

ポジティブフィードバックの回数が 「準」自然言語の単語の理解に対する自信度に与える影響

Impact of the number of positive feedback on confidence in guessing the words spoken in "semi"-natural language.

森下貴登^{1*} 福田聡子¹ 中島亮一² 大澤 正彦¹

Takato Morishita¹, Satoko Fukuda¹, Ryoichi Nakashima², Masahiko Osawa¹

¹ 日本大学

¹ Nihon University

² 京都大学

² Kyoto University

Abstract: 本研究では、「準」自然言語（自然言語の各音を「ド」と「ラ」に変換した音）を発するエージェントが行うポジティブフィードバックの回数が、ユーザの「準」自然言語の単語の理解に対する自信度に与える影響を調べた。実験参加者は、ロボットが「準」自然言語の単語を発する動画を視聴した後、それが意味する単語を二択から回答しそれに対する自信度を報告すると、ポジティブまたはネガティブなフィードバックが与えられる。結果、ポジティブフィードバックの回数が自信度に与える影響は小さいと考えられるが、回答時に2択の選択肢を与え、期待される割合よりも高い割合でポジティブフィードバックをすることで、自信度が向上する可能性を示した。

1 はじめに

近年、生成 AI を始めたとした自然言語を使用したエージェントが日常的に使用されるようになった。自然言語を用いるエージェントは人間との対等なコミュニケーションの実現が期待されるが、現代の技術レベルでも自然言語を使用したエージェントとのインタラクションには課題がある。小松らは [1]、元々期待していたエージェントの能力の大きさと実際にインタラクションすることで感じたエージェントの能力の大きさとの差を適応ギャップと定義し、前者と比べて後者が小さいとき（適応ギャップが負の値となる時）、ユーザはエージェン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。そのため、言語的インタラクションを破綻させずにエージェントとのコミュニケーションを成立させたときユーザに感じさせることは重要である。近年、「準」自然言語（自然言語の単語の各音を、アクセントを維持したまま「ド」と「ラ」に変換した音）を用いたシステムとの言語的インタラクションについて、ユーザとのコミュニケーションが成立した感じさせることができたという報告があった [2]。その際、システムとの言語的インタラクションを成立させるために、「しりとり」

という場面を想定させることや、内容の正誤に関わらずシステムがユーザの回答に対してポジティブなフィードバックを与えたことなどの様々な工夫をしていた。その結果、その研究では、実際にはシステム側が想定した内容とユーザ側が理解していた内容に相違があったにもかかわらず、ユーザにコミュニケーションが成立したと感じさせることができたと考えられる。本研究の目的は、先行研究で提案された「準」自然言語を用いたシステムとの言語的インタラクションについて、ユーザにインタラクションが成立したと感じさせた要因を具体的に明らかにすることである。

2 背景

2.1 「準」自然言語を用いたコミュニケーションの成立

清丸ら [2] は「準」自然言語を用いたエージェントとの言語コミュニケーションの場面を「しりとり」に限定することで成立する可能性を検証した。実験参加者は、事前に実験の注意点として以下の3点が説明された。

- エージェントはしりとり以外の通常のコミュニケーションも可能である。

*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: chta20070@g.nihon-u.ac.jp

- 実験参加者もまた、自由にコミュニケーションをとることが可能である。しりとり以外の質問や話しかけを行っても良い。必要がなければ、行わなくても良い。
- エージェントは幼稚園児程度の語彙しか持たない。ある文字から始まる語を持たない、または全て使い切ってしまった場合、答えられる語がないことを伝える。

これらの教示後、エージェントがしりとりに対応する「準」自然言語を発話してインタラクションが始まった。エージェントが「準」自然言語を用いて発話する単語（例えば「しりとり」という単語に対応する「ドラララ」という音声）を、実験参加者は音韻数やアクセントの情報から予想し、単語を聞き返した（例：「しりとり」と言ったの？）。それに対しエージェントは、実験参加者が聞き返した内容の正誤を問わず、ポジティブなフィードバック（例：エージェントが発話する音声「ドラドラ!!」）を与えた。エージェントからのポジティブなフィードバックを得てから、実験参加者はそれに続く単語を発話した（実験参加者がエージェントの発話の意味を「しりとり」と理解したのであれば、それに続く「りんご」などを発話する）。エージェントと実験参加者がお互いに1単語ずつを発話することを1往復とし、それを最低でも5回（つまり、しりとりを5往復）以上繰り返し、最大で5分が経過すると実験終了となった。実験終了後に、実験参加者は「準」自然言語を用いたエージェントとのコミュニケーションが成立した度合いについて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分以内に5往復が完了し、5名中4名が成立した度合いを5以上で評定したことから、清丸らは「準」自然言語を用いたエージェントとのコミュニケーションが成立する可能性があると主張した。また、5名の中で1名は成立度合いを7で評定した。しかし実際その参加者とのやりとりでは、9往復の「しりとり」のうち4回は実験参加者の発話の解釈とエージェントが発話した内容に相違があった（間違った回答をしていた）。このことから、エージェントが発した言葉を実験参加者が誤って推測したとしても、それが正解であることを示唆するフィードバックを与えれば、単語の推測に対する自信が向上し、結果的に実験参加者に「準」自然言語を用いたエージェントとのコミュニケーションができたと感じられることができる可能性がある。

2.2 先行研究での課題と本研究の調査

清丸らは、「準」自然言語を用いたエージェントとのコミュニケーションが成立する可能性を主張したが、彼

らの実験では言語的インタラクションを成立させるために様々な工夫や教示があった。言語的インタラクションの場を「しりとり」に限定したことや、エージェントは幼稚園児程度の語彙しか持たないことを教示することで、ユーザの「準」自然言語の単語の推測に手がかりを与えた。加えて、エージェントの発した言葉の意味を確認させ、確認した単語の正誤を問わずポジティブフィードバックを与えた。そのため、これらの工夫の中で、何がコミュニケーションが成立したと実験参加者に感じさせた要因かは明らかではない。そこで本研究では、清丸らの研究において、エージェントが実験参加者の回答に対して常にポジティブフィードバックを与えたことに着目し、ポジティブフィードバックが実験参加者の単語の理解に対する自信を向上させるかどうかについて検討する。エージェントが「準」自然言語の単語を発話した際、実験参加者は与えられた手がかりを元に単語を推測するが、その推測した単語の正誤を確認する手段はエージェントからのフィードバックしか存在しない。よって、本研究では、ポジティブなフィードバックを与えることで単語の理解に対する自信が向上しエージェントが発した単語をわかった気になり、エージェントとのコミュニケーションが成立する可能性を調査する。また、「準」自然言語に対して自然言語的な意味を読み取るということは、システムを擬人化しているからこそ生じる可能性があるため、ポジティブフィードバックの影響で「わかった気にさせた」場合のエージェントに対する印象（擬人化）についても調査する。

3 実験方法

3.1 目的

本研究の目的は、ポジティブフィードバックの回数によって、「準」自然言語の単語の理解に対する自信を向上させる度合いが異なるかを検討することである。実験では、実験参加者は、エージェントが発する「準」自然言語の単語を与えられた2択の選択肢から回答する。その際、2択の選択肢は、同じ音韻数かつ同じアクセントの単語、つまり「準」自然言語で発話すると全く同じ音声となる単語とした。これを13回繰り返すが、毎回の回答に対し、その都度フィードバックを与える。実験参加者を13回中13回のポジティブフィードバックを与える群（以後、13回群）、11回のポジティブフィードバックを与える群（以後、11回群）、9回のポジティブフィードバックを与える群（以後、9回群）に分け、それぞれの群における1回目の自信度評定値（エージェントからのフィードバックを0回受けている状態）と13回目の自信度評定値（エージェントからのフィードバックを12回受けている状態）を比較した。もし、ポジティブフィー



図 1: 実験で使用された「準」自然言語エージェントの動画画面

ドバックの回数が参加者の単語に対する自信度評定値を向上させることに有効であれば、13 回群、11 回群、9 回群の順に自信度評定値が高くなると予想される。また、自信度評定が向上した場合、そのエージェントに対しての印象（擬人化）も 13 回群、11 回群、9 回群の順に高くなるのが予想される。

3.2 実験設計

アンケートは web アンケート作成システムを用いて作成した。実験で使用した動画は、映像として表示する静止画（図 1）と音声データを組み合わせることで作成した。音声データは以下の手順で作成した。まず、清丸ら [2] が用いた音声のもととなった単語の中から、アクセントと音韻数の異なる 13 種類の自然言語の単語を選定した。次に選定したそれぞれの単語と、アクセント、音韻数、頭文字が同じ単語を用意した（表 1 に記載）。また、清丸ら [2] がしりとり以外のコミュニケーションのために、喜び、悲しみ、怒りの感情について記録した音声データから、喜びと悲しみに該当する音声データを用いた。前者はポジティブフィードバック、後者はネガティブフィードバックのために使用した。

3.3 方法

まず、実験参加者を 20 歳以上の日本語を母語とする男女をクラウドソーシングサービスで募集し、343 名（男性：128 名、女性：210 名、その他：6 名、年齢 39.8±10.7 歳）が Web アンケートを用いた実験に参加した。各実験参加者は、自分の持つ PC で web アンケートページにアクセスし実験を行った。その際、実験参加者は、ポジティブフィードバックを 13 回、11 回、9 回受ける群にランダムに振り分けられた。各群の参加者は、実験の

表 1: 実験で使用した 26 種類の「準」自然言語が表す単語

きた	きみ
みず	みぎ
もり	もも
くに	くち
むし	むら
まんが	まふゆ
ぼんだ	ぱんつ
うさぎ	うがい
ねずみ	ねいろ
ごりら	ごるふ
にわとり	にんにく
しまうま	しあわせ
あめりか	あきかん

前にアンケートで使用する動画の音声の問題なく聞こえることを確認するために、確認用動画で再生される音声（自然言語で「おにごっこ」という単語を提示）を聞き、それを回答欄に入力するように求められた。各群の参加者には、エージェントの発する「準」自然言語の単語に対して、文字数とアクセントを手がかりに 2 つの選択肢のうちどちらの単語を言っていたかを選ぶよう教示した。加えて、参加者の回答に対してエージェントが 2 種類の反応を行うことを教示し、ネガティブな反応とポジティブな反応について動画の再生ボタンを自分でクリックし、その動画を視聴しそれぞれの反応を確認した。動画の視聴後「次へ」ボタンを押し、次のページで動画の自動再生が正常に作動することを確認した。各群の参加者は、それらの情報を確認したうえで、実験を開始し動画を再生した。動画は一度再生すると非表示となり、2 択の選択肢が表示された。参加者は、動画で発せられた音声は何の単語を意味していたかを 2 択の選択肢から選択した。各群の参加者は「次へ」のボタンをクリックし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。そして「次へ」ボタンをクリックし、ネガティブまたはポジティブなフィードバックを行う動画を与えた。これを 1 回の試行として、13 種類の動画に対して、ランダムな順で 13 試行行った。13 試行終了後に、エージェントに対する擬人化の印象評価を実施した。実験参加者には、「実験中に見ていただいたロボットに対する印象について評価してください」と質問をし、実験で使用したエージェントについて 5 段階のリッカート尺度で回答した。エージェントに対する印象評価には、Godspeed Questionnaire[3] の「ANTHROPOMORPHISM（擬人化）」の要素である、偽物のような/自然な、機械的/人間的、意識を持たない/意識を持っている、人工的/生物学的、

ぎこちない動き/洗練された動きを用いた。またダミーとして、「LIKEABILITY（好ましき）」の要素である、嫌い/好き、親しみにくい/親しみやすい、不親切な/親切な、不愉快な/愉快な/, ひどい/良いを用いた。

3.4 データ分析

表 2 に示す基準をすべて満たす参加者を、実験の教示に従って回答した実験参加者としてデータ分析の対象とした。その結果、13 回群 86 名、11 回群 73 名、9 回群 84 名が分析対象となった。各群における 1 回目の回答に対する自信度評定値（フィードバックをまだ受けていない状態での自信度）と、13 回目の回答に対する自信度評定値（フィードバックを 12 回受けた後の自信度）を比較する。群を被験者間要因、フィードバック要因（1 回目と 13 回目の自信度評定値）を被験者内要因とし 2 要因分散分析を行った。また、擬人化傾向の値は群を被験者間要因とする分散分析を行った。

表 2: 分析対象とする参加者を選ぶ基準。全てを満たす参加者のデータを分析対象とする。

基準 1	年齢の質問に対して半角で 20 以上の数字を入力している
基準 2	動画再生確認の質問に対して正しく回答している
基準 3	自動再生確認の質問に対して「はい」を選択している
基準 4	アンケート項目まで解答している

3.5 分析結果

自信度評定値の結果を図 2 に示す。2 要因分散分析を行った結果、群（13 回群、11 回群、9 回群）の主効果は見られなかった ($F(2, 240)=0.681, p=0.507$)。また、フィードバック要因の主効果は有意であった ($F(1, 240) = 30.307, p < 0.001$)。Holm 法による事後比較の結果、13 回目の回答に対する自信度評定値が 1 回目の回答に対する自信度評定値よりも有意に高かった ($p < 0.001$)。また、フィードバック要因*群の交互作用は有意ではなかった ($F=2.889, p=0.057$)。

擬人化の印象評価値は、13 回群:2.50、11 回群:2.83、9 回群:2.61 となった。分散分析の結果、群間の主効果が見られた ($F(2, 240)=3.194, p=0.043$)。事後比較により、11 回群の数値が 13 回群よりも有意に高かった ($p=0.040$)。

群間の自信度評定値

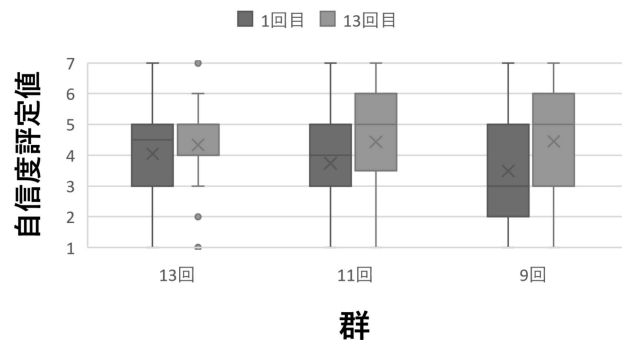


図 2: 自信度評定値の結果。箱中の線は中央値、箱中の×印は平均値を示している。ひげの外側にプロットされた点は外れ値を示しており、「第 3 四分位数+四分位範囲 × 1.5」以上のデータ、「第 1 四分位数-四分位範囲 × 1.5」以下のデータと定義する。

4 考察

自信度評定値に対して、群の主効果が見られなく、フィードバック要因における主効果が見られたことから、1 回目の自信度評定値から 13 回目の自信度評定値は何らかの影響で向上することは示されたが、ポジティブフィードバックの回数による影響があるかは確認できなかった。ここでいくつかの可能性がある。1 つは、ポジティブフィードバックの回数が、自信度に影響をしないという可能性である。例えば、実験参加者が実験の教示を信じておらず、実験参加者の回答に関係なくフィードバックを行っていると考えていた場合には、ポジティブフィードバックの回数が自信度に影響しない。また、複合的な要因によって、結果的に条件間に差が見られなかった可能性もある。例えば、ポジティブフィードバックが多ければ多いほど、自信度が向上する効果もありつつ、同時にフィードバックの信頼性を疑われやすくなるとしたら、結果的に条件間の差が見られないことが考えられる。

また、フィードバックの主効果が有意だった。これは、13 回のフィードバックのうちポジティブなフィードバックが 9 回、11 回、13 回のいずれの回数でも、自信度を向上させる効果があったといえる。ここで、本実験においては 2 択問題を出題しているため 5 割の確率で正解することとなるが、本実験ではいずれの条件でも 6 割以上の確率でポジティブフィードバックがされた。つまり、期待される割合よりも高い割合でポジティブフィードバックをすることが、自信度を向上させた要因と考えられる。

印象（擬人化）に対する分析結果より、11 回群と 13

回群を比較した場合に 11 回群が有意に高かったことが確認された。11 回群が有意に高かった要因は適度にネガティブなフィードバックを与えたことが考えられる。13 回群に関しては、全て同じ反応をするエージェントのフィードバックに正当性や信頼性が減少し、フィードバックの価値が下がったことが考えられる。すべて同じ反応をするエージェントに対して雑念や疑念が生じた可能性が考えられ、反対に、適度にネガティブなフィードバックを行ったことでフィードバックの信頼性が向上し、擬人化の数値に差が出たと考えられる。つまり、全てポジティブなフィードバック（同じ）反応をするエージェントよりも適度にネガティブなフィードバックを返すエージェントが印象（擬人化）を高める可能性がある。

5 おわりに

本研究では、ポジティブフィードバックの回数が準自然言語の単語の理解に対する自信度に与える影響について調査した。また、エージェントに対する印象（擬人化）についても調査した。もし、ポジティブフィードバックの回数が自信度と印象（擬人化）の向上に有効であれば 13 回群が最も高くなると予想した。結果、自信度の向上に関しては、1 回目の自信度評定値から 13 回目の自信度評定値では何らかの影響で自信度評定値が向上したが、ポジティブフィードバックの回数による影響は不明であった。一方で、2 択の選択肢から 5 割の確率で正解するため、全体の 6 割以上のをポジティブフィードバックを与えたことで自信度が向上する可能性を示した。また、エージェントに対する印象（擬人化）についても調査した結果、適度にネガティブなフィードバックを行うエージェントは数値が高くなる傾向を見せた。今後は、「適度」な回数を調査していくとともに、ポジティブフィードバックの回数が与える影響についても実験設計を改め、調査する必要がある。

参考文献

- [1] 小松孝徳, 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.
- [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] Christoph Bartneck, Elizabeth A. Croft, Dana Kulic, Susana Zoghbi. “Measurement Instruments for the Anthropomorphism, A Perceived Intelligence, and Perceived Safety of Robots.”,