AIは人間らしく話ができるか：ロボットと仲良くなるために

川原功司[1]　大道麻由[2]　高橋英之[3]
[1]名古屋外国語大学　[2,3]大阪大学
kkoji@nufs.ac.jp　omichi.mayu@irl.sys.es.osaka-u.ac.jp
takahashi@irl.sys.es.osaka-u.ac.jp

概要

ロボットが人間と遜色なく会話できるかという問題は，人工知能研究における大きな課題の一つである．本稿では，人間とロボットが快適に会話ができるかというプロジェクトで得られた報告を元に，人間とロボットの会話の特性の違いを明らかにする．特に，ロボットは会話を遂行する際に自分の誤りを修正する能力がなく，これが会話分析における他者修復という観点から分析することが可能であり，軌道修正をする力の有無が現在の人工知能と人間の会話を分ける大きな違いであると主張する．

1　はじめに

人工知能に知能があるかどうか検証する方法としては，チューリングテストやそれに対するジョン・サールによる反論の中国語の部屋が有名である．本稿では，ロボットが会話において人間らしくふるまえるかどうかについて考察するが，究極的な目標として人工知能における知性とは何かという問題意識を持っているということは指摘しておきたい．

1.1　チューリングテスト

チューリングテスト（イミテーションゲーム）は，「機械に考えることができるか」という知能に関する設問である [1]．この設問は「機械が人間と区別できないほど知的にふるまうことができるか」という問題に置き換えることができる．しかし，松原(2011)[2] が指摘するように，チューリングテストの位置づけを考えれば，知能の定義の指標として次のような問題がある．

- キーボードとディスプレイでのやりとりという設定は，インタラクションのチャネルとして細すぎる．
- 対話は知能の一部の側面しか反映していない．
- 話題を限定しない対話は不自然である．

1.2　中国語の部屋

このような問題をはらみながらも，チューリングテストは機械が知能を持つかどうかについて考えるきっかけになったことは間違いない．また，チューリングテストに関する反論としては，サールによる「中国語の部屋」の議論が有名である [3]．サールは人間の認知を説明できるような強い AI にさえ人間と同等の知能がないということを示すために，中国語の部屋という思考実験を提供している．中国語の部屋は以下のような設定である．ある部屋に中国語の分からないアメリカ人が一人おり，その部屋に中国語で書かれた質問が投げ込まれる．部屋にいるアメリカ人は英語のマニュアルを持っており，それを参考にして中国語の返事を部屋の外に返す．こうすれば，部屋の外から質問を投げ入れた人から見れば，部屋の中の人には中国語の知識があると判断するようになるが，その人はもちろん中国語のことが分かっているわけではない．つまり，チューリングテストが想定しているような，外から見た行動だけでは知能の有無は判断できないというものである．

これに関しては，既にいくつか反論がある．一つは，アメリカ人にマニュアルが与えられれば，つまり部屋全体というシステムで見れば中国語の知識があると判断できるため，その中国語の知識を構成するシステムの一部であるアメリカ人のみを見て判断することは適切ではないということ．もう一つは，計算量の問題を考慮に入れていないことである [4, 5]．

1.3　トータルチューリングテスト

この種のテストは人間の知能そのものの研究というよりは，いかに人間らしくふるまうかを試したテストであると考えられる．たとえば，人間と最もう

まく対話したシステムに栄誉が与えられるローブナー賞という大会では，人間が犯しがちなスペルミスをわざと入れてみるとうまく人がだませるようになったという事例が紹介されている [2]．また，この賞は結局，Chat bot のパターンマッチングの工夫に終始してしまい，2019 年を最後に開催されていない．こういう事例を聞けば，チューリングテストでは人間らしいふるまいとは何かをよく理解することができないとも思われる．しかし，相手の行為から判断してそこから人間性を発見するのだという視点に立てば，有効な活用方法もありそうである．そこで，テキストのやりとりだけではなく，人間のもつ全てのモダリティを拡張して行うトータルチューリングテストが提案された [6]．つまり，シンボル操作だけではなく，感覚や物理的な操作を統合したシステムを考慮すれば，より適切な「こころ」のモデルが形成できるというわけである．これを具体化した事例として，自律型アンドロイドを用いたトータルチューリングテストの報告もあり [7]，チューリングテストは基本方針として人間の「こころ」を探る手がかりにはなり得そうである．

1.4 ChatGPT の知能

チューリングテストやそれに準ずる基準を用いて人工知能，特に Transformer を実装した ChatGPT (3, 4) の知能を人間と比較する研究も行われた [8]．まず，性格診断に関わる，外向性，協調性，誠実性，神経症傾向，開放性について調査が行われた．結果，GPT-4 は人間の中央値に近かった．次に，行動経済学ゲームを通した調査をしたところ，GPT-4 は利他的で公平性を重視する傾向があり，独裁者ゲームや公共財ゲームでは人間の中央値を上回った．また，囚人のジレンマでは人間よりも大幅に協力的であった．総合すると，ChatGPT はかなり人間的な行動を示し，協力的で利他的な行動を示すということが示された．一方で，人間よりも行動が一貫しており，集中した戦略をとるという傾向もあった．

2 会話という行為

ヒューマンロボットインタラクションについて考えるために，本稿では会話という行為の特性をいくつか分析し，人工知能が人間らしくふるまうために必要となるであろう要素について考えていくことにする．本稿で注目するのは，あくまで言葉に関わる部分のみを通して，人工知能が人間らしくふるまえるかという点である．

2.1 順番交代

会話とは，人間の社会的行動の主要なものの一つである．会話では，発言のやりとりを行っているが，発言の順番が入れ替わることは順番交代と呼ばれている．順番交代は，「一人が話し，次にもう一人が話す」という規則性がある．この順番交代の間隔は，一般的には 200〜300ms 程度という非常に短い時間である [9]．一方，新しい発話を生成するには 600ms 以上の時間が必要である．この短い間隔で順番交代を成立させるためには，聞き手は話し手の発話が終わるタイミングを予測し，自分の発話の準備を事前にしている必要がある．Sacks et al. (1974)[10] によれば，発話の順番が終わる（可能性がある）点は移行適切場と呼ばれている．

Levinson & Torreira (2015)[11] が指摘するように，会話分析のモデルでは発話の終了を正確に予測する能力が必要であることが前提にある．それでは，短い順番交代の間隔で会話のやりとりができる予測能力があるということを示す根拠はあるのだろうか．Levinson & Torreira によれば，聞き手は話し手の発話の構造やイントネーションなどを頼りに発話の終了が予測でき，予測の過程は発話が終わる 400ms 前から始まっているということが実験で示されている．たとえば，質問があれば，聞き手は回答を準備する過程で関連する単語を予測している．こういった予測は，神経活動を計測する実験からも裏付けられている．また，順番交代の準備が発話理解と重なって進行していると考えざるを得ない．聞き手の予測や準備の並行性の他にも，話者同士で類似した単語や構文を使用することによる言語処理の効率化というプライミング効果も一役買っている [9]．一方で，発話の生成には 600ms 以上の時間が必要で，特に複雑な応答の場合には時間がかかる．つまり，人間は短い間隔の順番交代で会話を展開することができるが，これを可能にしているのは発話理解と生成のプロセスを並行して行うことで，この能力は予測能力に大きく依存しているということがわかる．また，このシステムは言語発達において重要な役割があり，幼児期から徐々に精緻化されて獲得される．

2.2 配列構造

Kendrick et al. (2020)[12] は，会話の隣接ペアにおける配列構造について，12 の異なる言語に基づいて

普遍的な特徴を記述している．隣接ペアは会話の基本単位であり，特定の発話に対して特定の応答が条件的に関連付けられている．そして，応答が欠けた場合，欠如として認識され，さらなる応答（追究）が発生することで会話が形成される．隣接ペアには質問と応答，挨拶と挨拶，申し出と許諾/拒否といったものがある．そして，隣接ペアは事前拡張という基本配列の前に補助的な発言がなされたり（Hey!という呼びかけや，名前を呼ぶ），挿入拡張という形で理解の確認や追加情報の要求があったり，事後拡張という形で応答の確認や追加の話題提示などが追加されることもある．

2.3 他者修復

会話では，順番交代が非常に短いタイミングで行われるため，聞き手が発話内容を理解できないことが頻発する．このために，他者修復として，聞き手が「What?」，「Sorry？」といった確認や聞き返しの作業が行われる．他者修復が発話されるまでの時間は，通常の順番交代よりも時間を要し，平均で700ms 程度のギャップが観察される．また，ギャップの長さは修復の内容や状況によっても変化し，「What?」や「Sorry?」といったどこにつまづいたのかを明示しない修復よりも（平均で 700〜800ms），「They're what?」や誰かが行くという情報が踏まえられている状況で「Who?」と尋ねるようなより特定的な修復の方がギャップが短い傾向にある（平均で400ms）[13].

他者修復における遅延の理由としては，以下の要因が考えられる．まず，聞き手が問題の認識に時間がかかること．つまり，聞き手は理解できなかった発話を事後的に遅れて認識する場合があったり，聞き逃しがあった場合にも記憶や文脈を手がかりにして内容を事後的に理解することがあることによるものである．他にも，聞き手が即座に他者修復をしないことで，発話者自身が自分のミスに気づき，自己修復する機会を与えるという側面もありうる (ibid.).

また，会話における修復のやりとりには，普遍的な傾向が観察されることも知られている．Dingemanse et al. (2015)[14] では，12 の異なる言語において調査が行われた．まず，修復という現象は，どの言語でも会話の 1.4 分毎に 1 回の頻度で発生する．また，修復とその返答のペアのやりとりの発話量は一定する傾向にあるということである．修復という現象の配列は，問題の発端となる問題が生じた発話，その問題を指摘するための修復の発起，そしてその問題を修正するための修復の解決という 3 つの基本要素から構成される．

- A: Oh Sibbie's sistuh had a ba:y　問題の発端
- B: Who?　修復の発起
- A: Sibbies sister.　修復の解決

そして，修復の発起としては，以下の 3 つのタイプがどの言語でも確認された．

- オープンリクエスト：問題があると指摘しつつ，どこなのか，何なのかを明示せず，明確化を求める．Huh?など．
- 制限付き要請：特定の問題点を指摘し，その明確化を求める．Who?など．
- 制限付き提案：何が言われたのかを確認したりする．She had a boy?など．

修復に当たっては，修復の発起の発言はできるだけ特定的なタイプを選択するという傾向がある．つまり，より多くの情報を把握しており，一部が分からないだけであれば制限付き提案や要請が優先され，広範囲に渡って不明確な部分がある場合にのみオープンリクエストが選択されるというわけである．また，修復配列は元の発話の長さに近い効率性を保つ傾向があり，修復プロセスが冗長にならず，会話の進行が妨げられないようになっている．また，労働量分担の原則があり，修復の負担は発起者と話し手で分担され，その労働量が反比例する傾向にある．つまり，制限付き提案は発起者の方により負荷がかかる分，話し手の負担が減り，オープンリクエストは発起者の負担が軽い分，話し手に負担がかかる傾向にあるということである．

3 ロボットとの会話

会話における特徴に注目して，人間とロボットの会話について概観した論文がある [15]．そこでは，ロボットは応答遅延や割り込みなど，順番交代における流暢さに課題が残っていることが既に指摘されている．人間の会話では，音声や視線，ジェスチャーなど複数のモダリティを駆使して順番交代を調整している．一方，ロボットの対話システムでは，サイレンスベースモデルという形で音声区間検出における沈黙の時間を手がかりにしたり，順番交代になりそうなポイントを音声区間検出から導き出したり，スピーチの連続体の中から探し出したりするモデルが採用されている．また，音声認識の技術

力そのものにまだまだ課題があったり，会話における複雑な信号処理が難しいという問題が指摘されている．今後の研究の方向性としては，音声認識技術を向上させることで順番交代のポイントを理解したり，タイミングを予測する精度を向上させる必要のほか，複数人の会話における順番交代をどのように理解させるのか，そしてロボットの身体性を向上させる必要性が指摘されている．

トータルチューリングテストにおける石黒のアンドロイド研究はまさにこの問題に取り組んでいるものであるし，マルチモーダルトランスフォーマーの研究にも注目する必要があるだろう [16].

3.1 現在のプロジェクト

今回の研究対象として，人間と日常的に雑談を通じてコミュニケーションをするロボットと人間の会話を分析する．ベースは ChatGPT である．

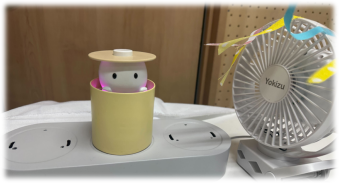

図 1 おしゃべりロボット

3.2 ロボットの順番交代

Skantze (2021)[15] で指摘されているとおり，ヒューマンロボットインタラクションにおいて，順番交代は人間同士の会話と同程度の時間間隔で進められているわけではない．このため，ロボットの返答が始まるまでに長いギャップが存在する．具体的にみると，人間が話しかけてからロボットが起動するのに 3〜5 秒，さらに応答の発話までに 3 秒以上かかっている．また，音声認識技術の不足のため，人間が通常の会話より明らかに意識して明瞭に発話する必要があり，流暢な順番交代ができていない．

この事実を踏まえると，今後は Levinson & Torreira (2015) の研究で指摘されているとおり，人間が行っている予測能力を明らかにし，それをどのように AI に実装していくかが鍵になりそうである．少なくとも，人間の発話を聞き終えてから処理をしているようでは，自然な順番交代にはならなさそうである．

3.3 ロボットの会話の配列構造

配列構造に関しては，かなり高いレベルで会話の基本構造が形成されていると言える．以下のように，質問に対しては応答を行っている．

- 人間：今何時
- ロボット：今は 13 時 48 分だよ．

また，挨拶をすると挨拶を返すこともでき，話し手と聴き手で発話が変わるダイクシス表現を使いこなすことができている．

- 人間：ただいま
- ロボット：おかえり！今日はどうだった？何か楽しいことがあった？

また，申し出に対しては許諾や拒否を行うことも問題はなさそうである．

- 人間：何かお話しして
- ロボット：ぼくは今日はぬくぬくの気分だよ！何か楽しいことあった？

3.4 ロボットの他者修復

人間は話が分からないまま会話を遂行することが比較的少なく，不明な点があればそれを修復して会話を遂行する．一方で，今回のプロジェクトでもロボットによる他者修復は見られなかった．この点は現行の人工知能の欠点とも言えるもので，より協力的な態度で会話を遂行しようとしてくれているとも言えるが，分からないことを分からないと言わないで辻褄が合いそうな返信をしてしまうため，的外れな回答をしてしまうことがある．ハルシネーションという言葉は日本語に浸透しつつあるほど有名であるし，会話の軌道修正を自力で行うようになることが今後の人工知能にとって大きな課題になるかもしれない．なお，言語学においては，自分が事実であると認識している情報のやりとりをするのが会話の遂行上必須であるという現象は，グライスによる質の格率という名前で広く知られている [17]．人工知能が質の格率を犯しがちという問題は，人工知能が協力的すぎるという側面の反映なのかもしれない．

なお，現在のプロジェクトでは，人間がロボットに質問されすぎることに嫌気がさして，話をやめてしまう傾向があることが分かった．話題を深く掘り下げることが苦手な上に，ロボットは会話を続けること自体が自己目的化しているため，すぐに新たな質問に移り，人間に余計な知的負荷を与えてしまっている．ロボットには，人間とは異なる動機が垣間見られるため，心地よい会話のやりとりをするためには目的を学習させる必要もありそうである．

謝辞

本研究は JSPS 科研費 24K00066 の助成を受けたものです.

参考文献

[1]Alan Turing. Computing machinery and intelligence. **Mind**, Vol. LIX, pp. 433–460, 1950.

[2]松原仁. チューリングテストとは何か. 人工知能, Vol. 26, No. 1, pp. 42–44, 2011.

[3]John R. Searle. Minds, brains, and programs. **Behavioral and Brain Sciences**, Vol. 3, No. 3, pp. 417–424, 1980.

[4]Hector J. Levesque. Is it enough to get the behavior right? In **Proceedings of the linternational Joint Conferences on Artificial lintelligence Organization**, pp. 1439–1444, 2009.

[5]中島秀之. 中国語の部屋再考. 人工知能, Vol. 26, No. 1, pp. 45–49, 2011.

[6]Stevan Harnad. Other bodies, other minds: A machine incarnation of an old philosophical problem. **Minds and Machines**, Vol. 1, No. 1, pp. 43–54, 1991.

[7]石黒浩. アンドロイドによるトータルチューリングテストの可能性. 人工知能, Vol. 26, No. 1, pp. 50–54, 2011.

[8]Qiaozhu Mei, Yutong Xie, Walter Yuan, and Matthew O. Jackson. A turing test of whether ai chatbots are behaviorally similar to humans. **Proceedings of the National Academy of Sciences**, Vol. 121, No. 9, p. e2313925121, 2024.

[9]Antje S Meyer. Timing in conversation. **J Cogn**, Vol. 6, No. 1, p. 20, 2023.

[10]Harvery Sacks, Emanuel Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. **Language**, Vol. 50, pp. 696–735, 1974.

[11]Stephen C Levinson and Francisco Torreira. Timing in turn-taking and its implications for processing models of language. **Front Psychol**, Vol. 6, p. 731, 2015.

[12]Kobin H. Kendrick, Penelope Brown, Mark Dingemanse, Simeon Floyd, Sonja Gipper, Kaoru Hayano, Elliott Hoey, Gertie Hoymann, Elizabeth Manrique, Giovanni Rossi, and Stephen C. Levinson. Sequence organization: A universal infrastructure for social action. **Journal of Pragmatics**, Vol. 168, pp. 119–138, 2020.

[13]Kobin Kendrick. Other-initiated repair in English. **Open Linguistics**, Vol. 1, pp. 164–190, 2015.

[14]Mark Dingemanse, Seán Roberts, Julia Baranova, Joe Blythe, Paul Drew, Simeon Floyd, Rosa Gisladottir, Kobin Kendrick, Stephan Levinson, Elizabeth Manrique, Giovanni Rossi, and N. J. Enfield. Universal principles in the repair of communication problems. **PLoS One**, Vol. 10, No. 9, pp. 1–15, 2015.

[15]Gabriel Skantze. Turn-taking in conversational systems and human-robot interaction: A review. **Computer Speech & Language**, Vol. 67, p. 101178, 2021.

[16]Kazuki Miyazawa and Takayuki Nagai. Survey on multimodal transformers for robots. **TechRxiv**, 2023.

[17]Paul Grice. Logic and conversation. In Peter Cole and Jerry Morgan, editors, **Syntax and Semantics: Speech Acts**, Vol. 3, pp. 43–58. Academic Press, New York, 1975.