[S4-P13]

雑談対話にキャラクタ性を付与するためのスタイル変換

近藤 里咲, 梶川 怜恩, 梶原 智之, 二宮 崇（愛媛大学）

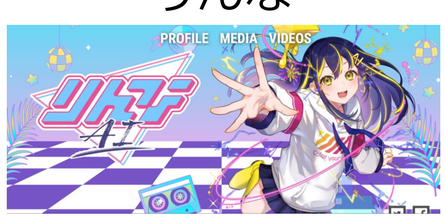
目的：大規模な応答生成モデルの恩恵を受けつつキャラクタ性を持つ雑談対話を実現したい

1．背景：応答の妥当性や流暢性を超えて、生成スタイル制御への関心が高まっている

近年、人間と自然に会話できる雑談対話システムが登場

キャラクタ性を付与することで
- 雑談対話システムがユーザに与える影響を変えられる
- 雑談対話システムがユーザに親しみを持ってもらえる

りんな



https://www.rinna.jp

なりきりAIプロジェクト

https://narikiri-qa.jp

2．課題：キャラクタ性を持つ雑談対話を実現するには応答生成モデルの追加学習が不可欠

先行研究として転移学習[1] や強化学習[2] ：応答生成モデルに対して追加学習を行っている
GPT-3のような**大規模モデル**やChatGPTのような**ブラックボックスモデル**は**追加学習が困難**

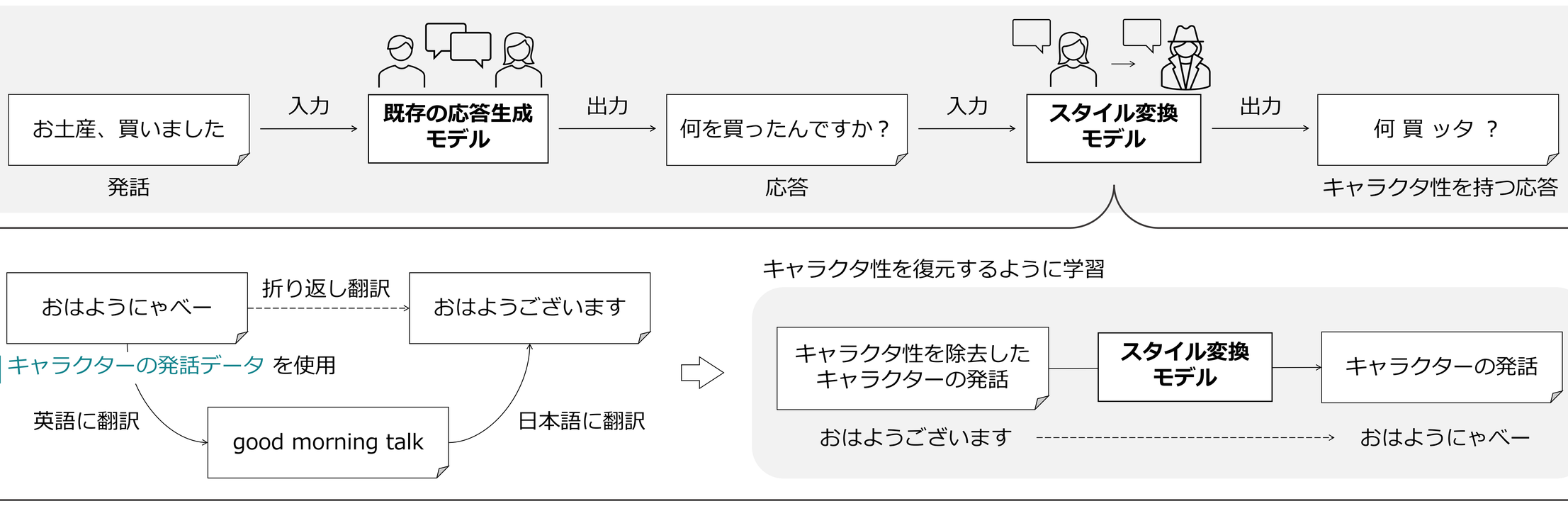
⇨ **応答生成モデルを追加学習することなく、キャラクタ性を付与したい**

[1] Akama et al. (2017) Generating Stylistically Consistent Dialog Responses with Transfer Learning

[2] 清水ら (2022) 強化学習を用いてキャラクタらしさを付与した雑談応答の生成

3．提案手法：応答生成とスタイル変換のパイプラインによってキャラクタ性を付与

発話に対する応答を**任意の応答生成モデル**で生成 → 生成した応答に対して**スタイル変換**



4．評価実験：人手評価の結果、キャラクタ性と総合評価において比較手法を上回った

4.1. 実験概要

以下の４キャラクターの Twitterデータを使用

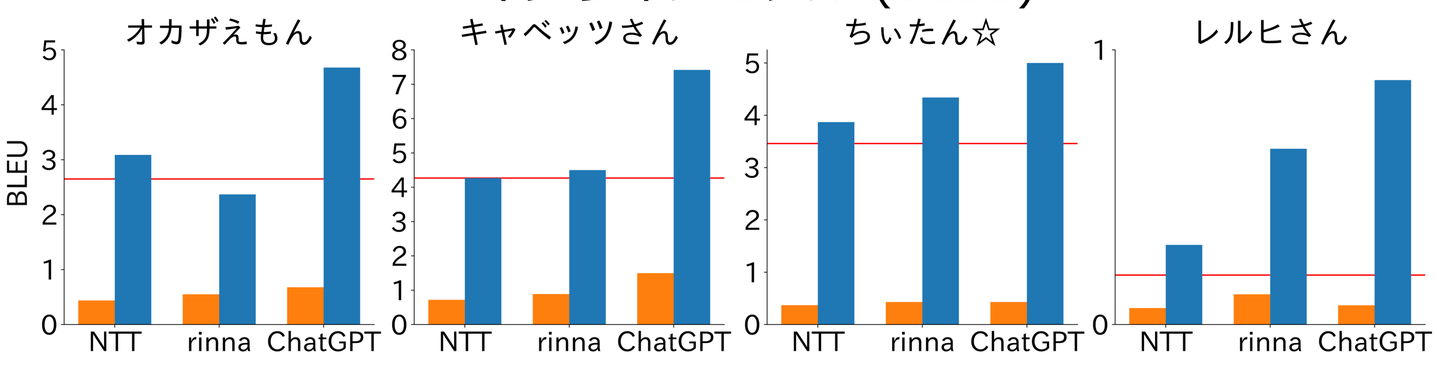
オカザえもん	ありがとうでござる! ご飯に肉が乗ってるだけでござる
キャベッツさん	今日は一緒に歌って踊れて楽しかったにゃべー♪
ちぃたん☆	お友達と追いかけっこしましたっ☆ちぃたん☆ですっ☆
レルヒさん	ぶろぐ 書イテタンヤ 秘密ニ シトケッテ 言ッタノニ

	発話数			発話応答ペア数
	スタイル変換			パイプラインモデル
	訓練用	検証用	評価用	評価用
オカザえもん	28,000	269	268	1,000
キャベッツさん	2,000	234	233	1,000
ちぃたん☆	2,000	138	138	1,000
レルヒさん	62,000	163	163	1,000

既存の応答生成モデル：NTT [3]，rinna [4]，ChatGPT　　**比較手法**：ChatGPT (3-shot)
スタイル変換モデル：BART-large

4.2. 実験結果

パイプラインモデル (BLEU)



パイプラインモデル (人手評価：ちぃたん☆)

	キャラクタ性	流暢性	妥当性	総合評価
NTT	1.12	4.30	3.22	0%
NTT+スタイル変換	3.15	4.14	3.16	5%
rinna+スタイル変換	3.53	4.28	3.59	24%
ChatGPT+スタイル変換	**3.62**	4.38	4.12	**34%**
ChatGPT (3-shot)	3.46	**4.90**	**4.24**	24%

自動評価では、スタイル変換を行うことによってBLEUが大幅に向上し、**ほとんどが比較手法を上回った**
人手評価では**妥当性・流暢性**で**ChatGPT(3-shot)に劣る**が、**キャラクタ性と総合評価では提案手法が優位**

[3] https://github.com/nttcslab/japanese-dialog-transformers

[4] https://huggingface.co/rinna/japanese-gpt-neox-3.6b-instruction-sft