

非自然言語を発するロボットとの しりとりにおけるユーザーの言語理解

Interaction with the robot speaking a non-natural language: Effect of "Shiritori" game on understanding the word

勝 将也¹ 中島 綾乃¹ 中島 亮一² 大澤 正彦^{1*}

Masaya Katsu¹, Ayano Nakajima¹, Ryoichi Nakashima², Masahiko Osawa¹

¹ 日本大学

¹ Nihon University

² 京都大学

² Kyoto University

Abstract: しりとりをすることで、非自然言語（自然言語の単語の各音を「ド」と「ラ」に変換した音）を発するロボットとの言語的インタラクションが成立する可能性が示されている。本研究では、しりとりをすることがロボットが発する非自然言語の理解に与える影響を調べた。実験参加者は、しりとりを想定する・しない状態で、ロボットが発する非自然言語を発する動画を視聴した後、ロボットの発した音声の意味する単語を回答し、それに対する自信度を報告した。しりとりを想定するときは、単語の正答率と自信度が有意に高く、頭文字が既知であることの影響が大きいと考えられる。

1 はじめに

国際ロボット連盟は、2020年にエンターテインメントロボットなどの個人用サービスロボットが、対前年6%増の1850万台が販売され、今後も順調に増加する見込みを報告した¹。それに伴い、人とロボットのコミュニケーション（特に言語的なコミュニケーション）も増加していくと考えられる。人とロボットのコミュニケーションにおいて、言語的インタラクションを破綻させずに継続させることは重要な課題である。近年、人とロボットの言語的インタラクションを継続させるために、自然言語（人間が生活の中で一般的に使っている言語を自然言語と定義する）を発話しないロボットを用いることが提案されており、ある特定の状況においてそれが有効であることも報告されている[1][2][3]。本研究の目的は、先行研究で提案された自然言語を発話しないロボットと言語的インタラクションが成立するために、どのような要素が重要かを具体的に明らかにすることである。

2 背景

2.1 適応ギャップ

ロボットとのインタラクションを継続するために重要な要素として、適応ギャップがある。小松らは、元々期待していたロボットの能力の大きさ(F_{before})と実際にインタラクションすることで感じたロボットの能力の大きさ(F_{after})との差を適応ギャップと定義し($F_{after} - F_{before}$)、 F_{before} と比べて F_{after} が小さいとき、すなわち適応ギャップが負の値となるとき、ロボットとのインタラクションを止める傾向があることを報告している[4]。このことから、適応ギャップが正の値であること（あるいは負の値であっても0に近いこと）が、ロボットと人の継続的なインタラクションにとって重要な要素である可能性が考えられる。負の値をとる適応ギャップを回避する一つの方法は、 F_{after} を大きくする、つまり期待以上のコミュニケーション能力を有するロボットを構築することである。しかし、現状の技術レベルで最大限能力の高いロボットを構築しても、コミュニケーションに失敗することも多い。その理由の一つとして、ロボットに対する人の期待が上がり、ロボットの振る舞いがその期待と乖離すると、人がロボットに裏切られたと感じてしまうことが挙げられる。つまり、 F_{after} を大きくするのに伴い、 F_{before} も大きくなるため、期待以上の能力を持たせること自体が非常に困難であり、負の値

*連絡先：日本大学文理学部

〒156-8550 東京都世田谷区桜上水 3-25-40

E-mail: osawa.masahiko@nihon-u.ac.jp

¹World Robotics 2021 Service Robots press release German.

<https://ifr.org/downloads/press2018/>

DE-2021-NOV-04-IFR_Pressemeldung_Service_Robots.pdf

をとる適応ギャップを回避できないことが多い。

負の値をとる適応ギャップを回避するための他のアプローチとして、人が元々ロボットに対して想定する能力の大きさ (F_{before}) を小さくすることが考えられる。実際に、人にロボットの能力を低く想定させることで、そのロボットとのインタラクションが継続しやすくなる可能性が指摘されている。例えば、佐田らの研究では、ゴミ箱ロボットが「ゴミを拾って！」などの自然言語を発話せずとも、ゴミに近づいたり、人の方に向かったりすることで、人がロボットに対して「助けてあげよう」という気持ちを示した [5]。西脇らの研究では、言葉足らずな発話をするロボットとインタラクションする場合には、人が自ら質問をしてロボットの足りない言葉を補うことで、会話を進めることに協力する傾向を示した [6]。つまり、人にロボットの能力が低いと想定させると、人がロボットに対して手助けをすることで、インタラクションが継続しやすくなる可能性がある。

2.2 非自然言語を発するロボットとのインタラクション

ロボットによる行動や言葉足らずな発話があれば、人とロボットとのインタラクションが可能であることが報告されてきたが、人がロボットに合わせてインタラクションの水準を下げる必要があった、ロボットの能力を限定しても、高い水準のインタラクションを維持することは可能だろうか。例えば、自然言語を発話しないロボットとの言語的なインタラクションは可能なのだろうか。ロボットが自然言語を発話しなくても、人にその言葉の意味を理解できていると感じさせることができるかを検討するために、清丸ら [2] は、実験参加者が自然言語を発話し、システムが自然言語を発話しない状態で、しりとりによる人とロボットのインタラクションをする実験を行った。システムは、自然言語の代わりに、自然言語の単語からアクセントと音韻数を維持して「ド」と「ラ」のみの音で読み上げたもの（これを非自然言語と呼ぶ）を発した。このしりとりにおいて、実験参加者には、直前にシステムが発した非自然言語による単語を予想し、それをシステムに聞き返し、システムからのポジティブな反応を得てから、それに続く単語を発話するという制約が設けられた。それを5回繰り返す（つまりしりとりを5往復する）か、5分が経過すると実験終了となった。結果として、実験参加者の5名中4名は5分以内に5往復のしりとりを完了した。また、しりとり終了後に、そのしりとりにおけるロボットとのコミュニケーションが成立した度合について7段階で評定した（1:完全に成立していなかった～7:完全に成立していた）。その結果、5名中4名は“5”以上と比較的高い評定値を回答し、自然言語を用いない言語的

なインタラクションが成立する可能性があることが示された。

大藤ら [3] は、清丸らの研究 [2] で使用されたりとりシステムをロボットに組み込み、非自然言語を発するロボットと人でしりとりを用いた言語的インタラクション実験を行った。ロボットが非自然言語を発する場合、聞き手側が相手の発する単語を予測することで言語的インタラクションが成立するという仮説を検証するために、この研究では、聞き手側の予測による言語理解に要する時間について調査した。そして、5秒以内でロボットに単語を聞き返しているときは人が予測に頼ってロボットが発する非自然言語を理解していて、5秒を超えたときは予測に頼った理解ができていないという基準が提案された。

これらの先行研究 [2][3] によって、非自然言語を発するロボットとの言語的インタラクションが成立する可能性が示されたが、実験では、さまざまな教示やロボットの振る舞いに関する工夫がされていた。例えば、システムが幼稚園児程度の語彙しか持たないと教示したことや、ロボットの発話した言葉の意味を確認しフィードバックを求めさせたこと、しりとりを言語的インタラクションの方法として用いたことなどがある。そのため、これらの要素のうち、何の影響で自然言語を用いない言語的インタラクションが成立したのかは明らかではない。本研究ではしりとりをするという要素に着目し、それが自然言語ではない音の意味の理解（つまり、何の単語を話しているかの理解）に与える影響について検討する。

2.3 非自然言語の理解に関する本研究の提案

清丸ら [2] の非自然言語を発するシステムとのしりとりによる言語的インタラクション実験において、“システムとのコミュニケーションが完全に成立していた”と報告した実験参加者がいた。しかし、その参加者は、実際には、しりとりシステムとのやり取りで9回中4回非自然言語を正しく理解していなかった（単語を間違えて理解していた）。このことから、人がロボットとの言語的インタラクションが成立したと感ずることは、非自然言語が示す単語について完全に理解したことではなく、理解しているという自信を持ったことかもしれない。

そこで本研究では、非自然言語を正しく理解しやすくなったり、理解しているという自信を持ちやすくなった要因が、言語的インタラクションとして用いたりとりという形式にあると考え、以下のような仮説を立てた。

仮説 1 しりとりをしている状態で非自然言語を聞くとその単語を正しく理解しやすくなる

仮説 2 しりとりをしている状態で非自然言語を聞くと

その単語を正しく理解しているという自信を持ちやすくなる

しりとりを行う際、自分が先に単語を発していれば、相手が発話する単語の頭文字が分かる。それにより相手が発するであろう単語の候補を絞り込める可能性がある。また、自身が発話した単語と、しりとりでよく現れる典型的な単語対のパターン（例：リンゴ、ゴリラなど）に基づき、次に発話されるであろう単語を予想できるかもしれない。これらの要因によって、非自然言語を正しく理解しやすくなる（単語を正しく推測する）ことと、正しく理解しているという自信を持ちやすくなるのではないかと考えた。また、後者の要因の影響を調べるため、本研究では、しりとりを行う条件だけではなく、ロボットが発する単語の頭文字を教示する条件も設定し、しりとり特有の効果があるかの検討も行った。

3 方法

3.1 目的・概要

本実験の目的は、2.3 節で示した仮説 1、仮説 2 を検証することである。実験参加者は、ロボットが発する非自然言語を発する動画を視聴した後、それが意味する単語は何かを回答し、その回答に対する自信度を報告した。その際、あらかじめ参加者をロボットとしりとりを行うことを想定させる群（以後、しりとり想定群と呼ぶ）、ロボットが発する単語の頭文字を知らせる群（以後、頭文字既知群と呼ぶ）、ロボットの発する単語について何も知らせない群（以後、無情報群と呼ぶ）に分け、各群の正答率と自信度を比較した。

3.2 参加者・動画・調査項目

20 歳以上の日本語を母語とする男女を対象として実験を行った。実験参加者はクラウドソーシングサービス²で募集され、Web アンケート作成システム³内で実験に参加した。この実験課題において、実験参加者は、まずロボットが発する非自然言語を発する動画を視聴し、その後二つの質問に回答した。一つ目の質問はロボットが発した非自然言語の単語が何であるか（自然言語の単語）、二つ目の質問は参加者自らの回答に対してどの程度の自信があるかであった。これを 12 種類の動画について繰り返した。

12 種類の動画は、映像として表示する画像と音声データを組み合わせることで作成した。動画で提示する画像として、1 枚のロボットの静止画を用意した（図 1）。ま



図 1: 非自然言語を発するロボットの動画

た、動画内の音声データは、以下のように作成した。まず、アクセントと音韻数の異なる 12 種類の自然言語の単語を用意した。それぞれの単語について、アクセントを維持したうえで各音を「ド」あるいは「ラ」の音に置き換えたものを、ロボットが発する音声とした。これらの音声データは、先行研究 [2][3] で使用されたものと同じであった。画像と音声を組み合わせる際、各動画の長さを 4 秒に統一するために、音声データの前に 1 秒、後に適切な空白時間（3 - 音声時間 [秒]）を設けた。

動画閲覧前に非自然言語の語彙に関して与えた情報の違いによって実験参加者を 3 群に分けた。それぞれの群の実験参加者が動画閲覧前に与えられた教示は以下の通りである。

しりとり想定群 このロボットとしりとりをしている状況を想定してください。そのやり取りの途中で、あなたは「あし」と言いました。

頭文字既知群 このロボットは今から「し」で始まる単語を言います。

無情報群 ロボットの発話する単語に関する教示なし

本実験は日本大学文理学部研究倫理委員会の承認を得て実施した。

3.3 実験方法

実験では、まず実験参加者の基本情報（年齢と性別）について質問した。また、動画が問題なく再生でき、参加者が音声を聞いていることを確認するために、動画再生確認の質問を行った。その後、実験課題において、参加者が回答する内容、および群ごとに動画視聴の状況についての教示を行った。参加者が教示を理解したら、動画視聴課題を行った。参加者は、動画の視聴後、動画内のロボットが発した音声は何という単語を意味していたかを推測し、それをひらがなで回答欄に入力した。その後、その回答に対する自信度を 7 段階（1: 全く自信がない～7: とても自信がある）で評定した。各設問について、回答した場合のみ次に進めるようにした。これを 12 回繰り返した。その際、12 種類の動画をランダムな順序で提示した。

² 「CrowdWorks」: <https://crowdworks.jp/>

³ 「SoSciSurvey」: <https://www.soscisurvey.de/>

3.4 統計解析

Web アンケート内の説明文や質問文に従って回答した実験参加者を解析対象とした。具体的には、以下の四つの回答基準を設け、どれか一つでも満たさない実験参加者を解析対象外とした。

基準 1 動画再生確認の質問に対して正しく回答している

基準 2 ひらがなでの指定に従って回答している

基準 3 12 回の単語の回答において一つ以上、広辞苑に記載されている単語を回答している

基準 4 指定した頭文字を用いて回答している
※しりとり想定群、頭文字既知群のみ

3.2 節で示した 3 条件を参加者間要因として、正答率と自信度それぞれについて、統計ソフトウェア⁴を用いて分散分析をした。

4 結果・考察

4.1 基本属性

実験参加者 392 名のうち 96 名が 3.4 節で示した基準を満たさなかった。解析対象であった実験参加者は合計 296 名であり、しりとり想定群は 93 名 (41.1±10.2 歳, 男性 43 名, 女性 50 名), 頭文字既知群は 102 名 (39.0±10.6 歳, 男性 51 名, 女性 51 名), 無情報群は 101 名 (39.0±10.3 歳, 男性 39 名, 女性 62 名) であった。

4.2 分析

正答率 (図 2 参照) に対して 1 要因分散分析を行った結果, グループの主効果が有意であった ($F(2,293) = 134.82, p < .001, \eta^2 = 0.479$). 多重比較の結果, 全ての群間に有意な差が見られた。しりとり想定群と無情報群 ($p < .001$), 頭文字既知群と無情報群 ($p < .001$), しりとり想定群と頭文字既知群 ($p = 0.005$). しりとり想定群と頭文字既知群は無情報群と比べて, しりとり想定群は頭文字既知群と比べて, 正答率が有意に高かった。

自信度 (図 3 参照) に対しても, グループの主効果が有意であった ($F(2,293) = 6.51, p = 0.002, \eta^2 = 0.043$). 多重比較の結果, しりとり想定群と無情報群に有意な差 ($p = 0.004$), 頭文字既知群と無情報群に有意な差が見られた ($p = 0.008$). しりとり想定群と頭文字既知群は無情報群と比べて, 回答に対する自信度が有意に高かった。しりとり想定群と頭文字既知群に有意な差は見られなかった ($p = 0.962$).

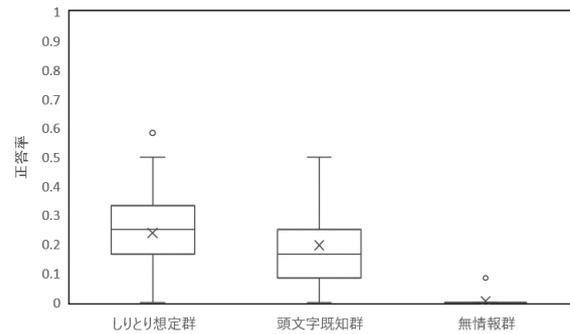


図 2: 正答率の箱ひげ図 (図における箱は四分位範囲, ひげの上端と下端は外れ値を除く最大値と最小値を示す。また, ドットは外れ値, ×印は平均値を示す)

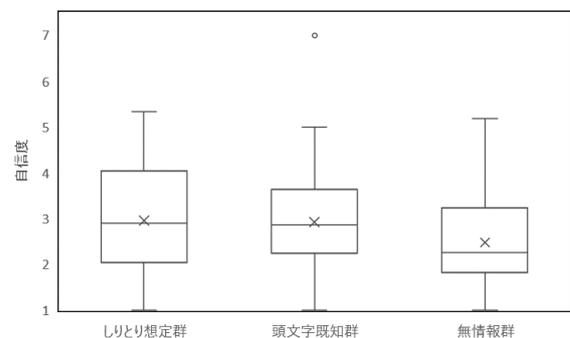


図 3: 自信度の箱ひげ図 (図における箱は四分位範囲, ひげの上端と下端は外れ値を除く最大値と最小値を示す。また, ドットは外れ値, ×印は平均値を示す)

4.3 考察

4.3.1 正答率について

頭文字既知群は無情報群と比べて, 正答率が有意に高いことから, ロボットが発する非自然言語が示す単語について, 頭文字を知っていることで, 非自然言語が意味する単語の候補が大幅に限定され, 非自然言語が意味する単語について正しく理解しやすくなったと考えられる。

また, しりとり想定群は頭文字既知群と比べて, 正答率が有意に高いと示されたことから, しりとりをしている状態で, 頭文字を知っていること以外の要素が影響して, 非自然言語が意味する単語について正しく理解しやすくなったと言える。ロボットが発話する非自然言語として用意した 12 種類の単語別に正答率を比較すると, 9 種類の単語 (例えば, 実験参加者が発話したとした単語が「あし」, その後ロボットが発話した単語が「しまうま」) は, しりとり想定群が頭文字既知群より高い正答率となり, 3 種類の単語 (例えば, 実験参加者が発話したとした単語が「うま」, その後ロボットが発

⁴ 「JASP」: <https://jasp-stats.org/>

話した単語が「まんが」)については,しりとり想定群が頭文字既知群より低い正答率となった.この結果は,通常のしりとりにて発話する頻度が高くなる単語と,そうではない単語があり,その頻度に基づいて,非自然言語が示す単語を回答した可能性が考えられる.しかしながら,通常のしりとりにおける単語の発話頻度については,本実験では検討していなかったため,この可能性については今後さらなる検討が必要である.

4.3.2 自信度について

頭文字既知群は無情報群と比べて,自信度が有意に高いことから,頭文字を知っているという要素の影響で,非自然言語が意味する単語について正しく理解したという自信を持ちやすくなったと言える.しりとり想定群は,無情報群と比べると自信度が高かったが,頭文字既知群との間には差が見られなかった.そのため,しりとりをしている状態で,頭文字を知っていること以外の要素が,回答の自信に影響しているとは言えない.

ただし,本実験では,参加者が単語回答後に,その自信度を問われるため,自身の回答以外に正解となる単語を想像させることを誘発したかもしれない.そのため,仮に2.3節で述べた,自身が発話した単語をもとに,次に発話されるであろう単語を予想していたとしても,それが自信度への影響として現れなかった可能性もある.これについては,今後詳細な検討が必要である.

4.3.3 今後の展望

先行研究 [2][3]において,非自然言語を発するロボットを用いた言語的インタラクションでは,システムが幼稚園児程度の語彙しか持たないと教示したことや,ロボットの発話した言葉の意味を確認しフィードバックを得ることなど,複数の工夫がなされている.そのため,本実験でしりとりという要素に絞って実験を設計したように,複数の工夫に対して,個別に実験を設計することで,言語的インタラクションが成立するために重要な要素について検討を行っていく予定である.

5 おわりに

本研究では,非自然言語を発するロボットと言語的インタラクションにおいて,しりとりをしていることで,非自然言語が表す単語の意味を正しく理解しやすくなったり,理解しているという自信を持ちやすくなるかを調査した.結果として,しりとりをしている状態の,頭文字を知っているという要素の影響で,非自然言語を正しく理解しやすくなったり,理解しているという自信を持ちやすくなったと言える.また,通常のしりと

りにて発話する単語の頻度に基づいて非自然言語が意味する単語の理解に影響がある可能性が示唆された.

参考文献

- [1] 福田聡子, 澤田志織, 川崎邦将, 奥岡耕平, 大澤正彦, 長田茂美, 今井倫太, 適応ギャップ理論を拡張したインタラクションデザインの提案, HAI シンポジウム, 2018.
- [2] 清丸寛一, 大澤正彦, 今井倫太, “予測的認知を用いた非自然言語による言語的コミュニケーション”, 第6回汎用人工知能研究会, 2017.
- [3] 大藤聖菜, 妹尾卓磨, 清丸寛一, 川崎邦将, 大澤正彦, 長田茂美, 今井倫太, “予測的認知と「間」の関係—言葉を話せないロボットによるしりとりを題材とした考察—”, HAI シンポジウム, 2017.
- [4] 小松孝徳, and 山田誠二. “適応ギャップがユーザのエージェントに対する印象変化に与える影響.” 人工知能学会論文誌 24.2 (2009): 232-240.
- [5] 佐田和也, 山際康貴, and 岡田美智男. “ゴミ箱ロボットにおける<弱さ>の表出について.” ヒューマンインタフェース学会論文誌 18.3 (2016): 219-228.
- [6] 西脇裕作, 板敷尚, and 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 21.1 (2019): 1-12.