

# コミックのセリフと発話者対応付けデータセットの構築と その困難性

阿部 和樹<sup>†,a</sup> 中村 聡史<sup>††,b</sup>

<sup>†</sup> 明治大学大学院 先端数理科学研究科 <sup>††</sup> 明治大学 総合数理学部

a) ai3ekzk@gmail.com b) satoshi@snakamura.org

**概要** 本研究では、我々がこれまで実現した Manga109 を対象としてセリフと発話者を対応づけるデータセット構築システムを利用し、147,918 件のセリフと発話者の対応付けデータセットを構築した。また構築したデータセットを Web 公開した。さらに、構築したデータセットを分析することにより、人手によるデータセット構築の困難性 (2 人による一致率 86.9% など) についても明らかにした。一方、セリフがあるコマとキャラクタおよび発話者の存在に関する分析を行うとともに、吹き出しの位置やしっぽの方向のみならず、一人称や語尾といった特性を利用した簡易的な自動推定手法を提案し、組み合わせた手法により 70% 程度で推定できたことが分かった。

**キーワード** セリフ, 話者, 自動推定, データセット構築, 公開

## 1 はじめに

全国出版協会による出版市場調査[1]によると、2017 年に初めて電子コミックの売り上げが紙媒体のコミックの販売を上回るという結果となった。また、同協会による 2019 年上半期の調査[2]では、電子コミックの売り上げは 1,133 億円にのぼっている。このことより、デジタル端末を利用して電子コミックとして漫画を楽しむことが一般的になっており、デジタルならではの楽しみも今後増えていくと期待される。

コミックのデジタルならではの楽しみを増やしたり、デジタルならではの処理を行ったりする研究は増加している。こうした研究を推進するものとして、プロの漫画家によって描かれた 109 冊の漫画からなる Manga109 データセットがアノテーションとともに公開[3][4]されており、様々な研究に活用されている。また、Guérin ら[5]は eBDtheque というデータセットを、Iyyer ら[6]は COMICS というデータセットを公開している。一方、4 コマ漫画に特化しているが、上野ら[7]は 4 コマ漫画のストーリーの構造をデータセットとして作成している。

ここで、コミックの翻訳[15]や、コミックの内容をもとにした検索[16]・推薦[17]、ネタバレ防止[18]などを可能にするには、そのコミック内のコマ推定[8]やコミック内の顔認識[9]、セリフ認識や吹き出しの検出[10]、セリフの吹

き出しの形からの感情推定[11]のみならず、セリフの話者が誰なのかということ推定する必要がある。Manga109 にはこの対応付けに関するアノテーションデータは存在していないが、これまでも話者推定に関する研究は行われてきており、Riguard ら[12]は吹き出しのしっぽからの距離による推定手法を、山本ら[13]はキャラクタの表示サイズやしっぽの方向などを用いた手法を提案し、その有用性を検証している。我々もセリフからの話者推定に向けて、これまで多くのコミックに関する分析を行い、その困難性について整理するとともにセリフと話者の対応付けを可能とするデータセット構築手法を実現してきた[14]。しかし、これまでの研究は小規模なデータセットをもとにしたものであり、また我々の研究も分析とデータセット構築手法を実現したにすぎなかった。

そこで本研究では、我々が過去に実現したセリフと話者の対応付けを可能とするデータセット構築手法を運用し、Manga109 データセットのすべてのコミックを対象としたセリフ・話者対応付けデータセットを構築する。また、そのデータセットをもとに、コミックの特性や、ひとによるアノテーションの難しさといった点について分析を行う。さらに、吹き出しと話者の位置や、吹き出しのしっぽの方向、一人称や語尾といった特性を利用した簡易的な手法を適用することにより、自動推定可能性について検討を行う。

Copyright is held by the author(s).

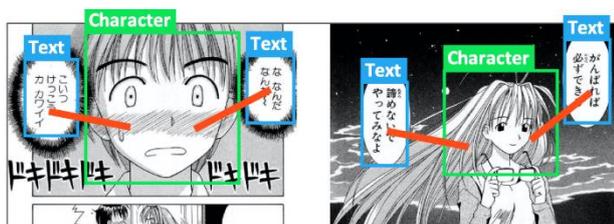
The article has been published without reviewing.

## 2 セリフ・話者対応付けデータセット構築手法

ここでは、まず我々が過去に提案・実現してきたセリフ・話者の対応付け手法[14]について説明する。

今回利用するコミック 109 冊からなる Manga109 だけで、その総発話数は 147,918 件あり、1 冊平均で 1,357 発話ある計算となる。1 冊あたり 1,357 件の発話について、1 発話ずつ手作業で登場人物リストから指定していくのは容易ではなく、時間を浪費してしまう。

ここでコミックは、絵とセリフが一体化したコンテンツであり、そのセリフの発話者(キャラクタ)が、そのコミックの登場人物のうち誰にあたるのかを読者が即座に理解できるよう工夫されていることが多い。セリフの発話者が誰であるかを視覚的にわかりやすくするには、ひとがあるオブジェクトの集合からグループを知覚する際に自然と使っており、デザイン原理としても有名な、ゲシュタルトの法則の近接の要因や閉合の要因が重要になる。実際、あるセリフとその発話者(登場人物)の絵は、近くに配置されることが多く(近接の要因)、またそのセリフと発話者は同じコマ(1 つのフレーム内)に存在していることが多い(閉合の要因)。

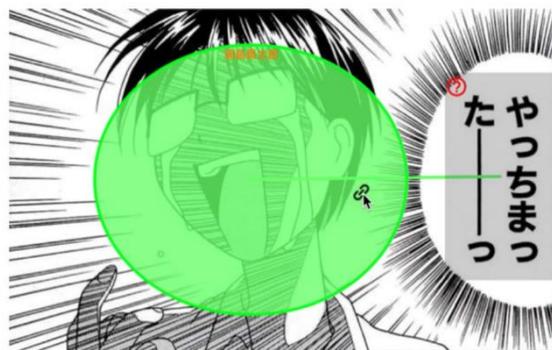


©赤松健「ラブひな」

図 1. Text(セリフ)と Character(発話者)の対応付け

この、セリフと発話者は近くに存在している可能性が高く、発話者はその識別性を高めるためある程度大きく提示されているという特性を利用し、我々はセリフをキャラクタの顔領域+体領域にマウスでドラッグ・アンド・ドロップすることにより、割り当てを行う手法を提案し、実現してきた(図 2)。この手法は、そのマウス操作における距離の近さや、マウスによる操作対象の大きさを考慮すると、フィッツの法則の特性からも効率的であるといえる。

なお、セリフと発話者が別のページに存在しているなど、セリフが発話者の近くに存在しない可能性もある。そのため、コミック内のセリフをクリックしたうえで、キャラクタリストから発話者を選択するインタフェースも合わせ実現している。ここでは、キャラクタが多く名前だけでは把握できないことも考慮し、キャラクタの顔画像も併せて提示している(図 3)。



©赤松健「ラブひな」

図 2. システムを用い、ドラッグ・アンド・ドロップでセリフを発話者に対応付けしている様子



©赤松健「ラブひな」

図 3. セリフのクリックとキャラクタリストからの選択による対応付けインタフェース

1 ページあたりに多数のセリフが存在しており、どのセリフについて対応付けを行ったのかわからなくなったり、同じセリフに複数人を付与してしまったりといった問題があったため、図 4 に示す通り、セリフと発話者が対応付けられているセリフは緑色のアイコン、複数の発話者に対応付けられているセリフにはその数、対応付けられていないセリフには赤色のアイコンを示すようにした。また、発話者がわからない場合は、「不明」のタグを付与してもらうようにし、オレンジ色のアイコンを示すようにした。

またコミックは、1 冊あたり約 200 ページからなるため、すべてを最初から順に、一度に対応付けすることは難しい。例えば、現時点で読んでいて発話者がわからないセリフは後回しにしておき、後で理解した後に戻ってきてセリフと発話者を割り当てるといったことが考えられる。そこで、ページ単位での管理インタフェースも用意し、それを利用して全体を把握しながらアノテーション付与可能とした(図 5)。



©八神健「密・リターンズ」

図 4. 操作 UI. アイコンの色によりアノテーション付与状況が提示(緑色: 対応付け済, 緑色と数字: 複数に対応付け済, 赤色: 未対応付け)

001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016
017	018	019	020	021	022	023	024	025	026	027	028	029	030	031	032
033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048
049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064
065	066	067	068	069	070	071	072	073	074	075	076	077	078	079	080
081	082	083	084	085	086	087	088								

図 5. 進捗状況を色により提示

### 3 セリフ・発話者対応付けデータセット

前章で説明したアノテーション付与システムを用いて、Manga109 のコミックに登場するセリフを発話者に対応付けるといったデータセットの構築を行う。また、データセットからセリフに対する発話者の正解データセットを構築するとともに、その特性について分析を行う。

#### 3.1 データセット構築手順

データセット構築の協力者は、我々が実現した Web システムから、1 冊ずつコミックを登録して構築作業(アノテーション付与)を行った。ここで、データセット構築者にアノテーション付与してもらおう対象のコミックは、ランダムに選定されることとし、1 冊完了することで新しいコミックを追加することができるものとした。また、アノテーションが難しいことを考慮し、アノテーション付与を完了しているものを除き、3 冊まで対象として登録可能とした。図

6 はそのアノテーション付与に関するトップページであり、ログインユーザに対してアノテーション対象が割り当てられている。また、完了したコミックはグレーアウトされ、画面左下の「作品を追加」ボタンを利用することで、作品を追加することができるようになっている。

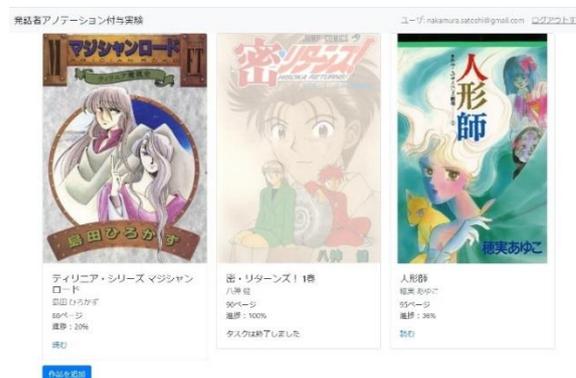


図 6. アノテーション対象選択 UI

ここで、キャラクターには「不明」という選択肢を用意しており、発話者がわからないものについては「不明」を選択するようにデータセット構築者に指示した。また、セリフの中には状況の説明や注釈など、明確にキャラクターの発話ではないものも存在したため、これらのセリフについては「ナレーション」というタグを選択するように指示した。

すべてのセリフについて発話者の対応付けが終わると、図 5 に示す進捗状況がすべて緑色かオレンジ色になり、作業完了を行うことができる。作業完了したコミックについては、図 6 中央のように表紙がグレーアウトされていく。

#### 3.2 データセット構築結果と分析

実験では Manga109 の、109 冊全てのコミックに登場するセリフに対して、2 名ずつにアノテーション付与を行ってもらった。その結果、218 冊分(109 冊×2 人分)のデータセットを構築することができた。データセット構築の協力者は、著者ら 2 名と、明治大学に通う学生 31 名の合計 33 名であった。なお、データセット構築において、データセット構築者の時間的な都合または難易度の理由で、途中で断念されたものが 28 冊、手を付けずに断念されたものが 27 冊あった。これは多くの協力者に依頼したことが原因であると考えられる。

構築したデータセットは中村聡史研究室のウェブページ<sup>1</sup>にて公開している。

表 1 に、全 147,918 発話について、それぞれどのようなアノテーションの付与がなされたかの結果を示す。

<sup>1</sup> <https://nkmr.io/comic/speaker-dataset/>

表 1. アノテーション付与結果

意見	内容	データ数	割合
一致	同じ人物を選択	128,502 件	86.9%
	「ナレーション」を選択	3,531 件	2.4%
	「不明」を選択	414 件	0.3%
不一致	異なる人物を選択	9,720 件	6.6%
	片方が「不明」を選択	2,433 件	1.6%
	その他	3,318 件	2.5%

この結果より、2名のデータセット構築者が、あるセリフについて同じキャラクターを発話者として選択された割合は86.9%であり、ナレーションを含めても89.3%であることがわかる。この数値は高いようにも考えられるが、Manga109 データセットにおける見開き2ページあたりの平均セリフ数は約14件であったため、1つの見開き当たり約1.8件程度の発話者の不一致があった計算になる。このことから、ひとによる判断でもセリフと発話者の割り振りは容易ではないことがわかる。

なお、2.4%がナレーションとなっていたが、これは図7に示すような4コマの話のタイトルやト書き部分がセリフとして事前にアノテーションされており、そうしたデータが多かったためである。こうしたタイトルやト書きをどのように扱うかは課題である。



図 7. ナレーションと付与された例

次に、2人のデータセット構築者の対応付けが一致している128,502件のセリフのうち、発話者が不明であるもの、発話者が複数であるものの計7,138件を除いた121,364件のセリフについて、そもそもセリフと発話者は同じコマにいたのかどうかを分析したものが表2である。

表 2. セリフがあるコマとキャラクターの関係

	件数	割合
コマにキャラクターがいない	10,317	8.5%
コマにキャラクターが1人だけいる	44,992	37.1%
コマにキャラクターが複数いる	66,055	54.4%
コマにキャラクターとして発話者のみ	39,098	32.2%
コマに発話者が存在している	103,482	85.3%

この結果より、あるセリフが与えられたときに、そのコマに1人のキャラクターしかいないというコマは37.1%しかなく、またセリフがあるコマの内8.5%もコマにキャラクターがいないことがわかる。このことから、コマ内だけの推定では難しいことがわかる。さらに、54.4%のコマにおいては複数のキャラクターがいるため、コマ内での対応付けも重要になる。一方、そのコマに1人しかいないときに、そのキャラクターが発話者である確率は86.9%であることがわかる。このことより、コマにセリフがあり、キャラクターが1人しかいない場合は、高精度に発話者を推定できることがわかるが、13.1%の確率で予想を外してしまうこともわかる。

また、セリフがあるコマに、発話者が存在している確率は85.3%の確率であるが、複数のキャラクターがいることが多く、ランダムに発話者を推定した場合の正答率は60.6%であることがわかった。

### 3.3 データセット構築者からのフィードバックと考察

データセット構築者に対してフィードバックを求めたところ、選択に迷った具体的な例として下記のようなものが得られた。

- ナレーションとモノローグが判別できない場合
- アナウンスなど発話者がその場に存在しない場合(図8)
- キャラクターが人形に乗り移っている場合(図9)
- キャラクターに他者が乗り移っている場合(図10のように、キャラクターAが死亡してキャラクターBに乗り移っており、その主体となる人物が入れ替わるなど)
- 1コマに多くの登場人物が存在する場合
- キャラクター名が明らかでない者が発話者の場合

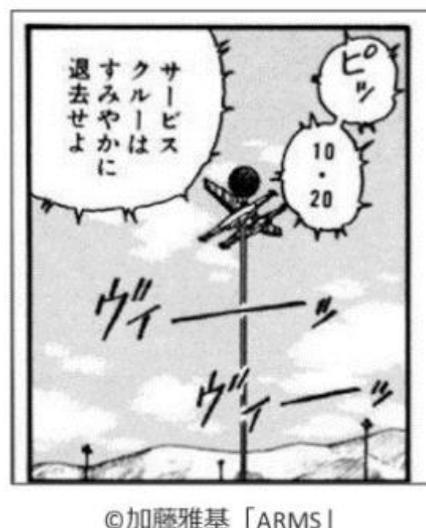


図 8. 発話者の選択に困った具体例



©穂実あゆこ「人形師」

図 9. 人形を遠隔操作し、話しかけているのはこの場にはいないキャラクター



©八神健「密・リターンズ!」

図 10. ある登場人物に、他者が乗り移っている

また、以下はフィードバックの中でも、データセットや実験環境の問題によって発生したものである。

- コミックの 1 巻目ではなく続巻から読む場合
- 異なるセリフが 1 つのセリフとして扱われている場合(図 11)
- 作者のコメントや効果音などセリフとは無関係のテキストである場合

この、コミックの 1 巻目ではなく続巻から読む場合というのは、Manga109 において 1 巻目ではない作品の総数が 18 巻あることが関係している。これらの 18 巻は、1 つの作品の 1 巻目・2 巻目と連続してデータセットに含まれているものもあれば、1 巻と最終巻の 2 つのみ、2 巻目のみなど様々なパターンが存在する。これらのコミックは 2 巻以降の続巻から読む場合、そのコミックの知識がないデータセット構築者にとっては難しいタスクになって

いたと考えられる。この、続巻から多人数で大規模なデータに対するアノテーションデータセットを構築する際にありがちな問題である。将来的には、コミックのセリフと発話者の対応付けについて、コマ単位に分解し、10 秒程度で実施可能なマイクロタスクなどの形で実現したいと考えているが、こうした問題はそのハードルを高めているともいえる。

また、図 11 のような「2 つの異なるセリフが 1 つのセリフとして扱われていた場合」については、本来であれば 2 つの異なるセリフとして扱われるはずのものが 1 つのセリフとして登録されていたことが原因となる。これらの多くは吹き出しが連結したものが多く、コミック画像からテキストの位置や文字列を自動で推定する研究であれば特に問題がないが、今回のようなセリフを発話者ごとに対応付ける場合は大きな障害となる。



©愛田真夕美「魔夜の赤い靴」

図 11. 吹き出しがくっついているが、複数の発話者による対話の例

## 4 自動推定手法

本章では本論文の目的であるセリフ発話者の自動推定について、実際に推定を行いその精度を評価する。また、今回は 1 つのセリフに対して発話者が 1 人の場合に限定し、発話者の推定を行う。

### 4.1 セリフと発話者の対応付けの手掛かり

我々のこれまでの研究[14]および 3 章の結果より、セリフを判断するための手がかりは大きく分けて「吹き出しの形状」「セリフとキャラクターの位置関係」「セリフの表現とキャラクターの特性」の 3 つが存在する。本論文はこのそれぞれの手がかりを利用し組み合わせることで発話者を推定する手法を提案する。今回は以下に示す 4 種類の手法によってセリフの発話者の推定を行う。

1. セリフとキャラクターの距離の情報による推定
2. 同じコマ内にいるキャラクターの情報による推定
3. 吹き出しのしっぽの方向の情報による推定
4. 一人称と語尾の情報による推定

これらのそれぞれの推定によって、発話者の候補となるキャラクタにスコアを割り振ることとする。以下ではそれぞれの手法について詳しく述べる。

#### 4.2 セリフとキャラクタの距離の情報による推定

「セリフとキャラクタの位置関係」の手がかりの1つとして、セリフとキャラクタが近い距離に存在する可能性が高いというものがある。Rigaudら[12]の発話者推定の研究についても、この手法によって発話者の推定を行なっている。

今回はそのセリフとキャラクタの距離の計算について、Manga109のアノテーション情報より、セリフおよびキャラクタの画像上の位置(矩形)の情報を利用する。具体的には、推定対象のセリフ領域の中心座標を求め、同じページに存在するキャラクタ領域の中心座標をそれぞれ求める。このセリフとキャラクタの距離をもとに、距離が近いほどスコアが高くなるように計算を行なった。そのため、セリフに近いほど発話者と推定される可能性が高くなる。

#### 4.3 同じコマ内にあるキャラクタの情報による推定

3.2節でも述べた通り、セリフを判断する手がかりの1つは、セリフとその発話者であるキャラクタが同じコマに存在する可能性が高い。そこで、セリフが所属するコマに対して、同一コマに存在するキャラクタを発話者として推定する。

具体的には、Manga109のコマ領域に関する情報をもとに、セリフと同じコマに所属するキャラクタを発話者の候補とし、発話者らしさを示すスコアを割り振る。スコアの合計は1になるように正規化しており、例えば同じコマに所属するキャラクタが1名の場合はそのキャラクタが1点、キャラクタが2名の場合はそれぞれ0.5点といった具合にスコアリングを行う。

ここで、この手法では同じコマに所属するキャラクタが複数人いる場合、発話者を1人に絞ることができないという問題が存在する。そのため、本手法は他の手法と組み合わせることで推定の役割を果たすことができる。例えば、4.2節で説明した距離による手法について、同程度の距離に2名の発話者候補がいる場合に、同一コマ内かといった情報が利用できると思われる。

#### 4.4 吹き出しのしっぽの方向の情報による推定

発話者を判別する手がかりの1つである「吹き出しの形状」より、吹き出しのしっぽの方向によって発話者を推定する。この吹き出しのしっぽの検出については、田中ら[11]の手法を参考に、吹き出しを検出する。ここでは、吹き出しの輪郭線を検出し、その曲率を求めることで吹き出しのしっぽとなる座標を求めた。

発話者の候補は、吹き出しのしっぽの先端から、セリ

フの重心とは逆方向に線分を伸ばし、線分とキャラクタの矩形が交差したものを発話者として推定することとした。この線分の長さについては、コミック画像の対角線の10%の長さに固定した。また、線分と交差するキャラクタが複数名いる場合、それぞれに対して均等にスコアを割り振ることとした。

#### 4.5 一人称と語尾の情報による推定

発話者を判別する手がかりの1つである「セリフの表現とキャラクタの特性」より、キャラクタの一人称と語尾を利用する。なお、今回は著者が選定した一人称と語尾のコーパスを利用し、その単語が含まれるセリフのグループから発話者の推定を行う。

具体的な方法としては、まずコーパスの単語と同じ一人称または語尾をもつセリフによってグループを作成する。このグループについて、4.2節の距離による手法によって推定された発話者の情報をもとに得点化を行う。例えば、「僕」という一人称によってグループ化されたセリフ群について、その中に距離による推定によってある特定のキャラクタが発話者として推定されていた場合、そのグループに所属するセリフはそのキャラクタが発話者である可能性が高いといえる。つまり、事前に別の手法によって推定された発話者の情報を利用し、本手法でさらに新しいスコアを生成することとした。

以上で述べた手法を組み合わせることによって、セリフの発話者の推定を行う。また、既存の研究[12]ではセリフとキャラクタの距離の情報をもとにした発話者の推定がなされている。そのため、このセリフとキャラクタの距離の情報のみによって推定する手法と、様々な情報を組み合わせた手法を比較することで、既存の研究との比較とする。

## 5 実験と結果

### 5.1 実験方法

前章で説明した推定手法を用いてセリフ発話者の自動推定の実験結果について述べる。具体的には、3章で構築したデータセットのセリフを対象に、4章の手法を利用して発話者を推定し、その推定精度を検証する評価実験を行う。

データセットのセリフの総数は128,502件であるが、これらのデータには1つのセリフに対して複数人の発話者が対応付けられたデータも存在する。また、データセットの発話者の中には「その他」として登録されているキャラクタも存在する。これは、コミックの画像上に存在するキャラクタではあるが、Manga109において特に名前が指定されていないキャラクタの総称となっている。「その他」のキャラクタは特定の人物を指さないため、今回の推定の対象から除外する。「その他」のセリフを除外し

たことにより、今回の対象となるセリフの総数は 121,364 件となった。

今回はこの 121,364 件のセリフに対して発話者となるキャラクタを推定し、そのキャラクタがどれだけ合っていたかを正解率として計算し、精度を評価する。

## 5.2 実験結果

4章で述べた4つの手法について、単独で推定を行なった場合と組み合わせで推定を行なった場合の正解率を表3に示す。表の「組み合わせの寄与率」については、手法を組み合わせた場合の正解率の変化を示している。例えば「同じコマ」「しっぽ方向」「一人称・語尾」の3つの手法の組み合わせによる正解率が44.3%の場合、4つの手法の組み合わせによる正解率の70.7%との差である26.4%が「距離」の手法の寄与率となる。

表3より、「セリフとキャラクタの距離の情報による推定」が最も正解率が高く、組み合わせた際の寄与率も高かった。逆に、「一人称と語尾の情報による推定」は正解率が低く、単独だと4.6%の精度であった。しかし、手法を組み合わせることで精度が下がることはなく、すべての手法を組み合わせた推定が最も精度が高い結果となった。

表3. 単独の手法と組み合わせによる正解率

手法	正解率	組み合わせの寄与率
距離	55.7%	26.4%
同じコマ	34.5%	9.0%
しっぽ方向	13.5%	2.2%
一人称・語尾	4.6%	0.6%
組み合わせ	70.7%	

ここで、精度が低い原因の1つとして発話者の候補が多すぎる問題が考えられる。Manga109におけるキャラクタの数は、コミック1冊に対して平均31.7人、最大124人が登場するコミックも存在していた。そのため、対象を全てのキャラクタではなく、出番の多いキャラクタに絞ることで、精度の向上が見込める可能性がある。

ここでは新たに発話者となるキャラクタを限定するため、Manga109のキャラクタのアノテーション情報をもとに、キャラクタの出現回数を計算した。この情報をもとに、各コミックにおいて出現回数が上位5位のキャラクタのみに絞った。つまり、それぞれのコミックにおいて主要なキャラクタのセリフのみを対象に発話者推定の実験を行う。

1冊ごとに5名のキャラクタに限定するため、発話者候補となるキャラクタの総数は109冊×5名=545名となり、推定対象となるデータセットのセリフについても、発話者がこの545名に含まれないものは除外した。その結果、推定に利用するセリフの総数は88,297件となった。

この新たに用意したデータセットをもとに、再度推定を行なった結果を表4に示す。

表4. 主要なキャラクタのみによる推定の正解率

手法の組合せ	正解率	組み合わせの寄与率
距離	66.9%	24.0%
同じコマ	46.1%	6.7%
しっぽ方向	14.6%	1.8%
一人称・語尾	6.0%	0.6%
組み合わせ	78.6%	

この結果より、距離のみの推定であっても66.9%の精度で推定可能であり、表3の結果より10%近く高い精度で発話者の推定が行えることがわかった。しかし、組み合わせた際の各手法の寄与率は表3の結果とそこまで大きな変化はなかった。

ここで、表3および表4の結果について、「吹き出しのしっぽ方向の情報による推定」および「一人称・語尾による推定」が低かった原因について、それぞれの手法が適用できるセリフが少なかったことが考えられる。対象の121,364件のセリフに対して、しっぽが検出できたセリフ数が22,645件(全体の18.7%)、一人称・語尾が検出できたセリフ数が11,025件(全体の9.1%)であり、検出できたセリフ数を母数としたときの正解率はそれぞれ72.4%、50.6%であった。このことから、「しっぽ方向」および「一人称・語尾」による推定はこれらの特徴が見られるセリフに対してはある程度の精度で推定できることがわかった。

## 5.3 考察

前節の結果より、「同じコマ内にあるキャラクタの情報による推定」を加えることで精度が向上したことより、セリフの発話者はコミックにおいて同じコマに存在する確率が高いと考えられる。その一方で、「一人称と語尾の情報による推定」によって精度がそれほど向上しなかった理由については、まず今回のManga109データセット内において、一人称や語尾が含まれる機会がそれほど多くなかったことが原因の一つと考えられる。また、今回用意したコーパスが、コミックにおけるキャラクタの特徴を網羅できていなかったことが原因として考えられるが、若干でも精度が上がっているものについては、これらが重要な手がかりとして推定に利用された可能性がある。

セリフ内容などの言語特徴を使った推定については、他にもセリフから得られる印象とキャラクタのイラストから得られる印象の一致をみるなどの方法が存在する。そのため、さらに新しい手法を組み合わせることで少しずつであっても精度を向上させることができ、最終的には高精度な発話者推定が可能であると考えられる。

## 6 おわりに

本稿では、コミック工学の研究に貢献することを目指して、Manga109 データセットを拡張する、セリフと発話者の対応付けデータセットを構築した。ここでは主に、セリフと発話者の対応付けデータセット構築手法を開発し、33名によりコミック109冊、147,918発話について2人ずつが発話者の対応付けを行った。その結果、ひとによる判断でも86.9%程度の一致率であること、またデータセット構築においては途中から読むだけでは厳しいことなどを明らかにした。さらに、自動推定に関する手法の検討を行い、すべての要因を組み合わせることで、70%程度で推定可能であることが分かったが、十分な精度ではなかった。

今後は、データセットの対象となるコミックの数を増やしていくとともに、本システムを用いて3名以上による対応付け情報を得ていく。そのうえで、ひとによる判定の困難性などについても明らかにしていく。さらに、三原ら[18]が取り組んでいるような、コミックのセリフと発話者の対応付けに関するマイクロタスクを実現し、より多くのアノテーションを集めていくことで、コミック工学分野へ貢献する予定である。

## 謝辞

本研究の一部は、JST ACCEL (グラント番号JPMJAC1602)によるものである。

## 参考文献

- [1] 公益社団法人全国出版協会: 2017年のコミック市場規模発表, 入手先 (<https://www.ajpea.or.jp/information/20180226/index.html>) (参照 2019-11-23).
- [2] 公益社団法人全国出版協会: 2019年上半期の出版市場を発表, 入手先 (<https://www.ajpea.or.jp/information/20190725/index.html>) (参照 2019-11-23).
- [3] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T. and Aizawa, K.: Sketch-based manga retrieval using manga109 dataset, *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21811-21838 (2017).
- [4] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T. and Aizawa, K.: Object Detection for Comics using Manga109 Annotations, arXiv:1803.08670 (2018).
- [5] Guérin, C., Rigaud, C., Mercier, A., Ammar-Boudjelal, F., Bertet, K., Bouju, A., Burie, J., Louis, G., Ogier, J. and Revel, A.: eBDtheque: A Representative Database of Comics, 12th International Conference on Document Analysis and Recognition, pp. 1145-1149 (2013).
- [6] Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daumé III, H. and Davis, L.: The Amazing

Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives, IEEE Conference on Computer Vision and Pattern Recognition (2017).

[7] 上野未貴: 創作者と人工知能: 共作実現に向けた創作過程とメタデータ付与4コマ漫画ストーリーデータセット構築, 人工知能学会全国大会論文集, Vol. JSAI2018, 4Pin116.

[8] Nguyen Nhu, V., Rigaud, C. and Burie, J.: What do We Expect from Comic Panel Extraction?, 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), Vol. 1, pp. 44-49 (2019).

[9] Chu, W. T. and Li, W. W.: Manga face detection based on deep neural networks fusing global and local information, *Pattern Recognition*, Vol. 86, pp. 62-72.

[10] Tolle, H. and Arai, K.: Method for Real Time Text Extraction of Digital Manga Comic, *International Journal of Image Processing* (2011).

[11] 田中孝昌, 外山史, 宮道壽一, 東海林健二: マンガ画像の吹き出し検出と分類, *映像情報メディア学会誌*, Vol. 64, No. 12, pp. 1933-1939.

[12] Rigaud, C., Le Thanh, N., Burie, J., Ogier, J., Iwata, M., Imazu, E. and Kise, K.: Speech balloon and speaker association for comics and manga understanding, 13th International Conference on Document Analysis and Recognition (ICDAR), pp.351-355 (2015).

[13] 山本和慶, 小川徹, 山崎俊彦, 相澤清晴: データドリブンなアプローチを用いた漫画画像中の吹き出しの話者推定, *映像情報メディア学会技術報告*, pp. 287-292 (2018).

[14] 阿部和樹, 中村聡史. 漫画における台詞発話者の自動判定に向けた技術的困難性による整理とデータセット構築手法の検討, 第2回コミック工学研究会発表会, pp.7-14, 2019.

[15] Mantra: マンガの超高精度な自動翻訳, (オンライン), 入手先 (<https://mangahot.jp/>) (参照 2019-11-23).

[16] 佐藤剣太, 牧良樹, 中村聡史. 続巻への動機付けのためのコミック読書進度に応じたシーンのデータセット構築と自動推薦に関する検討, 第11回データ工学と情報マネジメントに関するフォーラム (DEIM2019), Issue.G3-5, pp.1-7, 2019.

[17] Yoshiki Maki, Satoshi Nakamura. Do Manga Spoilers Spoil Manga?, The Sixth Asian Conference on Information Systems (ACIS 2017), pp.258-262, 2017.

[18] 三原鉄也, 石川夏樹, 豊田将平, 永森光晴, 杉本重雄: 画像認識とマイクロタスク型クラウドソーシングを組み合わせたマンガのコマ領域の判定, 人工知能学会全国大会論文集 (JSAI2018), No. 4M105 (2018).