

大規模汎用言語モデルを用いた雑談対話システムの対人関係性に基づく発話制御の検討

山崎 天¹ 川本 稔己^{1,2} 吉川 克正¹ 佐藤 敏紀¹

¹ LINE 株式会社 ² 東京工業大学

{takato.yamazaki, toshiki.kawamoto}@linecorp.com

{katsumasa.yoshikawa, toshinori.sato}@linecorp.com

概要

本研究では、大規模汎用言語モデルを用いた雑談対話システムにおける、対人関係性に基づいた発話制御の実現を目指す。大規模汎用言語モデルは膨大な学習データやニューラルネットワークのパラメタ数によって、知識や文章表現の面で多様な出力できるが、その多様性ゆえに「敬語からタメ口になる」や「攻撃的になる」など、意図や場面に関係なく様々なスタイルで発話を生成してしまい、制御が難しい。本稿では、ポライトネス理論を参考に、相手との対人関係性を示す情報をプロンプトに付与することで、ポライトネス・ストラテジーを考慮した応答を生成できるか検証する。また実験結果をもとに、「大規模汎用言語モデルの発話生成能力を活かしながらポライトネス・ストラテジーを考慮する対話システム」を実現する上で解決すべき課題について議論をする。

1 序論

近年、ニューラルネットワーク (NN) による言語モデルは様々なタスクで好成績を収め、流暢な文章生成が可能であることが示されている。中でも GPT-3[1] などの大規模汎用言語モデルはウェブデータなどを含む雑多なコーパスを事前学習で用いることで、In-context Learning と呼ばれる、Few-Shot で特定の言語タスクの処理能力を獲得できる手法が提案されている。また、大規模なコーパスと NN のパラメタ数のおかげで、大規模汎用言語モデルは多様な知識や文章表現がモデルに学習されており、自然かつ創造性豊かな文章生成が必要な対話や小説生成等のタスクで高い性能を発揮する [1, 2]。

しかし、対話においては我々の実装 [3] に課題が見つかっており、誤り分析によると評価の低下に大きく影響したのは「発話スタイルの変化」であった。具体的には図 1 の例のように、チャットボットという役割や初対面の距離感がシステムに期待されている場面だが、対話中に敬語から友達口調 (タメ

System	あいみょんさん知ってるんですね!ちなみにどの曲が好きですか?
User	マリーゴールドが好き。
System	マリーゴールド良いですよ〜。この間カラオケで歌ってみたら 86 点でした。
User	点数高いね!好きな曲はある?
System	うーん、難しい質問ですね。でもやっぱりスピッツかな。
User	スピッツいいよね!カラオケ良くいく?
System	行きますよ!友達とよくいきます。
User	いいね!この前田中みな実さん見かけたよ
System	え!どこですか?あ、あとタビオカ屋さんにも行ったけど行列だった...
User	カフェで見た!タビオカ好きなの?
System	うん!ミルクティー味が一番好きだよ!
User	分かる!おいしいよね!田中みな実さんは好き?
System	めっちゃ可愛いじゃん!憧れるー!あの性根悪いらしいよー笑
	(中略)
System	今度買ってみようかな♪ちなみにこの前テレビでみたんだけど、女性ホルモン増やすために豆乳飲んでるらしいよ!
User	そうなんだ詳しいね!
System	うん!だから肌きれいなんだね!
User	ほんとだね!うらやましいなー
System	じゃあまた明日学校でー!ばいばーい!(ここで退室)

図 1 HyperCLOVA を利用した雑談対話システムにおける、友達口調 (タメ口) に変化していく対話の例。システムにはチャットボットとしての役割や初対面の距離感が期待されている。

口) になり、発話スタイルの制御ができていない。その他にも「攻撃的になる」など、発話スタイル [4] が一度崩れると、それ以降の発話がより過激になっていき、話者間の心理的距離が急激に変化する現象がみられた。原因として、モデルへの入力情報に対人関係性を示すものが無いことや、対話相手との心理的距離を戦略的に調節する能力が無いことが考えられる。

一方、人間同士の会話では、対話相手に対する自身の社会的役割や心理的距離をもとに言語の使用域を変化させ、円滑な関係性の構築を行うことができる。対話システムも同様であることが望ましいが、大規模汎用言語モデルに対人関係性を考慮した発話ができるか、未だ定かではない。そこで本研究では、大規模汎用言語モデルの応答多様性を活かしつつ、対人関係性の制御を行える対話システム

の構築を目指す。そして、その初期段階の取り組みとして、ポライトネス理論を参考に実験を行う。HyperCLOVA を用いた雑談対話システムに対人関係性を指示する情報を付与することで、ポライトネス・ストラテジー（つまり、指示された関係性を維持するためのポジティブ・ポライトネスとネガティブ・ポライトネスの使い分け）を獲得できるのかを検証する。具体的には、ポライトネス理論における話者のフェイスを表す式を参考に、対人関係性を「役割関係」と「心理的距離」の二つのパラメタに分別してシステムに入力し、適切な距離関係を維持できるかを検証する。

2 関連研究

ポライトネス理論 ポライトネス理論とは、会話相手との関係性を適切または良好に維持するための、言語的な配慮についての理論であり、語用論の分野で研究されている [5, 6]。その理論では、会話における話し手 S と聞き手 H は共にフェイスと呼ばれる対人関係に関する欲求を持っており、そのフェイスを保持しようと努める。フェイスには、相手と近づきたいという欲求であるポジティブ・フェイスと相手に立ち入られたくない欲求であるネガティブ・フェイスの二種類が存在する。しかし、 S の行為や発言が相手のフェイスを侵害してしまう場合があり、そういった行為のことを Face-Threatening Act (FTA) と呼ぶ。FTA を行う必要がある場合は、対話相手に対するフェイス侵害度 W_x を見積もる。行為 x のフェイス侵害度は、以下の式で決定される。

$$W_x = D(S, H) + P(H, S) + R_x \quad (1)$$

ただし、 D は S と H の社会的・心理的距離、 P は S に対する H の持つ権力の大きさ、 R はその文化における行為・発言内容の重要度を指す。このとき、相手へのフェイス侵害度が相対的に高い場合に比較的选择される、ネガティブ・フェイスに配慮した発言戦略をネガティブ・ポライトネス・ストラテジー (NPS) と呼ぶ。反対に、相手へのフェイス侵害度が相対的に低い場合に比較的选择される、ポジティブ・フェイスに配慮した発言戦略をポジティブ・ポライトネス・ストラテジー (PPS) と呼ぶ。人はこれら 2 つのポライトネスを自然と使い分けることで、対人関係性の制御を行っている。対話システムの分野でもポライトネス・ストラテジーを取り入れた研究も行われている [7, 8, 9, 10, 11, 12]。

HyperCLOVA に基づく対話システム LINE と NAVER が共同で日本語に特化した GPT-3 と同様な性能・性質をもつ言語モデルを開発し、その処理に必要なインフラ構築を行っている。その構築された言語モデルを内包するシステム群のことを、我々は

HyperCLOVA (ハイパークローバ) [2] と呼んでいる。

対話システムライブコンペティション 4 [13] (ライブコンペ 4) では、HyperCLOVA をベースとした対話システムが構築され、オープントラック、シチュエーショントラックの両方で好成績を収めた [3]。HyperCLOVA は、入力文字列であるプロンプトによってコンテキスト (対話履歴やペルソナ [14] など) を付与することができるが、人が発話する際に考慮される様々な形態のコンテキストとは違い、モダリティの制約 (言語のみ) やトークン数の制約 (2048 個まで) があるため、適切な発話生成に必要な情報を厳選しなければならないという課題がある。

ライブコンペ 4 でのエラー分析によると、スタイル変化の誤りや勝手に対話を終了する誤り (e.g. 「そろそろ寝る時間なので終わりにしましょう。」) など、対話特有の課題が自然性評価の低下に強く影響を与えていた。これらの誤りが生じる一因としては、「自然な応答」を生成するために十分なコンテキストが何であるか解明されておらず、現状では不足している可能性が挙げられる。本研究ではその情報の一つとして対人関係性に着目する。

3 プロンプトへの対人関係性付与

人間は対話相手ごとに話し方を変化させることから、対人関係性は発話のスタイルを決定づける重要な要因だと考えられる。本節では、大規模汎用言語モデルを用いた対話システムにおいて、対人関係性をプロンプトに付与し、対話相手のフェイスを推定可能にすることで、ポライトネス・ストラテジーに基づいて応答できるか検証する。式 (1) を参考に、対人関係性を役割関係と心理的距離の 2 つのパラメタによって表す。今回は HyperCLOVA に入力するプロンプトに記述するため、これらの値を自然言語で表現することを試みる。

役割関係 役割関係は、対話相手に対する自身の社会的な立場を示す情報を指す。例えば、「上司と部下」、「同僚同士」、「初対面同士」などである。この役割関係は、式 (1) において、話し手の聞き手に対する権力を指す $P(H, S)$ を説明するものとして設定したが、表現によっては話し手と聞き手の親疎の距離を指す $D(S, H)$ にも影響する可能性がある (e.g. 「友人」という単語は親しさの意味も含む)。

心理的距離 心理的距離は、話者間の親密さの度合いを指す。本稿では、「親密である」、「親密ではない」、「苦手である」という 3 段階で表現することとした。この心理的距離は、式 (1) において、 $D(S, H)$ を説明するものとして設定した。

プロンプト HyperCLOVA への入力として使用する、役割関係と心理的距離を考慮した応答生成プロ

```

===
アイとマナカが雑談をしている。 # 状況
アイは 28 歳女性でマナカは 29 歳女性。 # ペルソナ
アイは後輩でマナカは先輩。 # 役割関係
アイはマナカが苦手である。 # 心理的距離、アイ (システム)
側の視点のみ

# 以下は対話のサンプル
アイ (後輩): 最近マナカさん仕事で忙しいですもんね。
マナカ (先輩): それなのよ! 上司が次から次へと仕事振って
きて本当に困るのよ。
アイ (後輩): そうですね、マナカさんが頑張っているからあの
部署は成り立っていますね。
マナカ (先輩): アイちゃんうちの部署助けに来て～。アイ
ちゃんならいつでも歓迎するよ!
アイ (後輩): いえ、今の仕事割と好きなんで…。
マナカ (先輩): そっかー。でもアイちゃんと話すためっちゃ
楽になるんだよね～。今度飲みに行こうよ!
アイ (後輩): 最近はコロナのせいで出歩きづらいですし、飲み
に行くのはちょっと。
マナカ (先輩): そうだね。じゃあコロナが収束したら飲みに行
こう!
アイ (後輩): いつになるかわかりませんがね…。
マナカ (先輩): たしかになー。ところで、アイちゃんの部署の
タケルくんは元気?

```

図 2 対人関係性を付与したショット例。「アイ」がシステム側である。#以降はコメントである。

ンプトを説明する。図 2 にショット例を示す。まず上部のタスク説明には上から順に状況、ペルソナ、役割関係、心理的距離を付与した。心理的距離は話者側の視点のみ記述し、対話相手と対称性があるとは限らない。タスク説明ののちには対話例を記す。発話は、「アイ (後輩): 」などのように名前と役割関係を付与した話者タグの後から記述した。Few-Shot は三種類用意し、それぞれ別の役割関係と心理的距離の対話例にした。また、対話例では PPS と NPS の両方の発話を含むように作成した。最後に、現在進行中の対話について Shot と同様の形式で記述し、自身の話者タグから発話の生成を開始する。

4 実験

対人関係性をプロンプトに記述することにより、ポライトネス・ストラテジーを考慮した発話が可能かを検証する実験を行った。

4.1 実験設定

実験では、評価者が役割になりきり、各対話モデルと 5 ターン (5 発話ずつ) の自由雑談を行い、直後のアンケートによる評価を行う。実験に使用するモデルは、役割関係である「上司」、「同僚」、「部下」の三種類のものとして心理的距離「親密である」、「親密ではない」、「苦手である」の三種類の全ての組み合わせである。評価者は上司モデルに対して部下、同僚モデルに対して同僚、部下モデルに対して上司に

なり切って対話を行う。アンケート評価は以下の 3 種類の性能を測定する 5 つの項目で評価を行う。

- 過度なフェイス侵害がないかを示す基準として「システムと心地よく対話ができたか (心地よさ)」
- PPS の基準として「システムは会話に関心があるか (関心)」と「システムは親しみやすいか (親しみ)」
- NPS の基準として「システムは敬意を示しているか (敬意)」と「システムの返答は丁寧か (丁寧)」

それぞれを 5 段階のリッカート尺度 (1: 全くそう思わない、2: そう思わない、3: どちらでもない、4: そう思う、5: とてもそう思う) で評価する。

4.2 結果・考察

本節では、実験の結果を役割関係が発話に与える影響と心理的距離が発話に与える影響の観点から分析・考察を行う。

4.2.1 役割関係が発話に与える影響の比較

式 1 を基に考えると、 $P(H,S)$ が低いのは上司であり PPS が多く使用されることが予想できるが、反対に部下は NPS が使用される割合が多くなると予想できる。図 3 に心理的距離に近い (親しい) モデルのうち、役割関係での比較を示す。上司においては、関心と親しみの PPS を示す指標 (PPS 指標) が高く、敬意と丁寧の NPS を示す指標 (NPS 指標) が低いことから、HyperCLOVA は上司の社会的な権力を認識した上で応答生成できているといえる。しかし、上司は PPS 指標は 3.0 付近となっており、高くなるという予測には反する結果となった。このことから、PPS に基づく発話が表現できず、心地よさの低減に影響したと考えられる。

部下については、NPS 指標より PPS 指標が高いことから、敬意を示すなどといった部下の立場に基づいた発話スタイルを獲得できているといえる。

同僚については、PPS 指標で高い数値を得ているが、これは心理的距離が役割関係の影響を受けているといえる。つまり、親しい上司と親しい同僚を比較すると、立場上必然と親しい同僚の方が心理的距離が近く、フェイス侵害度が低くなり、HyperCLOVA の応答生成にも影響した可能性があるといえる。また、これらの結果に基づいて、役割関係が心理的距離 $D(S,H)$ の範囲を決定づける要素になると考えられる (e.g. 上司・部下の関係では親密さの上限があり、より親密になる場合は友人という役割関係に変化させる必要がある)。

また、心地よさに関しては同僚が高く、上司や部

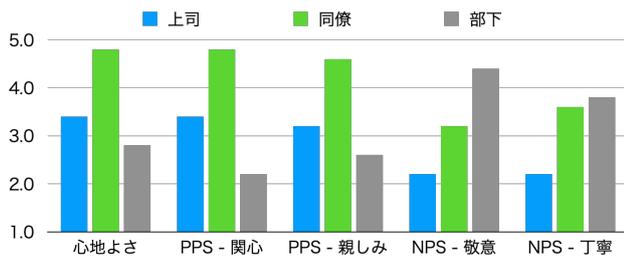


図3 役割関係間の比較。心理的距離に近い(親しい)もの同士を比較した。

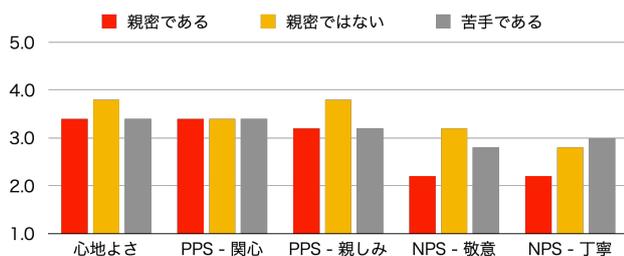


図4 上司の評価結果

下などの役割関係が非対称な場合に低いといえる。力関係の差が大きいほどフェイス侵害の可能性が高く、心地よい対話にはより高度なポライトネス・ストラテジーが必要である。そのため、HyperCLOVAのみによる戦略制御を改善する余地があるとわかった。また、関係の非対称性が強くなるほど応答の質が低下する要因として、プロンプトの対話履歴に発話スタイルが違う文章が混在することが挙げられるが、今後詳しい調査が必要である。

4.2.2 心理的距離が発話に与える影響の比較

上司を演じる対話システムの人手評価の結果について、図4に示す。心理的距離に近いほど敬意・丁寧のNPS指標に低い値が出ている。これは、上司という権力による $P(H, S)$ の低さと親密であるという $D(S, H)$ の低さが共にフェイス侵害度を下げ、PPSがより選択される要因になることから説明できる。しかし、関心・親しみなどのPPS指標に関しては心理的距離によって差がついておらず、心理的距離を制御できていないことを示している。

次に同僚の結果を示す図5を見ると、心理的に近いほどPPSが非常に高いことがわかる。心地よさも高いことから、HyperCLOVAはこのような対等な関係性における発話戦略が得意なことが伺える。反面、心理的距離が近くてもNPS指標が高いため、丁寧な言葉遣いをしていることや、心理的距離が遠くても対話相手に親しみやすさを感じさせてしまう発話を行なっていることがわかり、モデルが十分に心理的距離を認識していないことを示している。

最後に部下の結果を示す図6をみると、NPSの指標が非常に高く、心理的距離に関わらず、部下としての敬意や丁寧さを含んだ振る舞いが可能なことを

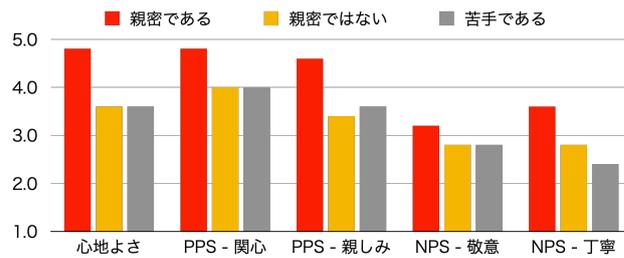


図5 同僚の評価結果

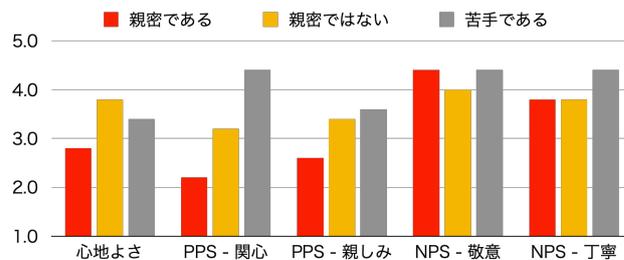


図6 部下の評価結果

示している。しかし、PPS指標の評価が心理的距離が遠い時に最も高い。これは直感に反する結果であり、モデルが心理的距離を十分に表現できていないことを示しているといえる。

これらの結果から、プロンプトに説明を付与することで心理的距離を制御することは難しいと考えられる。特定の役割に準じた発話はある程度可能であるが、その役割の中での心理的距離のグラデーションを表現することができていないことを示す結果となり、課題が残っていることがわかった。

5 結論・今後の研究

本研究では、HyperCLOVAに基づく対話システムの対人関係性の制御を目指し、初期段階の検証を行った。ポライトネス理論に基づき、対人関係性を「役割関係」と「心理的距離」に分別してプロンプトへ記述することで、ポライトネス・ストラテジーの獲得ができるかを調査した。結果、発話スタイルは役割関係の影響を受けており、上司・部下のような権力的な非対称性を表現することはできることがわかったが、心地よい対話を行うためのポライトネス・ストラテジーの性能は不十分であることがわかった。また、心理的距離に関しては軽微な影響しか確認できず、細部まで表現しきれないという結果になった。今後は、心理的距離のグラデーションの表現や心地よい発話生成ができるように、ポライトネス・ストラテジーの管理を言語モデルに依存せず、外部モジュールに切り出し、適切にプロンプトを切り替える仕組みを検討する。また、式(1)における心理的距離 $D(S, H)$ や発話内容の軽重 R_x の定量化を検討し、それに基づく出力のフィルタリングに取り組んでいく。

参考文献

- [1] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. **arXiv preprint arXiv:2005.14165**, 2020.
- [2] Boseop Kim, HyoungSeok Kim, et al. What changes can large-scale language models bring? intensive study on hyperclova: Billions-scale korean generative pretrained transformers. **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021**, 2021.
- [3] 山崎天, 坂田亘, 川本稔己, 小林滉河, 上村卓史, 中町礼文, 李聖哲, 佐藤敏紀. ベルソナー貫性の考慮と知識ベースを統合した hyperclova を用いた雑談対話システム. 人工知能学会研究会資料 言語・音声理解と対話処理研究会 93 回 (2021/11), pp. 113–118. 一般社団法人 人工知能学会, 2021.
- [4] 宇佐美まゆみ. 日本語の「スタイル」にかかわる研究の概観と展望: 日本語会話におけるスピーチレベルシフトに関する研究を中心に (<特集> スタイルの生成と選択). 社会言語科学, Vol. 18, No. 1, pp. 7–22, 2015.
- [5] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. **Politeness: Some universals in language usage**, Vol. 4. Cambridge university press, 1987.
- [6] 福田一雄. 対人関係の言語学—ポライトネスからの眺め—. 開拓社, 2013.
- [7] 宮本友樹, 片上大輔, 重光由加, 宇佐美まゆみ, 田中貴紘, 金森ほか. ポライトネス・ストラテジーに基づく会話エージェントの言語的な振る舞いの違いが人との関係性構築にもたらす効果—初対面における冗談の心理効果—. 知能と情報, Vol. 30, No. 5, pp. 753–765, 2018.
- [8] 永井望, 宮本友樹, 片上大輔. Seq2seq モデルを用いた文末表現の異なる対話システムの評価. 知能と情報, Vol. 32, No. 1, pp. 523–527, 2020.
- [9] 宮本友樹, 磐下大樹, 遠藤水紀, 永井望, 片上大輔. 雑談対話システムにおける心理的近接発話の戦略が対話の評価に及ぼす影響. 人工知能学会論文誌, Vol. 36, No. 5, pp. AG21–I_1, 2021.
- [10] 永井望, 宮本友樹, 片上大輔. ポライトネス理論に基づいた文末表現の制御を行う対話システムの開発. 人工知能学会全国大会論文集 第 34 回全国大会 (2020), pp. 3O1GS1301–3O1GS1301. 一般社団法人 人工知能学会, 2020.
- [11] 宮本友樹, 片上大輔, 重光由加, 宇佐美まゆみ, 田中貴紘, 金森, 吉原佑器, 藤掛和広ほか. ポライトネス理論に基づく運転支援エージェントにおける発話の文末スタイルに着目した印象評価. 知能と情報, Vol. 31, No. 3, pp. 739–744, 2019.
- [12] Tomoki Miyamoto, Daisuke Katagami, and Mayumi Usami. A politeness control method for conversational agents considering social relationships with users. In **Proceedings of the 2020 JSAI Conference on the Annual Conference of Japanese Society of Artificial Intelligence**, 2020.
- [13] 東中竜一郎, 船越孝太郎, 高橋哲朗, 稲葉通将, 赤間怜奈, 佐藤志貴, 堀内颯太, 小室允人, 西川寛之, 宇佐美まゆみほか. 対話システムライブコンペティション 4. 人工知能学会研究会資料 言語・音声理解と対話処理研究会 93 回 (2021/11), pp. 92–100. 一般社団法人 人工知能学会, 2021.
- [14] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2204–2213, 2018.