

Novel2DialCorpus: 小説を用いた対話コーパスの自動構築手法

岩本和真¹ 安藤一秋²

¹ 香川大学大学院 創発科学研究科 ² 香川大学 創造工学部

{s24g351, ando.kazuaki}@kagawa-u.ac.jp

概要

本研究は、小説テキストから雑談対話コーパスを自動で構築する手法 Novel2DialCorpus を提案する。提案手法では、小説テキストから各台詞を会話グループ単位で抽出した後、各台詞に対する発話者を特定することで、雑談対話コーパスを自動構築する。Novel2DialCorpus で構築した雑談対話コーパスの特徴を分析した結果、3人以上の会話や多様な口調、人間関係に応じた口調の変化など、従来手法では構築できない会話を含んでいることを確認した。提案手法のコードは、GitHub¹⁾で公開している。

1 はじめに

雑談対話システムの開発には、雑談対話コーパスが必要となる。従来の雑談対話コーパスは、人手による新規文章の作成 [1], 人同士の会話の録音・収集 [2], および SNS やクラウドソーシングを通じた会話収集 [3, 4] などによって構築されてきた。しかし、これらの方法は人手による工程が多く、構築コストが高いため、雑談対話コーパスを容易に構築することは困難である。この課題を解決する手法として、小説を用いた雑談対話コーパスの自動構築が考えられる。小説内の登場人物同士の台詞のやりとりを会話とみなし、これらを自動的に収集することでコーパスを構築できる。さらに、小説は日々出版されるため、継続的なコーパス構築や多様な会話を含む対話コーパスの構築が可能である。

小説を用いた雑談対話に関する研究として、言語資源の構築に関する研究 [5, 6, 7, 8] や対話システムに関する研究 [9, 10] などがある。小説を用いた言語資源の構築に関しては、人手による構築手法 [7, 6] や発話者特定手法 [11, 12] など、一部のプロセスの自動化に関する研究は進んでいる。しかし、小説を用いて対話コーパスを自動構築するプロセス全体を体系化した研究はほとんど存在しない。Du ら [5]

1) <https://github.com/ganbon/Novel2DialCorpus>

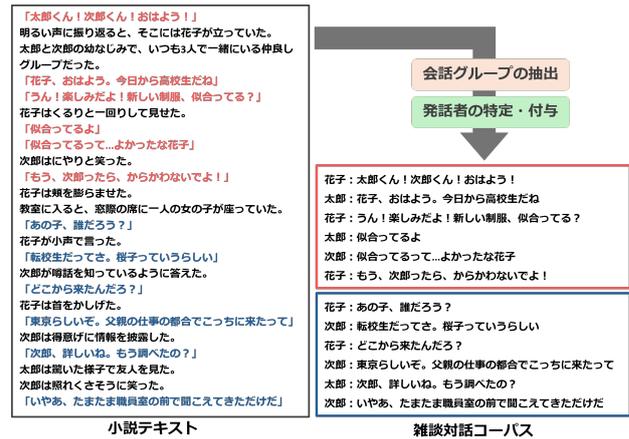


図1 Novel2DialCorpus を用いた対話コーパスの構築例

は、小説からルールベースで発話者特定および対話抽出することで、対話コーパスを自動構築する手法を提案している。しかし、ルールベースのみでは、文脈を考慮した対話抽出や発話者特定は困難である。

本研究では、小説テキストから文脈を考慮した対話（会話グループ）抽出手法と、複数の手がかりを用いた発話者特定・付与処理を用いて、対話コーパスを自動構築する手法 Novel2DialCorpus を提案する。本稿では、提案手法を説明するとともに、小説から構築したコーパスの特徴と課題を分析する。

2 Novel2DialCorpus

Novel2DialCorpus は、図1に示すように、小説テキストから会話グループを抽出する処理と、台詞の発話者を特定して付与する処理の2つで構成される。

2.1 会話グループの抽出

小説テキストから、台詞同士のやり取りが会話となる部分を抽出する。しかし、連続した台詞を1つの会話とみなす方法のみでは、1会話に含まれる発話数が平均2.98発話と少ない [13]。小説には、台詞同士が離れていても発話応答関係が成立する箇所が存在する。本研究では、文脈を考慮しながら離れ

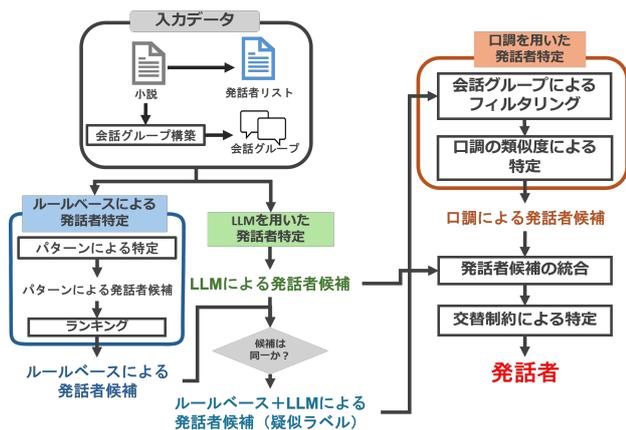


図2 発話者特定手法の概要

た台詞対の発話応答関係を判定するモデル [13]²⁾を用いる。具体的には、連続して記述されている台詞を1会話とみなす方法に加え、文脈的に発話応答関係があると判定された台詞対を連続した台詞として扱い、2台詞以上が連続している台詞群を1会話グループとして抽出する。発話応答関係判定モデルは、BERTモデル [14]の後半4層におけるCLSトークンのベクトルを結合したモデル構造に、「小説家になろう³⁾」の小説から構築したデータセットを学習させて構築する。

性能評価の結果、ルールベース手法のF1値が64%に対し、提案モデルは80%で発話応答関係を判定できることを確認した [13]。

2.2 発話者の特定・付与

会話グループ内の各台詞に発話者を付与するため、各台詞の発話者を特定する。発話者特定の手がかりとしては、台詞周辺の情報 [11]、人物の口調 [12]、人間関係などが挙げられる。これらの手がかりは台詞によって異なるため、本稿では、台詞周辺の情報と人物の口調に注目し、複数の手法を組み合わせた発話者特定手法を提案する。

提案する発話者特定手法を図2に示す。我々の先行研究 [15]では、ルールベースと口調に基づく発話者特定手法を提案したが、本稿では、これに加えてLLMを用いた手法を提案する。図2に示すパターンによる特定と口調を用いた発話者特定は先行研究 [15]で提案した手法を用いる。ルールベースによる発話者特定では、台詞内の情報と台詞周辺の地の文を用いた2種類のパターンを用いる。

LLMを用いた発話者特定は、特定対象の台詞とその前後10文を入力としたzero-shot推論で発話者を特定する。なお、人物名の特定が困難な場合は、発話者を「不明」とするよう指示をする。LLMには、Swallow-8B⁴⁾ [16]を用いる。LLMとルールベースの判定結果が一致した発話者候補を、ルールベース+LLMによる発話者候補とする。

口調を用いた発話者特定では、台詞の対照学習によって構築したSentenceBERT [17]ベースの口調ベクトルモデル⁵⁾を用いて台詞を口調ベクトルに変換し、ベクトル間の類似度から発話者候補を特定する。まず、ルールベース+LLMによる発話者候補を疑似ラベルとして利用し、疑似ラベルが付与された発話が α 以上存在する人物に対して、その台詞ベクトルの平均を算出し、平均ベクトルをその人物の口調特徴を表す代表ベクトルとする。本稿では、予備実験に基づき $\alpha = 20$ と設定する。次に、特定対象の台詞ベクトルと各代表ベクトルの類似度を計算し、類似度が閾値 β を超える人物を口調による発話者候補とする。閾値 β は、各人物の疑似ラベル付き台詞ベクトルと代表ベクトルの類似度分布における第1四分位数とする。最後に、LLMによる発話者候補と口調による発話者候補から、以下の3つのパターンを用いて、一意に発話者を特定する。

- LLMと口調による発話者候補の両方に同一人物が存在する場合、その人物を発話者とする
- LLMによる発話者候補が脇役人物⁶⁾である場合、その人物を発話者とする
- LLMによる発話者候補が不明であり、口調による発話者候補の口調類似度が閾値 γ を超える人物が1人の場合、その人物を発話者とする

閾値 γ は、各人物の疑似ラベル付き台詞ベクトルと代表ベクトルの類似度分布における第3四分位数とする。ルールベース+LLMによる発話者候補で発話者が特定できず、上記3つのパターンにも該当しない場合、発話者を「不明」とする。

発話者特定手法の性能を評価する。比較手法として、ルールベースのみを用いた手法とLLMのみを用いた手法を用いる。評価データは、「小説家になろう³⁾」に掲載されている10作品に対し、人手で発話者を付与した台詞を用いる。評価指標には、

- 4) <https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-Instruct-v0.2>
- 5) <https://huggingface.co/ganbon/novel-sentence-bert-base-tone-embedding>
- 6) 疑似ラベルが付与された発話が α 未満の人物

2) <https://huggingface.co/ganbon/novel-bert-base-relationship>
3) <https://syosetu.com/>

表1 発話者特定の評価結果

	Precision	Recall	F0.5
ルールベース	0.663	0.173	0.273
LLM	0.735	0.589	0.697
提案手法	0.780	0.532	0.707

表2 雑談対話コーパスの統計情報

	発話数	会話数	1 会話内の 発話数	特定 発話数	特定 会話数
合計	149,855	26,301	-	98,550	8,317
平均	7,492.7	1,315.0	5.73	4,927.5	415.9

Precision と Recall に加え、台詞が十分に存在することを考慮して Precision 重視の F0.5 を用いる。

評価結果を表 1 に示す。Recall は LLM が最も高いが、Precision と F0.5 は提案手法が最も高い。よって、複数の手がかりを用いることで発話者をより正確に特定できるといえる。

3 雑談対話コーパスの特徴分析

Novel2DialCorpus を用いて構築した雑談対話コーパスについて、構築可能なコーパスの規模、発話者の口調、会話内容の観点から特徴を分析し、構築したコーパスの言語資源としての活用可能性、課題について議論する。対象とする小説は、「小説家になる³⁾」の恋愛ジャンルに属する 20 作品である。

3.1 コーパス規模の分析

構築した雑談対話コーパスの規模と、1 作品から構築可能な対話規模を分析する。コーパスの統計情報を表 2 に示す。特定発話数は発話者が特定できた発話数、特定会話数は会話内の全発話が特定できた会話数を示す。1 作品あたり約 7,500 発話が存在する場合、約 1,300 会話の対話コーパスを構築できる。1 会話あたりの平均発話数は 5.73 であり、連続した台詞のみを会話とする場合より、2.8 発話増加した。また、1 作品あたりの最大発話数は平均 45.8 であり、現実世界では収集が困難な長い会話を収集できる。発話者が特定できた平均発話数は 4,927 発話であり全体の 65%、特定会話数は 415.9 会話で全体の 31%である。また、特定会話数の中で 3 人以上の会話が 1,224 会話存在した。

3.2 発話者の口調分析

構築した雑談対話コーパスに含まれる発話者の特性を口調と人間関係の観点から分析する。まず、雑談対話コーパスに含まれる口調の特徴を分析する。各人物の代表ベクトルを Ward 法で階層クラスタリ

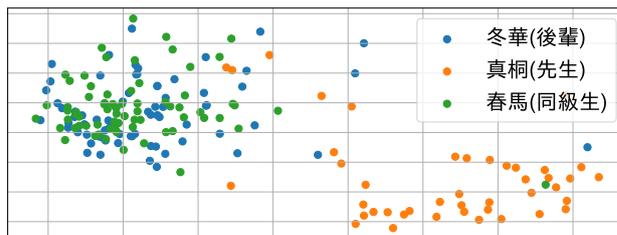


図3 話し手による発話者の口調分布

ングし、上位 20 クラスタを抽出する。各クラスタに対して、口調の特性が特に表れる助詞・助動詞の連語と人称代名詞を抽出する。助詞・助動詞の連語は tf-idf、人称代名詞は tf に基づき、上位 3 件を各クラスタの口調表現として抽出する。

表 3 に各クラスタの抽出結果を示す。表中の人数は各クラスタに含まれる人物数であり、口調タイプは抽出した助詞・助動詞の連語と人称代名詞に基づき筆者が付与したものである。口調タイプが全体で 11 種類存在することから、恋愛ジャンルのような現実世界の小説でも多様な個性を持つ人物が含まれていることがわかる。さらに、口調タイプとクラスタ内の人数分布から、敬語や丁寧な口調の人物が 20 人いるのに対し、明るい口調や優しい男性口調など、対人距離の近い口調が大半を占めている。一方、ID18 のように口調の判断が困難なクラスタも存在しており、口調に統一性のない人物や発話者の誤特定が原因であると考えられる。

実際の対話では、話し相手によって口調を変えることから、雑談対話コーパスにおいて、発話者の口調が話し相手との人間関係によって変化しているかを分析する。分析には、雑談対話コーパスで用いた小説 1 作品⁷⁾の主人公の発話を対象として、口調ベクトルを散布図で可視化し、各話し相手に対する発話ごとの口調の偏りを調査する。話し相手として、主人公との関係性が異なる後輩、同級生、先生の 3 人物を選定した。図 3 に話し相手ごとの口調の散布図を示す。先生との発話のみ、後輩と同級生との発話より離れていることがわかる。よって、小説に登場する人物も話し相手との人間関係や親密度によって口調を変化させるといえる。

3.3 コーパスの活用可能性と課題

2 つの分析結果に基づき、小説から構築した雑談対話コーパスの活用可能性と課題を述べる。構築したコーパスには、1 会話あたり平均 5 発話以上が含

7) <https://ncode.syosetu.com/n4634et/>

表3 各クラスターの助詞・助動詞の連語の tf-idf, 人称代名詞の tf 上位3件と口調タイプ (括弧は人手による補足)

ID	人数	口調タイプ	助詞・助動詞の連語	人称代名詞
1	10	優しい男性口調	ないかい, (し) たかい, となんか	君, 僕, 私
2	13	男性口調	けどな, だよなあ, だけどな	俺, お前, 君
3	2	老人口調	のでな, ではないか, たのだ	俺, 貴様, お前
4	13	男性口調	けどな, だよなあ, よな	俺, お前, 僕
5	18	女性口調1	ないかしら, ないわね, たわね	私, あなた, あんた
6	4	普通口調	んだね, だよねえ, だよ	僕, 私, 俺
7	11	明るい女性口調	はねえ, (し) たっしょ, ていうか,	あたし, 私, アタシ
8	6	女性口調2	ずと, ないもん, だよ	私, あたし, アタシ
9	6	女性口調1	じゃないのよ, じゃ無い, たわね	あんた, アタシ, 私
10	14	女性口調1	だからねえ, じゃないわよ, だよ,	私, あたし, あなた
11	9	丁寧口調	たのです, なのですが, なのです	私, 彼, 君
12	5	お嬢様口調	ましたわね, だけですわ, ませんの	私, 彼女, 彼
13	6	丁寧口調	ではないです, ですけど, たですし	私, わたし, あなた
14	10	後輩口調	ですなー, ですよなー, じゃないですかぁ	私, 彼女, あなた
15	10	丁寧口調	なかったです, ですか, ました	私, 俺, 僕
16	4	丁寧口調	はごさいません, ですか, はははは	私, 俺, ボク
17	8	乱暴口調	っすかね, (ないっ) つの, じゃねーから	俺, あんた, お前
18	9	不明	わわわ, にん, だよ	僕, わたし, 私
19	5	後輩口調	っした, じゃないっすか, っすよ	おれ, わたし, 私
20	6	普通口調	だよ, ですか, だけど	僕, わたし, 私

まれており, 最大発話数が45.8であることから, 会話の文脈を考慮した対話モデルの学習データとして活用できるといえる。また, 3人以上の会話が存在することから, 複数人対話に対応した対話モデルの学習にも活用できる。さらに, 表3から, 小説を用いることで多様な口調が含まれることから, キャラクター性をもつ対話システムの構築に有効である。また, 本研究で構築したコーパスは発話単体ではなく会話単位でコーパスを構築しているため, 図4に示すように, 口調だけでなく冗談や文句など, 親密な関係性を示す会話も収集できる。これにより, 対話レベルで親密性の高い対話モデルの実現可能性が示唆される。

しかし, 対話グループの作成と発話者特定の性能は, それぞれF1値80%, F0.5値70%であるため, コーパスにはノイズとなる会話が一部存在する。特に, 発話者特定はRecallが低く, 完全に特定ができた会話数が少ない。よって, ノイズとなる会話のフィルタリングや発話応答関係判定と発話者特定の性能向上が必要となる。また, 本コーパスは会話とその発話者情報のみで構成されており, 小説の背景知識や人物のペルソナなどが含まれていないため, 情報不足となる会話が存在する。よって, コーパス内の会話情報を補完するために, 小説のあらすじ要約, 人間関係情報や人物のペルソナなどの抽出手法を検討する必要がある。

千歳：二月と言えは？
 周：学年末考査
 千歳：ねえ、何で華の男子高校生がそんなしみつたれた発想になるの？
 周：男子高校生に華があるかは知らんが、学生なら当然の発想だと思うんだが……
 千歳：青春してる男子高校生ならバレンタインって言う筈でしょ？
 周：青春してないので分からんな
 千歳：またまたー
 周：で、相談って言うのは？

図4 小説から構築した雑談対話コーパスの会話例⁸⁾

4 おわりに

本稿では, 小説テキストからコーパスを自動構築する手法 Novel2DialCorpus を提案した。提案手法は, 会話の文脈を考慮した会話グループの抽出と, 複数の手がかりを用いた発話者特定・付与から構成され, 各処理の性能はF1値80%, F0.5値70%であった。また, 構築したコーパスの分析から, 3人以上の会話や多様な口調, 親密性の高い会話など, 実際の対話では確認できるが収集が困難な会話を含むコーパスを構築できることを確認した。

今後の課題として, 構築したコーパスを用いた雑談対話システムへの活用法, 小説テキストから場面や人間関係, 人物のペルソナといった情報を自動抽出し, コーパスに付与する手法を検討する。

8) 例は <https://ncode.syosetu.com/n8440fe/> の小説から構築した対話コーパスから抜粋した会話である

参考文献

- [1] 赤間怜奈, 磯部順子, 鈴木潤, 乾健太郎. 日本語日常対話コーパスの構築. 言語処理学会 第 29 回年次大会 発表論文集, 2023.
- [2] Itsuko Fujimura, Shoju Chiba, and Mieko Ohso. Lexical and grammatical features of spoken and written Japanese in contrast: exploring a lexical profiling approach to comparing spoken and written corpora. In **Proceedings of the VIIth GSCP International Conference. Speech and Corpora**, 2013.
- [3] Fumihiko Bessho, Tatsuya Harada, and Yasuo Kuniyoshi. Dialog system using real-time crowdsourcing and Twitter large-scale corpus. In **Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, 2012.
- [4] Sanae Yamashita, Koji Inoue, Ao Guo, Shota Mochizuki, Tatsuya Kawahara, and Ryuichiro Higashinaka. RealPersonaChat: A realistic persona chat corpus with interlocutors' own personalities. In **Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation**, 2023.
- [5] 白井清昭 Du Yulong. 小説からの自由対話コーパスの自動構築. 言語処理学会第 25 回年次大会発表論文集, 2019.
- [6] 上原隆一, 稲葉通将. オンライン小説サイトから収集したペルソナ対話データセットの構築. 人工知能学会全国大会論文集, Vol. JSAI2025, pp. 3R1OS4502–3R1OS4502, 2025.
- [7] Krishnapriya Vishnubhotla, Adam Hammond, and Graeme Hirst. The project dialogism novel corpus: A dataset for quotation attribution in literary texts. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 5838–5848.
- [8] Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 11836–11850, August 2024.
- [9] 小倉拓人, 谷津元樹, 原田実. Response generation in a dialogue system deep eve using lstm. 情報処理学会論文誌, Vol. 60, No. 3, pp. 967–975, 2019.
- [10] 秋山一馬, 稲葉通将. 小説から生成した疑似対話に基づくキャラクター対話システムの構築. 人工知能学会全国大会論文集, Vol. JSAI2022, pp. 2A6GS602–2A6GS602, 2022.
- [11] 銭本友樹, 古俣慎山, 宇津呂武仁. Bccwj を対象としたパターンマッチによる end-to-end 発話者分類. 言語処理学会 第 29 回年次大会 発表論文集, 2023.
- [12] 石川和樹, 小川浩平, 佐藤理史. 口調エンコーダを用いた小説発話の話者推定. 自然言語処理, Vol. 31, No. 3, pp. 894–934, 2024.
- [13] Kazuma Iwamoto and Kazuaki Ando. A method for determining utterance-response relationships between Japanese novel lines for constructing a daily dialogue corpus. In **2024 16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)**, 2024.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, 2019.
- [15] 岩本和真, 安藤一秋. 小説における台詞と口調, 地の文を活用した台詞の発話者特定手法. 言語処理学会 第 31 回年次大会 発表論文集, 2025.
- [16] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual llm adaptation: Enhancing Japanese language capabilities. In **Proceedings of the First Conference on Language Modeling**, COLM, p. (to appear), 2024.
- [17] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3982–3992, November 2019.
- [18] Ilya Gusev. Pingpong: A benchmark for role-playing language models with user emulation and multi-model evaluation. **arXiv [cs.CL]**, 2024.
- [19] Yen-Ting Lin and Yun-Nung Chen. LLM-eval: Unified multi-dimensional automatic evaluation for open-domain conversations with large language models. In **Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)**, pp. 47–58, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [20] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 110–119, San Diego, California, 2016.

A 対話モデルを用いた分析

自動構築した雑談対話コーパスを用いて対話モデルを構築し、対話モデルが生成した対話の特徴と課題を分析する。具体的には、雑談対話コーパスを学習していないモデル（モデル1）、コーパスの会話のみで学習したモデル（モデル2）、発話者情報を付与した会話で学習したモデル（モデル3）によるそれぞれの出力を比較、分析する。モデル2は、1会話あたり5発話以上の会話を学習に利用する。モデル3は、2者が交互に発話し、奇数または偶数のどちらかに発話者が付与された、1会話あたり3発話以上の会話を学習データとして用いる。事前学習モデルには、Swallow-8B⁴⁾を用いる。

評価方法は、Ilya [18] が提案した、2つの LLM 同士でマルチターン対話を生成し、その対話データを評価用 LLM で評価する手法を参考にした。評価用の雑談対話データは、User 側の LLM を日本語日常対話コーパス [1] で学習したモデル、Assistant 側には3つのモデルを用い、両モデルを交互に会話させることで自動生成する。Assistant 側のモデルには、小説内の任意の登場人物名、口調例として各5発話をシステムプロンプトに記述して発話を生成する。口調例としては、任意の登場人物の全発話の平均ベクトルとのコサイン類似度が高い発話の上位5発話を用いる。分析では、口調の異なる4人物を用いる。会話の冒頭には、学習データに含まれない日本語日常会話コーパスの会話の冒頭文を用い、1会話あたり10発話、合計50会話生成した。

Assistant 側のモデルが生成した発話に対して、LLM を用いた評価 [19]、口調ベクトルを用いた評価、distinct-1 [20] で評価する。評価用の LLM には、Gemini 2.5 flash⁹⁾を用いる。LLM を用いた評価では、以下の評価項目について5段階で評価する。

- 自然性：発話の意図が不明、応答が不適切など、会話に不自然な部分がないか
- 親密性：会話を通して親しみを感じられるか
- 口調特性：口調例の特徴を反映できているか
- 口調一貫性：口調が一貫しているか

口調ベクトルを用いた評価では、生成した発話と口調例で用いた5発話の平均ベクトルとのコサイン類似度を測る。

表4に各モデルの評価結果を示す。小説コーパス

9) https://ai.google.dev/gemini-api/docs/models?hl=ja#gemini-2.5-flash_1

表4 対話モデルを用いた評価。評価値はキャラクター4人全体の平均値を表す。

	distinct-1	口調類似度	自然性	親密性	口調特性	口調一貫性
モデル1	0.415	-0.033	4.70	3.10	1.34	3.94
モデル2	0.545	0.262	3.30	2.84	2.48	2.99
モデル3	0.559	0.500	4.16	3.49	3.50	3.69

表5 「健康のために豆乳を飲むようにしていましたが、少し飽きてきてしまいました。」に対する応答例。口調は「～があったね」のような口調例を与えて生成。

モデル1	モデル2
それは残念ですね。豆乳は栄養価が高く、健康に良いので続けることが大切です。ただし、毎日同じものを飲み続けると飽きてしまうこともあります。そこで、豆乳に混ぜるものを変えてみるのはいかがでしょうか？…	豆乳？その気持ちはわかるかな。味が薄すぎると飽きるよね

で学習した両モデル（モデル2と3）は、distinct-1が高く、口調類似性と口調特性も高い。一方、モデル1はいずれも低い。よって、プロンプトに口調例を含めるだけでは不十分であり、小説内の会話を学習することで、会話の多様性やキャラクター性が向上することがわかる。また、発話者情報を付与した会話で学習したモデル3は、発話情報なしの会話で学習したモデル2と比べて、口調の一貫性や会話の自然性が高いことから、発話者情報の付与は会話制御に有効であるといえる。しかし、自然性や口調の一貫性に関しては、雑談対話コーパスを学習していないモデル1が最も高い。これは、小説コーパスの構築プロセスにおいて、誤判定や誤特定した会話、情報不足となる会話の存在が起因していると考えられる。

表5に、モデル1とモデル3の応答例を示す。モデル1は、口調が丁寧語のままであり、説明的かつ1応答あたりの発話長が長い傾向がある。一方、モデル3は、口調例と類似した話し方を示し、発話長も一般的な範囲といえる。よって、小説テキストから自動構築した雑談対話コーパスを学習することで、口調だけでなく話し方などにも影響を与えていることがわかる。

今後の課題として、人手による評価の実施や、学習データに含まれるノイズの影響を詳細に分析し、Novel2DialCorpusの改善を図る。