

回答単位を小説登場人物とする 大規模言語モデルベース発話者分類

古俣慎山¹ 長谷川遼¹ 錢本友樹² 宇津呂武仁¹

¹筑波大学大学院 理工情報生命学術院システム情報工学研究群

²名古屋大学大学院 情報学研究科

概要

小説中の発話文の発話者がどの登場人物かを分類する発話者分類タスクは、小説や登場人物の分析において重要なタスクである。英語・中国語小説を対象とした発話者分類では深層学習モデルを用いた手法の有効性が確認されているが、日本語小説に対する先行研究では主に規則を用いた手法が検討されている。そこで本論文では、日本語小説を対象とした発話者分類タスクに大規模言語モデル (GPT-4o, Gemini, Claude) を適用した際の性能評価を行う。評価用データとして、複数のウェブ小説に発話者情報をアノテーションしたデータを用いる。本論文では、大規模言語モデルを用いて「小説登場人物名リスト作成」・「発話文への発話者対応付け」という二段階の処理を行い、小説登場人物を回答単位とした発話者の推論を行う。この結果、どの大規模言語モデルにおいても正答率が85%を超え、日本語小説を対象とした発話者分類における大規模言語モデルの利用が効果的であることを示した。

1 はじめに

大規模言語モデルの発展 [14] により、より高度で自然な応答をする対話システム [7] や、魅力的な小説を生成する小説生成システム [10] が提案されている。しかし、これらの対話システムや小説の登場人物は、発話相手や周囲の状況などの具体的な状況を考慮できていないという課題がある。この課題の解決には、様々なキャラクターの具体的な状況での具体的な言動データを収集することが必要である。

キャラクターの言動の特徴分析のために、文献 [8] と文献 [4] では、特定のキャラクターの発言を投稿するボットからキャラクターの発言を大量に収集している。しかし、このようなボットで収集可能なキャラクターの種類は限られており、多様なキャラ

クターの分析は不可能である。これに対して、小説中の発話文の発話者がどの登場人物の発話かを分類するタスクを解くことができれば、無数の様々なキャラクターの発話を自動的に大量に収集できるようになり、より多様なキャラクターの言動の詳細な分析が可能になる。そこで本論文では、様々なキャラクターの発話文の大規模な収集を目的として、小説中の発話文の発話者を分類する発話者分類タスクに取り組む。

この発話者分類タスクでは、小説中の発話文と、小説中の登場人物を対応付ける。先行研究では、この発話者分類タスクは複数のサブタスクに分解して取り組まれており、図 1 に示すように、1) 発話文の周囲の文章から、その発話文の発話者を表す人名¹⁾を対応付ける (以下、本論文では、「発話文に対して発話者を表すキャラクター・メンションを対応付ける」と呼ぶ) サブタスク、及び、2) 抽出したキャラクター・メンションのうち、同一の登場人物を指す人名をクラスタリングする (以下、本論文では、「同一人物へのキャラクター・メンションのクラスタリング」と呼ぶ) サブタスクの二つのサブタスクに分解されている²⁾。この発話者分類タスクについて、英語小説を対象とした先行研究では、複数の規則を用いた手法 [9] や深層学習を用いた手法 [2, 11] が提案されており、深層学習を用いた手法は規則を用いた手法よりも優れた性能を持っていることが示されている。日本語小説を対象とした先行研究 [12] でも、複数の規則を用いて発話者分類を行うシステムが提案されている。これに対して、本論文では、大規模言語モデルには複雑なタスクを解く能力が確認

1) 本論文では、小説中の登場人物を指し示す表現のことをキャラクター・メンションと呼ぶ。

2) 実際には小説本文中の発話を表す文字列を抽出する発話文抽出も必要である。しかし、発話文抽出の誤りを考慮すると発話者分類の評価が困難になり、文献 [12] から95%以上のF1値での発話文抽出が可能であることが明らかになっているため、本論文では発話文は既知情報として扱う。

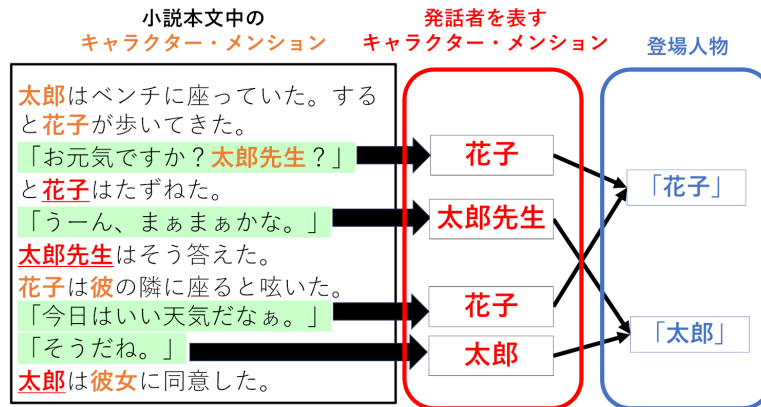


図1 発話者分類の全体像 (本論文では、「地の文」は小説本文中の発話文以外の文章のことであり、「キャラクター・メンション」は固有名詞(「田中太郎」や「田中」)や代名詞(「彼」や「私」)、職業(「先生」や「お巡りさん」)などの登場人物を指し示す表現のことである。同一人物に対する「田中太郎」、「田中先生」、「先生」、「彼」といったキャラクター・メンションは、「田中太郎先生」という単一の実体である登場人物を指す。)

されていることをふまえて 1) 小説登場人物名リスト作成、及び、2) 発話文への発話者対応付けの二段階の処理による発話者分類を行い、その性能を評価する。また、文献[15]の提案した、大規模言語モデルを用いた「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの結果を本論文の手法に適用した際の性能についても評価を行う。

本論文の貢献は以下のとおりである。

1. 日本語小説における発話者分類タスクにおいて大規模言語モデルを適用した手法を提案し、その性能を明らかにした。
2. 「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの結果を、大規模言語モデルを適用した発話者分類への入力に追加することで、発話者分類性能が改善される可能性を示した。

2 関連研究

日本語小説を対象とした発話者分類タスクのうち、1) 「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクまでの手法を提案している先行研究として、文献[3]がある。また、2) 「同一人物へのキャラクター・メンションのクラスタリング」サブタスクまでの手法を提案している研究として、文献[5, 12]がある。また、これら以外にも、キャラクター・メンションを経由せず、発話文の口調のみを用いて発話者分類を行う文献[6, 13]のように、上述の発話者分類タスク

の区分に当てはまらない手法を提案しているものもある。文献[3]では、青空文庫の小説四編を対象として、地の文を利用したパターンマッチによって、「発話文に対して発話者を表すキャラクター・メンションを対応付ける」ことを試みている。文献[6]と文献[13]では、ライトノベル中の発話文を対象として、口調の類似性を利用した発話者分類を試みている。しかしこれらの手法は、分類先となる発話者の口調が事前情報として必要であり、任意の発話者への分類には対応していない。また、文献[5]では、地の文を利用したパターンマッチや一人称、口調の類似性を併用した発話者分類を行っているが、特にパターンを利用した発話者分類においては、地の文において発話動詞の存在を仮定する点に限界がある。一方で文献[12]では、パターンを利用した発話者分類において、文献[5]のパターンを包含する体系を提案している。

表1 ウェブ小説データセット

小説タイトル	(a) 小説先頭から 500 発話文 を含む 「話」の数	(b) (a)中の 登場 人物 数	(c) (a)中の キャラ クター・ メンション 数	(d) 評価 対象 発話 文数
ずたぼろ令嬢は 姉の元婚約者に 溺愛される	19	13	1,231	400
無職転生 - 異世界行ったら 本気だす -	8	11	928	400
リビルドワールド	7	9	1,314	400
俺は星間国家 の悪徳領主!	9	23	1,013	400
合計	43	56	4,486	1,600

3 データセット

本論文で提案する発話者分類システムの評価には、文献 [15] と同様にウェブ小説投稿サイトである「小説家になろう」³⁾ に掲載されている小説から作成したウェブ小説データセットを使用する。このウェブ小説データセットは「小説家になろう」に掲載されている小説について、「恋愛」、「ファンタジー」、「文芸」、「SF」の各ジャンルから一作品ずつ選定した合計四作品から作成されている。このウェブ小説データセットは、四作品に対して、図 2 に示すように、「発話文」、「各発話文の発話者となる登場人物 (以下、参照発話者と呼ぶ)」、「各発話文の参照発話者を指すキャラクター・メンション (以下、参照キャラクター・メンションと呼ぶ)」を付与している。ウェブ小説データセットの小説タイトル、小説先頭から 500 発話文を含む「話」の数及びそれらの話数における登場人物数とキャラクター・メンション数⁴⁾、評価対象発話文数⁵⁾ を表 1 に示す。

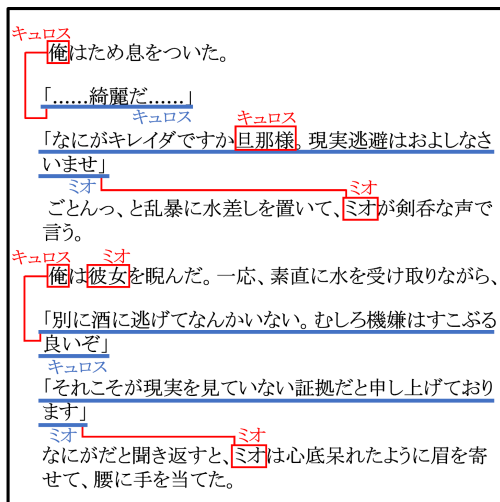


図 2 ウェブ小説データセットにおけるキャラクター・メンション、キャラクター・メンションが指す登場人物、発話文、発話文の発話者、および、それらの間の関係の例 (赤四角部はキャラクター・メンションを表し、上の赤字はそのキャラクター・メンションが指す登場人物を表す。青線部は発話文であり、下の青文字はその発話文の発話者を指している。また、各発話文の発話者を表すキャラクター・メンションが赤線で繋がれている。)

3) <https://syosetu.com/>

4) 表 1 (b) および (c) においては、発話者となった登場人物及びそれら登場人物のキャラクター・メンション数を示す。

5) 文献 [15] では、評価対象の発話文を、「ずたぼろ令嬢は姉の元婚約者に愛される」においては先頭から 101 番目の発話文から 500 番目の発話文の 400 発話文とした。また、それ以外の三小説については、各小説の先頭から評価対象の 400 発話文を評価対象とした。本論文においても、同様の設定において、各モデルの発話者分類性能を評価する。

4 発話者分類

ある発話文に対して、その発話文の発話者となる登場人物に対応付ける発話者分類タスクを、大規模言語モデルを用いて行うシステムを開発した。大規模言語モデルとしては、GPT-4o⁶⁾、Gemini⁷⁾、Claude⁸⁾ を利用した。大規模言語モデルは入力プロンプトにタスクの回答例を含める few-shot 学習を行うことで、モデルの性能が向上することが一般的に知られている [1]。しかし、大規模言語モデルは入力プロンプトが長くなるほど費用が高額になるため、本論文ではタスクの回答例を含めない zero-shot での推論を行う。

このシステムでは、ウェブ小説データセットの各小説に対して、以下の処理を行う。

(P1)各小説の本文全体を入力し、小説の「要約」・「小説登場人物名リスト」・「各登場人物の特徴 (口調・性格)」を回答させる。

(P2)分類対象の発話文・対象となる発話文が含まれる「話」の本文全体・手続き (P1) で得られた回答を入力とする。これらを入力として、分類対象の発話文の発話者が、手続き (P1) で得られた小説登場人物名リスト中のどの人物によるものかを回答させる。

図 4 に、手続き (P2) のプロンプト及び出力例を示す。また、付録 A 節に示すように、手続き (P1) への入力 (小説の本文全体) 及び手続き (P2) への入力 (対象となる発話文が含まれる「話」の本文全体) に対して、「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの結果を適用する実験を行った。

5 評価

3 節で述べた評価用データを用いて、発話者分類タスクにおける、GPT-4o、Gemini、および Claude の性能を評価する。各モデルは、4 節に示した手続き (P1)、手続き (P2) を行う。そして、対象の発話文に対して、手続き (P1) で得た「小説登場人物名リスト」から発話者を選択し、予測発話者として出力する。この時、付録 A 節に示すように、手続き (P1)、手続き (P2) それぞれに対して「発話文に対して発話者を表すキャラクター・メンションを対応付ける」

6) gpt-4o-2024-08-06

7) gemini-1.5-pro

8) claude-3-5-sonnet-20240620

表2 評価結果(%)(各小説のタイトルを以下のように略記する。ずたぼろ: ずたぼろ令嬢は姉の元婚約者に愛される, 無職: 無職転生 - 異世界行ったら本気だす -, 悪徳: 俺は星間国家の悪徳領主!)

モデル	キャラクター・メンション対応付け結果 [15] の使用有無		正答率				
	小説登場人物名リスト作成時	発話文への発話者対応付け時	ずたぼろ	無職	リビルドワールド	悪徳	全体
GPT-4o	無	無	88.8 ($\frac{355}{400}$)	86.3 ($\frac{345}{400}$)	95.8 ($\frac{383}{400}$)	85.0 ($\frac{340}{400}$)	88.9 ($\frac{1,423}{1,600}$)
	有	無	91.0 ($\frac{364}{400}$)	80.3 ($\frac{321}{400}$)	94.3 ($\frac{377}{400}$)	79.3 ($\frac{317}{400}$)	86.2 ($\frac{1,379}{1,600}$)
	無	有	86.3 ($\frac{345}{400}$)	87.3 ($\frac{349}{400}$)	96.3 ($\frac{385}{400}$)	85.8 ($\frac{343}{400}$)	88.9 ($\frac{1,422}{1,600}$)
	有	有	90.8 ($\frac{363}{400}$)	78.8 ($\frac{315}{400}$)	96.5 ($\frac{386}{400}$)	83.0 ($\frac{332}{400}$)	87.3 ($\frac{1,396}{1,600}$)
Gemini	無	無	84.0 ($\frac{336}{400}$)	82.5 ($\frac{330}{400}$)	96.0 ($\frac{384}{400}$)	87.3 ($\frac{349}{400}$)	87.4 ($\frac{1,399}{1,600}$)
	有	無	94.3 ($\frac{377}{400}$)	82.0 ($\frac{328}{400}$)	94.5 ($\frac{378}{400}$)	88.0 ($\frac{352}{400}$)	89.7 ($\frac{1,435}{1,600}$)
	無	有	86.3 ($\frac{345}{400}$)	85.8 ($\frac{343}{400}$)	96.5 ($\frac{386}{400}$)	87.8 ($\frac{351}{400}$)	89.1 ($\frac{1,425}{1,600}$)
	有	有	94.3 ($\frac{377}{400}$)	83.3 ($\frac{333}{400}$)	96.3 ($\frac{385}{400}$)	88.5 ($\frac{354}{400}$)	90.6 ($\frac{1,449}{1,600}$)
Claude	無	無	87.8 ($\frac{351}{400}$)	83.0 ($\frac{332}{400}$)	94.3 ($\frac{377}{400}$)	78.3 ($\frac{313}{400}$)	85.8 ($\frac{1,373}{1,600}$)
	有	無	88.0 ($\frac{352}{400}$)	83.8 ($\frac{335}{400}$)	93.3 ($\frac{373}{400}$)	85.5 ($\frac{342}{400}$)	87.6 ($\frac{1,402}{1,600}$)
	無	有	85.3 ($\frac{341}{400}$)	82.5 ($\frac{330}{400}$)	94.8 ($\frac{379}{400}$)	79.3 ($\frac{317}{400}$)	85.4 ($\frac{1,367}{1,600}$)
	有	有	84.8 ($\frac{339}{400}$)	85.0 ($\frac{340}{400}$)	92.5 ($\frac{370}{400}$)	86.8 ($\frac{347}{400}$)	87.3 ($\frac{1,396}{1,600}$)

サブタスクの結果を適用するか否かについて個別に性能を調べた。

5.1 評価手順

本論文の提案するシステムでは、出力する予測発話者が参照発話者と同じの文字列として得られない場合がある。そのため、モデルが各発話文に対して正しい発話者に対応付けているかを評価するために、4節に示した手続き(P1)においてモデルが出力する小説登場人物名リストの各要素と、参照発話者が同一の登場人物を指しているかを人手で照合し、対応関係表を作成する。予測発話者は、手続き(P1)によって得られる小説登場人物名リストから選択されるため、この対応関係表を用いて参照発話者と一致するか否かを判定する。そして、両者が一致する割合を発話者分類タスクの正答率とする。

5.2 評価結果

ウェブ小説データセットに対する GPT-4o, Gemini, 及び Claude の評価結果を表2に示す。表2から、発話者分類タスクにおいて、どの大規模言語モデルもウェブ小説データセット全体に対する正答率が85%を超えていることが分かる。また、Geminiにおいては、小説登場人物名リスト作成時・発話文

への発話者対応付け時のいずれにおいても「発話文に対して発話者を表すキャラクター・メンション対応付ける」サブタスクの結果を適用することで性能が向上した。この原因として、「発話文への発話者対応付け」を行う際に、発話文周辺に発話者を表すキャラクター・メンションを挿入したことが挙げられる。これにより、キャラクター・メンションを挿入しない場合には誤答した例に対しても、モデルは正答していた。一方で、発話文周辺に誤ったキャラクター・メンションを挿入することで、本来正答していた例に対して誤答する例も存在した。この改善例の数と悪化例の数はモデルによって異なり、GPT-4o, Gemini においては正答率維持・正答率改善という結果になったが、Claude においては悪化例の方が多く正答率がわずかに下がるという結果となった。

6 おわりに

本論文では、日本語小説に対して、GPT-4o, Gemini, 及び Claude を用いて、発話者分類タスクの性能評価を行った。その結果、どのモデルにおいても正答率が85%を超え、発話者分類における大規模言語モデルの利用が効果的であることが示された。

謝辞

本論文は、一部、科研費 21H00901、電気通信普及財団 2023 年度 研究調査助成、弥生株式会社共同研究の支援を受けたものである。

参考文献

- [1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners. In **Proc. 33rd NeurIPS**, pp. 1877–1901, 2020.
- [2] C. Cuesta-Lazaro, A. Prasad, and T. Wood. What does the sea say to the shore? a BERT based DST style approach for speaker to dialogue attribution in novels. In **Proc. 60th ACL**, pp. 5820–5829, 2022.
- [3] Yulong Du, 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.
- [4] R. Ishii, R. Higashinaka, K. Mitsuda, T. Katayama, M. Mizukami, J. Tomita, H. Kawabata, E. Yamaguchi, N. Adachi, and Y. Aono. Methods for efficiently constructing text-dialogue-agent system using existing anime characters. **Journal of Information Processing**, Vol. 29, pp. 30–44, 2021.
- [5] 石川和樹, 佐藤理史, 宮田玲, 小川浩平. 複数の手がかりを利用した小説発話の話者推定. 言語処理学会第 29 回年次大会発表論文集, pp. 2170–2175, 2023.
- [6] 石川和樹, 宮田玲, 小川浩平, 佐藤理史. 口調ベクトルを用いた小説発話の話者推定. 情報処理学会研究報告, Vol. 2022-NL-253, No. 14, pp. 1–8, 2022.
- [7] S. Liu, H. Cho, M. Freedman, X. Ma, and J. May. RECAP: Retrieval-enhanced context-aware prefix encoder for personalized dialogue response generation. In **Proc. 61st ACL**, pp. 8404–8419, 2023.
- [8] C. Miyazaki, T. Hirano, R. Higashinaka, and Y. Matsuo. Towards an entertaining natural language generation system: Linguistic peculiarities of Japanese fictional characters. In **Proc. 18th SIGDIAL**, pp. 319–328, 2016.
- [9] G. Muzny, M. Fang, A. Chang, and D. Jurafsky. A two-stage sieve approach for quote attribution. In **Proc. 15th EAACL**, pp. 460–470, 2017.
- [10] K. Yang, D. Klein, N. Peng, and Y. Tian. DOC: Improving long story coherence with detailed outline control. In **Proc. 61st ACL**, pp. 3378–3465, 2023.
- [11] D. Yu, B. Zhou, and D. Yu. End-to-end Chinese speaker identification. In **Proc. NAACL-HLT**, pp. 2274–2285, 2022.
- [12] Y. Zenimoto, S. Komata, and T. Utsuro. Large scale evaluation of end-to-end pipeline of speaker to dialogue attribution in Japanese novels. In **Proc. 37th PACLIC**, pp. 12–23, 2023.
- [13] Y. Zenimoto and T. Utsuro. Speaker identification of quotes in Japanese novels based on gender classification model by BERT. In **Proc. 36th PACLIC**, 2022.
- [14] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, and J.-R. Wen. A survey of large language models. **arXiv e-prints**, pp. 1–97, 2023.
- [15] 錢本友樹, 古俣慎山, 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の評価. 電子通信学会論文誌, Vol. J108-D, No. 04, , 2025.

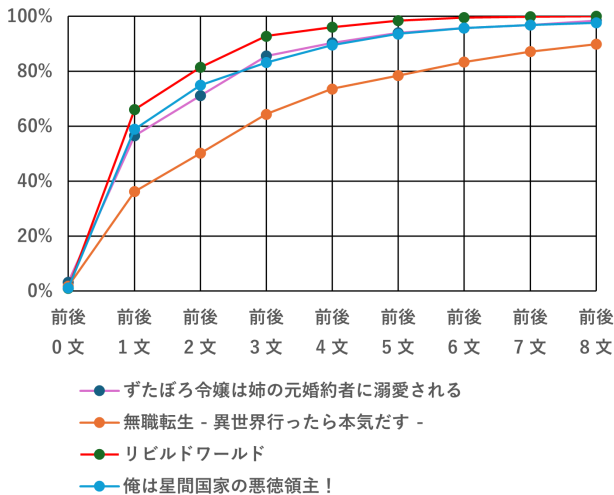


図3 発話文の前後 n 文に参照発話者を表すキャラクター・メンションが存在する発話文の割合 (表1に示す小説を対象とする。)

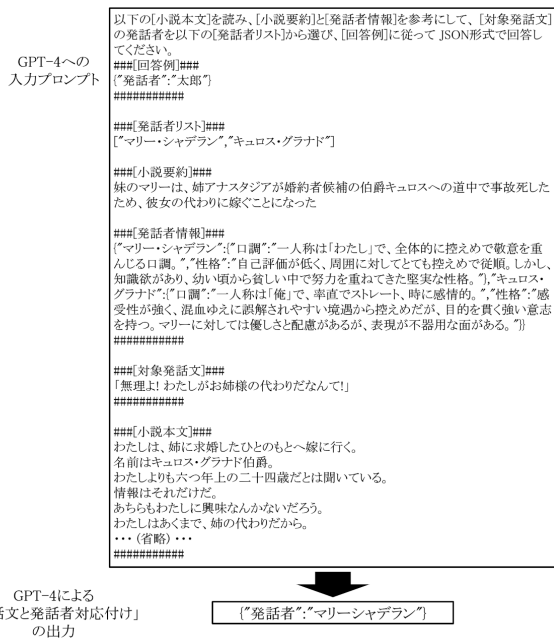


図4 発話文への発話者対応付け時の入力プロンプト例・出力例

A 発話者を表すキャラクター・メンションへの発話文対応付け

文献 [15] では、発話者分類のサブタスクとしてある発話文に対して、その周囲の文章からその発話文の発話者を表す人名を対応付けるサブタスク (「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスク) を提案し、このタスクを行うシステムを開発した。ここで本論文の発話

者分類において、このサブタスクの結果を利用することで、正答率が向上する可能性がある。したがって、4節に示した手続き (P1), 手続き (P2) に対して、「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの結果を適用した場合においても実験を行う⁹⁾。

B ウェブ小説データセット中の参照発話者を表すキャラクター・メンション

3節に示したウェブ小説データセット中の発話文について、その発話文の参照発話者を表すキャラクター・メンションが、発話文の直前および直後の1文、2文、3文といった範囲 (ただし、表1に示す数の「話」の範囲を超えてキャラクター・メンションを検出することはしない) に存在する割合を図3に示す¹⁰⁾。図3中の「前後0文」は、発話文中に参照キャラクター・メンションが存在する発話文を意味している。図3から分かるように、「無職転生-異世界行ったら本気だす-」を除く三つの小説では、前後6文までに参照キャラクター・メンションが存在する発話文の割合が90%以上であることが分かる。一方、「無職転生-異世界行ったら本気だす-」では、他の小説と比べて、参照キャラクター・メンションが周囲に存在する発話文が少ない。この割合は、付録A節に示す「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの難しさに大きく影響する。

9) 文献 [15] は、このサブタスクに対する大規模言語モデルの性能を評価した結果、Claude が最も高い正答率となった。正答率の低いサブタスクの結果が、本論文の発話者分類システムの性能向上に寄与するとは考えにくいため、「発話文に対して発話者を表すキャラクター・メンションを対応付ける」サブタスクの結果の利用には、最も正答率の高い Claude の結果のみを用いる。

10) 本論文では、地の文部分を「。!?」の記号で分割した結果得られる各文を一文として定義する。ただし、発話文の途中では分割を行わず、地の文中に発話文を含む場合でも、それらの発話文を含んだ形で一文が定義される。