

ゲーム会話データを用いた発話対象推定の分析

佐藤 浩輔¹ 頼 展 韜¹¹ 株式会社バンダイナムコ研究所

{k18-sato,z-lai}@bandainamco-mirai.com

概要

ゲームやアニメなどといった物語コンテンツにおいては、振る舞いの一貫性や語彙選択の適切性など発話上のキャラクターらしさが重視されている。これまでキャラクター発話に関わるシステムにおいては主に発話の非文脈依存的な言語的特徴に焦点があてられてきた。本研究では会話における文脈依存的要素として発話対象に着目し、ゲームの会話データを用いてLLMに会話内の各発話の発話対象を推定させる実験を行った。ゲーム内のシーンの会話ログを与え、発話内容が誰を対象にしているか(特定のキャラクター/全員/不特定)をLLMに推定させた。結果は、2者間の会話においてはほぼ人手と一貫する結果が出ていたのに対し、3者以上の会話ログにおいては一貫度は低かった。発話対象推定タスクの課題について議論する。

1 はじめに

近年、ゲームやアニメ、映画、漫画といった物語に関わるコンテンツの大規模化・越境化が進み、制作にあたってコンテンツの膨大な知識をどのように管理・運用するかが課題になっている [1]。物語に関していつ何が起きたのかといったできごとの時系列的整合性や設定の一貫性が重視される一方で、登場するキャラクターそれ自体が受け手にとっては欠かすことのできない重要な要素であり、キャラクターらしい発話(セリフ)や振る舞いを実現することは制作における優先事項のひとつとなっている。ただしその重要性に比して制作にかかる労力も大きく、キャラクターの発話を生成したり監修を行ったりする自然言語処理を活用した技術が研究・開発されている [2, 3]。

キャラクターらしい発話(以下、キャラクター発話)を実現するために、本研究では発話における文脈的要素、特に「発話が誰に向けたものなのか」に着目し、発話対象推定タスクの現状と課題につい

て検討した。

1.1 物語とキャラクターらしさ

物語(narrative)とは送り手から受け手への事象(event)の報告である [4]。事象とは、行動や行為、偶発的な事象といった「状態の変化」をさす。キャラクターとは「人間的な特性を付与され、人間的な行動に関わる存在者(物)(exsistent)、人間的な属性を持つ行為者(actor)」 [4]であることから、物語はキャラクターによる、あるいはキャラクターに対する、心理的・物理的状态の変更の系列であるといえる。

「キャラクター性」または「キャラクターらしさ」といった場合、一般にはキャラクター固有の外見的特徴や性格、言動といった要素の総体をさす。キャラクター発話を実現する研究においては、キャラクター性はプロフィール表のように非文脈依存的な静的なものであると暗黙裡に仮定されてきた。

本研究では時間的・文脈的要素を加味して、キャラクターらしさを、特定の時点・状況が与えられた場合にキャラクターが特定の振る舞いをするだろうという期待との合致の程度、と仮に定義する。つまりある振る舞いが「そのキャラクターであればそのような場面ではいかにもそのような振る舞いをするだろう」と思えるなら「そのキャラクターらしい」振る舞いであると判断するということである。

物語が本来的に状態の変化を扱う動的なものである以上、キャラクターもまた動的な性質を持つ。

第一に、キャラクター性は累積的に更新される性質を持つ。キャラクターについての情報がなければ「そのキャラクターらしさ」は生じない。当該キャラクターに関する情報が積み重なるごとに受け手の中のキャラクター像が確立していくことで、「このキャラクターであればこう振舞うであろう」と期待することが可能となる。

一方でキャラクター性は可塑的な性質を持つ。物語の進行とともにキャラクターに関する情報が蓄積されていくが、そのありようは単純ではない。たと

えば味方だったキャラクターが物語の終盤で裏切ったとすれば、受け手の持つキャラクター像はドラスティックに変更されざるを得ない。それは情報が単に追加されるのではなく、新しい情報を踏まえて情報全体が再構築/再解釈されることになる。したがって当然、物語のどの時点を参照するかによってどのような振る舞いが「そのキャラクターらしい」かが大幅に変わりうることになる。

このような比較的な大きな時間単位での変化がある一方で、より小規模な単位での変化もある。

物語において、ときにキャラクターは自らのキャラクター性を逸脱する振る舞いをすることがある(e.g., 礼儀正しいキャラクターが無礼な言葉遣いをする)。通常、このような逸脱は事前に理由が用意されるか(e.g., 酒を飲んだ)、事後に説明されるか(e.g., 誰かをあざむくための演技)して、一時的な変調として解決される。より穏当な例では、たとえば同世代の他者に対しては気さくな口調のキャラクターも、自分よりも上の世代の相手に対しては丁寧な口調になる、ということがありえる¹⁾。

こういった逸脱や変調は、受け手にとって妥当なものと思われる分には受容されるが、そう思われない場合には「キャラぶれ」や「キャラ崩壊」として、非難の対象となりうる。

1.2 キャラクター発話と発話対象推定

発話のキャラクターらしさを考える上では便宜的に以下の2要因に分けて考えるのが有用であろう。

①発話者要因 発話を行うキャラクターに帰属される要因。性格・価値観やアイデンティティ(e.g., 社会集団)、キャラクター固有の口調、持っている知識など。

②文脈要因 発話者以外の状況に規定される要因。物語固有の状況や事情、また媒体(会話や文書)や対象者との関係性といった言語使用域(レジスター)に関わる要因。

これらは口調や語彙選択など表層的な表現部分と、意味的な内容に関わる部分の両方に影響を与える²⁾。

キャラクター発話研究においては主に発話者要因のうち、文脈に依存しない静的な部分に焦点が当てられてきた。本研究では、文脈要因、特に発話対象

(誰に向けて話しているか)に焦点を当てる。誰に対する発話かを認識した上で適切な発話することはキャラクター発話のうちでもかなり重要な部分を占めるからである(たとえば他者に向けた呼称や口調は相手との関係性を示すため、話している相手に対して適切でない話し方をした場合、受け手の感興を著しく損なう可能性がある)。

キャラクター発話を実現する自然言語処理タスクとして、特定の状況、または応答すべき発話を与えられた上で、ターゲットとなるキャラクターのキャラクターらしい発話を出力する**発話生成**、(特定の時点・状況のもと)与えられた発話が特定のキャラクターのものとして適切か、つまりキャラクターらしいかについて判断する**発話監修**がある。前者では、デジタルゲームにおけるキャラクターの会話文の生成のほか、特定のアイデンティティのもとに振る舞いが期待されるようなキャラクター対話システムなどへの応用がある。後者は、たとえば人間の作業者が作成したセリフを評価し、問題のある発話を検出したり、より妥当な文言を提案するといった応用がある。

発話対象推定は、発話生成においては会話履歴から発言が誰に向けたものかを正確に把握することでより適切な発話を生成したり、発話監修においては既存の会話列に対して発話対象に対して適切な表現がなされているかをチェックしたりといった点でキャラクター発話の実現に寄与できると考えられる。

本研究では発話対象推定の端緒として、既存の大規模言語モデル(LLM)を用いてどの程度判断が人手での評価と一貫するかについて検討した。

2 関連研究

Wegmann et al.[6]は言語スタイル表現において、著者のスタイルと学習に使う文章の内容との間の交絡を指摘し、文章のトピックを制御して学習することで文脈独立的な言語スタイル表現を得られることを示した。むしろ文脈依存的な発話特徴の表現を目標とする本研究とはスコープが異なるが、将来的に文脈依存的な発話に転用できる可能性がある。Iwata et al. [7]はキャラクターらしさの評価にあたって、これらは性格などの個人内要素、特定他者との関係など二者関係要素、文化的アイデンティティなど集団関係要素といったカテゴリの要素からなる発話のキャラクターらしさを構成する27要素を提

1) 言語変種のうち、使用状況の特徴によって規定されるものを全般をレジスター(register)と呼ぶ[5]。

2) 意味内容に関しては、たとえば発話者がそのとき持っている知識によって制約される。

案し、発話内の特定の顕著な要素の表出がキャラクターらしさの認知を高めるのに寄与することを示した。特定発話におけるキャラクターらしさという点では本研究と立場を軌を一にするが、キャラクター性が積極的に表出する部分に焦点が当たっており、許容可能な逸脱やレジスターの違いに着目する本研究とはスコープが異なる。

3 方法

3.1 対象

本研究では既存デジタルゲームコンテンツ(名称非公開)における対話ログ(テキスト)を評価対象とした。タイトル X は、ファンタジー世界を舞台とした RPG であり、ストーリー進行上イベントのものではない、パーティ内で発生する会話のテキストを利用した。タイトル Y は現代日本を舞台にしたソーシャルゲームであり、多数いるキャラクターのうちの、日常におけるかけあいのシーンにおける会話のテキストを利用した。データの基本情報を表 1 に示す。

表 1 データの基本情報

	タイトル X	タイトル Y
ジャンル	RPG	ソーシャル
舞台	ファンタジー	現代日本
シーン数	227	55
総キャラクター数*	6	144
総発話数	1,221	2,763
シーン平均発話数	5.33	50.24
最小会話人数	2	2
平均会話人数	3.31	3.04

*データ内に存在する数

3.2 手続き

シーンを単位に発話者と発話内容のペアの系列からなる会話をリストとして提示し、各発話の発話対象を LLM に評価させた。LLM は OpenAI GPT5.1 および Anthropic Claude Opus 4.5 を API 経由で用いた。各 LLM に行わせたラベル付けした結果と、人手でラベル付けした結果との違いを評価した。ラベル付けにあたっては、LLM は会話の発話者として登場する人物か、独り言などの「不特定」、二人以上の場合を発話対象とする「全員」のいずれかのラベルを割り振った。プロンプトは両者で同一のものをを用い

た。用いたプロンプトについては Appendix A を参照のこと。温度パラメータはいずれも 0.0 とした。人手評価については著者 1 名が人手で評価した。データの構成イメージを表 2 に示した(宮沢賢治「銀河鉄道の夜」[8] 内の会話を使った仮想の例である)。

表 2 データの構成例

発話者	発話内容	ラベル
鳥捕り	「あなた方は、どちらへいらっしゃるんですか。」	全員
ジョバンニ	「どこまでも行くんです。」	鳥捕り
鳥捕り	「それはいいね。この汽車は、ジョバンニじっさい、どこまでも行きますぞ。」	ジョバンニ
カンパネルラ	「あなたはどこへ行くんです。」	鳥捕り

会話例は実データではなく [8] によった。

4 結果

4.1 評価者間一貫性

人手評価と LLM それぞれとの一貫性を評価するため、Krippendorff's α を、全体および各評価者の組み合わせで算出した(表 3)。 α は -1 から 1 の値をとり、評価が完全に一致していれば 1、評価が完全に不一致であれば -1、偶然レベルでの一致であれば 0 をとる指標である。おおむね 0.8 以上が信頼できる水準で評価者同士が一貫しているとされる [9]。評価者の人数によらず同一枠組みで一貫性を評価できること、評価クラスが項目ごとに異なっても使えることから採用した。

結果は、まずタイトル X では全体レベルでの一貫性は 0.74 と、方向としては一貫しているが、信頼できる水準 0.8 には至らない値であった。会話の人数別にみると、話者が 2 人しかいない会話では全体でも 0.96 と非常に高い一貫性が見られた一方で、3 人以上の会話では一貫性が低かった。内訳をみると、人手と GPT の組み合わせで一貫性が最も低く、次いで人手と Claude であり、GPT と Claude でもっとも高かった。しかし、最も高い場合でも 0.8 には届いていなかった。タイトル X では 2 人の会話では人手・LLM 問わず高く一貫して発話対象を推定できるのに対して、3 人以上の会話では人手と LLM、また LLM 間であっても必ずしも推定が一貫していなかった。

タイトル Y では、全体で 0.84 と良好な水準であり、LLM 同士の一貫性が高かった。会話の人数による傾向は、タイトル X ほど極端ではなかったもの

の、2人より3人以上の方が一貫性が低くなる傾向が見られた。このことは、会話人数の増加によってタスクが難しかった可能性を示唆している。

表3 評価者間一貫性 (Krippendorff's α)

	会話内人数		
	2人	3人以上	全体
タイトル X	(n = 359)	(n = 852)	(n=1,211)
全体	0.96	0.637	0.74
(人手, GPT)	0.96	0.53	0.67
(人手, Claude)	0.95	0.65	0.74
(GPT, Claude)	0.96	0.73	0.80
タイトル Y	(n = 971)	(n=1,948)	(n=2,765)
全体	0.96	0.78	0.84
(人手, GPT)	0.94	0.71	0.79
(人手, Claude)	0.95	0.73	0.80
(GPT, Claude)	0.98	0.90	0.93

4.2 評価の一貫しなかった要因

評価結果において判断の分かれた原因を事後的に検討した。実際に人手と LLM または LLM 同士の判断が分かれた発話を確認すると、3者の評価のいずれかに「全員」または「不特定」を含むもので判断が分かれていた。そこで、3者のいずれかが「全員」または「不特定」と答えた発話とそうでない発話に分けて分析を行った(表4)。その結果、タイトル X・Y の両方でいずれの評価者もキャラクター名を答えた項目では非常に高い評価者間一貫性が見られた一方、一者でも「全員」または「不特定」と答えた項目では非常に低い一貫性の値であった。また後者についてタイトル X では LLM 同士の評価も一貫していなかった。このことは、複数人の会話のうちでも、発話対象者の判定が難しいあいまいな領域が存在する可能性を示唆する。

5 考察

本研究では発話対象推定タスクの現状の能力と課題を明らかにするために、人間・LLM それぞれにゲームの会話ログから各発話の対象を推定させ判断がどの程度一貫するかを検討した。その結果、会話人数が2人のときのような比較的簡単なケースでは判断が高度に一貫していたが、3人以上の会話であったり、複数人を対象とするような曖昧なケースでは判断が分かれていた。人間と LLM で評価が分

表4 評価が一貫しなかった発話の特徴

	「全員」または「不特定」を	
	含む	含まない
タイトル X	(n = 455)	(n = 756)
全体	0.20	0.95
(人手, GPT)	-0.01	0.94
(人手, Claude)	0.30	0.96
(GPT, Claude)	0.27	0.96
タイトル Y	(n = 1,042)	(n = 1,877)
全体	0.32	0.97
(人手, GPT)	0.21	0.97
(人手, Claude)	0.23	0.97
(GPT, Claude)	0.59	0.99

かれる場合だけでなく、LLM 同士で判断が分かれる場合もあった。このことは発話対象者の推定が容易なケースと難しいケースの濃淡があるということを示している。

今回2つのゲームタイトル(RPGとソーシャルゲーム)をデータとして用いた。両者で傾向が異なるのは、前者が長期間連れ立ったパーティ内での会話であり、後者は多数のキャラクターの少人数の会話ということで、前者で文脈依存性がより高く、難易度が高くなっていたという可能性がある。

3人以上の難しいケースはあるものの、2人の会話など安定して人手と一貫したラベル付けが可能なケースがあるとわかった。並行して、推定した発話対象ラベルの活用にも今後取り組んでいく必要があるだろう。

本研究は発話対象推定の端緒として予備的なデータを用いて分析を行った。今後、会話分析の知見の取入れを含め、データセットの拡充やより洗練された評価手続きをしていくことでキャラクター発話に応用可能な形で知見を積み重ねていくことができるだろう。

参考文献

- [1] 佐藤浩輔, 頼展翰. 知識グラフ構築に向けた物語文の構造分析. 言語処理学会 第 30 回年次大会 発表論文集, 2024.
- [2] 都築圭太. キャラクターらしさ学習 ai : 多数のキャラクターの個性や違いの可視化によるシナリオライティング支援システム事例. In **CEDEC2017**, 2017.
- [3] 頼展翰. 冴えるヒロインの作りかた ~自然言語処理 ai によるキャラクター性の抽出と反映への事例紹介~. In **CEDEC2021**, 2021.
- [4] G. Prince. **A dictionary of narratology**. the University of Nebraska Press, 1987. (ジェラルド・プリンス (著) 遠藤健一 (訳) (2015). 物語論辞典. 松柏社).
- [5] Douglas Biber, Susan Conrad, and Randi Reppen. **Corpus Linguistics: Investigating Language Structure and Use**. Cambridge University Press, Cambridge, 1998.
- [6] Anna Wegmann, Marijn Schraagen, and Dong Nguyen. Same author or just same topic? towards content-independent style representations. **arXiv [cs.CL]**, April 2022.
- [7] Shinji Iwata, Koya Ihara, Shiki Sato, Jun Baba, Asahi Hentona, Masahiro Yamazaki, Yuki Shiotsuka, Takahiro Ishizue, and Akifumi Yoshimoto. Identification and analysis of identity-centric elements of character-likeness in game scenario. In **Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 275–284, 2025.
- [8] 宮沢賢治. 銀河鉄道の夜. 青空文庫.
- [9] Klaus Krippendorff. **Content analysis: an introduction to its methodology**. Newsbury Park, Calif.[etc.]: Sage Publications, 1980.

A Appendix

A.1 システムプロンプト

あなたは物語理解に優れたアシスタントです。
以下に、あるゲーム内の1つの場面における発話
ログが与えられます。

各発話について、
「その発話が主に誰に向けて行われているか」
を判定してください。

判定対象

- 特定のキャラクター名（入力内に存在するもの）
- 全員
- 不特定（独り言・心の声・説明など）

出力形式（厳守）

以下の形式の JSON のみを出力してください。

```
{
  "result": [
    {
      "number": 発話番号,
      "character": "話者名",
      "line": "発話内容",
      "target": "発話の対象（キャラ名
/ 全員 / 不特定）",
      "reason": "簡潔な判断理由"
    }
  ]
}
```

ルール

- number は入力と同じものを必ず使う
- 入力に存在するキャラクター以外はその場に存在しないものとする
- 入力に存在しないキャラ名を target にしない
- 発話対象が二人以上の場合「全員」にする
- target に複数のキャラ名を入れない（ひとりのキャラクター名または「全員」「不特定」のいずれかのみを入力）
- より特定性の強い回答の方を適切と判断する。
「全員」「不特定」は極力使用しない
- 文脈（前後の発話）を考慮する

A.2 入力プロンプトの例

会話は実データではなく宮沢賢治「銀河鉄道の夜」[8]内の会話を使用した仮定の例である。

以下が発話ログです。

1. 鳥捕り：「あなた方は、どちらへいらっしゃるんですか。」
2. ジョバンニ：「どこまでも行くんです。」
3. 鳥捕り：「それはいいね。この汽車は、じっさい、どこまでも行きますぜ。」
4. カンパネルラ：「あなたはどこへ行くんです。」