

# 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

古俣楨山<sup>1</sup> 錢本友樹<sup>1</sup> 宇津呂武仁<sup>1</sup>

<sup>1</sup> 筑波大学大学院 システム情報工学研究群  
知能機能システム学位プログラム

## 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。本論文では大規模言語モデル・規則に基づく手法・BERTをこの発話者分類タスクに適用した際の性能を『現代日本語書き言葉均衡コーパス』とウェブ小説の話者情報アノテーションデータを用いて評価する。この結果、大規模言語モデルの正答率は他手法より優れた値、あるいは同等の高い値となり、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により、自然な応答をする対話システムや、小説生成システムが提案されている。これらでは、キャラクターや対話システムの個性を制御するが、その個性の評価や分析には、実際に特定のキャラクターの言動を収集する必要がある。文献[1]と文献[2]では、キャラクターの特徴分析と対話システムの構築のために、大人数の協力や特定キャラクターの発言を投稿するBotを利用し、特定のキャラクターの発言を大量に収集している。しかし人手やBotによる収集は、コストが大きく、収集できるキャラクターの種類にも限界がある。そこで、小説中の発話文の発話者を特定するタスクを解くことで、様々なキャラクターの発言を自動的に大量に収集することが可能となり、より多様なキャラクターの特性と行動の詳細な分析が可能になる。

発話者分類タスクは、図1に示すように、1) 小説本文から発話を表す文章を抽出する「発話文抽出」、2) 発話文に対して、その周囲の地の文から、その発話文の発話者を表すキャラクター・メンションを割り当てる「発話文とキャラクター・メンションの対応付け」、3) 抽出したキャラクター・メンションの

うち、同一のキャラクター・エンティティを指すものをクラスタリングする「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」、という三つのタスクから構成されている。英語小説に対する研究では、Muznyら[3]は複数のルールを利用して、上に示した三つのタスクすべてを行う発話者分類システムを提案した。また、Cuesta-Lazaroら[4]は、深層学習を用いた初めての発話者分類システムを提案し、既存のルールベース手法よりも高精度な発話者分類結果を示した。

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されていることから、この大規模言語モデルの発話者分類への適用を試みる。まず、日本語小説に対して、「発話文とキャラクター・メンションの対応付け」と「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」を規則に基づいて行う手法を示し、これをベースラインとする。次に、「発話文とキャラクター・メンションの対応付け」に対して、大規模言語モデルとBERTを適用し、規則に基づく手法との比較を行う。

## 2 関連研究

日本語小説を対象とした発話者分類の先行研究には、地の文を利用した研究[5][6]と口調を利用した研究[7, 8][6]が存在する。文献[5]では、青空文庫の小説4編を対象として、地の文を利用したパターンマッチによる発話者抽出を試みている。この手法では0.72という高い正答率で発話者を抽出できているが、評価に利用した発話文の数は161文と少なく、同一人物を指す発話者のクラスタリングは行っていない。文献[7]と文献[8]の研究では、ライトレ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。

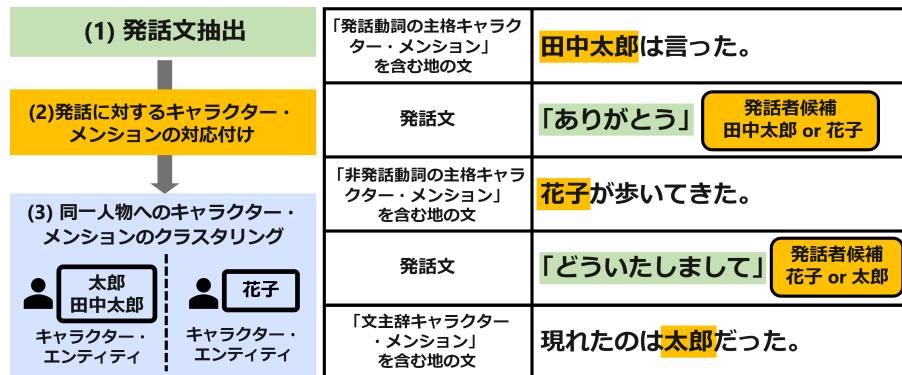


図 1 発話者分類の全体像(本論文では、「地の文」は小説本文中の発話文以外の文章を指す。また、「キャラクター・メンション」は固有名詞(「田中太郎」や「田中」)や代名詞(「彼」や「私」),あるいは「先生」等の役割語といった、地の文中でキャラクターを指し示す表現のことである。加えて、「キャラクター・エンティティ」はキャラクターの実体のことであり、同一人物に対する「田中太郎」、「田中先生」、「先生」、「彼」といったキャラクター・メンションは、「田中太郎先生」という単一の実体である「キャラクター・エンティティ」を指す。)

い。また、文献 [6] の研究では、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、地の文において発話動詞の存在を仮定するなど、用いられているパターンが本論文で利用する規則に基づく手法の一部に留まっている。

### 3 規則に基づく発話者分類

#### 3.1 データセット

『現代日本語書き言葉均衡コーパス』(BCCWJ) [9]の大規模な公開話者情報アノテーションデータセットを利用する。このデータセットには、1986年から2005年に刊行された計2,845個の小説と、167,917文の発話者がアノテーションされた発話文が含まれている。これら2,845小説のうち、60%を訓練データ、20%を検証データ、20%をテストデータとし、評価にはテストデータを用いる。

#### 3.2 発話者抽出

##### 3.2.1 人名抽出・動詞判定

人名抽出では、GiNZAの固有表現抽出器と、日本語WordNet [10]から作成した人名单語辞書を用いて、地の文から人名を抽出する。また、この抽出した人名の掛かり先である動詞が発話を意味する場合、その人名は発話者である可能性が高い。そこで、日本語WordNet [10]から発話動詞辞書を作成し、この辞書中の動詞を発話動詞と判定し、他の全ての動詞を非発話動詞と判定する。

#### 3.2.2 発話文とキャラクター・メンションの対応付け

事前分析から、キャラクター・メンションの多くが地の文の主格か主辞として文に存在することがわかった。そこで本論文では図 1 に示すような3種類の発話者タイプを定義する。

##### 発話動詞の主格位置でのキャラクター・メンション

地の文の発話動詞の主格が人名である場合に、その人名をキャラクター・メンションとして抽出する。

##### 非発話動詞の主格位置でのキャラクター・メンション

地の文の非発話動詞の主格が人名である場合に、その人名をキャラクター・メンションとして抽出する。

##### 文主辞位置でのキャラクター・メンション

地の文の主辞が人名である場合に、その人名をキャラクター・メンションとして抽出する。

特定の発話文の周囲には複数の地の文が存在し、その地の文それぞれでキャラクター・メンションが存在する場合は以下の二つの手続きを順に適用する。

(P1) 各発話文の前後1文までの地の文中に存在するキャラクター・メンションを検出する。その範囲に複数のキャラクター・メンションが存在する場合は、そのキャラクター・メンションと発話文の相対位置に着目し、発話文を含む地の文、発話文の後ろの地の文、発話文の前の地の文という優先順位でキャラクター・メンションを抽出する。

(P2) 手続き (P1)において、二つの発話文で同一の

キャラクター・メンションが抽出された場合、そのキャラクター・メンションを両方の発話文に対応させる。

続いて、手続き (P1) 及び手続き (P2) 後にキャラクター・メンションが対応付けされなかった発話文を対象として、以下の二つの手続きを適用する。

#### 話者交代制約

連続した発話文において、 $n$  番目の発話の発話者が不明で、 $(n \pm 2)$  番目の発話の発話者が判明している場合、 $(n \pm 2)$  番目の発話者を  $n$  番目の発話者とする。本論文では、連続した発話文を、発話文間に地の文が存在しない発話文の集合とする。

#### 発話者としての頻度が最大となる人名の適用

上記手続きで対応付けられなかった発話文は、当該小説内で対応付けられた頻度が最も大きい人名を対応付ける。

以上の手順による発話者抽出手法を「規則に基づく標準手法」とする。この手法の手順を変形した発話者抽出手法については A 節に記載する。

### 3.3 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

最後に、3.2.1 節で抽出した人名のうち、「固有名詞」か「役割語(先生・将軍など)」が一致した人名は同一人物を指しているとしてクラスタリングする。例えば、「田中先生」、「先生」、「田中太郎さん」、「ミスター太郎」という四つの人名が抽出された場合は、それぞれ共通部分を持つため、全て同一人物として分類される。

## 4 大規模言語モデルによる発話者分類

次に、「発話文とキャラクター・メンションの対応付け」タスクに対して、大規模言語モデルを適用する。大規模言語モデルとしては、OpenAI 社が提供する ChatGPT<sup>1)</sup> モデルである GPT-3.5<sup>2)</sup> および GPT-4<sup>3)</sup> を利用する。大規模言語モデルと規則に基づく手法の比較評価を行うために、ウェブ上に公開されている小説「ずたぼろ令嬢は姉の元婚約者に溺愛される」<sup>4)</sup> の第 1 話から第 97 話の小説本文に、発話文の発話者情報を付与したデータセット(以下

ウェブ小説データセットとする)を利用する<sup>5)</sup>。B.2 節の手順に従い、このデータセットにおける参照発話者の出現位置の統計を求めた結果においては、前後 1 文の主格・文主辞位置に参照発話者が存在する発話文が 929 (23.3%) 個、その位置以外に参照発話者が存在する発話文が 3,064 (76.7%) 個となった。このうち、前後 1 文の主格・文主辞位置以外に参照発話者が存在する 7 割以上の発話文に対しては、規則に基づく手法では、「話者交代制約」や「発話者としての頻度が最大となる人名の適用」でしか対応付けを行えない。しかし、この対応付けは発話文の前後の地の文からキャラクター・メンションを抽出する対応付けよりも、B.2 節に示す、発話者分類性能の評価指標である人名一致正答率が低くなる点が課題である。

これに対して、大規模言語モデルによる発話者対応付けの場合には、前後 1 文の主格・文主辞位置以外に参照発話者が存在する 7 割以上の発話文の場合でも、適切な対応付けが行えることが期待できる。一例として、図 2 の例の場合には、規則に基づく手法では適切な発話者対応付けが行えないのに対して、ChatGPT モデルを適用した場合には適切な発話者対応付けを行うことができる。図 2 においては、参照発話者である「ジュニア」が発話文の後ろの 2 文目に含まれている。ここで、規則に基づく手法では、誤った発話者「私」を発話文に対応付ける<sup>6)</sup> のに対して、大規模言語モデルでは、参照発話者である「ジュニア」を発話文に対応付ける。このように、大規模言語モデルを利用することによって、規則に基づく手法を上回る性能を達成できることが期待できると考えられる。

実験では、大規模言語モデルがキャラクター・メンションを抽出する範囲を、前後 8 文とした場合・前後 1 文とした場合についてそれぞれ発話者分類性能を調べる。この各条件でモデルが抽出したキャラクター・メンションを 3.3 節に示した手法でクラスタリングし、性能を評価する。

5) 利用規約に反するため、BCCWJ は ChatGPT の性能分析に利用できない。

6) 規則に基づく手法では、3.2.2 節で示す手法に従い、まず発話文の前後 1 文を参照する。この時、発話文の後ろには「メランダとは、私の呼び名だ。」という地の文が存在するが、この文からはキャラクター・メンションを抽出できない。次に「話者交代制約」による発話者の対応付け処理を行うが、この例では前後に他の発話文が無いため、発話者の対応付けができない。最後に、「発話者としての頻度が最大となる人名の適用」によって、発話者として「私」が対応付けられるが、この対応付けは誤りである。

1) <https://platform.openai.com/docs/models/>

2) gpt-3.5-1106

3) gpt-4-1106-preview

4) <https://ncode.syosetu.com/n1860fv/>

表1 BCCWJ テストデータ・ウェブ小説データセットに対する各モデルの人名一致正答率

手法	BCCWJ			ウェブ小説データセット		
	人名一致正答率(%) ( $\frac{\text{正答数}}{\text{該当発話文数}}$ )			人名一致正答率(%) ( $\frac{\text{正答数}}{\text{該当発話文数}}$ )		
	(A)	(B)	(C)	(A)	(B)	(C)
規則に基づく標準手法	45.2 ( $\frac{15,082}{33,336}$ )	87.1 ( $\frac{10,427}{11,978}$ )	73.2 ( $\frac{11,268}{15,396}$ )	45.2 ( $\frac{1,806}{3,993}$ )	90.2 ( $\frac{838}{929}$ )	72.9 ( $\frac{1,170}{1,606}$ )
発話動詞の 主格キャラクター・メンション優先	45.3 ( $\frac{15,091}{33,336}$ )	86.3 ( $\frac{10,343}{11,978}$ )	72.7 ( $\frac{11,200}{15,396}$ )	45.5 ( $\frac{1,815}{3,993}$ )	91.2 ( $\frac{847}{929}$ )	73.4 ( $\frac{1,179}{1,606}$ )
発話文の前の地の文中の キャラクター・メンション優先	44.0 ( $\frac{14,667}{33,336}$ )	83.1 ( $\frac{9,949}{11,978}$ )	70.2 ( $\frac{10,801}{15,396}$ )	46.3 ( $\frac{1,847}{3,993}$ )	94.1 ( $\frac{874}{929}$ )	75.2 ( $\frac{1,207}{1,606}$ )
キャラクター・メンションと 発話文を1対1対応	43.7 ( $\frac{14,554}{33,336}$ )	79.8 ( $\frac{9,560}{11,978}$ )	68.1 ( $\frac{10,489}{15,396}$ )	43.1 ( $\frac{1,720}{3,993}$ )	83.9 ( $\frac{779}{929}$ )	69.1 ( $\frac{1,110}{1,606}$ )
BERT	29.2 ( $\frac{9,721}{33,244}$ )	66.2 ( $\frac{7,914}{11,953}$ )	58.2 ( $\frac{8,934}{15,349}$ )	16.2 ( $\frac{647}{3,993}$ )	40.7 ( $\frac{378}{929}$ )	33.6 ( $\frac{540}{1,606}$ )
GPT-3.5	—	—	—	48.5 ( $\frac{1,935}{3,993}$ )	51.7 ( $\frac{481}{929}$ )	54.6 ( $\frac{877}{1,606}$ )
GPT-3.5(前後1文)	—	—	—	33.5 ( $\frac{1,337}{3,993}$ )	89.1 ( $\frac{828}{929}$ )	79.8 ( $\frac{1,281}{1,606}$ )
GPT-4	—	—	—	59.5 ( $\frac{2,374}{3,993}$ )	63.8 ( $\frac{593}{929}$ )	66.6 ( $\frac{1,070}{1,606}$ )
GPT-4(前後1文)	—	—	—	35.4 ( $\frac{1,415}{3,993}$ )	93.0 ( $\frac{864}{929}$ )	84.2 ( $\frac{1,352}{1,606}$ )

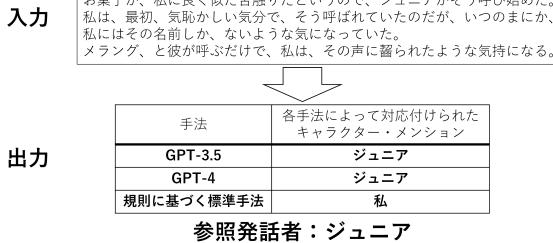


図2 規則に基づく標準手法と大規模言語モデルの比較

## 5 BERT モデルによる発話者分類

最後に、「発話文とキャラクター・メンションとの対応付け」タスクに BERT [11] を適用し、大規模言語モデル・規則に基づく手法との比較を行う。この比較にあたり、事前学習済モデルである東北大版BERT<sup>7)</sup>を利用する。BERT の訓練には3.1節に述べた BCCWJ の訓練・検証データを使用する。発話文の前後2文に対してキャラクター・メンションの抽出を行い、4節と同様にキャラクター・メンションをクラスタリングし、発話者分類性能を評価する。

## 6 評価

表1に BCCWJ テストデータ・ウェブ小説データセットに対する大規模言語モデル・規則に基づく手法・BERT による発話者分類の評価結果を示す。評価基準にはB.2節に示した人名一致正答率を用いる。表1について規則に基づく手法と BERT を比較

7) <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>

すると、BCCWJ・ウェブ小説データセットのどちらにおいても、規則に基づく手法が BERT より優れていることが分かる。次に、ウェブ小説データセットでは、人名一致正答率(A)、人名一致正答率(C)では GPT-4 が最も高い値であるのに対して、人名一致正答率(B)では規則に基づく手法が最も高い値であることが分かる。この原因として、規則に基づく手法は、キャラクター・メンションを抽出できなかった発話文に対して、「話者交代制約」と「発話者としての頻度が最大となる人名の適用」によって適切な人名を対応付けられる場合があることが挙げられる。一方で、同様の場合でも、大規模言語モデルによる発話者分類ではそうした対応付けを行っていないため、適切な人名を対応付けられない。したがって、GPT-4 の結果に対して同様の処理を追加することで、人名一致正答率(B)においても規則に基づく手法に劣らない発話者分類を行えると考えられる<sup>8)</sup>。

## 7 おりわりに

本論文では、日本語小説に対して、大規模言語モデル・規則に基づく手法・BERT による発話者分類の性能評価を行った。この結果、大規模言語モデルの正答率は他手法より優れた値、あるいは同等の高い値となり、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。

8) 人名一致正答率(B)の範囲で、GPT-4(前後1文)は10個の発話文に対して「不明」という結果を出した。この10個の発話文に対して、人手で「話者交代制約」と「発話者としての頻度が最大となる人名の適用」によって人名を対応付けた結果、8件の発話文に適切な人名が対応付けられた。この結果、GPT-4(前後1文)の人名一致正答率(B)は93.9%となり、規則に基づく手法と同等の正答率となった。

## 参考文献

- [1] 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.
- [2] 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.
- [3] G. Muzny, M. Fang, A. Chang, and D. Jurafsky. A two-stage sieve approach for quote attribution. In **Proc. 15th EACL**, pp. 460–470, 2017.
- [4] 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.
- [5] D. 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.
- [6] 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.
- [7] 石川和樹, 宮田玲, 小川浩平, 佐藤理史. 口調べ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.
- [8] Yuki Zenimoto and Takehito Utsuro. Speaker identification of quotes in Japanese novels based on gender classification model by BERT. In **Proc. 36th PACLIC**, 2022.
- [9] K. Maekawa, M. Yamazaki, T. Ogiso, T. Maruyama, H. Ogura, W. Kashino, H. Koiso, M. Yamaguchi, M. Tanaka, and Y. Den. Balanced corpus of contemporary written Japanese. **Language Resources and Evaluation**, Vol. 48, No. 2, pp. 345–371, 2014.
- [10] F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi, and K. Kanzaki. Enhancing the Japanese WordNet. In **Proc. 7th ALR**, pp. 1–8, 2009.
- [11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proc. 14th NAACL-HLT**, pp. 4171–4186, 2019.
- [12] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. **Information retrieval**, Vol. 12, pp. 461–486, 2009.

## A 「規則に基づく標準手法」およびその変形版の説明

3.2.2 節に述べた手法(規則に基づく標準手法)に対して、以下の点を変形することで手法の変形版が得られる。

- (1)発話・非発話動詞の主格位置・文主辞位置のキャラクター・メンションをどの優先順序で抽出するか。
- (2)発話文を含む地の文・発話文の後ろの地の文・発話文の前の地の文に存在するキャラクター・メンションをどの順序で抽出するか。
- (3)一つのキャラクター・メンションに、一つの発話文のみを対応付けるか、または複数の発話文を対応付ける例を受け入れるか。

これらの点の変形版について以下に示す。

### A.1 規則に基づく標準手法

この手法は 3.2.2 節に述べた手法に従ってキャラクター・メンションと発話文の対応付けを行う。この時、上に示した変更可能な点を以下のように設定している。

- (1)優先順序を付けずに抽出する。
- (2)発話文を含む地の文、発話文の後ろの地の文、発話文の前の地の文、の優先順序で抽出する。
- (3)一つのキャラクター・メンションに複数の発話文を対応付ける例を受け入れる。

### A.2 発話動詞の主格キャラクター・メンション優先

この手法は、「規則に基づく標準手法」の設定のうち(1)を変更し、発話動詞の主格位置、非発話動詞の主格位置、文主辞位置、の順でキャラクター・メンションを抽出する。

### A.3 発話文の前の地の文中のキャラクター・メンション優先

この手法は、「規則に基づく標準手法」の設定のうち(2)を変更し、発話文を含む地の文、発話文の前の地の文、発話文の後ろの地の文、の順にキャラクター・メンションを抽出する。

### A.4 キャラクター・メンションと発話文を1対1対応

この手法は、「規則に基づく標準手法」の設定のうち(3)を変更し、一つのキャラクター・メンションに一つの発話文のみを対応付ける。

## B 規則に基づく手法の評価手順

発話者分類はテストデータを対象として行い、クラスタリング評価と人名一致正答率の2種類の評価を行う。

### B.1 クラスタリングの評価手順

3.3 節における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結果に対して、まず、一つの人名のクラスタを  $N$ 、人名のクラスタの集合を  $\mathbb{N}$  と表記する。そして、発話者が  $N$  内の人名と予測された発話文の集合を予測クラスタ  $P$  とする。この処理を  $\mathbb{N}$  全体に適用し、得られた予測クラスタ  $P$  の集合を  $\mathbb{P}$  と表記する。同様に、アノテーションデータから得られた参照クラスタ集合を  $\mathbb{R}$  とし、 $\mathbb{R}$  中の一つのクラスタを  $R$  と表記する。ここで、

本論文では、評価指標として、Cuesta-Lazaro ら [4] と同様に  $B^3$  の適合率、再現率、F1 値 [12] を用いる。 $B^3$  では、発話文一つ一つに対して以下の二つの指標を計算する。

1. 適合率: 一つの予測クラスタ  $P$  内の発話文のうち、参照クラスタ集合  $\mathbb{R}$  において発話者が同一であるとされている発話文の割合。
2. 再現率: 一つの参照クラスタ  $R$  内の発話文のうち、予測クラスタ集合  $\mathbb{P}$  において発話者が同一であると予測された発話文の割合。

そして、予測クラスタ集合  $\mathbb{P}$  および参照クラスタ集合  $\mathbb{R}$  の全体に対して発話文全体における適合率・再現率のミクロ平均を算出し、これをクラスタリングにおける適合率・再現率とする。

### B.2 人名一致正答率の評価手順

BCCWJ では各発話文に対して一意の参照発話者名を与えており、そこで、予測発話者が正しく参照発話者を指す割合を評価するため、人名一致正答率という評価指標を定義する。人名一致正答率の算出においては、まず発話文の予測発話者と参照発話者との間で 3.3 節の手法を適用し、予測発話者と参照発話者が同一人物かどうかを判定する。そして、この同一人物判定を全ての発話文に対して行い、予測発話者と参照発話者が同一人物であると判定された発話文の割合を「人名一致正答率(A)」として評価する。ここで、テストデータ中の 21,358 (64.1%) 個の発話文は、前後 1 文の主格・文主辞位置に参照発話者を指すキャラクター・メンションが存在しないため、規則に基づく手法によって前後 1 文の範囲からは参照発話を対応付けることはできない。そこで、地の文を利用した規則に基づく手法の有効性を分析するため、それらの文以外の 11,978 (35.9%) 個の発話文のみを対象とした場合について、予測発話者と参照発話者が同一人物であると判定された発話文の割合を「人名一致正答率(B)」として評価する。また、前後 1 文の任意の位置に参照発話者を指すキャラクター・メンションが存在する発話文はテストデータ中に 15,396 (46.2%) 個存在した。これらの発話文のみを対象とした場合について、予測発話者と参照発話者が同一人物であると判定された発話文の割合を「人名一致正答率(C)」として評価する。

発話者分類の最終的な目標は、人名一致正答率を 100% にすることである。しかし、規則に基づく手法では、予測発話者が代名詞や愛称のときに正しく評価できないため、B.1 節で述べたクラスタリングによる評価も行う。