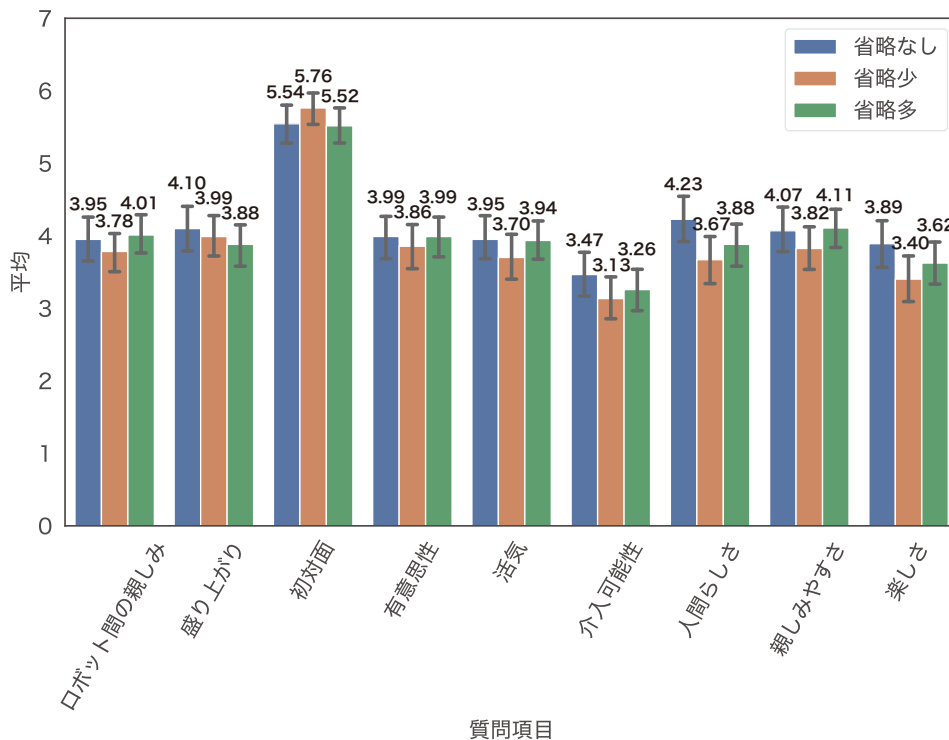


図 5 対話の雰囲気に関する質問への回答の平均値。エラーバーは 95%の信頼区間を表す。

提案手法により生成される発話文の整合性を改善する方法としては係り受け関係を考慮できるようにしたり、代名詞に置き換えることで文の構造を保つようにしたりすることが考えられる。また、指示語はお互いの共通認識を明示的に示すものであるため、指示語への置き換えによって観察者にとっての発話内容の理解度と話者にとっての発話内容の理解度が異なると判断されやすくなる効果が期待できる。

一方で、提案システムが行う省略と同程度に文の整合性を失う省略は人間同士の発話内では起こりうるものである。それにもかかわらず本実験で提示した対話が違和感を生じる理由が、話者がロボットであるからなのか、それとも省略の仕方に人間の省略と何らかの相違があるからなのかは今後検討の必要がある。

「対話の雰囲気に関する質問項目」では、「初対面」、「介入可能性」を除き省略箇所が多くなるほど評価が大きくなるという予測に反して、結果としては「人間らしさ」の項目を除いた全ての項目において条件間で有意な差がなかった。この理由として、元となった対話例が人間と Wizard of Oz[16] 形式のエージェントとの間の雑談であったために、元の対話例の時点で適度に省略されていた可能性が挙げられる。したがって、元の対話文として対話コーパスからゼロ省略などを復元した過度に冗長な対話文を採用することで雰囲気に関する項目をより詳細に分析できると考えられる。

最後に、対話内容に関する質問項目では多くの箇所でも

有意な差が出ていたのにも関わらず対話の雰囲気に関する質問項目では有意な差がないことから、内容の理解は雰囲気の評価に影響を与えない可能性が考えられる。これには観察者にとって内容が理解できない場合でも話している人間同士には共通のコンテキストがある可能性を了承しているということが理由として考えられる。

5.2 アンケートで得られた意見

アンケートでは最後に、対話内容について気になる点を自由記述してもらった。ここではその中で注目すべき意見を 2 つ取り上げる。

まず、対話の雰囲気に関して、全ての条件において面接のようであるという意見が多数挙げられた。これは元の対話が一問一答形式に近いためだと考えられる。

次に、会話の整合性について、特に省略された会話に関して噛み合っていないという意見が多く見られた。また、噛み合わない会話に違和感を感じたという意見も挙げられていた。

以上の 2 つの意見が多かったことから、一問一答形式の対話では、省略した発話に違和感が存在すると噛み合っていない会話とみなされてしまうのではないかと考えた。一問一答形式の対話を成立させるには質問と応答が対応している必要があり、さらに次の質問は前の回答を踏まえて投げかける必要がある。そのため、省略により実験参加者が発話内容を理解できなかった場合に、回答者が質問の意図を汲み取っていない、あるいは質問者が回答の内容を理解せず

別の質問を投げかけているという印象になるのではないかと考えられる。

6. おわりに

本論文ではロボットが人間との関係性や共通体験に応じて親密感を表出するために、対話の文脈に基づき言語モデルを用いて省略・代名詞化した発話を生成する手法を提案した。省略・代名詞化の基準にBERTのマスク予測機能の確信度を利用することでパラメータによる省略・代名詞化の度合いの調節を実現した。実験では提案手法を適用したロボット同士の音声対話が観察者にどのような印象を与えるかを調査した。実験の結果、省略・代名詞化によって対話内容の理解が困難になっても、観察者が知覚するロボット間の親しみは有意に下がらないことが確認された。

謝辞 本論文は、JST, CREST, JPMJCR19A1とJSPS KAKENHI JP21J13789の支援を受けたものである。また、本研究では、国立情報学研究所のIDRデータセット提供サービスにより大阪大学から提供を受けた「大阪大学マルチモーダル対話コーパス Hazumi」を利用した。

参考文献

- [1] 宮下善太, 神田崇行, 塩見昌裕, 石黒浩, 萩田紀博: 顧客と顔見知りになるショッピングモール案内ロボット, 日本ロボット学会誌, Vol. 26, No. 7, pp. 821–832 (2008).
- [2] Sperber, D. and Wilson, D.: *Relevance: Communication and cognition*, Vol. 142, Citeseer (1986). (D. スペルベル, D. ウィルソン, 内田聖二, 栄南先, 中達俊明, 田中圭子 (訳), “関係性理論—伝達と認知—”, 研究者出版, 1993.).
- [3] Grice, H. P.: Logic and conversation, *Speech acts*, Brill, pp. 41–58 (1975).
- [4] 山崎良幸: 代名詞の機能について, 日本文学, Vol. 7, No. 1, pp. 29–37 (1958).
- [5] 西脇裕作, 板敷尚, 岡田美智男: ロボットの言葉足らずな発話が生み出す協調的インタラクションについて, ヒューマンインタフェース学会論文誌, Vol. 21, No. 1, pp. 1–12 (2019).
- [6] 板敷尚, 西脇裕作, 大島直樹, 岡田美智男: なぜスマートスピーカーはよそよそしいのか? ロボットとの親近感を生み出す代名詞の役割, ヒューマンインタフェース学会論文誌, Vol. 22, No. 2, pp. 65–76 (2020).
- [7] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [8] 佐藤雄一: 話しことばの文の諸相から (はなしことばの現在)–(はなしことばの構造), 国文学解釈と鑑賞, Vol. 59, No. 1, pp. p70–77 (1994).
- [9] Matsumori, S., Fukuchi, Y., Osawa, M. and Imai, M.: Do others believe what I believe? Estimating how much information is being shared by utterance timing, *Proceedings of the 6th International Conference on Human-Agent Interaction*, pp. 301–309 (2018).
- [10] Iida, R., Torisawa, K., Oh, J.-H., Kruengkrai, C. and Kloetzer, J.: Intra-sentential subject zero anaphora resolution using multi-column convolutional neural network, *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 1244–1254 (2016).
- [11] Yamashiro, S., Nishikawa, H. and Tokunaga, T.: Neural japanese zero anaphora resolution using smoothed large-scale case frames with word embedding, *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation* (2018).
- [12] 橋本さち恵, 乾健太郎, 白井清昭, 徳永健伸, 田中穂積, 東京工業大学, 九州工業大学: 日本語文生成における照応表現の選択, 情報処理学会自然言語処理研究会, NL-143 (2001).
- [13] 飯田龍, 徳永健伸ほか: 日本語書き言葉を対象とした参照表現の自動省略-人間と機械処理の省略傾向の比較, 研究報告音声言語情報処理 (SLP), Vol. 2012, No. 15, pp. 1–10 (2012).
- [14] 工藤 拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842 (2002).
- [15] 駒谷和範, 岡田将吾: 大阪大学 マルチモーダル対話コーパス Hazumi. 国立情報学研究所情報学研究データリポジトリ. Hazumi1902 (2021). <https://doi.org/10.32130/rdata.4.1>.
- [16] Fraser, N. M. and Gilbert, G. N.: Simulating speech systems, *Computer Speech & Language*, Vol. 5, No. 1, pp. 81–99 (1991).