

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

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

雑談対話コーパスの構築コストが高いという課題を解決するため，本研究では小説の台詞を用いた対話コーパスの自動構築に取り組んでいる．小説には様々な人物が登場するため，単に台詞を抽出するのみでは口調の一貫性を保ったコーパスが構築できない．そこで本稿では，台詞と口調，台詞周辺の地の文から得られる手がかりを用いて，台詞の発話者を特定する手法を提案する．実験の結果，ルールベースで発話者を特定する手法と比べて，提案手法は Precision をほとんど減少せずに，Recall を向上できることを確認した．

1 はじめに

雑談対話システムを実装する際の課題の1つに，対話コーパスの構築が困難であることが挙げられる[1]．この課題を解決するため，本研究では小説の台詞を用いた雑談対話コーパスの自動構築を目的とする．小説には，登場人物の台詞が存在しており，登場人物間における台詞のやり取りを会話とみなすことができる．したがって，台詞，つまり，登場人物の発話を収集することによって，大規模な対話コーパスが構築できる．また，小説は日々出版されるため，比較的容易に対話コーパスの量的・質的な拡張が可能であるといえる．

我々の先行研究[2]において，小説を用いた対話コーパスの構築時に，2つの課題が存在することを確認した．1つ目の課題は，小説内で連続して出現する台詞のみを会話として抽出した場合，1会話に含まれる発話数が既存コーパスと比べて少なくなることである．この課題は，対話モデルを構築する際，文脈理解の学習不足を引き起こす．2つ目の課題は，小説には様々な人物が登場するため，単に台詞を抽出するのみでは口調の一貫性を保ったコーパスが構築できないことである．この課題は，生成する発話・応答文の口調にも影響を与える．

本稿では，2つ目の課題に注目する．口調の一貫性を保つためには，台詞に発話者を明記する方法が考えられる．発話者を明記するためには，台詞に対する発話者を特定する技術が必要になるが，発話者を特定する手がかりは，小説や台詞によっても大きく異なる．主な手がかりとしては，台詞周辺の地の文，台詞，話者の口調，人間関係などが考えられるが，単体の手がかりのみで発話者を特定することは困難である[3, 4]．

本稿では，発話者を特定する主な手がかりの中から，台詞周辺の地の文，台詞，話者の口調の3つを用いて，台詞の発話者を特定する手法を提案する．

2 関連研究

小説における台詞の発話者特定を目的とした研究[3, 4, 5, 6]は，幅広く取り組まれている．その中でも，日本語の小説を対象とした研究[5, 6]もいくつか実施されている．また，発話者を特定して雑談対話コーパスを構築することを目指した研究[1, 7]も存在している．

石川ら[5]は，日本語小説に対して，BERTによる口調エンコーダモデルを用い，ルールベースで抽出した発話者候補の中から，口調を手がかりに発話者を推定する手法を提案している．評価実験により，主要人物の発話を対象に，64～80%の macro-F1 で発話者を特定できることを示している．石川らの実験では，人手で選択した主要人物の台詞のみを利用している．しかし，小説には，主要人物以外の台詞も多数存在しているため，ノイズが少ない状況での評価に留まっているといえる．また，主要人物同士の口調が類似している小説の場合，口調のみで発話者を特定することが困難である．これらの課題に対して，本稿では，口調ベクトルの類似度だけでなく，台詞や地の文から得られる手がかりも活用することで，発話者を特定する手法を提案する．

口調の分析や類似性評価に関する研究[8, 9]が存在している．銭本ら[9]は，口調の類似性を評価する

ため、小説内の台詞を用いて、対照学習により文ベクトルを構築する手法を提案している。本稿では、対照学習を用いて構築した文ベクトルモデルによる弁別判定の性能についても検証する。

3 発話者特定手法

本研究における発話者特定手法の構成を図1に示す。まず、小説本文と登場人物の人物名リストを入力とする。人物名リストは、登場人物ごとに、登場人物を表す人物名集合（完全表記や略称など）とその性別で構成されている。人物名リストは、固有表現抽出器 GiNZAⁱを用いて小説本文から人物名を抽出し、人手で同じ人物を表す人物名集合とその登場人物の性別を付与することで構築する。

提案手法は、ルールベースと口調の特徴を用いた2つ手法で構成される。まず、ルールベースで発話者を特定した後、特定できた台詞を用いて、人物ごとの口調ベクトルを構築し、構築した口調ベクトルの類似度を用いて発話者候補を特定する。そして、ルールベース手法と口調の特徴を用いた手法の結果を統合して、最終的な発話者を特定する。

我々の先行研究[1]において、地の文が存在することで台詞同士が離れているが、それらの台詞間に発話応答関係が成立する台詞対に注目し、離れた台詞間の発話応答関係を判定するモデルを提案した。本研究では、発話応答関係判定モデルが関係ありと判定した台詞対を連続した台詞とみなし、2つ以上の台詞が連続している台詞群を1会話として会話グループを形成する。そして、会話グループ内の各台詞に対して発話者を特定する。

以下、図1に示す流れにしたがって、提案手法の詳細を説明する。

3.1 ルールベース手法

小説には、台詞ではなく、地の文に発話者を特定する情報が書かれている場合や、台詞自体に書かれている場合がある。ルールベース手法は、それらの情報を定めたパターンに基づいて発話者または発話者候補を特定する。その後、台詞の口調による性別判定により、発話者候補を絞り込み、その候補の中で最も多い人物名をルールベースで特定した発話者とする。

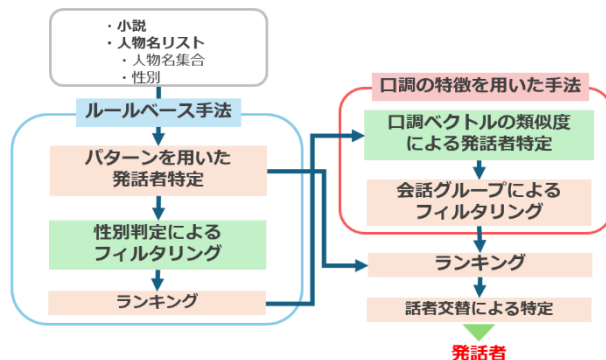


図1 発話者特定手法の構成

3.1.1 パターンを用いた発話者特定

パターンを用いた発話者特定では、周辺情報を用いたパターンと台詞内容を用いたパターンを用いる。周辺情報を用いたパターンは、会話グループ内の発話者を特定する台詞（対象台詞）の前2文または後ろ2文が地の文であり、対象台詞の直前の文の主語が登場人物である場合、その人物を対象台詞の発話者候補とする。台詞内容を用いたパターンは、対象台詞の直前または直後の台詞に「太郎、～」といった人物の呼びかけが含まれる場合、その人物を発話者候補とする。なお、自己紹介といった発話者自身の名前を含む台詞は存在するが、出現する割合は非常に少ないといえる。よって、対象台詞に出現した登場人物は発話者候補から除外する。

3.1.2 性別判定によるフィルタリング

台詞の口調は、性別に依存する傾向があるため、話者を特定する情報として有用である[10]。前述の2つのパターンによって得られた発話者候補から可能性の低い候補者を取り除くため、台詞の口調から発話者の性別を判定するモデルを構築して利用する。具体的には、対象台詞に対して性別を判定し、パターンを用いた発話者特定による発話者候補から性別が異なる人物を候補から外す。

性別判定モデルは、対照学習に基づく文ベクトルモデルに、256層の線形層を加えた2値判定モデルである。なお、文ベクトルモデルは、銭本らの研究[9]を参考に Sentence BERTⁱⁱを用いて構築する。文ベクトルモデルの学習データは「小説家になろうⁱⁱⁱ」に掲載されている小説190件から抽出した台詞ペア443,412件のTripleデータ（比較元の台詞、正例の台

ⁱ ja_ginza_electra を使用

ⁱⁱ tohoku-nlp/bert-base-japanese-v3 を使用

ⁱⁱⁱ <https://syosetu.com/>

詞（比較元と同じ発話者）、負例の台詞（比較元と異なる発話者）からなる。なお、学習データは、台詞が 4 文字以上でかつ助詞、助動詞が含まれている台詞のみで構成されている。性別判定モデルは、小説 10 件内の台詞に人手で性別ラベルを付与したデータセット（男性：2,011 件、女性：1,949 件）を用いて、線形層のパラメータのみ学習する。

性別判定モデルの性能を評価した結果、F1 値が 78%[11]となり、利用可能であると判断した。

3.1.3 ランキング処理

発話者候補から一人の発話者を特定するために、得られた発話者候補の中で最頻出の人物を発話者とする。最頻出の人物が複数存在する場合は、一意に発話者を特定することが困難なため、発話者を該当者なしとする。

3.2 口調の特徴を用いた手法

ルールベース手法は、対象台詞の前後に発話者となる人物名が存在する場合には適用できるが、その割合は 25.3%しかない。そこで、登場人物ごとの口調を特徴量に利用することで、特定できる台詞数を増やす。関連研究[5]において、口調の特性によって発話者を分類する手法が提案されている。しかし、小説では、類似した口調の人物が複数登場する場合もある。その場合、口調の特性のみで一意に発話者を特定することは困難である。

本稿では、口調の特徴によって発話者を一意に定めるのではなく、対象台詞と口調が類似している人物を発話者候補として特定する手法を検討する。

3.2.1 口調ベクトルの類似度による発話者候補の特定

この処理では、性別判定で用いた文ベクトルモデルを口調ベクトルモデルとして利用する。対象台詞を口調ベクトルモデルでベクトル化し、PCA によって 100 次元に圧縮する。ここで、口調ベクトルモデルでベクトル化した台詞を口調ベクトルとよぶ。また、ルールベース手法で特定した発話者を疑似ラベルとして利用する。疑似ラベルの数が α 以上の人物に注目し、疑似ラベルが付与されている口調ベクトルの平均を算出する。得られた平均口調ベクトルを注目人物の代表口調ベクトルとする。すべての口調ベクトルと代表口調ベクトルとの類似度を算出した

後、類似度が閾値 β を超えた代表口調ベクトルに対応する人物をその台詞の発話者候補とする。

3.2.2 会話グループによるフィルタリング

口調ベクトルの類似度によって特定した発話者候補を絞り込むために、会話グループとその周辺文章に出現する人物名を抽出し、その会話グループの発話者候補とする。つまり、口調ベクトルによる発話者候補から会話グループの発話者候補に存在しない人物名を削除する。

会話グループによる発話者候補の抽出範囲は、対象の会話グループの 1 つ前の会話グループの直後の地の文から対象の会話グループの 1 つ後の会話グループの直前の地の文までとする。

3.3 二手法の結果を統合した発話者特定

ルールベース手法の性別判定によるフィルタリングは、大まかな口調の特徴を捉えることができるため、可能性の低い候補者を削減する処理として有効である。しかし、性別判定の性能が F1 値で 78%であるため、本来は正しく特定できる発話者候補を排除してしまう事例が存在する。また、口調の特徴を用いた手法も、会話内に類似した口調を持つ人物が複数登場する場合は一意に特定できない。

これらの状況を改善するため、ルールベース手法内のパターンを用いた発話者特定と口調の特徴を用いた手法によって得られたそれぞれの発話者候補を統合する。統合した発話者候補を 3.1.3 で説明したランキング処理を用いて発話者を一意に特定する。これにより、複数の観点から発話者を特定できると考えられる。

最後に、口調とパターンによる発話者候補から一意に特定できた発話者と、小説内で連続した台詞は話者が交互に発話している傾向があることを利用し、話者交替の制約を用いて発話者を特定する。

4 弁別判定の評価

まず、対照学習で構築した口調ベクトルの弁別能力を評価する。2 つの小説から口調が異なる 3 人の登場人物に対して評価する。人手でアノテーションした 3 人物の台詞からランダムに抽出した 20 台詞をベクトル化し、平均したベクトルを代表ベクトルとして用いる。判定方法は、本稿で述べた手法と同じ、判定対象の台詞と代表ベクトルの類似度が最も

高い代表ベクトルに対応する人物を発話者とする．評価指標は Precision, Recall, F1 を用いる．

表 1 に評価結果を示す．各人物の口調は，関連研究[5]をもとに分類した．各小説の Macro-F1 が約 84% であることから，対照学習によって構築した口調ベクトルは，発話者特定にも有効であるといえる．誤判定した台詞を調査した結果，同じ口調や同じ性別である人物同士の誤判定が多いことを確認した．

表 1 弁別判定の評価結果

	人物	性別 / 口調	台詞数	Pre	Rec	F1
小説 1	A	女性/お嬢様	128	0.933	0.875	0.903
	B	男性/乱暴	85	0.840	0.929	0.882
	C	男性/優しい	65	0.734	0.723	0.728
	Macro			0.836	0.842	0.838
小説 2	A	男性/普通	219	0.861	0.821	0.841
	B	女性/敬語	170	0.851	0.911	0.880
	C	女性/普通	154	0.802	0.792	0.797
	Macro			0.838	0.841	0.839

5 発話者特定の評価

提案した発話者特定手法の性能を評価する．関連研究[4]では，主要人物の台詞のみを特定対象として評価していた．本研究は，対話コーパスの自動構築を目的とするため，主要人物のみではなく，主要人物以外の台詞も特定対象に設定して，提案手法の性能を評価する．

比較対象として，提案手法のルールベース手法と交替制約条件を用いた手法，口調の類似度と会話グループのフィルタリングのみを用いた手法を用いる．後者の手法は，各台詞と代表ベクトルの類似度を計算し，最も類似度が高い代表ベクトルに対応する人物かつ会話グループのフィルタリングを適用した発話者とする．代表ベクトルは，疑似ラベル数 α が 20 件超えた人物に対して構築する．提案手法において，発話者候補とする閾値 β は 0.7 とする．なお，発話者候補の閾値は，0.5～0.9 の閾値別の性能から定めた．閾値別の性能は付録 B に記載する．

評価指標は，Precision, Recall, F1 を用いる．評価データは「小説家になろう」に掲載されている小説の中で恋愛ジャンルに属する小説 10 件を用いる．各小説で会話グループを構築し，各小説 100 件の会話グループ内の台詞全てを特定対象とする．評価に

用いるデータに対して，人手で台詞に発話者ラベルを付与する．評価データの詳細情報は付録 A に示す．

表 2 に評価結果を示す．各評価指標の値は 10 小説の Micro 平均である．評価の結果，口調の類似度のみを用いて発話者を特定する方法の F1 が最も高いが，Precision については最も低い結果となった．またルールベース手法は，Precision が最も高くなったが，Recall については最も低い結果となった．一方，ルールベースおよび口調の特徴を併用する提案手法は，ルールベース手法の Precision の性能を維持しつつ，ルールベース手法よりも Recall が向上する結果となった．よって，小説内に類似した口調を持つ人物同士の会話や主要人物以外を特定対象とする場合，口調を手がかりに特定した発話者候補が有効であったといえる．

提案手法の Recall が約 37% に留まった原因として，類似した口調を持つ登場人物同士の会話の場合，一意に特定することが困難な台詞があることが考えられる．また，発話者が「クラスメイト」といった不特定の人物による台詞や代表ベクトルを構築できなかった人物の台詞に対する誤りの存在を確認した．本稿の評価では，閾値を人手で一意に定めたため，小説ごとに最適な閾値が設定できていない可能性がある．

表 2 発話者特定の評価結果

	Precision	Recall	F1
ルールベース	0.757	0.173	0.273
口調	0.496	0.524	0.502
口調＋ルールベース	0.718	0.375	0.484

7 おわりに

本研究では，雑談対話コーパスの構築に向け，台詞と口調，地の文から得られる 3 つの手がかりを用いた発話者特定手法を提案した．評価の結果，提案手法は，ルールベース手法より Recall が約 2 倍向上し，口調の類似度のみを用いる手法よりも Precision が向上することを確認した．

今後の課題として，閾値の設定方法があげられる．現状は人手で各閾値を設定しているため，小説ごとに最適な閾値設定ではないと考えられる．よって，小説ごとの特定性能を分析し，分析に基づいた閾値の自動設定法を検討することで，提案手法の特定性能および汎用性の向上を目指す．

参考文献

1. Du Yulong, 白井清昭. 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. pp.623-626, 2019.
2. Kazuma Iwamoto, Kazuaki Ando. A Method for Determining Utterance-Response Relationships Between Japanese Novel Lines for Constructing a Daily Dialogue Corpus. Proc. of 16th IIAI International Congress on Advanced Applied Informatics. pp.391-396. 2024.
3. Zhenlin Su, Liyan Xu, Jin Xu, Jiangnan Li, Mingdu Huang¹. SIG: Speaker Identification in Literature via Prompt-Based Generation. Proc. of the AAAI Conference on Artificial Intelligence. 38(17). pp.19035-19043. 2024.
4. Yue Chen¹, Zhen-Hua Ling¹, Qing-Feng Liu. A Neural-Network-Based Approach to Identifying Speakers in Novels. Proc. of Interspeech 2021, pp.4114-4118. 2021. doi: 10.21437/Interspeech.2021-609
5. 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. Journal of Natural Language Processing, 第 31 巻. 3 号. pp.894-934. 2024.
6. 古俣慎山, 銭本友樹, 宇津呂武仁. 日本語小説の発話者分類における大規模言語モデルおよび規則の評価. 出版地不明. 言語処理学会第 30 回年次大会発表論文集, pp.2090-2094. 2024.
7. 小倉拓人, 谷津元樹, 原田実. 小説対話システム Deep EVE における LSTM を用いたキャラクター性のある応答生成. 情報処理学会論文誌. Vol.60 No.3. pp.967-975. 2019.
8. 川北雄大, 石川和樹, 夏目和子, 小川浩平, 佐藤理史. 口調弁別評価データセットの作成と口調エンコーダの評価. 研究報告自然言語処理 (NL) . 2024-NL-259. pp.1-10, 2024.
9. Yuki Zenimoto, Shinzan Komata, and Takehito Utsuro. Style-sensitive Sentence Embeddings for Evaluating Similarity in Speech Style of Japanese Sentences by Contrastive Learning. Proc. of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop. pp.32-39. 2023.
10. Yuki Zenimoto, Takehito Utsuro. Speaker Identification of Quotes in Japanese Novels based on Gender Classification Model by BERT. Proc of the 36th Pacific Asia Conference on Language, Information and Computation. 2022.
11. 岩本和真, 安藤一秋. 台詞間の関係と登場人物の性別を用いた小説台詞の発話者特定方法の検討. 情報科学技術フォーラム講演論文集. 第 2 分冊. pp. 387-388. 2024

A. 評価データの詳細情報

5 節の評価に用いた 10 小説の詳細なデータを表 1 に示す. 表 1 の「話者が明確な台詞数」は, 台詞に対して任意の人物の台詞であることが特定可能な台詞の数であり, 「代表ベクトルを構築できた台詞数」は, 疑似ラベルが 20 件より多く付与された人物における台詞の数である.

表 1 から, 各小説において, 代表ベクトルを構築できた人物数にばらつきがあることから, 小説で書かれる会話形式が多様であることがわかる.

表 3 評価データに用いた各小説のデータ

	台詞数	話者が明確な台詞数	代表ベクトルを構築できた人物のみの台詞数	登場人物数	代表ベクトルを構築した人物数
1	543	461	313	12	6
2	613	609	608	4	4
3	426	400	303	7	3
4	683	512	343	10	4
5	569	542	542	5	5
6	576	497	497	5	5
7	469	391	243	8	4
8	476	336	246	8	2
9	336	275	238	7	3
10	635	630	613	6	5

B. 閾値 β を変化させた場合の性能評価

口調ベクトルの類似度による発話者候補の特定で利用する, 発話者候補とする閾値 β を 0.5~0.9 に変動させたときの性能の変化を示す. 図 1 は閾値ごとの precision, 図 2 は閾値ごとの recall のグラフを示す. その他の実験設定は, 5 節の実験と同様である.

Precision において, 閾値を上げると向上する小説, 大きな変化がない小説, 減少する小説など様々であることがわかる. 一方, Recall については, ほとんどの小説が閾値 0.8 あたりから大きく下がっているといえる. この 2 つの結果から, 最適な閾値は小説ごとに異なることがわかる.

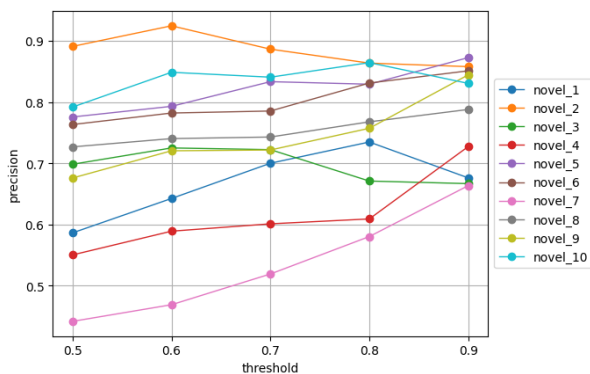


図 2 閾値ごとの各小説の Precision

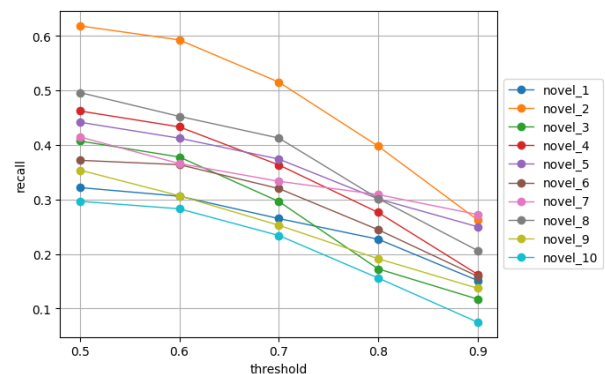


図 3 閾値ごとの各小説の Recall