

条件推論における AI の発話の解釈と推論抑制： ポライトネス理論に基づく検討

Interpretation and inference suppression of AI utterances in conditional inference: a study based on politeness theory

松本 和紀[†], 高橋 達二[†], 中村 紘子[‡]

Kazunori Matsumoto, Tatsuji Takahashi, Hiroko Nakamura

[†] 東京電機大学, [‡] 日本学術振興会

Tokyo Denki University, Japan Society for the Promotion of Science

24rmd39@ms.dendai.ac.jp

概要

本研究では、条件文の発話者の性格や追加情報を提供する主体（人間または AI）が、条件推論の抑制に与える影響をポライトネス理論に基づき検討した。先行研究では、気難しい相手に対する曖昧な発言は、訂正を意図したものと解釈されやすく、推論が抑制されることが示されている。本研究では、AI による曖昧な発言が推論に及ぼす影響を実験的に検討し、AI による曖昧な発言は相手の性格によらず、人間の場合よりも推論の抑制がされにくいことを示した。

キーワード：条件推論, ポライトネス理論, AI

1. はじめに

人のコミュニケーションにおいては曖昧な表現が用いられることがあり、曖昧な表現の解釈や推論は文脈の影響を受ける。近年では生成 AI などの普及、発展により、AI とコミュニケーションをする機会も増えている。そのため、AI による情報を人がどのように解釈し、推論するかを明らかにすることは重要だと考えられる。そこで、本研究では AI との対話に基づく条件推論過程を人間同士の対話に基づく条件推論過程と比較し、AI とのコミュニケーションに対する解釈過程を明らかにする。

2. 条件推論

「もし p ならば q 」といった形式の文を条件文といい、 p を前件、 q を後件と呼ぶ。条件三段論法とは、大前提となる条件文と小前提から結論を導く推論のことである。大前提となる条件文を「もし p ならば q 」としたとき、小前提 p から結論 q を導く推論を MP 推論という (Manktelow, 2012 服部・山沢 2015)。

Byrne は、大前提である条件文「もし p_1 ならば q 」に対し、追加条件文「もし p_2 ならば q 」が提示されると条件推論が抑制される場合があることを示した (Byrne, 1989)。例えば、「もしホワイト社のクリームを使えば、お菓子は美味しくなる」という条件文に対

し、「もしブラウン社のクリームを使えば、お菓子は美味しくなる」という追加条件文が提示された場合、追加条件文が訂正「ホワイト社のクリームを使っても美味しくなく、ブラウン社のクリームを使えば美味しくなる」と解釈された場合、「ホワイト社のクリームを使えば、お菓子は美味しくなる」という MP 推論の抑制が生じる。一方、追加条件文が単なる情報の追加「ホワイト社のクリームを使っても、ブラウン社のクリームを使ってもお菓子は美味しくなる」と解釈されると、MP 推論の抑制は生じない。

3. ポライトネス理論

Demeure らは、この曖昧な追加条件文の解釈が、ポライトネス理論で説明される社会的要因に影響されると主張した (Demeure et al., 2009)。ポライトネス理論とは、円滑なコミュニケーションを進めるための理論で、Brown & Levinson によって提唱された (Brown & Levinson, 1987)。この理論によれば、人々は対人コミュニケーションにおいて 2 種類のフェイスを持つ：ポジティブフェイス（他者に受け入れられたいという欲求）とネガティブフェイス（他者に邪魔されたくないという欲求）である。コミュニケーションの際、話者は聞き手のフェイスを脅かす可能性（フェイス脅威度）を考慮し、適切なポライトネス・ストラテジーを選択する。フェイス脅威度には社会的距離、社会的地位、要求量が影響を与えるとされる。

Demeure らは条件文の解釈や推論の抑制に、条件文の話者の性格や話者間の関係性が影響を与えることをポライトネス理論に基づき示した。気難しい性格の話者や関係性が悪い話者の発した条件文に対して、別の話者が曖昧な追加条件文を発言した場合、その発言はより訂正を意図したものとして解釈される傾向があり、MP 推論の抑制が生じやすかった。これは、気難しい性格の話者に対して訂正を行うことはフェイス脅威度が高く、より間接的な表現（つまり曖昧な条件文）

が用いられると解釈されるためだと考えられている。また、日本人参加者においても関係性による推論の抑制への影響がみられることが小倉他 (2023) や松本他 (2023) によって示されている。

4. 人間と AI の比較

近年では、AI の普及や発展により、人々が AI とコミュニケーションを行う機会が増えている。しかし、AI は誤った出力を行うこともある。その中で AI による情報を人間がどのように扱うか知ることが重要だと考えられる。このような状況下で、AI システムのポライトネスの利用についても知見が蓄積されてきている。例えば、音声アシスタントが間接的な表現や敬語を用いたり、ユーザーへの共感を示したりすることで、ユーザーの信頼感や受容性が高まることが報告されている (Ribino, 2023)。人間同士のコミュニケーションにおけるポライトネス・ストラテジーに関する Demeure らの研究では、人々は相手の性格や関係性に応じて適切なポライトネス・ストラテジーを選択するという信念を持っており、この信念が対人コミュニケーション場面の解釈や推論に影響を与えることが示されている。このような解釈や推論過程が人間と AI の対話の場合にも生じるかは不明である。そこで本研究では、人々が AI とのコミュニケーションにおいて、人間同士のコミュニケーションで見られるようなポライトネス・ストラテジーに関する信念を適用するかどうかを検討する。具体的には、AI の発する追加条件文に基づく推論が、人間が発する場合と同様のパターンを示すかどうかを明らかにする。また、AI による情報の捉え方には個人の差が生じると考えられる。今回の研究では AI がどのようなコミュニケーションをするかという信念に、擬人化傾向や AI への信頼感が関わるかを擬人化傾向尺度日本語版 Japanese version of the Individual Differences in Anthropomorphism Questionnaire : IDAQ-J (中村他, 2024) や対 AI 信頼感尺度 (片瀬, 2021) を用いて検討する。

5. 実験

5.1 参加者と手続き

クラウドソーシングサイト (CrowdWorks) を用いて、300 名の参加者を 110 円の謝金で募集した。分析に用いるデータは、途中離脱、教示操作チェック違反、回答時間が参加者全体の平均の ± 2.0 SD を超えるデータを除いた 232 名 (男性: 114 名, 女性: 117 名, その他: 1 名, 年齢範囲: 19~77 歳, 平均年齢: 37.8 歳, 標準偏差: 9.9 歳) のものとした。実験はオ

ンラインアンケート調査ツール (Qualtrics) を用いて、Web ブラウザ上で実施した。

5.2 実験計画と実験材料

実験計画は 2 (第 1 話者の性格: 気さく・気難しい) * 2 (第 2 話者: 人間・AI) の 2 要因参加者間計画とした。実験シナリオは Demeure et al. (2009) を参考に作成し、製菓会社で働く 2 人の会話についてのものとした。条件文の発話者の性格 (気さく・気難しい) と、追加条件文の発話者 (人間・AI) を操作した、4 パターンのシナリオを作成した (表 1)。

表 1 シナリオの例

性格と説明	X さんは気さくで公平な人柄です。 X さんは、相手を尊重し、他の人の意見によく耳を傾けます。 Y さんは X さんの同僚で、過去 5 年分の商品に関するデータや、社員の性格や行動についてのデータを持っています。 X さんと同僚の Y さんが、新商品について次のように会話しています。
条件文	社員の X さんが 「もし私たちがホワイト社のクリームを使えば、お菓子はおいしくなるでしょう」と言いました。
追加条件文	同僚の Y さんが 「もし私たちがブラウン社のクリームを使えば、お菓子はおいしくなるでしょう」と返しました。

参加者には次の課題を提示した

1. MP 推論課題: シナリオをもとに「ホワイト社のクリームを使った」場合、「お菓子はおいしくなる」と思うかを、「1. 結論は間違っていると思う」から「5. 結論は正しいと思う」までの 5 段階で評価するよう求めた。
2. 確率判断課題: 条件文「もしホワイト社のクリームを使えば、お菓子はおいしくなるでしょう」と追加条件文「もしブラウン社のクリームを使えば、お菓子はおいしくなるでしょう」が真だと思われる確率を、それぞれ 7 段階で評価するよう求めた。
3. 対 AI 信頼感尺度: AI の社会有益性への信頼感・AI の忠実性への不信感の質問を 5 問ずつ提示し、それぞれ 7 段階で評価するよう求めた。
4. 機械に関する擬人化尺度の測定課題: IDAQ-J の機械に関する質問 5 問を提示し、それぞれ 11 段階で評価するよう求めた。

6. 結果

条件文の発話者の性格および追加条件文の発話者ごとの結果を表2に示す。条件文の発話者の性格（気さ

表2 性格および発話者ごとの結果

personality	speaker	MP	$P(\text{if } p_1 \text{ then } q)$	$P(\text{if } p_2 \text{ then } q)$
		$M(SD)$	$M(SD)$	$M(SD)$
friendly	human	3.67 (0.785)	4.73 (1.12)	5.40 (0.955)
	AI	3.78 (0.658)	5.18 (0.964)	4.91 (0.967)
touchy	human	2.84 (0.934)	3.70 (1.05)	5.38 (0.778)
	AI	3.31 (0.814)	4.52 (1.25)	4.86 (0.870)

く・気難しい）と追加条件文の発話者（人間・AI）を独立変数とし、(1) MP 推論の評価、(2) $P(\text{if } p_1 \text{ then } q)$ の評価、(3) $P(\text{if } p_2 \text{ then } q)$ の評価をそれぞれ従属変数とする2要因分散分析を実施した。

(1) MP 推論の評価：性格の主効果 ($F(1, 228) = 37.10, p < .001$) と話者の主効果 ($F(1, 228) = 8.427, p < .001$) が有意であった。条件文の発話者の性格が気さくな場合、気難しい場合よりもMP推論が導かれやすく、追加条件文の発話者が人間の場合、AIの場合よりもMP推論が導かれやすいことが示された。

(2) $P(\text{if } p_1 \text{ then } q)$ の評価：性格の主効果 ($F(1, 228) = 33.34, p < .001$) と話者の主効果 ($F(1, 228) = 19.79, p < .001$) が有意であった。条件文の発話者の性格が気さくな場合、気難しい場合よりも条件文の確からしさを高く評価し、追加条件文の発話者がAIの場合、人間の場合よりも条件文の確からしさを高く評価することが示された。

(3) $P(\text{if } p_2 \text{ then } q)$ の評価：話者の主効果 ($F(1, 228) = 18.76, p < .001$) が有意であった。追加条件文の発話者が人間の場合、AIの発言よりも追加条件文の確からしさを高く評価することが示された。

各尺度が推論の抑制や条件文の確からしさの評価に与える影響を重回帰分析によって分析する。

擬人化傾向尺度の影響：発話者の性格、追加条件文の発話者、擬人化尺度得点を独立変数とし、(1) MP 推論の評価、(2) $P(\text{if } p_1 \text{ then } q)$ の評価、(3) $P(\text{if } p_2 \text{ then } q)$ の評価をそれぞれ従属変数とする重回帰分析を行った。

(1) MP 推論の評価：性格と擬人化傾向の交互作用が有意であった。単純傾斜分析の結果、擬人化傾向が低い場合、発話者が気難しいとMP推論が抑制されることが示された ($b = -1.220, p < .001$)。擬人化傾向が高い場合には発話者の性格による推論への影響はみられなかった ($b = -0.224, n.s.$)。

(2) $P(\text{if } p_1 \text{ then } q)$ の評価：性格の主効果 ($b = -0.904, p < .001$) と話者の主効果 ($b = 0.405, p < .05$) が有意であり、擬人化傾向の条件文の確率への影響はみられなかった。

(3) $P(\text{if } p_2 \text{ then } q)$ の評価：性格と話者、擬人化傾向の交互作用が有意となった。単純傾斜分析では有意差はみられなかった。

AIの社会有益性への信頼感の影響：発話者の性格、追加条件文の発話者、実験参加者のAIの社会有益性への信頼感を独立変数とし、(1) MP 推論の評価、(2) $P(\text{if } p_1 \text{ then } q)$ の評価、(3) $P(\text{if } p_2 \text{ then } q)$ の評価をそれぞれ従属変数とする重回帰分析を行った。

(1) MP 推論の評価：性格と有益性への信頼感の交互作用と話者と有益性への信頼感の交互作用が有意であった。単純傾斜分析の結果、話者と有益性の交互作用では有意差はみられなかった。性格と有益性の交互作用では、有益性を高く評価している場合、性格が気難しいと、推論の抑制が生じやすかった ($b = -1.344, p < .001$)。有益性の評価が低い場合、性格による推論の影響はみられなかった ($b = -0.390, n.s.$)。

(2) $P(\text{if } p_1 \text{ then } q)$ の評価：性格と有益性の交互作用が有意であった。単純傾斜分析の結果、有益性の評価が低い場合 ($b = -0.644, p < .05$) も高い場合 ($b = -1.459, p < .001$) も性格が気難しいほうが条件文の確からしさを低く評価することが示された。

(3) $P(\text{if } p_2 \text{ then } q)$ の評価：話者の主効果 ($b = -0.520, p < .01$) と有益性への信頼感の主効果 ($b = 0.414, p < .05$) が有意であった。有益性への信頼感が低い場合、高い場合よりも追加条件文の確からしさを低く評価することが示された。

AIの忠実性への不信感の影響：発話者の性格、追加条件文の発話者、実験参加者のAIの忠実性への不信感を独立変数とし、(1) MP 推論の評価、(2) $P(\text{if } p_1 \text{ then } q)$ の評価、(3) $P(\text{if } p_2 \text{ then } q)$ の評価をそれぞれ従属変数とする重回帰分析を行った。

(1) MP 推論の評価：話者の主効果 ($b = -0.845, p < .001$) が有意であった。AIの忠実性への不信感による推論への影響はみられなかった。

(2) $P(\text{if } p_1 \text{ then } q)$ の評価：性格の主効果 ($b = -1.038, p < .001$) と話者の主効果 ($b = 0.458, p < .05$) が有意であった。AIの忠実性への不信感の条件文の確からしさへの影響はみられなかった。

(3) $P(\text{if } p_2 \text{ then } q)$ の評価：話者の主効果 ($b =$

-0.495, $p < .01$) が有意であった。AI の忠実性への不信感の追加条件文の確からしさへの影響はみられなかった。

7. 考察

本実験の結果、性格による MP 推論や条件文の確からしさへの影響がみられ、Demeure らの先行研究の知見と部分的に一致する結果となった。先行研究と同様に条件文の発話者の性格が気さくな場合、気難しい場合に比べて MP 推論の抑制が生じにくいことが示され、気難しい相手に対する曖昧な発言がより訂正として解釈されやすいことが示唆された。発話者の違いに関しては MP 推論や条件文の確からしさ、追加条件文の確からしさへの影響もみられた。AI による発言の場合、人間による発言と比べて、追加条件文の確からしさが低いことで、条件文の確からしさが低下しづらく、MP 推論の抑制効果が低くなるのではないかと考えられる。この結果は、人々が AI の発言を人間の発言とは異なる方法で解釈している可能性を示唆している。

個人差要因に関して、擬人化傾向の影響は MP 推論でみられた。擬人化傾向が低い場合、発話者の性格が気難しい場合に MP 推論が抑制されることが示された。擬人化傾向が低い参加者は、発話者の性格（気難しいか気さくか）に敏感に反応し、Demeure らの研究で示されたような人間同士のコミュニケーションにおけるポライトネス解釈をより強く適用している可能性がある。社会有益性への信頼感の影響は MP 推論、条件文の確からしさ、追加条件文の確からしさに影響することが示された。AI の社会有益性を高く評価している参加者において、発話者の性格が気難しい場合に MP 推論の抑制が生じやすく、また、追加条件文の確からしさを高く評価することが示された。AI への信頼が高い人ほど、第 1 条件文の発話者の性格的特徴により敏感に反応し、人間か AI かに関わらず他者の発言をより信頼性の高いものとして評価する傾向があると考えられる。AI の忠実性への不信感の影響は MP 推論、条件文の確からしさ、追加条件文の確からしさのいずれでもみられなかった。対 AI 信頼感尺度は対人信頼感尺度をもとにしているため、人間と AI の間に差が出にくく、影響がみられなかったと考えられる。

8. おわりに

本研究では、ポライトネス理論に基づき、条件文の発話者の性格と追加条件文の発話者（人間か AI か）が条件推論の抑制に及ぼす影響を検討した。また、個人差要因として擬人化傾向、AI の社会有益性への信

頼感・忠実性への不信感の影響も調査した。その結果、発話者が気難しい場合に MP 推論の抑制が生じやすいことが確認された。一方、AI が追加条件文の発話者の場合、人間の場合と比べて推論の抑制効果が低くなることが明らかになった。個人差要因の分析からは、擬人化傾向が低い参加者や AI の社会有益性への信頼感が高い参加者は、発話者の性格による影響を強く受ける傾向が見られた。これらの結果は、人間と AI のコミュニケーションに対する推論プロセスが、個人の認知的特性や態度に影響される可能性を示している。今後の研究では、AI に関する知識や経験が推論に与える影響、条件文の確からしさの程度による推論抑制の変化の検討が必要だといえる。

文 献

- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (No. 4). Cambridge university press.
- Byrne, R. M., (1989). Suppressing valid inferences with conditionals. *Cognition*, 31(1), pp. 61–83. [https://doi.org/10.1016/0010-0277\(89\)90018-8](https://doi.org/10.1016/0010-0277(89)90018-8)
- Demeure, V., Bonnefon, J. F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), pp. 260–266. <https://doi.org/10.1037/a0013799>
- 片瀬 拓也. (2021). 人工知能 (AI) に対する信頼感尺度の作成と信頼性・妥当性の検討. 日本教育工学会研究報告集 2021(3) pp. 172–179. <https://doi.org/10.15077/jsetstudy.2021.3.172>
- Manktelow, K. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgement and decision making*. (邦訳: 服部 雅史, 山 祐嗣 (監訳) (2015). 思考と推論: 理性・判断・意思決定の心理学 北大路書房 pp. 69–91.)
- 松本 和紀・小倉 那央・高橋 達二・中村 紘子. (2023). 敬語表現が条件推論の抑制に及ぼす影響: ポライトネス理論に基づく検討 人工知能学会全国大会論文集, 3Xin4-22. <https://doi.org/10.11517/pjsai.JSAI2023.0.3Xin422>
- 中村 紘子・松尾 朗子・眞嶋 良全. (2024). 擬人化傾向尺度日本語版の作成 心理学研究. Advance online publication. <https://doi.org/10.4992/jjpsy.95.22217>
- 小倉 那央・高橋 達二・中村 紘子. (2023). 発話者間の関係性が条件推論の抑制に及ぼす影響: ポライトネス理論に基づく検討 情報処理学会第 85 回全国大会論文集, 7Q-02.
- Ribino, P. (2023). The role of politeness in human-machine interactions: a systematic literature review and future perspectives. *Artificial Intelligence Review*, 56(1), pp. 445–482. <https://doi.org/10.1007/s10462-023-10540-1>