

2019年度 修士学位請求論文

漫画におけるセリフと発話者の  
対応付け手法の研究

明治大学大学院先端数理科学研究科  
先端メディアサイエンス専攻

阿部 和樹

Master's Thesis

A Research of the Method to Recognition  
the Speaker of Comic Text

Frontier Media Science Program,  
Graduate School of Advanced Mathematical Sciences,  
Meiji University

Kazuki Abe

## 概要

漫画はイラストと文字によってストーリーを表現したマルチメディアコンテンツであり、日本のみならず世界中で鑑賞される人気のコンテンツである。また、近年ではスマートフォンやタブレットの普及に伴い、漫画をデジタル化した電子コミックとして鑑賞される機会も多くなっている。こうした漫画のデジタル化により、漫画を機械に認識させることで様々な利用方法を提案するサービスが登場しつつある。例えば、漫画の文章を外国語に自動翻訳して閲覧可能にするサービスや、ユーザの好みに合わせた漫画の推薦などが存在する。しかし、こうしたサービスを実現するためには、漫画の画像からキャラクタやテキストといった要素を抽出し、それらの情報をもとに漫画コンテンツの内容について機械が認識する必要がある。このような、漫画画像からキャラクタやセリフのテキストといった要素を自動で抽出し、漫画を機械で処理可能な形式へと変換するといった漫画の自動解析の研究も盛んに行われている。

このような機械による漫画の自動解析の1つとして、漫画内に登場するセリフの発話者を自動で推定する手法が必要とされている。漫画のセリフにもとづいたコマの検索や漫画のシーンの理解のためには、こうしたセリフの自動的な解析が必要であり、OCR（工学文字認識）などの技術によってある程度そのセリフの内容は認識できる。しかし、そのセリフがどのキャラクタの発言であるかについての情報は読み手が判断する必要があり、機械により自動で推定する手法はまだ確立されていない。

また、自動推定手法の妥当性を評価するためには、漫画についての大量のデータが必要となる。漫画のデータセットはいくつか存在するものの、セリフの発話者についての情報を持ったデータセットは存在しないという問題がある。

そこで本論文では、セリフの発話者であるキャラクタを自動推定する手法の実現に向け、まずは漫画におけるセリフと発話者の正解データを収集したデータセットを構築する。その際、セリフに対応したキャラクタのアノテーション付与を効率的に行うためのアノテーション付与システムを実装した。また、収集したデータの信頼性を高めるため、1つのセリフに対して2名の協力者がアノテーション付与を行なった。これにより、109冊の漫画に登場する147,918件のセリフに対して発話者であるキャラクタの対応付けを行なった。

次に、収集したデータをもとに人手によるセリフと発話者の対応付けについて分析を行う。1つのセリフに対して2名のデータ収集協力者が存在するため、この2名の意見が一致しているかを見ることによって収集したデータの正確さを評価するとともに、人が発話者を判断する際の難易度についても議論する。収集したデータを分析した結果、86%のセリフは意見が一致したが、その他のセリフについては意見が一致しないため、人にとっても発話

者の判断が困難なセリフが存在することが明らかになった。これらを踏まえ、実際の漫画の事例をもとにセリフと発話者の推定における課題を「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つの要因に整理した。

最後に、機械によってセリフの発話者を自動で推定する手法を考案し、データセットをもとに手法の精度を評価した。発話者を推定する手法として、「同じコマ内にいるキャラクターの情報」「セリフとキャラクターの距離の情報」「吹き出しのしっぽの方向の情報」「一人称と語尾の情報」の4つの情報を組み合わせる方法を提案した。推定の結果、全体のセリフに対して70%の精度で発話者を推定することができるという結果が得られた。また、発話者の対象を主要なキャラクターに絞って推定を行なったところ、最大で78%の精度で発話者を推定可能となった。これにより、発話者の手がかりとなる情報を組み合わせることで精度が向上することが明らかとなった。また、キャラクターの特徴とセリフの特徴の一致を見るなど、セリフの内容から得られる手がかりを用いることで推定の精度を向上させられることが示唆された。

# Abstract

Comics are multimedia contents expressing a story with illustrations and texts, and are popular content not only in Japan but also all over the world. Also, along with the popularization of smartphones and tablets in recent years, many people started reading digital comics. The digitalization of comics has led to appearance of services which propose various utilization method by making machines recognize comics. For example, some service automatically translates texts in comics into foreign languages, and other services make recommendations of comics according to the preference of the users. However, in order to realize such services, it is necessary for machines to extract elements such as illustrations of characters and texts from images of comics, and to recognize contents of the comics based on the information. Then, much research has been conducted on the automatic analysis of comics where comics are converted into a form which can be processed by machines.

As one of the automatic analysis by machines, a method to automatically estimate the speakers of texts is required. For understanding of the scenes, a method which estimates the speakers of the dialogue is necessary. The text strings can be recognized to some extent by the method such as OCR (Optical Character Recognition). However, it is still necessary for readers to judge which character says which utterance, and the automatic estimation method by machines has not been established yet.

Moreover, a large amount of data of comics is required to evaluate the validity of the automatic estimation. Also, there is no dataset of information about speakers of texts in comics.

To realize the automatic estimation by machine, the present study at first collected the correct data of speakers of utterances in comics. For the data collection, we implemented an annotation system to efficiently annotate the speakers. Also, in order to increase the reliability of the data, two users were assigned to each text. Using this system, we collected data of the speakers from 147,918 texts in 109 comic books.

Next, we analyzed features of the speakers of comic texts based on the collected data. Since there were two data collection collaborators for each text, we evaluated the accuracy of the collected data by seeing whether the two collaborators agreed or not and discussed the difficulty of human judgement on the speakers. Analysis of the collected data revealed that 86% of the texts were assigned to the same characters, but assignment of the other

texts were different between the annotators. This suggests that some texts were difficult for humans to judge regarding who their speakers are. Based on these, problems in the estimation of comic text' s speakers were classified into 3 factors of " speech balloon shape" , " position relationship between texts and characters" , and " matching the features of texts and characters" .

Finally, we proposed a method to automatically estimate the speakers of texts by a machine and evaluated the accuracy of the method based on the collected data. As an estimation method, 4 pieces of information of " characters in the same frame" , " distance between text and character" , " direction of the tail of the speech balloon" , and " information on first person utterances and endings of texts" were acquired and combined. The result found that speakers could be estimated with 70% accuracy for all collected texts in comic. Also, when the estimation target was limited to main characters, the estimate gained the accuracy of 78% at the maximum. This shows that the accuracy was improved by combining information of the speakers. It was also indicated that the accuracy of the estimation could be improved by using clues from contents of texts such as correspondence of features of the character and features of the text.

# 目次

第 1 章	はじめに	1
1.1	近年の漫画コンテンツと電子コミック	1
1.2	情報処理による漫画コンテンツの活用	1
1.3	セリフに対応する発話者の推定	2
1.4	漫画データセット	3
1.5	目的	3
第 2 章	関連研究	4
2.1	漫画のデータセットに関する研究	4
2.2	漫画の構造分析と内容理解に関する研究	5
2.3	漫画の活用に関する研究	7
第 3 章	セリフ発話者アノテーション付与システム	9
3.1	Manga109 データセット	9
3.2	セリフ発話者のアノテーション付与の問題点	9
3.3	アノテーション付与手法	10
3.4	実装	11
第 4 章	セリフと発話者の人手による対応付けとデータセット構築	12
4.1	データ収集	12
4.2	収集したデータの分析	13
4.3	収集したデータに対する考察	15
第 5 章	セリフと話者の対応付けの技術的困難性	20
5.1	要因の分析	20
5.1.1	吹き出しの形状	20
5.1.2	セリフとキャラクタの位置関係	22
5.1.3	セリフの表現とキャラクタの特性	23
5.1.4	特殊な例	25
5.2	技術的困難性にまつわる要因の整理	26
第 6 章	セリフ発話者の自動推定	27

6.1	推定手法の提案 . . . . .	27
6.1.1	セリフとキャラクタの距離の情報による推定 . . . . .	27
6.1.2	同じコマ内にいるキャラクタの情報による推定 . . . . .	28
6.1.3	吹き出しのしっぽの方向の情報による推定 . . . . .	28
6.1.4	一人称と語尾の情報による推定 . . . . .	28
6.2	評価方法 . . . . .	29
6.3	結果 . . . . .	29
6.4	考察 . . . . .	32
第7章	総合考察	36
第8章	まとめ	39
参考文献		41

# 第 1 章 はじめに

## 1.1 近年の漫画コンテンツと電子コミック

日本における漫画とは、平面上にイラストと文字を巧みに配置することでストーリーを表現するコンテンツであり、世界に誇る日本文化の 1 つである。一般的な漫画の鑑賞方法としては、漫画雑誌に掲載された漫画を読むほかに、1 つの作品が複数話まとめて収録された単行本や、Web 漫画やアプリ漫画として毎週決まった日時に更新される話を読むなど様々である。その規模は巨大であり、例えば、漫画雑誌の 1 つである週刊少年ジャンプ [1] の 2019 年 1 月～3 月における発行部数は約 170 万部（日本雑誌協会調べ [2]）となっている。また、同誌で連載中である「ONE PIECE」[3] の累計発行部数は 4 億 6,000 万部であるが、そのうちの 7,000 万部以上は海外における発行部数となっている。また、2019 年にはイギリスの大英博物館にて、日本の漫画をテーマにした展覧会である「The Citi exhibition Manga」[4] が開催されるなどの事例も存在し、日本の漫画は海外でも人気を誇っているといえる。

また、近年では紙媒体だけでなく、電子媒体で漫画を鑑賞する機会が増加している。全国出版協会による出版市場調査によると、2017 年に初めて電子コミックの売上が紙媒体の漫画の販売を上回るという結果となった [5]。さらに、同協会による 2019 年上半期の調査 [6] では、電子コミックの売上は 1,133 億円（前年同期比 27.9% 増）にのぼっている。このことから、近年ではタブレットやスマートフォンといったデジタル端末を利用し、電子媒体で漫画を読むことが一般的になっている。

## 1.2 情報処理による漫画コンテンツの活用

前節で述べたような電子コミックの普及に伴い、漫画を機械に認識させることで様々な利用方法を提案するサービスが存在している。そのサービスの 1 つとして漫画翻訳サービスが挙げられる。「Mantra」\*1[7] は、漫画内のテキストを外国語に翻訳して再配置することで、外国人も他国の漫画を読めるサービスを提供している。また、株式会社コアミックスの漫画配信アプリである「マンガほっと」\*2[8] では、読者がその時々の気分をアプリに入力することで、気分に適した漫画が推薦される機能が備わっている。

こうしたサービスの他にも、漫画の利用方法に関する研究が複数存在する。例えば、Narita ら [9] は、ユーザがスケッチを描くことで、スケッチに類似したキャラクターが登場す

---

\*1 <https://mntr.jp/>

\*2 <https://mangahot.jp/>

る漫画を検索可能とするシステムを提案している。また Park ら [10] は、漫画に登場するキャラクターの類似度を計算することで、同様のキャラクターが登場する漫画を探す方法を検討している。

また、漫画を工学的に扱い、活用できるようにするため、漫画に登場する絵やテキストを機械で自動的に解析する研究も盛んに行われている。例えば Ogawa ら [11] は、漫画におけるコマ・キャラクター・テキストの位置を高精度に推定する手法を提案している。また田中ら [12] は、漫画に登場する吹き出しの検出と、吹き出しの形状を自動で推定する方法を提案している。

このように、絵やテキストによって多様な表現が作り出される漫画の研究においては、画像処理・自然言語処理といった様々な分野の技術を用い、漫画の内容を機械に理解可能な形として処理する必要がある。

### 1.3 セリフに対応する発話者の推定

漫画を構成する要素の代表例として、キャラクターのセリフが存在する。セリフにもとづく漫画のコマ検索や漫画のシーンの理解のためには、こうしたセリフの自動的な解析が必要であり、OCR（工学文字認識）などの技術によってある程度そのセリフの内容は認識できる。しかし、そのセリフがどのキャラクターの発言であるかについての情報は読み手が判断する必要がある。また、これまで漫画のセリフに関するいくつかの研究 [13, 14] は行われてきたが、機械により自動推定する手法はまだ確立されてはいない。

ここで、セリフ発話者の自動推定が可能になると、以下のような応用が可能になると考えられる。

- 発言をもとにしたキャラクターの性格の自動抽出
- 会話の履歴にもとづいたキャラクターの関係性やストーリーの推定
- 漫画の音声読み上げにおけるキャラクターの自動割当
- 外国人のためのキャラクターの個性を反映したセリフ翻訳の実現

上記で挙げた応用については、すでに取り組みされている研究 [10, 15] が存在するものの、セリフに対する発話者の情報は人手で付与されているなど、膨大な数の漫画を手動で処理することは現状では困難である。よって、機械による漫画の認識・自動解析には、セリフに対する発話者の自動推定が必要不可欠であると言える。

また、漫画の読者（特に、漫画を読むことに慣れている読者）は、セリフの発話者が誰であるかについてはそれほど意識せずに自然に理解し、読書を進めることができる。これは、セリフの吹き出しの形状やキャラクターとの位置関係、一人称や方言、語尾の違いといったセリフの特徴、登場するキャラクターをユーザが理解していることや、前後の会話の文脈などから判断できるためであると考えられる。つまり、機械的にセリフの発話者を正確に推定するためには、吹き出しの形状や、セリフの内容、前後の会話の流れを考慮することが重要であると考えられる。

このように、機械によって漫画のセリフの発話者を自動で推定するためには、上記にあげ

たような特性を考慮する必要がある。そのため、画像に対する分析と文字情報に対する分析を組み合わせ、画像処理と自然言語処理の技術を横断的に利用することが必要であり、自動推定の難易度は非常に高い。そこで本論文では、これらの特性について分類および課題の整理を行った上で、機械による自動推定の可能性について模索していく。

## 1.4 漫画データセット

本論文では漫画のデータセットとして、漫画の画像データと様々なアノテーション情報が提供されている Manga109[11, 16] を利用する。Manga109 は日本のプロの漫画家によって描かれた 109 冊の漫画で構成されており、1 冊ごとに各ページの画像と、そのページにおけるコマや人物の情報がアノテーション情報として付与されている。しかし、Manga109 には本論文の目的であるセリフの発話者に関する情報は付与されていない。そこで本論文では、セリフに対応した発話者というアノテーション情報を手軽に付与するシステムの提案についても行う。さらに、集めた情報をもとに漫画に登場するセリフと発話者の特性についても分析を行い、自動推定における手がかりとする。

## 1.5 目的

本論文では、漫画の自動解析における課題の 1 つである、漫画内に登場するセリフに対応する発話者を機械によって自動で推定することを目的とする。そのために、漫画のセリフにおける特徴を分析し、それらを機械学習の手法によって発話者を推定する手法を提案する。また、手法の有用性を確認するため、セリフに対応する発話者の正解となるデータセットを作成し、推定結果と照らし合わせることで精度を検証する。

なお、本論文の目的はセリフに対する発話者の自動推定に絞り、漫画について事前にある程度の情報を取得した上で推定することを前提とする。つまり、セリフの位置や文字情報・キャラクターの位置・キャラクター名といった情報の推定は行わず、あくまでこれらの情報を機械が把握している上で、どのように発話者の推定を行っていくかについて議論する。これらの漫画を構成する要素を推定する方法については、第 2 章にて既存の研究をベースに述べる。

本論文は、本章を含む全 8 章から構成される。まず本章で漫画におけるセリフ発話者の自動推定における問題定義について述べた。以降では、第 2 章では、漫画コンテンツの自動解析および情報処理による利用の観点から関連する研究について述べ、本論文の位置づけについて述べる。第 3 章では、自動推定における精度の検証のために利用するデータセットの構築について述べ、第 4 章ではそれらの集めたデータセット自体の正確さや傾向について分析を行う。後半の第 5 章では、一般的な漫画におけるセリフとその発話者の事例をもとに、漫画において、どのような要因が発話者の自動推定の難易度に影響を与えるのかについて議論する。第 6 章では、本論文の目的であるセリフ発話者の自動推定の手法について述べ、データセットをもとに推定精度を評価した結果について議論する。第 7 章にて全体の結果に対する考察を述べ、最後に第 8 章で本論文をまとめる。

## 第 2 章 関連研究

### 2.1 漫画のデータセットに関する研究

漫画に関する研究では、漫画の画像とその漫画に関するメタ情報が必要となる。そのため、漫画の画像とメタ情報をデータセットとして作成した上で、他者が利用可能な形で公開している研究が存在する。Matsui ら [11, 16] が作成したデータセットである Manga109<sup>\*1</sup>は、日本の漫画家によって描かれた 109 冊の漫画で構成されるデータセットである。データセットにはアノテーション情報として、コマ・キャラクタ・テキストの画像上の位置やキャラクタ名、テキストの文字列といった情報が、各漫画の全てのページに対して付与された大規模なデータセットである。また、Guérin ら [17] は、フランス・アメリカ・日本で出版された漫画の画像 100 枚から構成される eBDtheque<sup>\*2</sup>というデータセットを構築している。Matsui らの Manga109 とは異なり、様々な作品から 1 ページあるいは複数のページに対してアノテーションを付与したものであり、コマや吹き出しの位置に加えて、吹き出しの順番や吹き出しのしっぽと呼ばれる部分が指す方向といった情報が存在する。他にも Iyyer ら [18] は、アメリカの漫画における 120 万以上の画像からなるデータセットである COMICS<sup>\*3</sup>を公開している。また、データの公開はされていないが、上野 [19] は日本の作家が描く 100 作品の 4 コマ漫画で構成されるデータセットを構築している。上野の 4 コマ漫画データセットは、他のデータセットとは異なり研究者が構築したストーリーをもとにプロの作家が漫画として描いたものである。そのため、ストーリーの流れを記述したプロットがメタデータとして用意されたデータセットとなっている。これらの漫画画像とそれに付与されたアノテーションの情報を構築するには膨大な労力が伴うため、公開されている漫画データセットは後述する漫画に関する様々な研究で利用されている。しかし、これらのデータセットには本論文で対象とするセリフの発話者に関する情報は含まれていない。よって、本論文では既存のデータセットである Manga109 に対して新たにセリフ発話者というアノテーション情報を収集することを目的の一つとする。

また、近年ではディープラーニングをはじめとした機械学習に関する研究の隆盛に伴い、学習に利用する大量のデータをどのように収集するかに焦点を当てたデータセットの構築方法に関する研究が存在する。漫画のデータセットに関して、三原ら [20] は漫画のコマが存在する領域の正解データを効率的に収集するため、事前に機械が自動で推定した領域の正

---

\*1 <http://www.manga109.org>

\*2 <http://ebdtheque.univ-lr.fr/>

\*3 <https://github.com/miyyer/comics>

誤判定をマイクロタスク型のクラウドソーシングによってデータを収集する手法を提案している。この研究では、一から人が画像上の領域を指定するのではなく、機械の予測が合っているかを判断するという簡単なタスクにすることでデータ収集の効率化を図っており、6 時間以上の時間がかかるタスクが 2 分程度で終了できるという結果が示されている。このようなデータ収集の方法に関する研究は漫画に限らず様々なデータに対して行われている。Papadopoulos ら [21] は、画像において物体が表示される領域のアノテーション付与において、表示領域の指定方法を物体周辺の上下左右 4 点をクリックするタスクにすることで従来の約 5 倍の速度でアノテーションを付与できる手法を提案している。また、中村ら [22] は画像に対する印象データを収集するタスクについて、音ゲーの中にマイクロタスクとして組み込む手法を提案している。また、クラウドソーシングによって不特定多数のユーザによってデータを収集できる反面、そのデータの信頼性や品質が問題になることがある。データの信頼性を高める方法としては、同じデータ収集のタスクを複数人が実行し、その多数決によって決定することが一般的である。これに対し、Su ら [23] は画像に対するアノテーション付与の方法として、実際にアノテーションを付与するタスク、付与されたアノテーションが正しいか確認するタスク、画像に対して必要な全てのアノテーションが付与されているか確認するタスクの 3 つに分解する方法を提案している。これにより、高品質なデータの収集を可能とするとともに、作業をマイクロタスクとして分解することで時間や人数といったコストを削減してデータ収集ができることを示している。このようにデータセット構築を効率よく行うため、作業の手間を省く工夫やマイクロタスクとして実行可能にするなどの様々な方法が提案されている。本論文では膨大な数のセリフに対して発話者となる情報を効率的に収集するため、Web システムによって漫画画像を操作することによる効率的なデータ収集を検討する。また、データの信頼性についても、セリフに対して複数人のユーザがアノテーションを付与することにより対応する。

## 2.2 漫画の構造分析と内容理解に関する研究

漫画コンテンツを機械によって処理するには、漫画の画像をコマ・テキスト・キャラクタといった要素に分解した上で、それぞれの要素から得られる情報をもとに漫画の内容について処理する必要がある。このような、漫画を機械に処理可能な形に変換するための様々な研究が存在する。

漫画を構成する主な要素としては、コマ・テキスト・キャラクタといった要素が上げられる。漫画の画像をコマごとに分割する研究の 1 つに Nguyen ら [24] の研究がある。Nguyen らは、画像の各ピクセルをコマの下地となる背景、コマ領域、コマの境界線の 3 つに分類することで画像上の複雑なコマの領域を推定しており、約 78% の精度でコマを抽出する方法を提案している。また、漫画内のテキストの検出に関する研究も複数存在するが、これらは直接テキストを抽出するのではなく、漫画の吹き出しを抽出したのちにテキストを抽出することがほとんどである。このような、テキスト抽出のために漫画画像から吹き出しを抽出する研究として、Correia ら [25] の研究がある。Correia らは、画像の勾配情報をもとに吹き出しを検出する方法を提案し、複雑な形状であっても柔軟に対応できる吹き出し抽出の手

法を提案している。また田中ら [12] は、吹き出しの検出と同時に吹き出しを分類する手法を提案している。吹き出しの輪郭から特徴量を抽出し、SVM によって吹き出しを分類しており、これらの吹き出しの形状に関する情報がセリフの感情や状況を示す手がかりになることを示唆している。他にも Kovanen ら [26] は、吹き出しの読まれる順序に着目し、漫画画像上に存在する吹き出しがどういう順序で読まれるかを機械が自動で推定する手法を提案している。このように吹き出しを抽出することで、その吹き出しの上に存在するセリフなどの文字を抽出することができる。Tolle ら [27] は、吹き出しの抽出からそこに描かれた文字列の認識までを全て自動で行う手法を提案している。研究の結果、吹き出しの抽出を高精度に行いつつ、日本語の文字列を 90% 以上の精度で抽出できることを示している。これらの研究は漫画の吹き出しに着目し、吹き出しから得られる情報を用いてテキストの抽出に活用する試みとなっている。一方で、近年ではディープラーニングによる画像処理が高度に発展したことにより、画像から直接テキストを抽出することも可能になりつつある。その例の 1 つとして Aramaki ら [28] の研究では、漫画画像上でテキストが表示されている領域を 85% の精度で自動で発見することが可能となっている。これらのコマやテキストは漫画を構成する重要な要素であるが、キャラクタなどのイラストを対象にした研究もまた存在する。Chu ら [29] は、機械によって漫画に登場するキャラクタの顔を認識するとともに、画像上の異なる位置に描かれたキャラクタを同定する手法を提案している。また、キャラクタを対象とした研究として、Stricker ら [30] は目や鼻といった顔のパーツ情報を推定する手法を提案している。他にも、Khungurn ら [31] はアニメや漫画のキャラクタの絵から骨格情報を抽出し、キャラクタの姿勢を推定する手法を提案している。このように、コマ・テキスト・キャラクタといった要素に分解することで漫画を機械に利用可能な形式に変換する手法が多数存在する。Ogawa ら [11] の研究は、こうした要素の分割を一度にまとめて行うため、漫画画像を入力することでコマ・テキスト・キャラクタの体および顔の位置を推定結果として出力する機械学習のモデルを構築している。また、漫画において色や絵柄の表現に用いられるスクリーントーンはこうした要素の抽出においてノイズとなるため、Li ら [32] は漫画画像を線画とスクリーントーンに分解する手法を提案している。このように、漫画の画像から要素を抽出するとともに、それぞれの要素について分析した研究が多数存在する。本研究はこれらの研究のように要素を抽出するのではなく、こうして抽出された要素の情報をもとに、セリフに対応した発話者を推定する手法を提案する。

これまでは漫画の画像から要素を抽出する研究を挙げたが、この他にもより漫画の内容を認識するための研究が存在する。例えば Daiku ら [33] は漫画のストーリーを機械によって自動認識させる手法を試みている。この研究では、漫画のジャンルの推定を作品に対してではなく作品の 1 ページごとに行っており、ページが進むごとのジャンル情報の変化がストーリーを認識する手がかりになることを示唆している。また、ストーリーを機械によって認識・管理するため、井上ら [34] は漫画のシーンを言語によって表すための手法を提案している。この他にも、Hiroe ら [35] は漫画内に登場するテキストの感嘆符を数えることで、その漫画の興奮度合いやテンションを推定する研究を行なっている。Park ら [10] の研究は漫画の内容ではなくキャラクタの理解を深めるため、キャラクタの性格を推定し、それをもとにキャラクタの分類を行なっている。さらに、Park らの研究は漫画から得

られる情報とは別に、キャラクターの性格の手がかりを Wikipedia などの Web 上の情報から手に入れている。このような Web 上の情報をもとに漫画を分析している研究は他にも存在し、Yamanishi ら [36] は漫画の読者が印象深いと感じるシーンを「感性的ハイライト」として抽出する手法を提案している。これらの研究は先述の漫画の要素の抽出とは異なり、より漫画の内容やユーザの認知に近い情報を自動で処理するための試みとなっている。本研究はここで述べたようなストーリーやキャラクターの性格といった高度な情報を抽出する手がかりとして、セリフに対応した発話者を推定する研究を行う。

本節の最後に、本研究と同様にセリフの発話者に着目した研究について述べる。Rigaud ら [13] の研究は本研究の目的と同じく、セリフに対応した発話者であるキャラクターを機械によって自動で認識させる研究である。Rigaud らはセリフとその発話者であるキャラクターは漫画画像上において近い位置に存在することに着目し、セリフにもっとも近い位置に存在するキャラクターを発話者として推定している。その結果、フランスの漫画に対しては 90% 以上の精度で推定可能であり、日本の漫画に対しても 87% の精度で推定可能なことを示している。他にも、山本ら [14] も同様に漫画におけるセリフの発話者を推定する手法を提案している。山本らは表示されているキャラクターのサイズやセリフとキャラクターの位置関係の特徴量とし、ディープラーニングの手法によって発話者を推定することを試みている。実験の結果、75%~80% の精度で推定が可能であるという結果が得られている。これらの研究は本研究と同様の目的であるが、これらの手法では近くに存在する発話者に対してのみ推定可能であり、他のコマやページに存在するキャラクターが発話者であった場合に推定できないという問題がある。また、これらは画像から得られる特徴にのみ着目しており、セリフ内容などの言語的な特徴に着目した研究は行われていない。本研究ではこれらのセリフの内容から得られる特徴を利用するとともに、セリフから離れた場所に存在する発話者に対しても推定可能な手法を目指す。

## 2.3 漫画の活用に関する研究

漫画を機械により処理可能な形にすることで、それらを情報技術の分野で活用するための様々な研究が存在する。その中でも漫画を検索することや、ユーザの好みに合わせた漫画の推薦に関する研究が盛んに行われている。Narita ら [9] は、ユーザがスケッチを描くことで、そのスケッチと一致した漫画のコマを検出する方法を提案している。また Nguyen ら [37] は、漫画に関するメタ情報をタグとして保存することで探索可能なシステムを提案している。他にも、Vie ら [38] はアニメや漫画のキービジュアルから意味的な情報を抽出し、推薦に利用する方法を提案している。これにより、情報が少ない漫画に対しても漫画に関するイラストのみで推薦できるため、推薦のコールドスタート問題にも対応できるとしている。また Park ら [39] も漫画の表紙情報をもとに、ユーザが好みそうな漫画を推薦する手法を提案している。

こうした検索や推薦といった活用とは別に、漫画コンテンツを電子ならではの利用方法を考える研究も存在する。Wang ら [40] は、漫画の音読を自動で生成するシステムを提案している。また、Saito ら [41] は漫画のフォントに着目し、多様なフォントを漫画に挿入する

ことでコンテンツの体験を拡張することを試みている。また、漫画コンテンツを他のコンテンツに利用する研究もふくすう存在する。Shimazaki ら [42] は、ユーザに健康について意識付けをさせるため、日々の健康に関する情報を漫画によって伝えることの効果を検証している。Koutnikov ら [43] は科学に関する勉強の教材に漫画を利用することで、学習の効率が上がることを調査によって明らかにしている。

また、漫画の利用のみならず、漫画や漫画の要素を自動で生成するための研究も存在する。Furusawa ら [44] は、漫画の画像を自動で着色する手法を提案しており、カラー漫画を楽しみたい読者が手軽に利用できるシステムを提案している。また Jin ら [45] は、二次元のキャラクタを自動で生成するツールを作成しており、誰でも簡単に好みのキャラクタを生み出すことを可能にしている。また、Cao ら [46] は実際に漫画を描く作者を支援するために、漫画の最適なレイアウトを自動で生成する手法を提案している。この他にも、Wu ら [47] は実際の風景写真を漫画風のイラストに変換することで、漫画の作成を支援する手法を提案している。これらの漫画の自動生成については、漫画の絵を自動で生成することが中心となっているが、迎山ら [48] の研究は 2 コマで表現される漫画について、そのストーリーも含めて自動で生成する方法を検討している。

このように、漫画を利用する研究や自動で生成するための研究が多数存在しており、これらは前節でも述べた漫画の要素を抽出することで実現可能となっているものも多い。よって、本研究で提案するセリフと発話者の推定が可能になることで、これらの電子による漫画の利用方法がさらに増えることが期待される。

## 第 3 章 セリフ発話者アノテーション付与システム

### 3.1 Manga109 データセット

本研究では，漫画に登場するセリフの発話者を自動で判別するためのデータとして，Matsui ら [11, 16] が構築した Manga109<sup>\*1</sup>を利用する．Manga109 は，日本の漫画家が描いた 109 冊の漫画で構成されており，各漫画のそれぞれのページに対して以下の 4 種類のアノテーションが付与されている．

- コマの位置
- キャラクタの体の位置とキャラクタ名
- キャラクタの顔の位置とキャラクタ名
- セリフの位置とセリフの文字列情報

これらの情報により，発話者の判別を利用するためのセリフの位置や内容，セリフの周辺に存在するキャラクタを取得することが可能である．しかし，Manga109 にはセリフに対し，その発話者であるキャラクタの情報は付与されていない．

そこで，Manga109 に存在する全てのセリフに対し，それらの発話者であるキャラクタの情報を人手によって付与するシステムを提案および実装する．なお，Manga109 を利用した研究では漫画内のキャラクタやセリフの内容についてそれぞれ機械的に推定することが試みられているため，将来的にはこうした手法と組み合わせることで新たな漫画に対応可能になると期待される．

### 3.2 セリフ発話者のアノテーション付与の問題点

漫画におけるセリフの数は漫画によって大きく異なるが，Manga109 には合計 147,918 個のセリフがあり，平均すると漫画 1 冊あたり 1,357 個のセリフが存在することになる．Manga109 には様々なジャンルの漫画が存在しているため，他の漫画においても同程度のセリフ数があると考えられる．また Manga109 では，こうした膨大な数のセリフに加え，漫画 1 冊に登場するキャラクタの人数が，平均 31.7 人，最大 124 人となっている．そのため，1 冊の漫画に登場する全ての人物を記憶することは容易ではなく，1 つ 1 つのセリフに対してその発話者であるキャラクタを選択するタスクはアノテーション付与の協力者に対して大きな負担になることが予想される．

---

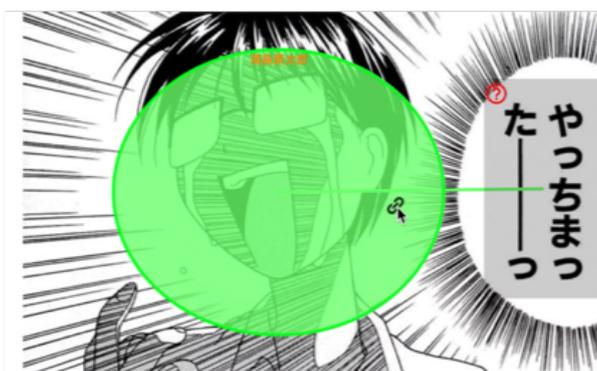
<sup>\*1</sup> <http://www.manga109.org>

一方、アノテーション付与においてはその漫画を何度も読んだ読者ではなく、初めてその漫画を読む読者も多いと考えられる。前述の通り、漫画の登場人物数は少なくないため、もし全ての登場人物を覚えていなければアノテーションを付与できないとすれば、タスクのハードルがかなり高くなってしまう。ここで、人はセリフの発話者を吹き出しの形状やキャラクターの位置関係によってセリフの発話者を判断できる。つまり、見た目によって判断しつつ、アノテーションの付与が行える仕組みが実現されれば、より多くのユーザにアノテーション付与の協力をあおぐことができると期待される。

なお、漫画は前後のページ・コマの情報が連続するものであり、協力者がそれらを単体で見ただけでは正しい発話者の判断ができない可能性もある。そのため、セリフ発話者の判断がつかないものについてはその判断を保留とする仕組みも重要となる。

### 3.3 アノテーション付与手法

前節の問題点をもとに、セリフ発話者のアノテーション情報を効率的に協力者が付与する手法を提案する。提案する手法のイメージを図3.1に示す。



©赤松健「ラブひな」

図 3.1: アノテーション付与手法を利用し対応付けを行っている様子

提案手法では、協力者はセリフをマウスによってドラッグし、隣接するキャラクターに向けてドロップすることにより、セリフと発話者の結びつけを行う。第5章でも説明するが、この方法はセリフとそれに該当する発話者は漫画のページ画像上において近い位置にすることが多いという特性を利用したものである。セリフの吹き出しに「しっぽ」が存在する場合や、セリフがキャラクターの付近、もしくはキャラクターに重なる形で表示されている場合は、セリフに隣接した位置に存在するキャラクターが発話者である可能性が高い。そのため、セリフを付近のキャラクターにドラッグアンドドロップするだけでセリフとその発話者の結びつけを可能にすることにより、1つ1つのセリフに対してキャラクターを選択する方法に比べ、手軽にアノテーションを付与することができると期待される。またその漫画に対する理解があまり深くなくとも、協力者がセリフと発話者となるキャラクターをある程度正確に対応付けできると期待される。

### 3.4 実装

提案手法をもとに、漫画を読みながらアノテーションを付与することができるシステムを Web システムとして実装した。本システムを利用することで、協力者はオンラインでアノテーションの付与を行うことができる。ここで、本提案手法は効率よくアノテーションを付与できる一方で、発話者がセリフと同じページに存在する場合にしか適用できないという問題がある。そのため、図 3.2 に示すように、セリフをクリックすることでキャラクターの一覧から発話者を選べるようにする機能も同時に実装し、どちらの方法によってもアノテーションの付与を可能にするシステムとした。



©赤松健「ラブひな」

図 3.2: キャラクター一覧から発話者を選択している様子

また、発話者の選択肢の中には「不明」「ナレーション」という 3 つの項目を用意した。「不明」は協力者の主観による判断が難しい場合に選択する項目である。また、漫画には登場するキャラクターのセリフとは別に、その状況を説明する文章やト書きが存在するため、これらは「ナレーション」としてアノテーションを付与するようにした。

この他にも、アノテーションの付与をより便利にするための機能として、複数のセリフをまとめてドラッグ&ドロップする機能も実装した。この機能はキーボードの Shift キーを押すことで複数選択する仕様としており、世の中のアプリやサービスにおいて複数のコンテンツを選択する際によく用いられる操作の 1 つである。また、「不明」や「ナレーション」を選択する際には必ずキャラクターの一覧から探す必要があり手間であることが予想されるため、キャラクターをマウスでダブルクリックすることで「不明」に、キーボードの Alt キーを押しながらダブルクリックすることで「ナレーション」にアノテーション付与できる仕組みとした。これら実際に使用されたシステムについては、一部の機能を再現したデモページ<sup>\*2</sup>を公開している。

<sup>\*2</sup> <https://exam-speaker.nkmr.io/demo/>

## 第 4 章 セリフと発話者の人手による対応付けとデータセット構築

### 4.1 データ収集

前章で説明したアノテーション付与システムを用いて、漫画に登場するセリフと発話者を人手によって対応付けするデータ収集を行う。また、それらのデータをもとに、セリフに対する発話者の正解データセットを構築する。このデータセットを用いることで、機械による発話者の自動推定において精度の検証に利用することができる。

データ収集は明治大学に通う学生 32 名および教員 1 名に依頼し、3.4 節で説明した Web システムによって発話者のアノテーションを付与してもらった。また、データ収集に使用する漫画は Manga109[11, 16] の漫画を対象とし、109 冊全ての漫画に登場するセリフに対してアノテーション付与を行った。

データ収集の協力者はアノテーション付与の Web システムから、漫画を登録してデータ収集を行う。また、協力者にアノテーション付与してもらった対象の漫画はランダムに選択されるものとした。ここで、データ収集では最大 3 冊までの漫画に対して並行してデータ収集のタスクに取り組むことを可能とした。これは、ランダムに選定された漫画が本人にとって馴染みがないものであったときに、その 1 冊のみを続けて読むことが負担になると考えたためである。また同様の理由でタスクを負担に感じた場合には「リタイア」を表明することで他の協力者へタスクを回すことが可能な仕組みも用意した。その際のペナルティについては、同時に読める漫画の数を 1 冊減らすこととした。こうしたタスクの取り組み方に自由度を設けることにより、協力者の負担を減らし円滑にデータ収集が行えるように試みた。

これらの条件のもと、1 冊の漫画に対して 2 名の協力者がアノテーションの付与を行うまでデータ収集タスクに参加可能とした。1 冊に対して 2 名を割り当てた理由としては、アノテーション付与者が 1 名であった場合そのデータが信頼できるかについて判断が難しいという問題が存在するためである。このデータの信頼性については次節で詳しく述べる。また、データ収集に参加した協力者には達成した漫画の冊数に応じて報酬を支払った。そのため、データ収集のタスク数に上限は設けておらず、1 冊のみに参加した協力者もいれば、最大で 35 冊もの漫画に対してアノテーションの付与を行なった協力者も存在する。これにより、収集開始から約 1 ヶ月で 109 冊のそれぞれの漫画に対して 2 名がアノテーションの付与を完了した。これらのデータ収集によって構築したセリフ発話者のデータセットは明治

大学中村研究室のウェブページ<sup>\*1</sup>にて公開している。

## 4.2 収集したデータの分析

本節ではデータ収集によって集まったデータについて、データの信頼度およびデータの傾向について分析を行う。

前節でも説明したとおり、データ収集では1冊の漫画に対して2名の協力者がセリフに対する発話者のアノテーションを付与している。ここで、収集したデータの信頼性を確認するため、セリフに対して2名の意見が一致していた場合と不一致であった場合について分析を行う。表4.1は、収集したデータについて一致していた件数、不一致だった件数とその内訳を示したものである。表が示す割合は、アノテーション付与の対象である全147,918件のセリフの数を分母にした際の割合である。

表4.1: 収集したデータの一致・不一致の件数と割合

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%



©新居さとし「あくはむ」

(a) 話のタイトル



©佐藤正「燃える！お兄さん」

(b) ト書き

図4.1: 「ナレーション」と判断された例

表4.1より、データセットに存在するセリフのうちの86.9%は発話者であるキャラクターが正確に対応付けられたと考えられる。また、一致していた中でも「ナレーション」と判断されたセリフが2.4%存在した。これは図4.1(a)のように4コマ漫画の話のタイトルで

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

あったり、図 4.1(b) のような場所や状況を説明するト書きであったりが含まれる。特に Manga109 における 4 コマ漫画には必ず話のタイトルにアノテーションが付与されているため、これらのほとんどが「ナレーション」として判断されたものであると考えられる。

また、図 4.2 は 109 冊の漫画ごとに一致率を求め、その分布を示したものである。図 4.2 の横軸は 5% ごとの一致率の階級値を示しており、縦軸はその階級の一致率である作品の数を示している。図 4.2 より、最も作品数が多いのは一致率 90%~95% の区間の 37 作品であり、最も精度が低い一致率 55%~60% の区間には 1 作品、最も精度が高い一致率 95%~100% の区間には 5 作品が存在することがわかった。

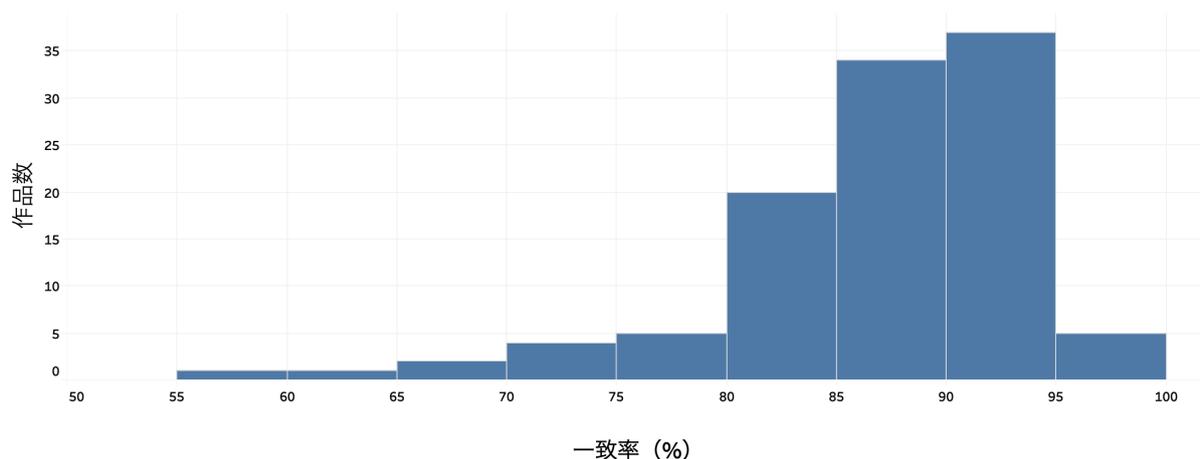


図 4.2: 作品ごとの一致率の分布

ここで、Manga109 の 109 冊の漫画にはジャンル・ターゲット層・出版年代の 3 種類のカテゴリが存在し、各作品をジャンル等によって分類することができる。表 4.2 はそれぞれのカテゴリにおける項目とその作品数を示したものである。

表 4.2: ジャンル・ターゲット層・出版年代の項目と作品数

ジャンル	ターゲット層	出版年代
4 コマ	少年	1970 年代
SF	少女	1980 年代
ギャグ	青年	1990 年代
サスペンス	女性	2000 年代
スポーツ		2010 年代
バトル		
ファンタジー		
ホラー		
ラブコメ		
時代物		
動物		
恋愛		

この各カテゴリの項目は漫画の特徴を表していると考えられるため、発話者の意見の一致率もこれらの項目ごとに異なる可能性がある。図 4.3 は漫画のジャンル・ターゲット層・出版年代のカテゴリごとに、一致率の平均を項目によって比較したものである。図 4.3(a) より、漫画のジャンルとしては「恋愛」「ラブコメ」の作品の一致率が比較的高く、「動物」「4コマ」の作品の一致率が比較的低かった。また図 4.3(b) より、漫画のターゲット層によってそこまで大きな差はなかったが、「青年」「女性」の方が一致率が高い傾向が見られた。また図 4.3(c) より、漫画の出版年代としては「2010 年代」の作品の一致率が最も高く、「2000 年代」の作品の一致率が最も低かった。また、2010 年代を除けば、出版年代が新しいほど一致率が下がる傾向が見られた。

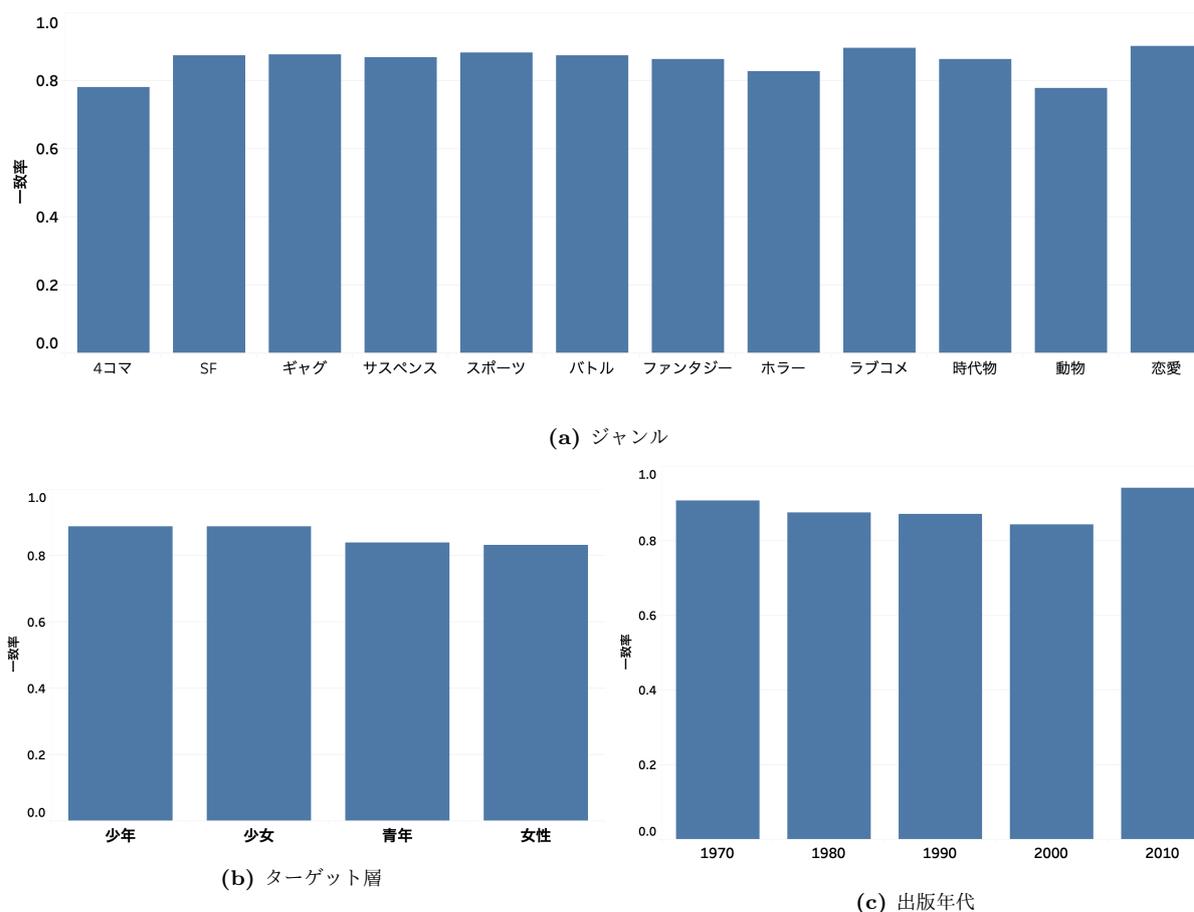


図 4.3: 漫画のカテゴリごとの一致率の比較

### 4.3 収集したデータに対する考察

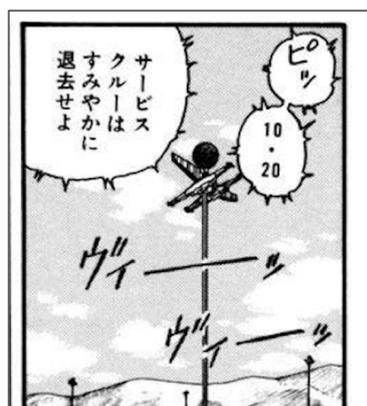
本節ではデータセットを分析した結果をもとに、人手によるセリフ発話者の対応付けについて考察する。

まず、アノテーションを付与した 2 名の意見が一致していたセリフの割合は 86.9% であったことから、漫画に登場するセリフの多くは人によって判断が可能であることがわかる。しかし、Manga109 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つの見開きあたり約1.8件程度の発話者の不一致があったこととなる。よって、人によって判断する場合も発話者が判断しづらい状況がある程度存在していることがわかる。

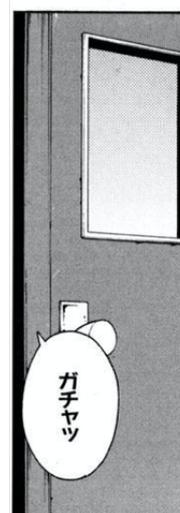
次に、2名の意見が「不明」であったものや、2名の意見が不一致だったものについて分析する。

図4.4は、2名の意見が「不明」であったものの例である。図4.4(a)はスピーカーから発せられるセリフであり、その場にキャラクターが存在しないため誰が発話者であるかわからなかった例である。このようなスピーカーや電話など遠隔地にキャラクターが存在するセリフの発話者を判断することは困難であることがわかる。また、図4.4(b)は扉を閉めた際の効果音であり、セリフとは異なるテキストとなっている。この他にも、「コンッコンッ」など、キャラクターではなく無機物などから出てくる音はデータセットにある程度存在するため、これらを「ナレーション」同様、発話者の存在するセリフと区別する必要があると考えられる。



©加藤雅基「ARMS」

(a) スピーカーの音声



©たくじ「メテオさんストライクです！」

(b) 無機物の効果音

図4.4: 2名とも「不明」であった例

次に、2名の意見が不一致だった例について述べる。図4.5は2名の意見が不一致だったものの例の1つである。青線で囲まれたセリフについて、片方のアノテーション付与者は下の男性とセリフが近いと判断していた。しかし、これは上の女性のセリフの続きであり、実際の発話者は上の女性の方が正解である。このように、前後のセリフの連続性の考慮などが発話者の判断には必要であり、セリフとキャラクターの距離だけを見ると発話者を間違えてしまう状況が存在する。

また、図4.6は2名のキャラクターが交互に会話している例であるが、このそれぞれがどちらのキャラクターのセリフであるかについて意見が別れていた。また、これらは吹き出しが連結している上に、発話者の方向を示す吹き出しの「しっぽ」も見当たらないため発話者の手がかりが極端に少ない例となっている。



©中貫えり「どんぶらっこ」

図 4.5: 距離が近いキャラクターと勘違いしている例



©愛田真夕美「魔夜の赤い靴」

図 4.6: 交互に会話してどちらかわからない例

また、図 4.2 の作品ごとの分布で最も精度が低い一致率 55%~60% の区間に存在した作品については、図 4.7 の青線のセリフのようにコマ内にキャラクターが存在せず、似たようなキャラクターの会話であるため判別が難しかった。図 4.7 の漫画では常に上の二名のメインキャラクターが会話してるうえに、コマ内にキャラクターがいないことが多いため、特に一致率が低い作品となっていた。

これらは人によっても発話者の判断が難しいものであると考えられ、セリフの発話者を判断する上での課題となると考えられる。また、セリフの対象の 1 コマだけで発話者を判断することは難しく、前後の会話の文脈や、セリフの内容からキャラクターの特徴を見る必要があるため、ある程度その漫画を読みストーリーやキャラクターについて理解していないと判断できないと考えられる。

また、図 4.3 のカテゴリごとの一致率の比較より、ターゲット層によっては大きな違いは見られなかったが、出版年代は新しくなるほど一致率が低くなる傾向が見られた。出版年代



©国樹由香「エヴリデイおさかなちゃん」

図 4.7: もっとも一致率が低かった作品の例

では 2010 年が最も一致率が高いが、これは 2010 年代の作品が 1 作品しか存在しないため、その作品ならではの一致率であると考えられる。出版年代が古いほど一致率が高い原因としては、新しい作品ほどより表現が多様になり、セリフと発話者の状況について複雑なものが増えた可能性が考えられる。

ここで、データ収集に参加した協力者に対して、データ収集のタスクやアノテーション付与システムに対する意見として表 4.3 に示す事後のアンケートをとった。回答は任意であったため、協力者 33 名のうち、25 名から回答が得られた。

表 4.3: 事後アンケートの設問と回答形式

	設問	回答形式
Q1	全体を通して発話者の選択はどの程度迷いましたか？	(全く迷わなかった) ~ (とても迷った) の 5 段階
Q2	選択に迷った具体的な例を教えてください。	自由記述
Q3	実験システムでどちらの機能を多く利用しましたか？	(ドラッグ&ドロップ)・(キャラクター一覧から選択)・(どちらも同程度に使用) の 3 項目から選択

アンケートの Q1「全体を通して発話者の選択はどの程度迷いましたか？」については、全く迷わなかったの評価値を 1，とても迷ったの評価値を 5 としたところ、評価値 2 が 12 人、評価値 3 が 4 人、評価値 4 が 9 人であり、評価値の平均は 2.88 であった。これより、発話者の選択にやや迷った協力者とそれほど迷わなかった協力者の数は同程度存在することがわかる。また、Q2「選択に迷った具体的な例を教えてください」にて回答された例を以下にまとめる。

- 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

- キャラクターの人格が入れ替わる場合
- 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

また、以下は Q2 によって得られた意見の中でも、データセットや実験環境の問題によって発生した例をまとめたものである。

- 漫画の 1 巻目ではなく続巻から読む場合
- 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
- 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

これら 3 つの意見は、セリフ自体に問題があるわけではなく、データセットの不備や実験システムの都合のためセリフと発話者の対応付けが困難になった例である。特に「漫画の 1 巻目ではなく続巻から読む場合」については、Manga109 において 1 巻目ではない作品の総数が 18 巻であることが関係する。これらは 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。これらの漫画は 2 巻以降の続巻から読む場合、その漫画の知識がない協力者にとっては難しいタスクになっていたと考えられる。また、「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、漫画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。

## 第 5 章 セリフと話者の対応付けの技術的困難性

セリフと発話者の機械的な自動対応付け手法の実現に向け、ここではセリフの発話者の判定における技術的困難性の分析を行う。以下の節では、その詳細について具体例を示し、最後に整理する。

### 5.1 要因の分析

#### 5.1.1 吹き出しの形状

漫画に登場するセリフの多くには、セリフの領域を囲んで表示される吹き出しが存在することが一般的である。ここで吹き出しには、図 5.1 のように発話者の方向を示す「しっぽ」と呼ばれる形状が存在することが多い。この図では、吹き出しのしっぽが同じフレームに存在する 1 人のキャラクターの方向を指し示しており、発話者が誰であるかを判断しやすくなっている。また、図 5.2 の例のように、フレームに複数のキャラクター・セリフが存在する場合でも、しっぽの方向によってそれぞれの発話者を判断することが可能である。



©石岡ショウエイ「ベルモンド Le VisiteuR」

図 5.1: 吹き出しのしっぽ方向による発話者の判別 (1 人)



©吉田秋生「海街 diary」

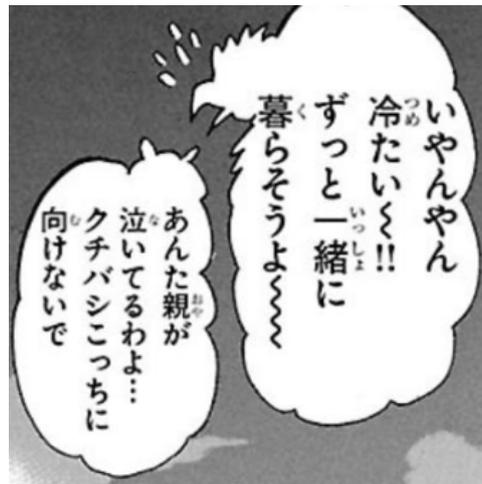
図 5.2: 吹き出しのしっぽ方向による発話者の判別 (複数人)

しっぽ以外にも、吹き出しから発話者を判断する手がかりを得ることができる。例えば図 5.3 は、人間と機械が対話しているシーンであるが、機械のセリフは角張った吹き出しで表現され、人のセリフは丸みを帯びた吹き出しで表現されるなど、吹き出しの形状で区別されている。また、図 5.4 のように、吹き出しの形にキャラクタを反映したものも存在する。このように、吹き出しのしっぽや形状は、発話者を判断するための大きな手がかりになると考えられる。



©山田胡瓜「AI の遺伝子」

図 5.3: 機械の音声と人間の音声による吹き出し形状の違い



©板垣巴留「BEASTARS」

図 5.4: 吹き出しの形状がキャラクタを表す例

しかし、吹き出しの形状のみで確実に発話者を判定することは限らない。例えば図 5.5 では、吹き出しのしっぽがコマの外側に向けて伸びているうえ、このコマはページ内の最後のコマであるため、その方向にキャラクタがおらず（この状況では発話者は子供であり、身長が低いいためこのような表現となっている）、吹き出しの形状のみでは発話者を判断することができない。また、図 5.3 の左下の吹き出しのようにしっぽが存在しないこともあるうえ、しっぽがあってもその向きが不明確であったり、しっぽが向いている方向に複数のキャラクタがおり判断がつかないこともある。さらに、図 5.6 のように吹き出しが存在し

ないセリフも多く登場する。なお、吹き出しがないセリフは、発話者の内面のセリフ（非肉声のセリフ）として表現されることが多い。



©三原和人「はじめアルゴリズム」

図 5.5: 吹き出しのしっぽがコマ外に向く例



©平雅巳「黒井戸眼科」

図 5.6: セリフの吹き出しが存在しない例

### 5.1.2 セリフとキャラクタの位置関係

セリフには、前節で述べた吹き出しのしっぽが存在しないものや、吹き出しが存在しないものがある。これらのセリフの発話者を判断する手がかりの1つとして、漫画のコマにおけるセリフと発話者の位置関係がある。



©佐々木倫子「チャンネルはそのまま！」

図 5.7: キャラクタに近い位置にセリフが存在



©出口竜正「ドールガン」

図 5.8: 顔の横にセリフが提示されている例

図 5.7 の例では、吹き出しのしっぽがないものの、セリフが発話者の近くに配置されたり、発話者のキャラクタの体に重畳するように配置されることで位置関係が示されている。また、図 5.8 の例のように、吹き出しのないセリフについても、キャラクタの顔の横に提示することで、そのキャラクタが発話者であることを示すものも多く存在している。

図 5.9 は 2 人のキャラクタが会話しているシーンであるが、最初の発話はキャラクタの近くに配置し、それ以降のセリフについては読者が読む順番に合わせて交互にキャラクタのセ

リフが入れかわるように配置されている。この工夫により読者は誰のセリフかを判断できるが、機械的に判断することは容易ではないだろう。



©緑山のぶひろ「畏ガール」

図 5.9: セリフの発話者が交互に入れかわっている例



©あきづき空太「赤髪の白雪姫」

図 5.10: セリフとキャラクターの配置が近くても発話者が異なるキャラクターである例

一方、図 5.10 では右側のキャラクターの近くにセリフがあるものの、そのセリフは左側のキャラクターから右側のキャラクターに向けて放たれたセリフであり、実際の発話者はセリフから遠い位置にいる。そのため、セリフとキャラクターの位置関係は手がかりの1つであるものの、それのみで発話者を判断することは難しい。

そこで、次節で述べるセリフの内容および表現から得られる情報にも着目する必要がある。

### 5.1.3 セリフの表現とキャラクターの特性

セリフ自体には、発話者が誰であるかといった情報が含まれており、読者はセリフの特徴から発話者を推定している。

図 5.11 のように、同じコマの中に複数のセリフと複数のキャラクターが存在する場合、吹き出しの形状や配置、しっぽの向きあいまいさによって発話者を判断することが難しい。この場合、セリフの内容が発話者を判断する手がかりの1つとなる。この図 5.11 の場合では、セリフに方言（関西弁）が使われていたり、語尾が特徴的だったり、言葉づかいが丁寧だったり、セリフの特徴から発話者を絞ることができる。ここで、図 5.11 の左端のセリ

フは吹き出しが一体となっているため、吹き出しの形状のみを見ると 1 人の発話者による連なったセリフに見えてしまう可能性がある。しかし、セリフの内容を見ると、吹き出しの左側のセリフのみ丁寧な言葉づかいとなっており、それぞれ別の発話者によるセリフだと判断することができる。この例以外にも、「俺」「わし」「私」「僕」といった一人称や、他者に対する呼び方「あなた」「お前」「貴様」なども発話者を判定する情報の 1 つとなる。



©赤松健「ラブひな」

図 5.11: セリフの内容によって発話者が判断できる例

一方、セリフの表現が発話者の手がかりとなる場合もある。図 5.12 では、AI のキャラクターのセリフのフォントが、女子高生のキャラクターのセリフのフォントと異なることで、フォントの違いにより発話者のキャラクターが異なることを表現している。また図 5.13 では、この漫画においてほとんどのセリフが縦書きで表現されているが、一部のセリフは横書きで表現されている。このように縦書き・横書きといった文字の向きの違いに関する表現は、外国語や存在しない未知の言語をしゃべるキャラクターのセリフの表現として用いられることが多く、登場するキャラクターの中から発話者の候補を絞ることが可能になる。



©かっぴー/うめ「アイとアイザワ」

図 5.12: フォントによる人と AI のセリフの区別



©高浜寛「ニクスの角灯」

図 5.13: 縦書きで日本語、横書きで日本語表記だが外国語を意味する表現

また、セリフから得られる特徴とキャラクターの見た目の一致から、発話者を判断することもできる。図 5.14 において、コマの左端にあるセリフの発話者は右側の女性であるが、キャラクターとの位置関係のみでは判断が難しい。ここで、左端のセリフは、セリフの文章や特徴的な吹き出しの形、吹き出しの中に汗を表す絵を入れるなどの表現により、発話者の

「焦り」の感情が示されている。また、セリフの感情と一致する表情のキャラクタが、発話者である可能性が高いことから、この図では右側の女性のセリフであると判断することができる。また先述の図においても、位置関係による判断は容易ではないものの、2人のキャラクタのうち、片方のキャラクタのみ口を開けた描写であるため、セリフの発話者はその口を開けたキャラクタであると判断できる。



©草水敏/恵三郎「フラジャイル 病理医岸京一郎の所見」

図 5.14: セリフの感情と発話者の表情の一致により発話者が判断できる例

一方、図 5.15 のようにセリフの近くや吹き出しの中にその発言者を示すマークを置くことで、発話者の情報を補足している場合も存在する。この例では、それぞれのキャラクタのデフォルメした顔を示す絵を吹き出しの中に描くことで、セリフの発話者を判断することが可能となっている。



©ろびこ「僕と君の大切な話」

図 5.15: セリフの発話者のマークが提示されている例

このように、セリフの内容や表現から得られる情報と、キャラクタに関する情報を照らし合わせることで、その発話者を判定できると考えられる。

#### 5.1.4 特殊な例

これまでに説明した例以外にも、様々なセリフとその発話者のパターンがある。例えば図 5.16 は、1つのセリフに対してその発話者が複数人いる場合である。このように、発話者が1人とは限らないため、複数人による発言についても考慮する必要がある。また、図 5.17 における吹き出しは、中央のキャラクタが持つ携帯電話から発せられるセリフであり、発話者はその場所にいないため自動判別が困難であるといえる。

このように、セリフとその発話者についての様々な例があるとともに、その判定の難易度



©赤松健「ラブひな」

図 5.16: 1 つのセリフの発話者が複数人である例



©赤松健「ラブひな」

図 5.17: 携帯電話からセリフが発せられる例

にも違いがある。また、これまで述べたように、特定の要素のみで発話者を断定することは難しく、吹き出しの形状やセリフの内容といったそれぞれの要素による判断を組み合わせ、統合的に判断する必要があると考えられる。

## 5.2 技術的困難性にまつわる要因の整理

表 5.1 は、5.1 節の具体例をもとに、発話者の判定において、その難しさに影響を与える要素を整理したものである。それぞれの要素の組み合わせにより、技術的困難性は変化していくことになり、全ての要素が Easy であれば簡単であるが、全ての要素が Hard だと困難になる。

表 5.1: 判定難易度の要因例

	要素	Easy	Hard
吹き出しの形状	吹き出しの有無	あり	なし
	吹き出しのしっぽ	あり	なし
	吹き出しの形	キャラクタに合わせた形状	同一
セリフとキャラクタの位置関係	発話者の存在	コマ内	コマ外
	コマ内のキャラクタ数	1人	複数人 or 0人
	コマ内のセリフ数	1つ	複数
	発話者とセリフの距離	近い	遠い
セリフの表現とキャラクタの特性	セリフのフォント	キャラクタ固有	同一
	セリフの方向	キャラクタ固有 (日本語は縦書き, 英語は横書きなど)	同一
	セリフの口調	キャラクタ固有 (一人称・方言・語尾など)	同一
	セリフとキャラクタの見た目の関係	同期している (感情と表情の一致・発話中の口の形など)	同期していない
	発話者のヒント	あり (キャラクタのマークや名前の頭文字)	なし

## 第 6 章 セリフ発話者の自動推定

### 6.1 推定手法の提案

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。

前章で述べたように、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（「距離」の手法）
2. 同じコマ内にあるキャラクターの情報による推定（「同じコマ」の手法）
3. 吹き出しのしっぽの方向の情報による推定（「しっぽ方向」の手法）
4. 一人称と語尾の情報による推定（「一人称・語尾」の手法）

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ー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。

#### 6.1.1 セリフとキャラクターの距離の情報による推定

前節で述べた「セリフとキャラクターの位置関係」のうちの手がかりの1つとして、セリフとキャラクターが近い距離に存在する可能性が高いといえる。Rigaud ら [13] 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 Manga109[11, 16] のアノテーション情報より、セリフおよびキャラクターの画像上の位置（矩形）の情報を利用する。具体的には、推定対象のセリフの画像上の中心座標を求め、同じくページに存在するキャラクターの中心座標をそれぞれ求める。このセリフとキャラクターの距離をもとに、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、より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。

以降では本手法を「距離」の手法と記述する。

### 6.1.2 同じコマ内にあるキャラクターの情報による推定

セリフを判断する手がかりの1つである「セリフとキャラクターの位置関係」より、セリフとその発話者であるキャラクターは同じコマに存在する可能性が高いと言える。よって、セリフが所属するコマに対して、同じコマに存在するキャラクター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ー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ーが1名の場合はそのキャラクターが1点、キャラクターが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ーが複数人いる場合、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、前項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推定が困難となる可能性がある。その際に、片方が本手法によってスコアを加算することで推定を容易にすることが考えられる。

以降では本手法を「同じコマ」の手法と記述する。

### 6.1.3 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手法を提案する

吹き出しのしっぽの検出については、田中ら [12] の手法を参考に、吹き出しを検出した。また、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、線分とキャラクター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、漫画画像の対角線の10%の長さに固定した。また、線分と交差するキャラクターが複数名いる場合、それぞれに対して均等にスコアを割り振ることとした。

以降では本手法を「しっぽ方向」の手法と記述する。

### 6.1.4 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ーの特性」より、キャラクターの一人称と語尾を利用する方法を提案する。

今回は著者が選定した一人称と語尾のコーパス（表 6.1）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、まずはじめに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、これまでに説明した「距離」「同じコマ」「しっぽ方向」による手法によって推定された発話者の情報をもとにスコアリングを行うこととする。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3つの手法による推定によってある特定のキャラクターが発話者として推定され

表 6.1: 利用した一人称と語尾のコーパス

一人称	語尾
私, わたくし, 僕, 俺, 俺様, 俺, 我輩, 我, 妾, 拙者, 某, あたし, あたくし, あたい, うち, わえ, わて, わい, オイラ, オラ, おい, おいどん, おれっち	アル, おす, かしら, ござる, じゃ, じゃき, じゃけん, ずら, づら, ぜ, ぜよ, だぜ, だべ, だよ, っしょ, であります, である, デース, でげす, でございます, でごわす, ですぞ, ですよ, ですわ, でやんす, のじゃ, やで, やん, わい

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ー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。

以降では本手法を「一人称・語尾」の手法と記述する。

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。組み合わせの際には, それぞれの手法によるスコアの合計値により, 発話者の推定を行うこととする。

## 6.2 評価方法

前節で説明した推定手法を用いてセリフ発話者の自動推定を試みる。具体的には, 第 4 章で構築したデータセットのセリフを対象に発話者を推定し, その推定精度を検証する評価実験を行う。

セリフ発話者の正解データセットには, アノテーション付与者 2 名の意見が一致していた 128,502 件のセリフを利用する。ここで, 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ーも存在する。これは, 漫画の画像上に存在するキャラクターではあるが, Manga109 において特に名前が指定されていないキャラクターの総称となっている。そのため, これら「その他」のキャラクター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必要がある。これら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ーを推定し, そのキャラクター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。

以上の評価方法により, 自動推定の評価実験を行った結果を次節にて説明する。

## 6.3 結果

6.1 節で述べた 4 つの手法について, 単独の手法によって推定した場合とそれらを組み合わせて推定した場合の結果を表 6.2 に示す。表 6.2 の正解率は推定対象の 121,364 件のセリフについて, 推定した発話者が正解していた割合を示している。表 6.2 の結果より, 単独の手法で最も精度が高いのは「距離」による推定の正解率 55.7% であり, 最も精度が低いのは

は「一人称・語尾」による推定の正解率 4.6% であった。また、4 つの手法を組み合わせた場合の正解率は 70.7% であった。

また、表 6.2 の組み合わせの寄与率は、その手法を組み合わせた際の正解率の上昇を示している。例えば「距離」の組み合わせの寄与率 26.4% は、それ以外の「同じコマ」「しっぽ方向」「一人称・語尾」の 3 つを組み合わせた際の正解率が 44.3% であり、4 つの手法を組み合わせた際の正解率である 70.7% との差の値を示している。表 6.2 より、単独の手法による正解率が大きいものほど組み合わせの寄与率も大きいことがわかった。また、組み合わせの寄与率がどの手法においても正の値であることから、いずれも組み合わせることによって推定精度が向上していることがわかった。

表 6.2: 手法ごとの正解率

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	2.3
一人称・語尾	4.6	0.6
<b>組み合わせ</b>	<b>70.7</b>	

ここで、第 3 章でも説明したように、Manga109 の漫画 1 冊に登場するキャラクターの人数は平均 31.7 人、最大 124 人となっており、1 冊の漫画において数回しか登場しないキャラクターも発話者の候補となっている。そのため、発話者の候補を出番の多いキャラクターに絞ることで、より高い精度で発話者の推定ができる可能性がある。そこで、発話者の候補となるキャラクターを限定するため Manga109 のアノテーション情報をもとに、各漫画において出現回数が上位 5 位のキャラクターに絞ることとした。つまり、それぞれの漫画において主要なキャラクター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 名のキャラクターに限定するため、発話者候補となるキャラクターの総数は 109 冊 × 5 名 = 545 名となった。またこれに伴い、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、推定に利用できるセリフの総数は 88,297 件となった。以降の結果については、この 88,297 件のセリフを対象とした結果について述べる。

表 6.3: 主要なキャラクターのみによる推定の正解率

手法	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
<b>組み合わせ</b>	<b>78.6</b>	

表 6.3 は、新たに作成したデータセットを対象に表 6.2 と同様の実験を行なった結果である。表 6.3 の結果より、単独の手法で最も精度が高いのは「距離」による推定の正解率 66.9% であり、最も精度が低いのは「一人称・語尾」による推定の正解率 6.0% であった。また、4 つの手法を組み合わせた場合の正解率は 78.6% であった。今回の実験では 1 つのセリフに対する発話者の候補は 5 名であるため、ランダムに発話者を選択した際の正解率の期待値は 20% となる。よって、本手法による発話者の推定はランダムに発話者を選ぶことに比べ 58% 以上の精度で推定できることがわかった。

また、表 6.2 の結果に比べると「距離」のみの推定の正解率は約 10% 向上し、組み合わせた際の正解率は約 8% 向上していた。しかし、組み合わせの寄与率については表 6.2 の結果より若干低い結果となった。よって、発話者の候補を主要なキャラクタに絞ることで正解率は向上するが、手法を組み合わせることの効果に大きな変化はないことがわかった。

また、図 6.1 は 109 冊の漫画ごとに手法を組み合わせた場合の正解率を求め、その精度の分布を示したものである。図 6.1 の横軸は 5% ごとの正解率の階級値を示しており、縦軸はその階級の正解率である作品の数を示している。図 6.1 より、最も作品数が多いのは正解率 80%~85% の区間の 30 作品であり、最も精度が低い正解率 55%~60% の区間には 2 作品、最も精度が高い正解率 90%~95% の区間には 6 作品が存在することがわかった。

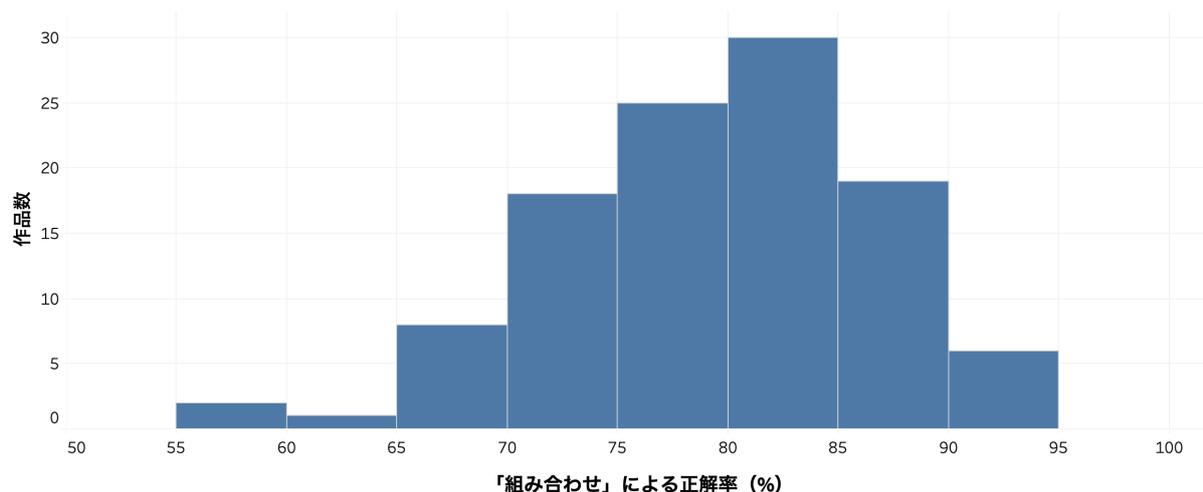


図 6.1: 作品ごとの正解率の分布

また、図 6.2 は第 4 章の表 4.2 で述べた Manga109 の漫画のジャンル・ターゲット層・出版年代のカテゴリごとに、正解率の平均を比較したものである。図 6.2 の各グラフの棒グラフは手法を組み合わせた場合の正解率の平均を示しており、棒グラフに重畳して表示されたマークの位置は「距離」「同じコマ」「しっぽ方向」「一人称・語尾」の 4 つの手法を単独で推定した場合の正解率の平均を示している。図 6.2(a) より、漫画のジャンルとしては「ホラー」「SF」の作品の正解率が比較的高く、「動物」の作品の正解率が比較的低かった。また図 6.2(b) より、漫画のターゲット層としては「青年」の作品の正解率が比較的高かった。また図 6.2(c) より、漫画の出版年代としては「1970 年代」の作品の正解率が高く、「2010 年代」の作品の正解率低くなっており、出版年代が新しいほど正解率が下がる傾向が見られた。

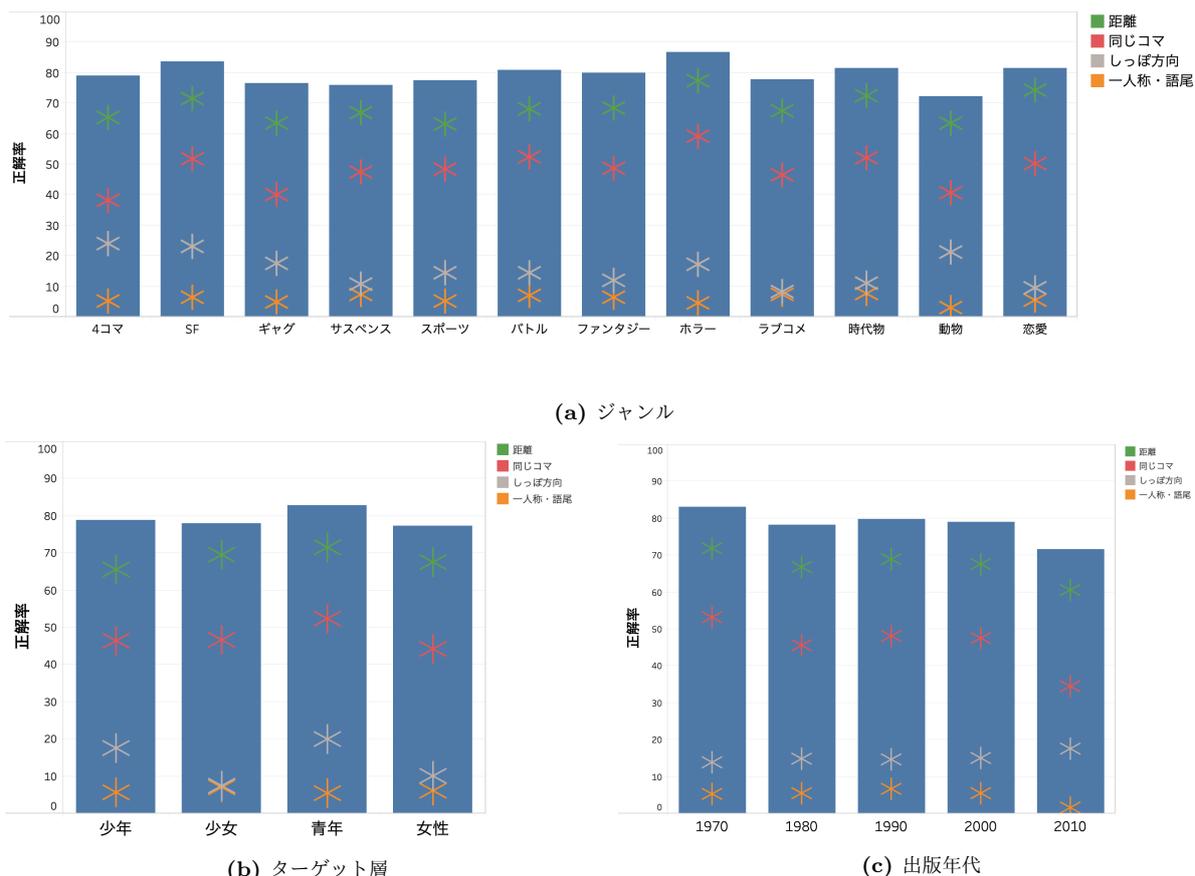


図 6.2: 漫画のカテゴリごとの正解率の比較

## 6.4 考察

前節の結果より，提案した「距離」「同じコマ」「しっぽ方向」「一人称・語尾」による発話者推定の手法について議論する。

主要なキャラクタを対象とし単独の手法によって推定した場合，「距離」による推定の正解率が 69.9% と最も高かった．このことより，セリフの約 70% については最も近い位置に存在するキャラクタが発話者であるということがわかった．また「同じコマ」の手法のみで推定する場合，セリフと同じコマに存在するキャラクタが 1 名の場合に限り，そのキャラクタを発話者として推定できる．よって，「同じコマ」の手法による正解率が 46.1% であることから，セリフの約 40% は同じコマに発話者であるキャラクタ 1 名のみが存在するという状況であることがわかった．これらの結果より，「距離」および「同じコマ」の手法のみであっても，ある程度高い精度で発話者の推定が可能であることが明らかになった．

ここで，「しっぽ方向」「一人称・語尾」の手法による正解率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．推定に利用した 88,297 件のセリフに対して，「しっぽ方向」によってスコアが与えられたセリフの数は 16,251 件（全体の 18.4%），「一人称・語尾」によってスコアが与えられたセリフの数は 7,925 件（全体の 9.0%）であった．よって，「しっぽ方向」については 20% 未満，「一人称・語尾」につ

いては 10% 未満のセリフのみを対象に推定を行っていたことが明らかになった。

「しっぽ方向」の手法を単独で推定に利用した際の精度について、スコアが与えられた 16,251 件のセリフを対象とした場合の正解率は 79.4% であった。「しっぽ方向」によってスコアが与えられないセリフは、吹き出しの輪郭が抽出できないセリフや、しっぽの先端が検出できないセリフとなる。よって、吹き出しの輪郭を抽出し、しっぽの先端が検出できたセリフに対しては約 80% の高い精度で発話者の推定が可能であることがわかった。また、図 6.3 のように吹き出しのしっぽではない箇所をしっぽとして判定しているものも存在した。そのため、「しっぽ方向」による推定の精度を向上するためには、しっぽを検出する手法自体の精度も改良する必要があると考えられる。



©八神健「ありさ2」

図 6.3: しっぽの検出が失敗した例

「一人称・語尾」の手法を単独で推定に利用した際の精度について、スコアが与えられた 7,925 件のセリフを対象とした場合の正解率は 66.7% であった。「一人称・語尾」によってスコアが与えられないセリフは、表 6.1 に示した一人称および語尾が含まれないセリフとなる。よって、これらの一人称・語尾が登場したセリフに対しては 66% の精度で推定が可能であることが明らかになった。ここで、「一人称・語尾」を組み合わせることにより、精度が改善した例を図 6.4 に示す。図 6.4(a) は、「距離」の手法により画像左の女性が発話者と推定されていたが、一人称である「俺」の情報を推定に加えることで、右の男性が発話者であると推定することが可能となっていた。また、図 6.4(b) の左上のセリフは同じコマや周囲にキャラクターが存在せず、しっぽの向いた先にもキャラクターが存在しないため推定が難しいという問題があった。しかし、語尾となる「じゃ」を検出したことで同じように語尾「じゃ」が含まれるセリフの発話者であるが右下の男性を選択することが可能となっていた。これらは 6.1.4 項で説明したように、「距離」「同じコマ」「しっぽ方向」の推定結果からキャラクターごとのセリフを一旦収集し、そこからセリフ内容から見られる特徴を抽出して新しい推定を行なっている。ここで、一人称「俺」や語尾「じゃ」はそれぞれ「男性的」や「老人的」といったキャラクターの属性を捉えていることが考えられる。

しかし、「一人称・語尾」を組み合わせることで推定が失敗した例も存在する。図 6.5 は、右下の女の子の一人称が「オレ」として推定されたため不正解であった例である。これについては、漫画全体において「オレ」を含むセリフが右下の女の子の近くに配置されていることが多かったため、そのキャラクターの一人称が「オレ」になってしまったと考えられる。



©南澤久佳「魔法使い養成専門マジックスター学院☆☆☆」

(a) 一人称による判断



©八神健「密・リターンズ!」

(b) 語尾による判断

図 6.4: 一人称・語尾によって正解した例



©うえだ美貴「爆裂! かんふ一娘」

図 6.5: 一人称の推定が失敗していた例

これらの結果より、「しっぽ方向」および「一人称・語尾」による手法は全体の正解率は低いものの、それらの特徴が出現するセリフに対してはある程度推定が可能であり、有用な手法であると考えられる。特に「一人称・語尾」の手法はセリフの内容からキャラクターの情報を抽出することが可能であると考えられる。よって、「一人称・語尾」以外にもセリフの内容から得られる特徴を利用することにより、より精度が高い推定が可能になると考えられる。

また、作品ごとの正解率の分布を示した図 6.1 の結果より、精度が低い正解率 55%~60% の区間には 2 作品、精度が高い正解率 90%~95% の区間には 6 作品が存在することがわかった。この精度が低い区間の 2 作品については、「犬」や「魚」など人とは異なる生物が多く登場しており、これらの生物が発話者として推定されていたことが精度が低い原因と考えられる。特に、ある生物についてキャラクターが会話する際、コマの中心にその生物を配置し会話のセリフのみが置かれている状況が多く存在する。そのため、これらは「距離」や

「同じコマ」による推定ではその生物が発話者であると誤った推定をしてしまう。これについては、その生物が会話が可能かといったキャラクターとしての特徴を推定することで、発話者の候補から除外するなどの手法が必要となる。また、これら 2 作品はどちらも「動物」というジャンルになっており、図 6.2(a) において「動物」のジャンルの作品の精度が低い原因となっている。

また、正解率 90%~95% の精度が高い区間に存在した 6 作品については、「距離」および「同じコマ」の推定が他の作品よりも高いという特徴が見られた。6 作品における「距離」の手法による正解率の平均は 89.1%、「同じコマ」の手法による正解率の平均は 71.5% であり、表 6.3 における全体の正解率と比較すると 25% 以上高い精度で推定できていることがわかる。

図 6.2(b) のターゲット層ごとの比較については、組み合わせた際の正解率についてはそれほど大きな差は見られなかった。しかし、ターゲット層が「少女」の作品に対する「しっぽ方向」の精度は他の作品に比べて低い結果となっていた。これは、少女漫画のセリフにはモノログ（心の声）が多く、それらに吹き出しが存在しないことが多かったからであると考えられる。また、図 6.2(c) の年代別の比較では、1970 年代が最も高く、2010 年代が最も低いことがわかった。また、年代が新しくなるにつれ正解率が低くなる傾向が見られた。この原因については、古い作品ほど「距離」や「同じコマ」によって推定できるわかりやすい場所に発話者を配置した作品が多く、新しい作品ほど発話者の配置方法が多様になっていることが考えられる。

以上より、こうしたカテゴリごとに精度が異なることより、そのカテゴリの項目に合わせた手法を適用する必要も考えられる。例えば少女漫画では「しっぽ方向」の精度が悪いことから、「距離」や「同じコマ」の精度を重視した推定を行うなどが考えられる。

## 第 7 章 総合考察

本論文の全体の結果をふまえ、セリフと発話者の対応付けに関する知見をまとめる。

まず本論文では、Manga109 に登場する 147,918 件のセリフに対し 2 名以上が発話者のアノテーションを付与したデータセットを構築した。このセリフ発話者データセットは、本論文と同じく漫画のセリフの発話者を自動推定するタスクの正解データとして利用できるとともに、Manga109 のアノテーション情報と同様に他の漫画の研究において利用可能な情報となっている。例えば、キャラクターの発話履歴から人物相関図を作成する上野ら [15] の研究のように、セリフと発話者の情報をあらかじめ入手することで、機械による漫画の認識に関する研究に活用することができる。

構築したデータセットの信頼性については、全体の 86.9% のセリフは 2 名の意見が一致しており、ナレーションも含めると 89.3% のセリフについては発話者がある程度正確に対応付けられていると考えられる。しかし、残りの約 10% のセリフについては発話者が不明であったり、2 名の意見が一致しないなど発話者の正解データとして利用しづらいデータとなっている。よって、構築したデータセットに対しこれらの 10% のセリフを対象に新たにアノテーション付与を行うことや、よりデータの信頼性を高めるためにアノテーション付与者を増やすなどの改善が考えられる。その際、第 3 章で提案したアノテーション付与システムを利用するほかにも、Su[23] らの研究のように、既に付与されたアノテーション情報が正しいかを Yes/No の二択によって回答するなど、今回のデータセットをもとにより簡単なタスクとして情報を収集することが考えられる。ただし、第 4 章でも述べたように、セリフの発話者の判断には前後の会話や漫画全体のストーリーの理解が必要となる。そのため、マイクロタスクなど短時間で実行可能なタスクに変換するには、発話者の判断のためにセリフの前後のストーリーについてどの程度の理解が必要かについての検証が課題となる。また、収集したデータをもとに人にとってセリフと発話者の判断が容易な場面・困難な場面について分析を行なった。分析の結果、人にとっても発話者の判別が難しい場面は多数存在することや、人によって発話者の判断が異なることがわかった。

さらに、本論文の目的であるセリフ発話者の自動推定について、「距離」「同じコマ」「しっぽ方向」「一人称・語尾」の情報を用いて推定を行うことで 78.6% の精度で推定可能であることがわかった。ここで、セリフ発話者推定の先行研究である Rigaud ら [13] の研究ではフランスの漫画に対しては 93%、日本の漫画に対しても最大で 87% の精度で推定可能であるなど、本論文の結果より高い精度で推定が可能となっている。これらは対象とする漫画作品が異なることや、Rigaud らは人手によってしっぽの先端位置をアノテーション付与しているなどの違いがあるため単純な比較は難しい。しかし、セリフとキャラクターの距離が発話者

の判断に有用であるなど、本論文の結果と同様の意見が得られている。これらの既存の研究と本論文が提案する推定手法の大きな違いとしては、「一人称・語尾」といったセリフの内容から得られる情報を利用している点である。今回は「一人称・語尾」の情報を加えることで大幅な精度の向上は見られなかったものの、対象とする単語が出現するセリフに対しては66.7%で推定可能な有用な情報であることがわかった。このことより、セリフ内容から得られる情報を更に増やすことで、より正確なセリフ発話者の推定が可能となると考えられる。

第5章でも述べたように、セリフの内容には発話者を判断するための様々な手がかりが存在する。例えば、図5.14で示したようなセリフの感情とキャラクターの感情の一致に関する情報を加えることで推定精度を向上させることが考えられる。セリフなどのテキストから感情を抽出する方法は、単語の形容詞に着目する研究[49]など様々な手法が自然言語処理の分野で研究されている。また、漫画からキャラクターの表情や感情を抽出する方法についても、キャラクターの顔のパーツを推定する[30]ことで表情を認識することが考えられる。他にも、漫画にはキャラクターの感情の表現として漫符と呼ばれる記号が画像上に存在し、これらの情報もキャラクターの感情を抽出する手がかりとなる。よって、セリフという文章から得られる感情と、キャラクターというイラストから得られる感情の一致を見るための情報は現状でもある程度抽出可能であり、これらの情報をセリフ発話者の推定に利用することが可能であると考えられる。また4つの手法を組み合わせた推定について、今回はそれぞれの推定のスコアの合計によって推定を行っていたが、それぞれの手法に対して重み付けを行なったうえで合計するなどの改善も考えられる。例えば、「距離」によるスコアと「しっぽ方向」によるスコアでは、今回の推定の結果より「距離」のスコアの方が精度が高いため、重視すべき情報と考えられる。このようなそれぞれの手がかりの重要度を考慮することで、組み合わせた際の精度を向上させることが可能であると考えられる。

最後に、本論文の結果を活用することによる、漫画の自動認識の展望について語る。本章の前半でも述べたように、構築したセリフ発話者のデータセットを利用することで、漫画の内容理解に関する様々な研究に活用できる。例えば、あるキャラクターのセリフを収集し、それらを発話履歴として分析することでそのキャラクターの性格の推定に利用することが考えられる。また、セリフから会話相手に対する呼び名（呼び捨てや敬称など）を抽出できれば、その相手との関係性を推定する手がかりとなる。このようなキャラクターに関する推定の他にも、キャラクター同士の会話の場面を推定することで漫画のストーリーの推定に活用することも考えられる。例えば、特定の男性キャラクターと女性キャラクターの会話が多く、会話内容から「好意」などの感情を抽出できれば、その2名のキャラクターを中心とした恋愛ストーリーであるなどの推測ができる。これらのセリフの発話者の情報をもとにした漫画内容の理解が可能となれば、本論文のセリフ発話者の自動推定と組み合わせることで機械による漫画の自動認識に大きく貢献することが可能であると考えられる。そのためには、今回のセリフ発話者の推定に必要な情報であるキャラクターの位置やセリフの文字列についても自動で抽出する必要があり、既存の研究[11, 28, 29]と組み合わせることで実現できると考えられる。これにより、最終的には1枚の漫画の画像を入力することで漫画の内容に関する様々な情報を自動で抽出可能となる手法を生み出すことが、本論文を含めた漫画に関する研究の大きな目標となる。また、これらの漫画に対する高度な自動認識が、よりユーザの好みにあっ

た漫画の推薦や新しい鑑賞方法へと活用される未来が期待される。

## 第 8 章 まとめ

本論文では漫画のセリフ発話者の自動推定の課題に対して、セリフと発話者を対応付けたデータセットの構築と人手によるセリフと発話者の対応付けの分析を行ったうえで、自動推定手法を提案し推定精度を評価した。

発話者のアノテーションのデータを効率よく収集するシステムの実装により、147,918 件のセリフに対する発話者のデータセットを構築した。

人手によるセリフと発話者を対応付けを分析した結果、86.9% のセリフはアノテーション付与者の意見が一致したが、それ以外は不一致であった。この結果より、人にとっても発話者の判断が難しい状況が存在することが明らかになった。

また、そのうちの 121,364 件のセリフを対象に、発話者を推定する手法として「セリフとキャラクターの距離の情報による推定」「同じコマ内にいるキャラクターの情報による推定」「吹き出しのしっぽの方向の情報による推定」「一人称と語尾の情報による推定」の 4 つの推定手法を提案した。データセットを用いた評価実験の結果、手法を組み合わせることで 70.7% の精度で発話者の推定が可能であり、発話者の候補を主要なキャラクターに絞ることで 78.6% の精度で推定できることがわかった。これらの結果より、より精度の高い発話者の推定にはセリフ内容から得られる他の特徴を組み合わせた推定が必要であることが示唆された。

また、これらの結果を用いて、さらに深い漫画の理解を機械が獲得することで、より多様な漫画サービスへの利用に貢献することを検討している。

## 謝辞

多くの人の支援により、本論文を執筆することができました。まずは中村聡史先生には、学部3年から修士2年までの4年間で論文執筆・学会発表といった様々な研究活動をご指導いただき、大変お世話になりました。この4年間は成長の4年間であり、中村研究室で培った能力は今後の人生の大きな糧になると信じています。貴重な時間・経験をいただいたことに深く感謝を申し上げます。また、同じく4年間共に研究室で学んだ山浦くん、斉藤くん、樋川くん、徳久くん、そして先輩・後輩の皆さんがいたことで、辛く厳しい研究も乗り越えることができたと感じています。特に土屋先輩、大野先輩、松井先輩には論文執筆にはじまり大変お世話になりました。深く感謝しています。先生および先輩方はまさに師匠であり、自分の目指すべき目標です。また、自分がときに指導し、ときにサポートしてもらった後輩の皆さんがいたからこそ、本論文を執筆することができました。後輩の又吉くんとはプログラミングや色んな知識について語る時間が多く、とてもいい刺激を与えてくれる存在でした。本当に3年間ありがとう！そしてなにより、6年間の昼食事情を支えてくれた中野のお弁当や「つるや」さんに深く感謝します。そして最後に、この25年間の人生をサポートしてくれた父、母、妹の家族の皆様にも深く深く感謝いたします。

## 参考文献

- [1] 集英社：週刊少年ジャンプ.
- [2] 一般社団法人日本雑誌協会：印刷部数公表（2019年1月～3月）少年向けコミック誌，一般社団法人日本雑誌協会，入手先〈<https://www.j-magazine.or.jp/user/printed/index/44/14>〉（参照 2020-2-18）.
- [3] 尾田栄一郎：ONE PIECE.
- [4] Museum, B.: The Citi exhibition Manga.
- [5] 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，公益社団法人全国出版協会，入手先〈<https://www.ajpea.or.jp/information/20180226/index.html>〉（参照 2020-2-18）.
- [6] 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，公益社団法人全国出版協会，入手先〈<https://www.ajpea.or.jp/information/20190725/index.html>〉（参照 2020-2-18）.
- [7] "Mantra": マンガの超高精度な自動翻訳，入手先〈<https://mangahot.jp/>〉（参照 2020-2-18）.
- [8] 株式会社コアミックス（編集）/イマジニア株式会社（運営）：マンガほっと - 人気マンガや名作マンガが基本無料で毎日読めるマンガアプリ，株式会社コアミックス（編集）/イマジニア株式会社（運営），入手先〈<https://mangahot.jp/>〉（参照 2020-2-18）.
- [9] Narita, R., Tsubota, K., Yamasaki, T. and Aizawa, K.: Sketch-Based Manga Retrieval Using Deep Features, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 03, pp. 49–53, DOI: 10.1109/ICDAR.2017.291 (2017).
- [10] PARK, B., IBAYASHI, K. and MATSUSHITA, M.: Classifying Personalities of Comic Characters Based on Egograms, *International Symposium on Affective Science and Engineering*, Vol. ISASE2018, pp. 1–6, DOI: 10.5057/isase.2018-C000029 (2018).
- [11] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T. and Aizawa, K.: Object Detection for Comics using Manga109 Annotations, *CoRR*, Vol. abs/1803.08670, available from 〈<http://arxiv.org/abs/1803.08670>〉 (2018).
- [12] 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, DOI: 10.3169/itej.64.1933 (2010).

- [13] 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, *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 351–355, DOI: 10.1109/ICDAR.2015.7333782 (2015).
- [14] 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).
- [15] 上野高士, 風間一洋: コミック中の発話タイミングに基づく人物相関図の作成手法, ARG 第 11 回 Web インテリジェンスとインタラクション研究会, pp. 1–6 (2017).
- [16] 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, DOI: 10.1007/s11042-016-4020-z (2017).
- [17] 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, *2013 12th International Conference on Document Analysis and Recognition*, pp. 1145–1149, DOI: 10.1109/ICDAR.2013.232 (2013).
- [18] 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*, , DOI: 10.1109/CVPR.2017.686 (2017).
- [19] 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, pp. 4Pin116–4Pin116, DOI: 10.11517/pjsai.JSAI2018.0\_4Pin116 (2018).
- [20] 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, Vol. JSAI2018, pp. 4M105–4M105, DOI: 10.11517/pjsai.JSAI2018.0\_4M105 (2018).
- [21] Papadopoulos, D. P., Uijlings, J. R. R., Keller, F. and Ferrari, V.: Extreme Clicking for Efficient Object Annotation, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4940–4949, DOI: 10.1109/ICCV.2017.528 (2017).
- [22] 中村聡史, 十文字優斗: 音楽理解技術を利用したマイクロタスク埋め込み型音楽ゲームの自動生成, 技術報告 7, 明治大学 / JST CREST, 東京農工大学 (2015).
- [23] Su, H., Deng, J. and Li, F. F.: Crowdsourcing annotations for visual object detection, pp. 40–46 (2012).
- [24] 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, DOI: 10.1109/ICDARW.2019.00013 (2019).
- [25] C. Correia, J. a. M. and P. Gomes, A. J.: Balloon extraction from complex comic

- books using edge detection and histogram scoring, *Multimedia Tools and Applications*, Vol. 75, No. 18, pp. 11367–11390, DOI: 10.1007/s11042-015-2858-0 (2016).
- [26] Kovanen, S. and Aizawa, K.: A layered method for determining manga text bubble reading order, *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4283–4287, DOI: 10.1109/ICIP.2015.7351614 (2015).
- [27] Tolle, H. and Arai, K.: Method for Real Time Text Extraction of Digital Manga Comic, *International Journal of Image Processing* (2011).
- [28] Aramaki, Y., Matsui, Y., Yamasaki, T. and Aizawa, K.: Text detection in manga by combining connected-component-based and region-based classifications, *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2901–2905, DOI: 10.1109/ICIP.2016.7532890 (2016).
- [29] 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, DOI: <https://doi.org/10.1016/j.patcog.2018.08.008> (2019).
- [30] Stricker, M., Augereau, O., Kise, K. and Iwata, M.: Facial Landmark Detection for Manga Images, *CoRR*, Vol. abs/1811.03214, available from <http://arxiv.org/abs/1811.03214> (2018).
- [31] Khungurn, P. and Chou, D.: Pose Estimation of Anime/Manga Characters: A Case for Synthetic Data, *Proceedings of the 1st International Workshop on CoMics ANalysis, Processing and Understanding*, New York, NY, USA, Association for Computing Machinery, , DOI: 10.1145/3011549.3011552 (2016).
- [32] Li, C., Liu, X. and Wong, T.-T.: Deep Extraction of Manga Structural Lines, *ACM Trans. Graph.*, Vol. 36, No. 4, DOI: 10.1145/3072959.3073675 (2017).
- [33] Daiku, Y., Augereau, O., Iwata, M. and Kise, K.: Comic Story Analysis Based on Genre Classification, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 03, pp. 60–65, DOI: 10.1109/ICDAR.2017.293 (2017).
- [34] 井上須美, 安尾 萌, 松下光範: 登場人物の行為に着目したシーンの言語化に関する一検討: 少女漫画を対象として, 第 2 回コミック工学研究会発表会, pp. 15–18 (2019).
- [35] Hiroe, S. and Hotta, S.: Histogram of Exclamation Marks and Its Application for Comics Analysis, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 03, pp. 66–71, DOI: 10.1109/ICDAR.2017.294 (2017).
- [36] YAMANISHI, R., SUGIHARA, K., INOUE, R. and MATSUSHITA, M.: Extraction of Kansei Highlight from Comic Using Social Data, *Transactions of Japan Society of Kansei Engineering*, Vol. 14, No. 1, pp. 155–162, DOI: 10.5057/jjske.14.155 (2015).
- [37] Nguyen, N.-V., Rigaud, C. and Burie, J.-C.: Digital Comics Image Indexing Based on Deep Learning, *Journal of Imaging*, Vol. 4, No. 7, p. 89, DOI: 10.3390/

- jimaging4070089 (2018).
- [38] Vie, J., Yger, F., Lahfa, R., Clement, B., Cocchi, K., Chalumeau, T. and Kashima, H.: Using Posters to Recommend Anime and Mangas in a Cold-Start Scenario, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 03, pp. 21–26, DOI: 10.1109/ICDAR.2017.287 (2017).
- [39] Park, B. and Matsushita, M.: Estimating Comic Content from the Book Cover Information Using Fine-Tuned VGG Model for Comic Search, *MultiMedia Modeling* (Kompatsiaris, I., Huet, B., Mezaris, V., Gurrin, C., Cheng, W.-H. and Vrochidis, S., eds.), Cham, Springer International Publishing, pp. 650–661 (2019).
- [40] Wang, Y., Wang, W., Liang, W. and Yu, L.-F.: Comic-Guided Speech Synthesis, *ACM Trans. Graph.*, Vol. 38, No. 6, DOI: 10.1145/3355089.3356487 (2019).
- [41] Saito, J. and Nakamura, S.: Fontender: Interactive Japanese Text Design with Dynamic Font Fusion Method for Comics, *MultiMedia Modeling* (Kompatsiaris, I., Huet, B., Mezaris, V., Gurrin, C., Cheng, W.-H. and Vrochidis, S., eds.), Cham, Springer International Publishing, pp. 554–559 (2019).
- [42] Shimazaki, T., Matsushita, M., Iio, M. and Takenaka, K.: Use of health promotion manga to encourage physical activity and healthy eating in Japanese patients with metabolic syndrome: a case study, *Archives of Public Health*, Vol. 76, No. 1, DOI: 10.1186/s13690-018-0273-5 (2018).
- [43] Koutníková, M.: The Application of Comics in Science Education, *Acta Educationis Generalis*, Vol. 7, No. 3, pp. 88–98, DOI: 10.1515/atd-2017-0026 (2017).
- [44] Furusawa, C., Hiroshiba, K., Ogaki, K. and Odagiri, Y.: Comicolorization, *SIG-GRAPH Asia 2017 Technical Briefs*, ACM Press, , DOI: 10.1145/3145749.3149430 (2017).
- [45] Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H. and Fang, Z.: Towards the Automatic Anime Characters Creation with Generative Adversarial Networks, *CoRR*, Vol. abs/1708.05509, available from <http://arxiv.org/abs/1708.05509> (2017).
- [46] Cao, Y., Chan, A. B. and Lau, R. W. H.: Automatic Stylistic Manga Layout, *ACM Trans. Graph.*, Vol. 31, No. 6, DOI: 10.1145/2366145.2366160 (2012).
- [47] Wu, Z. and Aizawa, K.: MangaWall: Generating manga pages for real-time applications, *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 679–683, DOI: 10.1109/ICASSP.2014.6853682 (2014).
- [48] 迎山和司: 既存の漫画に基づいた 2 コマ漫画の自律生成, 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 1K1OS2a02–1K1OS2a02, DOI: 10.11517/pjsai.JSAI2018.0\_1K1OS2a02 (2018).
- [49] 上西隆平, 阿部和樹, 大野直紀, 土屋駿貴, 中村聡史: 主観特徴と物理特徴の融合による音楽動画印象推定手法の検討, 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM), pp. 1–7 (2017).

## 研究業績

- [1] 阿部和樹, 土屋駿貴, 大野直紀, 中村聡史, 山本岳洋: 音楽動画に対するソーシャルコメントと音響・映像特徴量を用いた印象推定手法の検討, ワークショップ 2016 (GN Workshop 2016) 論文集, Vol. 2016, pp. 1-7 (2016).
- [2] 前島紘希, 阿部和樹, 中村聡史: スマートフォン上のスクリーンキャプチャ探索支援手法, 技術報告 38 (2017).
- [3] 上西隆平, 阿部和樹, 大野直紀, 土屋駿貴, 中村聡史: 主観特徴と物理特徴の融合による音楽動画印象推定手法の検討, 技術報告 40 (2017).
- [4] 阿部和樹, 大野直紀, 土屋駿貴, 前島紘希, 中村聡史: プライミング効果を用いた音楽の印象変容に関する基礎調査, 技術報告 13 (2018).
- [5] 細谷美月, 山浦祐明, 阿部和樹, 中村聡史: ポップアウトによるユーザの選択行動変容に関する分析, 技術報告 2 (2018).
- [6] 白鳥裕士, 牧 良樹, 阿部和樹, 中村聡史: ネタバレ確信度を考慮した試合実況データセット構築と分析手法の検討, ARG 第 12 回 Web インテリジェンスとインタラクション研究会 (2018).
- [7] 古市冴佳, 阿部和樹, 中村聡史: ヒップホップダンスにおける骨格情報を用いた個性抽出の検討, 技術報告 23 (2018).
- [8] 前島紘希, 阿部和樹, 中村聡史: スクリーンキャプチャを用いた並列したゲーム世界ライフログ閲覧手法, 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) (2019).
- [9] 阿部和樹, 田村柁優紀, 中村聡史, 山中祥太.: スマートフォンのセンサ情報を用いたユーザのストレス検知手法の検討, 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) (2019).
- [10] 白鳥裕士, 牧 良樹, 阿部和樹, 中村聡史: ネタバレ判定可能なスポーツとは?, 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) (2019).
- [11] 梶田美帆, 杉本知佳, 阿部和樹, 中村聡史, 山中祥太: スマートフォンのセンサ情報による UI 評価手法の検討, 電子情報通信学会ヒューマンコミュニケーション基礎研究会 (HCS), Vol. 119, No. 38, pp. 51-56 (2019).
- [12] Furuichi, S., Abe, K. and Nakamura, S.: The Possibility of Personality Extraction Using Skeletal Information in Hip-Hop Dance by Human or Machine, *Human-Computer Interaction – INTERACT 2019* (Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M. and Zaphiris, P., eds.), Cham, Springer International Publishing,

pp. 511–519 (2019).

- [13] 梶田美帆, 阿部和樹, 中村聡史, 山中祥太: 親指の長さがスマートフォンの操作に及ぼす影響の調査と親指の長さの自動推定手法の検討, 技術報告 20 (2019).
- [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).