

コミックのあらすじ文の共通要素に
着目したコンセプト文の生成と
その活用に関する研究

総合情報学研究科
知識情報学専攻

インタラクションデザインの理論と実践

23M7138

藤川 雄翔

論文要旨

本研究は、計算機を用いたコンセプト文の自動生成を目的とする。本研究では、例えば“敵対する組織に属する男女二人が恋に落ちる話”のように、ストーリーの内容を端的に表現した文章をコンセプト文と定義する。本研究では、コンセプト文は複数の物語作品の共通のストーリーを説明した文章であると考えられることから、漫画作品同士の共通点からコンセプト文の生成を試みる。その端緒として、本稿は類似する複数の物語作品のあらすじ文から抽出した共通する単語を使用することでコンセプト文は生成可能か、生成したコンセプト文は物語作品を説明可能か、を明らかにする。

コンセプト文の生成にあたって、本研究では2つの予備実験を実施した。1つ目は、同一物語作品のストーリー内容を短く説明した文章に読者ごとに差異が現れるのかを観察することにより、同一物語作品を指すそれらの文章に統一性はあるか、どのような特徴を有しているのかについて分析をするために、コンセプト文の利用可能性について検証する予備実験を実施した。2つ目は、あらすじ文と人手により収集したコンセプト文のそれぞれに記述される特徴を比較することで、あらすじ文からコンセプト文の作成に必要な要素を調査するために、コンセプト文とあらすじ文の特徴を比較する予備実験を実施した。

これらの予備実験から得た知見を踏まえて、本研究では類似する複数の漫画作品のストーリー内容の共通点からコンセプト文の生成を試みる。そこで、漫画作品のストーリー内容の共通点の抽出のために、コンセプト文の素材データとして漫画作品のストーリー内容がテキストベースで記述されるあらすじ文を使用する。コンセプト文の生成は主に(1)形態素解析によるあらすじ文からの特定の品詞の単語の抽出、(2)文章ベクトルの算出、(3)クラスタリング、(4)コンセプト文の作成に使用する単語の選定、(5)大規模言語モデル(LLM)によるコンセプト文の作成の5つのステップを経る。(1)は各漫画作品に対する文章ベクトル算出のための前処理であり、あらすじ文から各漫画作品に登場するキャラクターの特徴を表す名詞と形容詞、キャラクターのストーリー上の主軸となる行動を表す動詞を抽出する。(2)では、類似する漫画作品ごとに分類するために、抽出した単語から各漫画作品の文章ベクトルを算出する。(3)では、各漫画作品の文章ベクトルをもとにクラスタリングを行う。(4)では、クラスターごとにそのクラスターを特徴づける単語をコンセプト文の作成に使用する単語として選定する。(5)では、クラスターを特徴づける単語を変数として、LLMを用いてコンセプト文を生成する。

この一連のステップからは、クラスタリングにより共通点を持つ漫画作品同士に分類でき、類似する複数の漫画作品間に共通する話題を抽出することが可能であることが示唆され、コンセプト文の作成に有効であることが示唆された。実験では、生成されたコンセプト文の物語作品のストーリー内容の説明妥当性の評価実験を行った。アンケート評価実験の結果から、本研究で生成されたコンセプト文はそれぞれの当該クラスターに属する漫画作品の多くを大まかに説明できていることが示唆された。しかし、コンセプト文の生成に使用する単語の数が少なくなってしまうことから、複数の物語作品のあらすじ文に共通して出現する特徴的な単語の抽出には改善の余地があることがわかった。

目次

1	序論	1
1.1	本研究の背景	1
1.2	コミックの検索に関する取り組み	2
1.3	コンセプト文を使用したコミック検索の可能性	3
1.4	コンセプト文収集の困難性	4
1.5	本研究の目的	4
2	関連研究	6
2.1	物語作品の内容情報抽出に関する研究	6
2.2	物語作品の分類・類型に関する研究	7
2.3	本研究の位置付け	7
3	予備実験	9
3.1	コンセプト文の利用可能性の検討に関する予備実験	9
3.2	コンセプト文とあらすじ文との比較に関する予備実験	11
4	デザイン指針	15
4.1	コンセプト文の再定義	15
4.2	提案手法	15
5	コンセプト文の自動生成	18
5.1	あらすじ文データの収集	18
5.2	あらすじ文の文章ベクトル算出とクラスタリング	18
5.3	各クラスターのクラスター代表語の抽出	22
5.4	LLMを用いたコンセプト文の生成	24
6	実験	27
6.1	実験の狙い	27
6.2	コンセプト文の評価実験	27
6.3	コンセプト文の評価結果	27
6.4	コミックの既知性がコンセプト文の評価に影響を与えるかについての検証	30
7	考察	39
7.1	生成されたコンセプト文に関する考察	39
7.2	コンセプト文の活用に関する今後の展望	41
8	結論	43

1 序論

本章では、本研究に至った背景を述べ、本研究で取り扱う物語作品におけるコンセプト文について概観し、本研究の目的を述べる

1.1 本研究の背景

全国出版協会出版科学研究所 [19] によると、2023 年の紙媒体および電子媒体を合わせた出版市場全体における推定販売金額は前年比から 2.1%減の 1 兆 5,963 億円であり、2022 年に引き続いて 2 年連続で前年割れとなった。しかしながら、コミック市場においては紙媒体と電子媒体を合わせた推定販売金額は前年比から 2.5%増の 6,937 億円と 6 年連続で増加し続けていることに加え、4 年連続で過去最高の値を更新し続けている。出版市場全体におけるコミックの推定販売金額が 43.5%と出版物の半数に迫っていることや、電子媒体市場全体における推定販売金額のコミックの占有率が 90.3%とコミックが大部分を占めていることから、現代社会においてコミック市場の規模は隆盛を迎えていると言える。コミック市場の好調が際立つ背景としては、ゲーム会社やテレビ局、雑誌出版社といった多くの企業がコミックの可能性に注目しているためであり、アニメ化や実写化といった映像化が行われたり、SNSなどで話題に上がることで売り上げを上げるコミックもある。

コミックは個人だけでなく多くの企業からも注目を集めている。実際に、コミックに対して各企業は毎年様々な賞を授与したり、読者からの投票によってトレンドのコミックを調査している。例えば、KADOKAWA グループが運営する「次にくるマンガ大賞¹」では、読者が「次にくる」と思うコミックに対して投票によって、人気の高いコミックを決定するマンガ賞がある。この賞には、紙媒体をメイン媒体として連載しているコミックス部門と Web をメイン媒体として連載している Web 漫画部門があり、2024 年に行われたその投票数は両部門合わせて 522,257 票と多くの読者からの投票がなされ、注目を集めている。コミックシーモアが運営する「電子コミック大賞²」では、出版社からエントリーされた 100 作品の電子コミックから読者による投票によって人気の高いコミックを決定するマンガ賞もあり、男性部門、女性部門、異世界部門、ラノベ部門、BL 部門、TL 部門という多岐にわたる観点で決定される。2024 年に行われた電子コミック大賞では投票数は 300 万票以上にもものぼった。さらに、一般社団法人アニメジャパンが運営する「アニメ化してほしいマンガランキング³」という読者が「アニメで見たい！」と思うマンガを決定する上記の賞とは異なる観定の賞であり、この賞もまた 2024 年に行われた時には総投票数が 103,473 票と多くの注目を集めている。このようにコミックに対して様々な観定で人気のあるコミックが決定され、その投票にも多く読者、読者が注目していることが伺える。また、これらの賞には電子コミックに対しても注目が集まっていることから、電子媒体でコミックを読むことは当たり前になっていると言える。

¹<https://tsugimanga.jp/> (2025/2/13 存在確認)

²https://www.cmoa.jp/comic_prize/ (2025/2/13 存在確認)

³<https://www.anime-japan.jp/activities/ajranking/> (2025/2/13 存在確認)

1.2 コミックの検索に関する取り組み

コミックは電子媒体で読むことが主流になっている昨今において、多くの電子書籍配信プラットフォームにおいて読者に閲覧されている。多種多様なコミックはWeb上の多くのプラットフォームで配信・蓄積されている。コミックに関する制作活動は日々多くの人間により行われており、これからも多くの閲覧可能なコミックが増えていく。実際、書籍扱いも含んだ紙コミックスの新刊点数は2023年で14,761点である[19]。このように多くのコミックに触れることが容易になった現在、選択可能なコミックが膨大にある中から自分の嗜好に合ったコミックを探し出すことは難しい。

現在のコミックの検索手法は、漫画を対象とした場合、書誌情報（e.g., 著者, 出版社, 出版年）やジャンル情報（e.g., アクション, ラブコメディ, ホラー）, タグ情報（e.g., 忍者, 三角関係, 復讐）など、提供するプラットフォーム側によって事前に付与されたメタ情報をキーワードとして用いている。例えば、LINEヤフー株式会社が運営するebookjapan⁴では、作品名や作者名, キーワードを検索クエリとしてコミックを探索する方法や、“少女・女性向けまんが”, “少年・青年向けまんが”といったジャンルからコミックを探索する方法がある。

電子書籍配信プラットフォーム以外にも、コミックに関連する検索を目的としたWebサイトもある。読者自身の好きなキャラクターの属性を選択していくことで理想のキャラクターを発見することを目的としたキャラ属性王国⁵やデータベースに登録されたアニメやゲームのキャラクターを属性, フリーワード, タグからの検索を可能としたNeoApo⁶などがある。類似するコミックを検索するWebサイトとして、類似漫画検索⁷やアル株式会社が運営する「近いマンガ」がわかるマンガ新検索⁸がある。類似漫画検索は、“なんだか似ていると思うコミックを検索可能とした”Webサイトであり、インターネットから収集したコミックに対する感想などの情報を用いて、独自の解析により類似するコミックの検索結果を出力している。「近いマンガ」がわかるマンガ新検索では、X（旧 Twitter）において「#私を構成する5つのマンガ」というハッシュタグが付けられた62万件以上の投稿で選ばれているコミックのデータを元に機械学習を行い、読者の好きなコミックと「雰囲気似ているマンガ」を検索できるWebサイトである。

このように、昨今では様々な検索方法を用いてコミックを探索することが可能となっている。しかし、メタ情報のみの検索や読者の感想や投稿を元にしたデータからの検索では、「主人公が挫折を乗り越えて強くなる物語」や「物語コンテンツAのような物語に恋愛要素を加えた物語」などのコミックのストーリー内容に踏み込んだ検索や推薦は難しく、ユーザーは自身の嗜好に合致するコミックに出会うまでに多大な労力と時間を要してしまう。

⁴<https://ebookjapan.yahoo.co.jp/>（2025/2/13 存在確認）

⁵<https://chara-zokusei.jp/explore>（2025/2/23 存在確認）

⁶<https://neoapo.com/characters/>（2025/2/13 存在確認）

⁷<http://ruijianime.com/comic/>（2025/2/13 存在確認）

⁸<https://alu.jp/MangaNearestMap/>（2025/2/13 存在確認）

1.3 コンセプト文を使用したコミック検索の可能性

コミックをはじめとする物語作品におけるストーリーには基本的なコンセプトが存在し、そのコンセプトにしたがってキャラクターが行動することにより、ストーリーは駆動していく。コンセプトとはストーリーの土台となるアイデアであり、物語作品を創作する際の起点としてだけでなく、他者に物語作品を推薦するための簡単な説明や、漠然としたストーリーの表現から物語作品を探索する際のクエリとして活用することができる。例えば、“敵対する組織に属する男女二人が恋に落ちる話”といったストーリーの内容を端的に表現した文章がそうであり、本研究ではこのような文章をコンセプト文と述べる。実際に、マンガ総合情報サイトであるマンバ⁹では、コミックのストーリー内容の特徴を簡単に説明した短い文章とその他の詳細を伝えることで、ユーザーが思い出せないコミックを他のユーザーに教えてもらうサービスを提供している。X (旧 Twitter) などの SNS においても、コンセプト文のような投稿者自身の好きな物語作品のストーリーの漠然な表現に該当する新たな物語作品を他者に教えてもらう趣旨の投稿がなされることがある。

コミックをはじめとする物語作品は、その物語作品解釈は一意に定まらず、読者によって異なり、多数の解釈を行うことが可能である [25]。例えば、タッチ (© あだち充, 小学館) は“双子の兄弟バッテリーが甲子園を目指す野球漫画”でもあるが、“双子の兄弟とその幼馴染との三角関係の恋愛模様を描いたストーリー”でもある。このように物語作品は多面的な解釈が可能であり、これらの解釈を把握することによりその物語作品の理解がより深まると考えられる。コンセプト文は物語作品の説明文であり、その物語作品の解釈であるとも考えることができる。したがって、端的に物語作品のストーリー内容を表現するコンセプト文は、物語作品を多面的に捉えるための足がかりとして有用であると考えられる。

コンセプト文と同様にストーリーの内容を伝達するものとして、Wikipedia などの Web 百科事典や Amazon や楽天といった EC サイトに記載されるあらすじ文が挙げられる。あらすじ文は物語作品の要約と捉えることができ、作品全体の要約や 1 巻ごとの要約、作品の“世界観”に言及した要約、作品のメタコンテンツが付与された要約 [26] といったように多様な粒度で存在しているが、その中には、売上や掲載紙の情報といったようなストーリーとは無関係のノイズが含まれていることもある。ストーリーの進行に伴ってあらすじ文が長くなることもあり、そういったあらすじ文の場合はその物語作品の内容情報を理解することは難しい。

あらすじ文はあらすじ文そのものを読むことで、どの物語作品のストーリー内容を説明しているのかは一意に特定可能である。一方で、物語創作のためのアイデアであるコンセプト文は、キャラクターや舞台設定などを変化させても変わることのない複数の物語作品に当てはまる共通のストーリーの根幹であるとも言える。そのため、コンセプト文は複数の物語作品の説明文としても使用することができる。

コンセプト文は複数の物語作品に当てはまる説明文であることから、物語作品のストーリー内容という観点において類似する物語作品を探索するための有用な手がかりになると考える。そこで本研究は、物語作品のストーリーを漠然と表現することのできるコンセプト文に着目することにより、ユーザーが探し求める物語作品へのアクセスが可能になると

⁹<https://manba.co.jp/> (2025/2/13 存在確認)

考えた。コンセプトに基づいた物語作品への情報アクセスが実現すれば、従来の物語作品の検索に用いられるジャンルや著者などといった書誌情報よりもストーリーの内容を踏まえた検索が可能になることが期待される。1.2節でも述べた通り、コミックの新刊点数が1年で1万件以上あることから、コミックの作品数はこれからも増えると予想され、ユーザーが読みたいストーリーを有するコミックに出会うことがますます困難となる。したがって、本研究では対象とする物語作品をコミックとする。

1.4 コンセプト文収集の困難性

コンセプトに基づいたコミックをはじめとする物語作品への情報アクセスを可能にするためには、個々の物語作品に対するコンセプト文が必要になる。コンセプト文は端的にその物語作品のストーリーの内容のみを伝えることを目的にしているため、重要な情報だけが記述される文章であるといえる。ただし、コンセプト文は、見どころを伝えるハイライトやストーリーの流れを示すダイジェストと異なり、コンセプト文そのものを読む行為のエンタテインメント性は高くないことから、全ての作品に対して必ずしも用意されているものではない。そのため、コンセプト文が記述されたりソースは少なく、物語作品についてのコンセプト文を収集することは容易ではない。

個々の物語作品に対するコンセプト文を収集することは容易ではないことから、コンセプト文の作成が求められる。しかし、コンセプト文を人手で作成するとなると、膨大な数の物語作品それぞれにコンセプト文を割り当てなくてはならない。加えて、物語作品はさまざまな解釈を行うことが可能である [25] ことから、コンセプト文は複数の物語作品に当てはまる観点を変えた分だけ存在可能となる。したがって、人手によるコンセプト文作成も非常に困難な作業であると考えられる。

1.5 本研究の目的

本研究は、計算機を用いたコンセプト文の自動生成を目的とする。コンセプト文の自動生成が可能となることで、コミックをはじめとする各物語作品に該当するコンセプト文を付与することができ、物語作品のストーリー内容を踏まえた情報アクセスの実現に近づくと考えられる。

本研究では、コンセプト文は複数の物語作品の共通のストーリーを説明した文章であると考えられることから、コミック同士のあらすじ文に記述される共通点からコンセプト文の生成を試みる。その端緒として、本稿は類似する複数のコミックのあらすじ文から抽出した共通する単語を使用することでコンセプト文は生成可能か、生成したコンセプト文はコミックを説明可能か、を明らかにする。具体的に以下の4点の手続きに則って確認する。

- (1) コンセプト文が複数の物語作品の説明文にもなり得ることから、ストーリーの内容情報の1つであるあらすじ文の類似度によってコミックを分類する。
- (2) 類似する複数のコミックに共通する単語を抽出し、LLMを用いることでコンセプト文の生成を試みる。

- (3) 実験では，生成されたコンセプト文がその生成の元となったコミックらを説明できるかアンケート評価により検証する．
- (4) (1)～(3)の得られた知見を用いて，コンセプト文の情報アクセス以外の活用方法について検討を行う．

2 関連研究

本章では、本研究の関連研究として、物語作品の内容情報の抽出に関する研究と物語作品の分類に関する研究について述べる。物語作品の内容情報の抽出に関しては、コンセプト文を生成するにあたって、生成するためのリソースとして何を使用すべきかを明確にする。物語作品の分類に関しては、これまでに行われた物語作品を分類・類型化することでトピックを抽出する、検索に用いる研究を紹介し、本研究との違いや提案手法の方針について明らかにするために示す。これらの関連研究の特徴について述べ、本研究の位置付けを明らかにする。

2.1 物語作品の内容情報抽出に関する研究

直接的なアプローチによるコミックをはじめとした物語作品の内容情報の抽出を試みた研究がある。コミック画像を用いてコマの抽出 [16] やコマの順序の推定 [12]、キャラクターのセリフを表す吹き出しの検出とその分類 [15]、セリフの順序の推定 [20]、キャラクターの顔抽出とその認識 [2]、キャラクターの姿勢推定 [28] などの内容情報の抽出がこれまでに行われてきている。これらの研究は、現状では内容情報の抽出までにとどまっており、内容情報を用いたストーリー情報の抽出までは行われていない。

一方で、間接的なアプローチによるコミックをはじめとした物語作品の内容情報の抽出を試みた研究もある。服部ら [6] は、新規ユーザの価値基準に沿ったコンテンツを推薦するために対象コンテンツそのものの内容を用いず、ユーザのレビューサイトにおける評価履歴のようなコンテンツの外部情報を用いた手法を提案した。山下ら [27] は TF-IDF と hLDA を利用することでレビュー文から抽出した名詞と形容詞をコミックの内容情報とし、その内容情報に基づいたユーザのコミックへの情報アクセス支援を試みた。白石ら [13] はコミック作品の概要と読者の好みとが反映されたそのコミックを表す語句の抽出の自動化を試みた。作品の概要を表す語句は Wikipedia に記載されている概要やあらすじで 3 回以上出現した名詞を“概要キーワード”と定義し、読者の評価に関わる語句はレビューサイトにおけるコミックのレビュー文から形容詞の基本形を“評判キーワード”と定義した。これらの語句の抽出には、形態素解析器は MeCab を、辞書は mecab-ipadic-NEologd を利用した。実際にコミックの特徴を抽出した結果、“概要キーワード”にはコミック作品の内容やジャンルを表す語句が抽出され、“評判キーワード”には読者が抱くコミック作品への印象や感想等が抽出された。この抽出された 2 種類のキーワードを見ることで、直感的にそのコミック作品の特徴を把握可能であることが示唆された。しかし、Wikipedia は作品によって記述される量に差が生じているため、他の概要情報からも抽出すべきであると白石らは述べた。朴ら [29] はコミックのメタ的な情報からはユーザの要求に合わせた柔軟なコミックの推薦や検索が難しいと考え、コミックの表紙画像から内容情報を推定する手法を提案した。コミックの表紙からは、作者の絵柄やコミック作品のおおまかなストーリーが読み取れることから、ストーリーの中でもコミック作品のコンセプトが表れる、“キャラクターの衣服や持ち物、背景や周囲のオブジェクトなどの描画要素”に着目した。Amazon から収集したコミックの表紙画像にタグの付与を行い、Fine-Tuning を活用することで表紙からコミック作品のコンセプトに関わる情報の抽出を試みた。涌井ら [17] は登場人物の関係を考慮したエピソード推

薦システムを構築することを目指し、テキスト形式の漫画のエピソードごとのあらすじから登場人物を抽出し、登場人物らの関係抽出と関連度の算出を行っている。これらの関連研究もコミックの内容情報の抽出にコミックのストーリーを表現するキャラクターのセリフやイラストを用いた直接的なアプローチを行わず、コミックのレビュー文や表紙画像といった間接的なアプローチをとっている。

2.2 物語作品の分類・類型に関する研究

物語作品の分類・類型に関する研究は、人文学や自然言語処理の分野において行われてきている。

まず、人文学の分野で行われた研究について紹介する。Propp[10]は、100篇の魔法昔話を分析し、それらの物語に共通して観察可能な登場人物の行為である31の機能を明らかにし、それらの機能を用いて8つの類型にロシアの魔法昔話を分類した。Snyder[11]は、映画を対象に、ストーリーの中核を成す3つの構成要素からストーリーを“家のなかのモンスター”、“金の羊毛”、“魔法のランプ”、“難題に直面した凡人”、“人生の岐路”、“相棒愛”、“なぜやったのか”、“おバカさんの勝利”、“組織のなかで”、“スーパーヒーロー”の10のストーリー・タイプに分類した。

つづいて、自然言語処理の分野で行われた研究について紹介する。Lisaら[4]は、ジョークに関して、登場人物や場所、その他の詳細が異なっても同じジョークであると認識することがある点から、ジョークを意味ベースで検索するための新しい手法を提案した。ジョークのオチを比較するモデルと可変が起りやすい特定のカテゴリの単語(e.g., 職業, 国名)にタグ付けしたモデルを提案し、ジョークの識別の改善を示した。Dongら[8]は、オランダの民話を対象とした物語作品の類似性の認識について、クラウドソーシングを用いて大規模な調査を行った。安尾ら[21]は、絵本を対象に、絵本の中に出現する単語の評価極性から物語作品のストーリー展開を把握し、バタチャリア係数を用いて物語作品の評価極性の推移の類似度を算出した。Parkら[30]は、絵本同士が持つ共通の話題を絵本本文から潜在的トピック推定を行い、物語作品のストーリー展開を把握し、バタチャリア係数を用いて物語作品のトピックの推移の類似度を算出した。Snigdhaら[1]は、人間は異なる物語作品同士に共通点を見つける能力が備わっているという考えのもと、映画のリメイクを対象に、Wikipediaから収集した映画のプロット要約を用いて物語作品の類似性を識別する手法を提案した。秋山ら[22]は、コミックの直感的な情報アクセスの実現を目指し、コミックの内容や特徴を捉えるために、BERTopicとLLMを用いて各コミックのあらすじ文からアトリビューットの生成を試みた。

2.3 本研究の位置付け

2.1節では、物語作品の内容情報の抽出に関して、直接的なアプローチを用いた研究と間接的なアプローチを用いた研究について概説した。本研究が対象とするコミックは、絵と文字が相補的かつ協調的に利用されているクロスモーダルなコンテンツであるため[14]、直接的なアプローチによるストーリー内容の取得は、いくつもの内容情報を抽出し、それらを組み合わせて処理を行う必要があり、計算量が膨大となる。また、間接的なアプローチ

に用いられるレビュー文 [13, 27] には, “面白い” や “かわいい” といった読者の感想も含まれ, 感想に使用された語彙はストーリー内容を把握するという点においては, 妨げとなってしまふ恐れがある. そのため, 本研究では, まず間接的なアプローチに用いられるストーリー内容を表現するあらすじ文を使用することにより, ストーリー内容をコンセプト文で表現可能かを検証する.

2.2節では, 物語作品の分類・類型に関する研究について概説した. いずれの研究も物語作品をストーリーの内容から分類することにより, ストーリー内容を考慮した物語作品の検索手法の実現を目指している. 単話で完結する短いテキストベースの物語作品や絵本のような物語作品は物語作品本文をそのまま使用することが可能であるが, コミックは基本複数巻に渡って話が展開するため, 本文を利用することは難しく, 本文のみを利用するとイラストの情報も損なわれることから Snigdha ら [1] や秋山ら [22] と同様にあらすじ文を使うことが好ましい.

Park ら [30] は, 絵本同士が持つ共通の話題を抽出するために, 絵本のトピックの推移の類似度を算出し, 共通の話題の観点から絵本の分類を試みた. 本研究でも, コミック同士が持つ共通点をあらすじ文から抽出するために, コミックを分類し, コミックの共通点からコンセプト文の作成を試みる.

物語作品は, キャラクタやストーリーが繰り広げられる舞台などは変化するが, ストーリー展開などの共通点が存在する [4, 1]. コンセプト文は複数のコミックの共通点から作成することから, コンセプト文には物語作品を特定できるような固有表現は用いないこととする. しかし, 本研究で使用するあらすじ文には, その物語作品に登場するキャラクタ名や地名といった固有表現が用いられる. このことから, あらすじ文から固有表現を排除した上で, コンセプト文を作成することとした.

秋山ら [22] が提案したアトリビュートと本研究で提案するコンセプトは類似した考えではある. しかし, アトリビュートは単語に近い説明であり, コンセプトは文章での説明という点で異なる. また, アトリビュートは特定の特徴が物語作品にあれば, そのアトリビュートに付与できる点に対して, コンセプト文はたとえ特定の特徴があっても類似した物語作品でなければコンセプト文が作成できないといった点がある. 例えば, “高校生の男女たちが覚醒した能力により世界を救う物語” と “高校生の男女の甘酸っぱい恋の物語” といったような説明がある場合, “高校生の男女” といったようなアトリビュートは付与することができるが, コンセプト文はこれらの物語作品を別物として特徴づけるという点にある.

以上のことから, 本研究はコミックをあらすじ文から分類し, あらすじ文からコミック同士の共通点を抽出することでコンセプト文の生成が可能か, 生成されたコンセプト文の妥当性を検証する. コンセプト文の生成にあたって, 本研究では2つの予備実験を実施した. 1つ目はコンセプト文の利用可能性について検証する予備実験である. これは, 同一物語作品のストーリー内容を短く説明した文章に読者ごとに差異が現れるのかを観察することにより, 同一物語作品を指すそれらの文章に統一性はあるか, どのような特徴を有しているのかについて分析するために実施した. 2つ目はコンセプト文とあらすじ文の特徴を比較する予備実験である. これは, あらすじ文と人手により収集したコンセプト文のそれぞれに記述される特徴を比較することで, あらすじ文からコンセプト文の作成に必要な要素を調査し, コンセプト文を作成可能とするために実施した.

3 予備実験

本章は、コンセプト文の利用可能性を検証する予備実験と、コンセプト文とあらすじ文のそれぞれの特徴を比較する予備実験について述べる。それぞれの予備実験では、それぞれ同一物語作品を説明した文章に統一性の有無とその特徴について、あらすじ文とコンセプト文の作成に必要な要素についての確認をする。本章では便宜上、コンセプト文の利用可能性の検討に関する予備実験 [23] を予備実験 (1)、コンセプト文とあらすじ文との比較に関する予備実験 [24] を予備実験 (2) と呼称する。

3.1 コンセプト文の利用可能性の検討に関する予備実験

予備実験 (1) は、人手によりコンセプト文のデータセットを構築し、構築したデータセットを基にコンセプト文の特徴を分析することにより、分析結果として得られる特徴を用いたコンセプト文の利用可能性を検討することを目的として実施した。

3.1.1 コンセプト文の収集

予備実験 (1) ではコンセプト文を収集するために、関西大学総合情報学部にも所属する学部生を対象としたアンケートを行った。自身の好きな物語作品を他者に推薦するという状況設定のもと、物語作品の内容が記述された 287 件の短文を収集した。短文の記述に際し、(1) 140 字以内で要約すること、(2) 文の末尾を「～物語」で締めくくることが、(3) その物語作品の固有名詞 (e.g., キャラクタ名, 組織名) を使わないこと、の 3 つの条件を課した。アンケート収集の際、短文に加えてその物語作品のタイトルとジャンルについても回答を課した。収集された 287 件の短文のうち、同一の物語作品について 3 件以上得られた物語作品は 19 作品あり、1 作品あたり最大 11 件の短文が得られた。予備実験 (1) では、これらのうち物語作品に固有の名詞が含まれていたものを除いた 97 件を対象として特徴分析を行った。

3.1.2 短文のベクトル化による分析

3.1.1 節の方法で収集した短文は、その物語作品に対する捉え方が回答者によって異なる場合がある。そこで予備実験 (1) では、収集した短文各々の文章ベクトルを算出し、同一物語作品の文章ベクトルの類似度から、物語作品が要約された短文が物語作品としてのコンセプトとして利用可能であるか検討する。

文章ベクトル算出のために、収集したそれぞれの短文に対して形態素解析などの前処理を行った。形態素解析器には MeCab (ver. 0.996)¹を採用し、形態素解析用辞書は ipadic-NEologd²を用いた。Slothlib³に含まれる単語と、1 文字のひらがなとカタカナをストップワードとして設定した。また、どのコンセプト文も文末は「～物語」で締めくくることが、 “物語” の単語も除外することとした。形態素解析後、名詞と動詞 (それぞれ非自立を除く)、形容詞、副詞のみを抽出した。

¹<https://taku910.github.io/mecab/> (2023/10/28 確認)

²<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (2023/10/28 確認)

³<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (2023/10/28 確認)

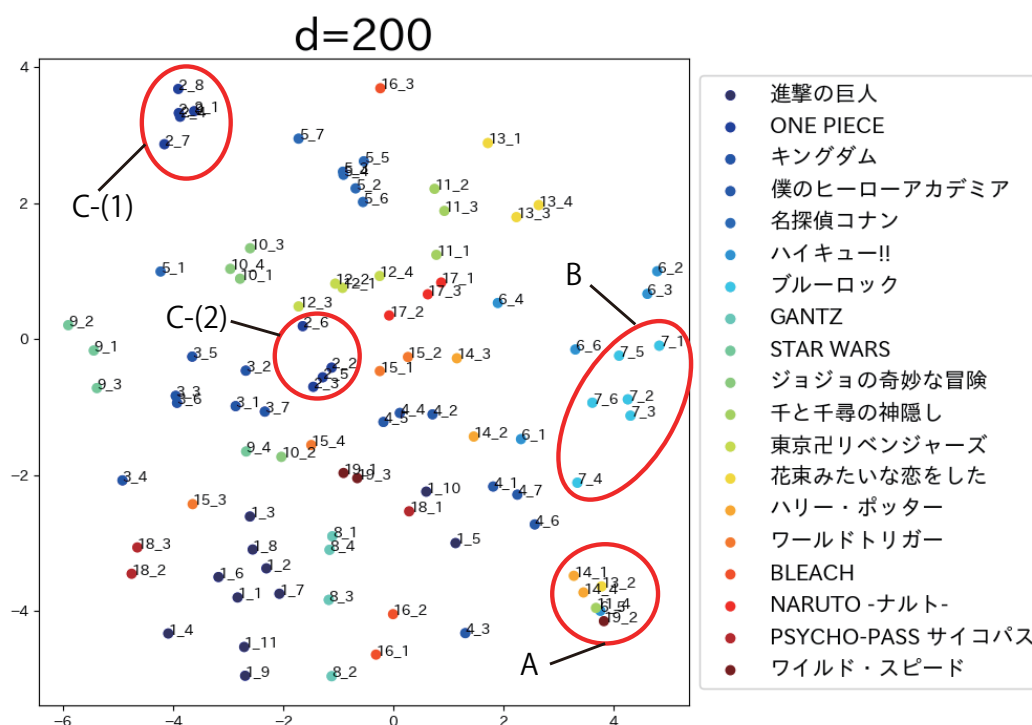


図 3.1: 3.1.2 節で得られた 200 次元の文章ベクトルを 2 次元に圧縮し t-SNE を用いてプロットした散布図. それぞれのラベルは作品番号と文章 id を示している.

予備実験 (1) では, 短文の文章ベクトルの算出に word2vec[7] を用いた. 鈴木ら [18] が日本語版 Wikipedia の本文全文から作成した Wikipedia エンティティベクトルモデル (次元数は 200 次元)⁴を用いて, 短文に含まれる単語の分散表現ベクトルを得た. このとき, “余儀ない” “謎解き” “ずば抜ける” の 3 単語は分散表現ベクトルを得ることができなかつたため, 除外することとした. 予備実験 (1) では, 1 つの短文に含まれる全単語の分散表現ベクトルの平均を文章ベクトルとして算出した.

3.1.3 考察

図 3.1 は, 3.1.2 節で得られた 200 次元の文章ベクトルを Scikit-learn (ver. 1.3.2) の TSNE を用いて 2 次元に圧縮し, 散布図にプロットしたものである. 図 3.1 より, 同一物語作品の短文群にまとまりがあることが確認された. これらのコンセプト群の文章を確認したところ, 同一もしくは類似した単語が複数含まれていた. ただし, 同一物語作品の群から外れる短文もいくつか確認された. これらの内容を確認したところ, 短文の長さが他の短文に比べて短い, 他の短文には含まれていない単語が含まれている, という傾向が確認された. また, 異なる物語作品の短文であるにも関わらず, 散布図上で群を形成しているように見えるものが観察された (図 3.1 中 A). これは同一物語作品群から外れていた物語作品同士の短文が, 200 次元の文章ベクトルを 2 次元に圧縮したために近くにプロットされた可能性がある.

⁴<https://github.com/singletonue/WikiEntVec/releases> の jawiki.all_vectors.200d.txt.bz2 (2019 年 5 月 20 日版) モデルを使用 (2023/10/28 確認)

表 3.1: アンケートによって収集されたコンセプト文（一部）

id	タイトル	コンセプト文
2.1	ONE PIECE	夢見る少年が仲間を集めながら海の王を目指す物語
2.2	ONE PIECE	主人公が航海に出て信頼できる仲間を集め、数々の敵と戦いながら、夢を追い求める姿や仲間との友情、強くなるための努力を描いた物語。
2.3	ONE PIECE	「海賊王になる」という夢を持ち、目指す主人公がその夢の実現のために航海し、その過程でそれぞれ「世界一の剣豪」や「世界地図を描く」といった夢や野望を持った多種多様な仲間たちを集めながら航海していき、またその中で国や人を救うために悪者と戦いながら心身ともに成長していく物語
2.4	ONE PIECE	子供の頃に憧れた海賊を追いかけて仲間を集めながら冒険し、海賊王を目指す物語。
2.5	ONE PIECE	とある果実を口にしたことで不思議な能力を得た主人公が仲間とともに海賊王を目指す物語
2.6	ONE PIECE	自分の夢を叶えるため主人公が海に出て旅をしていく中で仲間を集め強敵や困難な出来事また悲しい出来事や自分の中での葛藤を乗り越え海賊王を目指す物語。
2.7	ONE PIECE	麦わら帽子がトレードマークの海賊である主人公が多くの海賊が蔓延る時代に、伝説の大秘宝を目指して仲間と共に航海に出て、冒険を繰り広げる物語。
2.8	ONE PIECE	一人の青年が仲間と共に海賊王を目指す物語
2.9	ONE PIECE	麦わら帽子を被った少年が仲間を集めて海賊の王を目指す物語
7.1	ブルーロック	「日本一のストライカー」を目指し、強豪選手たちとの過酷なトライアウトに挑む物語
7.2	ブルーロック	主人公のエゴと才能を武器に、周りと競い合って世界一の点取り屋（ストライカー）を目指す物語。
7.3	ブルーロック	集められた若きサッカー選手たちが、極限の環境の中自分のエゴを磨き、最強のストライカーを目指す物語。
7.4	ブルーロック	日本がW杯優勝するために日本サッカーで欠けている絶対のエースストライカーを生み出す物語。
7.5	ブルーロック	無名の高校のサッカー部の主人公が全国から集められた高校生のフォワード達と勝負や切磋琢磨し、世界一のストライカーのストライカーを目指す物語
7.6	ブルーロック	日本中から集められた高校生のストライカーの中から、世界一のストライカーを誕生させる物語

表 3.1 に、アンケートによって収集された短文の一部を示す。ここでは特に、作品単位で群としてのまとまりが見られた“ブルーロック”（図 3.1 中 B）と作品内で 2 つの群に分かれて分布した“ONE PIECE”（図 3.1 中 C）の 2 作品について考察する。“ブルーロック”では、どの短文にも“ストライカー”という単語が含まれることや、ストライカーを“目指す”、“誕生させる”などのその物語作品のキャラクターたちの目的が共通して記述されている。このことは、本作品を象徴する短文では、同様の単語が用いられて類似した短文によって、作品のある特定の 1 側面について記述されることが示唆される。“ONE PIECE”も同様に、“仲間”という単語や“海賊王を目指す”等のキャラクターの目的が共通して記述されている。しかし、キャラクターの目的を達成するための行動として、“冒険をする”（図 3.1 中 C-(1)）と“敵と戦う”（図 3.1 中 C-(2)）という 2 種類の異なる見解が記述されたことで、短文群が二分したと考えられる。本作品は複数の側面から作品が評価され、本作品を象徴する短文が記述されることを示唆する。

3.2 コンセプト文とあらすじ文との比較に関する予備実験

予備実験 (2) は、コンセプト文とあらすじ文、各々の文章ベクトルを作成し、物語作品間の類似度からあらすじ文とコンセプト文の差異を観察することで、あらすじ文からコンセプト文を生成できるかについて検討することを目的として実施した。

3.2.1 あらすじ文の収集

コンセプト文との比較に用いるあらすじ文は、漫画全巻ドットコム⁵内の「作品概要」とし、歴代発行部数ランキングの上位 150 作品を対象として、人手で収集した。作品概要の

⁵<https://www.mangazengan.com/>（2024/1/10 存在確認）

表 3.2: コンセプト文同士のコサイン類似度上位 5 作品 (一部). 各 ID に対応するコンセプト文は表 3.4 を参照.

ID	SD1	SD3	SD5	SD6	SD7
1	SD8	SO1	SD8	HH5	HH5
	0.8184	0.6436	0.8087	0.6970	0.6641
2	SD4	KO1	SD4	HK2	KT1
	0.8044	0.5829	0.7906	0.6257	0.6529
3	SD9	YK2	SD9	SD7	SD6
	0.7715	0.5598	0.6973	0.6137	0.6137
4	SD2	GT1	SD2	SD9	HI1
	0.7404	0.5509	0.6958	0.6002	0.6070
5	SD5	KJ1	SD1	YH1	BH1
	0.6679	0.5464	0.6679	0.5930	0.5947

一部に, 商品説明や同名作品の他シリーズの作品説明が記述されているものはコミックのストーリー内容に直接的な関係がないため, 該当箇所を除いた. また, 予備実験 (2) ではコミックを対象としていることと, 物語作品のストーリーを要約したコンセプト文との比較を行うことから, 作品概要の大部分に商品説明やメタ情報が記述されているものやライトノベルの作品を除いた. この処理の結果, 条件に当てはまった 135 作品のコミックのあらすじ文を対象とした.

3.2.2 コンセプト文の収集

予備実験 (2) ではコンセプト文を収集するために, Yahoo!クラウドソーシング⁶ (以下, クラウドソーシング) を用いて, 自身の好きなコミックを他者に推薦するという状況設定のもと, 195 名の回答者から 3 作品ずつ, 計 585 文のコミックを要約した短文を収集した. 短文の記述に際し, (1) 100 字以内で要約すること, (2) 文の末尾を「～物語」で締めくくることが, (3) そのコミックの固有名詞 (e.g., キャラクター名, 組織名) を使わないこと, の 3 つの条件を課した. アンケート収集の際, コミックを要約した短文に加えてそのコミックのタイトルとジャンルについても回答を課した.

収集された 585 文の短文のうち, 上記の条件を満たし, かつ収集したあらすじ 150 作品と同一作品の短文である 126 文を予備実験 (2) の分析対象とした. このとき収集した短文を手で確認し, 明らかな誤字は修正し, 読み仮名や, 短文の先頭に記述されたそのコミックのタイトル, 末尾の「です。」「。」は削除を行った.

3.2.3 収集した短文の分析

前節の方法で収集した短文は, その物語作品に対する捉え方が回答者によって異なる場合がある. そこで予備実験 (2) でも, 収集した短文各々の文章ベクトルを算出し, 先行研究 [23] と同様に同一物語作品の文章ベクトルの類似度から, 物語作品が要約された短文が

⁶<https://crowdsourcing.yahoo.co.jp/> (2024/1/30 存在確認)

表 3.3: 同一コミックにおけるあらすじ文とコンセプト文の類似度（降順一部）. 各 ID に対応するコンセプト文は表 3.4 を参照.

タイトル	ID	類似度
ジョジョの奇妙な冒険	JJ2	0.9036
はじめの一步	HI1	0.8214
...
ONE PIECE	OP6	0.1430
ビー・バップ・ハイスクール	BH1	0.0133

物語作品としてのコンセプトとして利用可能であるか検討する.

予備実験 (2) では, sentenceBERT[9] を用いて収集した短文のベクトル化を行い, それぞれの短文同士のコサイン類似度を算出した. モデルには, 日本語用の sentenceBERT モデルである sonoisa/sentence-bert-base-ja-mean-tokens⁷ を採用した. 表 3.2, 3.4 は “SLAM DUNK” を代表とした結果の一部とそれぞれの ID に対応する短文である. 以下, 個々のコンセプト文に対しての結果, 考察を述べる際にも, 表 3.4 の ID を用いて説明を行う.

SD1, SD5 はコサイン類似度の上位 5 作品すべてが同一のコミックの短文となった. これは短文内にそのコミックの主題である “バスケット” や “バスケットボール” が含まれていることと, 主人公を中心としたキャラクターがバスケットボールを軸にどのようなようになるかが記述されていることによると考えられる.

一方で, SD3, SD6, SD7 は上述の 2 文とは異なり, コサイン類似度が最も高い値でも, 0.70 を下回る結果となった. SD3 はコサイン類似度の上位 5 作品には同一のコミックの短文は含まれなかった. SD3 では, そのコミックの内容を示すタグ情報となるような単語を中心に記述しており, “根性” や “学園”, “友情” といった単語と類似する文脈で用いられる単語が含まれる短文が上位に入る傾向にあった. SD6 では, コミック内で主人公が成長することを軸に記述されていることから, 同様に主人公が成長することを軸に記述された短文の類似度が高くなったと考えられる. SD7 では, スポーツを通じてキャラクターが成長することを軸に記述されていることから, 同様にスポーツが絡んだキャラクターの成長を軸に記述された短文の類似度が高くなったと考えられる.

3.2.4 コンセプト文とあらすじ文の類似度測定

あらすじ文も前章のコンセプト文と同様に, sentenceBERT を用いてベクトル化を行い, コンセプト文とあらすじ文のコサイン類似度を算出した. その結果の一部を表 3.3 に示す.

同一コミックにおいて最も類似度が高い上位 2 作品のあらすじ文とコンセプト文は “ジョジョの奇妙な冒険 (JJ2)” と “はじめの一步 (HI1)” であった. どちらのコンセプト文も, あらすじ文と類似したキャラクターの特徴や, そのキャラクターがストーリーにおいてどのような行動をとるのか, またどのような変化が生まれるのかが記述されていたことから, あらすじ文との類似度が高くなったと考えられる. また, JJ2 のコンセプト文と類似度の高いあらすじ文のコミックを確認したところ, “DEATH NOTE”, “北斗の拳”, “るろうに剣

⁷<https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens> (2024/1/30 存在確認)

表 3.4: クラウドソーシングによって収集されたコンセプト文（一部）

ID	タイトル	コンセプト文
BH1	ビー・バップ・ハイスクール	仲間同士で学校を守ったり仲間を助ける物語
GT1	銀魂	主人公が日常を面白く過ごしたり、ある時は真剣に悪人を倒したり、ギャグがある物語
HH5	HUNTER × HUNTER	少年が仲間とともに成長していく冒険物語
HI1	はじめの一步	いじめられっ子がある日ボクサーに助けられ、ボクシングに魅了され成長していくスポーツ物語
HK2	ハイキュー!!	低身長主人公がバレーの世界で成長していく物語
JJ2	ジョジョの奇妙な冒険	吸血鬼や未知の能力を持った者が死闘を繰り広げる物語
KJ1	寄生獣	恐ろしい生物が出てくるもユーモラスな表現とキャラクターで魅了される社会派な物語
KO1	今日から俺は！	ツッパリ少年たちの笑いあり涙ありの青春物語
KT1	キャプテン翼	少年時代から仲間やライバルと共にサッカーを通じて成長していく物語
OP6	ONE PIECE	仲間を大切にす物語
SD1	SLAM DUNK	バスケット初心者赤髪男がチームメイトと協力して成長していく物語
SD2	SLAM DUNK	バスケットボール少年たちのバイブルである主人公の成長物語
SD3	SLAM DUNK	単純なスポコンでなく学園物でもあり、かといってなまなま友情物でもない、熱い男の物語
SD4	SLAM DUNK	不良の高校生がバスケットに出会い、色々な仲間と協力し合いながら全国大会出場を目指す物語
SD5	SLAM DUNK	高校のバスケットボール部で、色々なキャラクターがバスケットボールに打ち込む物語
SD6	SLAM DUNK	初心者の落ちこぼれが仲間とともに成長して大活躍する物語
SD7	SLAM DUNK	個性豊かな若者達が一つのスポーツを通じて先走ったり、その場に留まったりしながらも少しずつ成長する過程に親近感がわく物語
SD8	SLAM DUNK	不良少年がバスケットボールで仲間達と活躍する物語
SD9	SLAM DUNK	バスケットボールが未経験だった主人公が、努力して成長していく物語
SO1	魁!!男塾	男しか登場しない、根性や友情、死闘を繰り広げる物語
YH1	幽☆遊☆白書	子供を助けるために事故死をしてしまった主人公が生き返る為に事件を解決しながら仲間とともに成長する物語
YK2	有閑倶楽部	お金持ちなどの男女の学生の冒険や事件に巻き込まれたりしたりしたエピソードを面白おかしく描写した物語

心—明治剣客浪漫譚—”，“バカボン”が上位5作品に含まれた。これらの作品は全てライバルたちと死闘を繰り広げていくコミックであることがあらすじ文から読み取れることから、キャラクターのストーリーにおける主軸となる行動について記述されたコンセプト文は、物語のコンセプトに基づく情報アクセスに有用であると考えられる。

一方で、あらすじ文とコンセプト文の最も類似度が低かった“ビー・バップ・ハイスクール（BH1）”と“ONE PIECE（OP6）”は表3.4より他のコンセプト文と比較すると、キャラクターの特徴に関する記述がないことがわかる。さらに、ストーリーにおいてキャラクターの主軸となる行動についての記述も、多くの作品に当てはまるような抽象的な記述であるため、あらすじ文との類似度が低くなったと考えられる。このことから、物語のコンセプトに基づく情報アクセスに有用であるコンセプト文の記述にはキャラクターの特徴に関する記述も含めることが望ましいといえる。

以上のことから、キャラクターに関する情報やキャラクターのストーリー上での主軸となる行動についての記述があるコンセプト文が、物語のコンセプトに基づく情報アクセスに有用であると示唆された。あらすじ文にはこれらの記述が含まれていることから、これらの情報を抽出しつなぎ合わせることで、あらすじ文からコンセプト文が生成可能であると考えられる。

4 デザイン指針

本章では、3章で得られた結果を用いて、本研究で生成するコンセプト文について改めて定義し、コンセプト文生成の提案手法について述べる。

4.1 コンセプト文の再定義

本節では、3章で得られた知見を踏まえて、コンセプト文生成に必要な条件を再定義する。

1.4節で、どんな物語作品にもコンセプトが存在し、コンセプトは物語作品を創作する上でのアイデアや物語作品のストーリーを簡単に説明として活用できると述べた。このような物語作品のストーリー内容を端的に表現した文章を、本研究ではコンセプト文と定義している。

コンセプト文は、物語作品のストーリー内容を端的に表現した文章であるが、それは1つの物語作品だけを説明するのではなく、複数の物語作品に対しても説明可能な文章である。これは2.3節で述べた通り、コンセプト文に特定の物語作品の固有表現を含まないことに起因する。また、3.1.2項の予備実験(1)の結果の1つであるC群より、コンセプト文は複数の側面から物語作品を評価できる。そのため、1つの物語作品に対してコンセプト文は1つとは限らない。加えて、3.2節の予備実験(2)の結果より、コンセプト文は“キャラクターに関する情報”と“キャラクターのストーリー上での主軸となる行動”が有用であることが示唆された。そのため、生成するコンセプト文にもこれらの情報を入れることとする。最後に、コンセプト文は複数の物語作品を説明可能な文章であることから、コンセプト文に該当する物語作品は類似する共通点を持つ可能性が高い。したがって、異なるコンセプト文同士は互いに異なる物語作品の特徴を持つ。

以下に、本研究で生成するコンセプト文についての条件をまとめる。

- 類似する複数の物語作品内で共通するストーリーの内容を説明する
- 特定の物語作品の固有表現を含まない
- 1つの物語作品にはコンセプト文は1つ以上存在する
- “キャラクターに関する情報”と“キャラクターのストーリー上での主軸となる行動”からなる
- コンセプト文は他のコンセプト文とは異なる物語作品の特徴を持つ

4.2 提案手法

本節では、本稿で提案する提案手法(図4.1参照)について述べる。提案手法は、(1)形態素解析によるあらすじ文からの特定の品詞の単語の抽出、(2)文章ベクトルの算出、(3)クラスタリング、(4)コンセプト文の作成に使用する単語の選定、(5)大規模言語モデルによるコンセプト文の作成の5つの機能で構成される。

本研究で使用するデータはコミックのあらすじ文とする。コンセプト文は物語作品の内容を説明する文章である。物語作品の内容を説明するリソースとしては、あらすじ文が挙

げられる。あらすじ文は単行本の裏表紙や電子書籍販売サイトの各コミックのページに採用として書かれるか、自由参加型 Web 百科事典に書かれることが多い。自由参加型 Web 百科事典では決まった書き方がないことから、要約の範囲が全体や「～編」といったストーリーとして切りのいい区切りごとになっており、物語作品によってストーリー内容の情報量に差が生じる。このことから、本研究では使用するデータとして情報量の差が少ない電子書籍販売サイトの各コミックのページに書かれるあらすじ文を使用する。

本研究ではコンセプト文を生成するために必要な、類似する複数のコミックの共通単語をあらすじ文から抽出する。予備実験 (2) の結果より、コンセプト文の作成には、“キャラクターに関する情報”と“キャラクターのストーリー上での主軸となる行動”が必要な情報であることが示唆された。そこで本研究ではコンセプト文を生成するために、あらすじ文から名詞と形容詞を抽出することで“キャラクターに関する情報”を取得し、動詞を抽出することで“キャラクターのとり行動”を取得することを試みる。その品詞特定のために、形態素解析器に MeCab、固有表現に強い形態素解析用辞書である ipadic-NEologd 辞書を使用する。

コンセプト文は類似する複数の物語作品の共通するストーリー内容の説明である。そこで、コンセプト文を自動生成するためには複数の類似する物語作品群が必要となる。本研究では、コミックをクラスタリングすることで複数の類似するコミック群を得る。クラスタリングを行うためには、それぞれのあらすじ文の文章ベクトルを算出する必要がある。そのため、本研究では 3.2 節で述べた予備実験 (2) と同じ複数の文章におけるクラスタリングに有用な sentenceBERT を用いる。また、物語作品 1 作品につきコンセプト文は 1 つとは限らず、物語作品の類似点の観点が異なることに伴って類似する物語作品の数が増減し、コンセプト文の具体化、抽象化が起こると考えられる。こういった観点から自由にクラスタ数を決定することが可能な k-means 法を採用する。

コンセプト文は類似する複数の物語作品の共通の要素を含む文章であり、該当しない作品群とは異なる特徴を持つと考えられる。そのため、コンセプト文作成に使用するための単語（以下、クラスター代表語と記す）は、そのクラスターに頻出し、かつ他のクラスターには出現しづらい単語であることが望ましい。そこで、各クラスターにおけるクラスター代表語の候補単語（以下、候補単語と記す）の決定は、sentenceBERT の埋め込み表現を利用したトピックモデリング手法である BERTopic[3, 5] で採用されている c-TF-IDF (class-based TF-IDF) を参考に行う。通常の TF-IDF は、文書内での任意の単語の出現頻度と、文書集合の中で任意の単語が文書に出現した割合を逆数にし対数をとった値を掛け合わせることで、単語の重みづけを行う。一方、c-TF-IDF は各クラスターの全ての文書を結合させ 1 つの文書としてみなすことで、クラスターでの任意の単語の出現頻度と、クラスター集合の中で任意の単語がクラスターに出現した割合を逆数にし対数をとった値を掛け合わせることで、単語の重みづけを行う。本稿は、類似する物語作品群の特徴を表す代表語をコンセプト文の作成に使用することから、c-TF-IDF を採用する。

しかし、クラスターに属する 1 作品の単語が TF-IDF 値に影響を及ぼす場合を考慮しなければならない。同一クラスターに属する他の物語作品には出現しない単語が 1 作品内のあらすじ文において高頻度で出現した場合、その単語の TF-IDF 値が高くなり、コンセプト文として使用される単語として選出される可能性がある。このことから、コミックのあらすじ文を 1 作品だけ削除し、同様の方法で TF-IDF 値を求めた時に、同一クラスターに

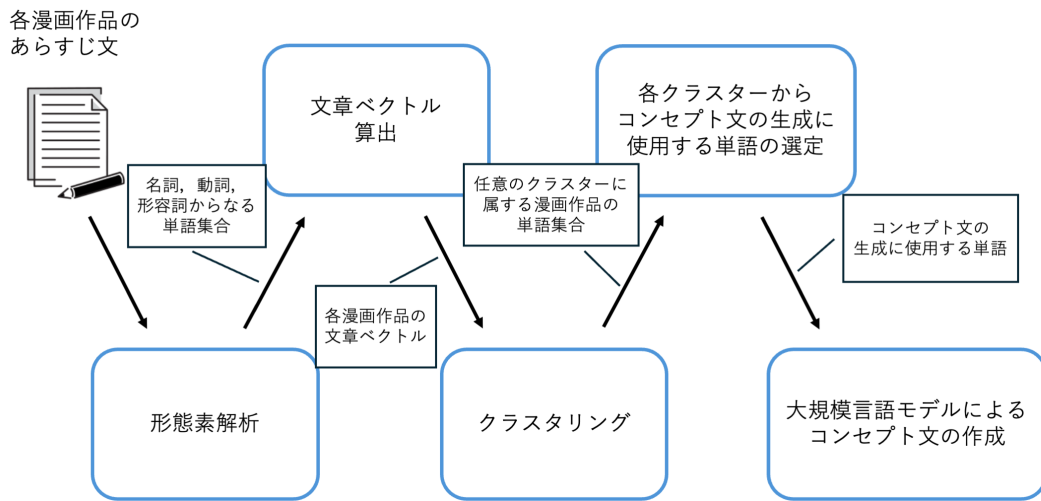


図 4.1: 本稿で提案する提案手法の流れを示す。角丸の四角で囲まれているものが提案手法のそれぞれの機能である。それぞれの矢印に付与されている吹き出しは、それぞれの機能に対する入力データと出力データを表す。

属するどのコミックのあらすじ文を削除しても共通して含まれる TF-IDF 値の高い単語をクラスター代表語として決定する。

4.1 で述べた通り、コンセプト文は条件の多い文章である。このことから、クラスター代表語からコンセプト文を生成する際に、条件を考慮しながら文章を生成する必要がある。本稿では、文章生成タスクに有用な大規模言語モデルである ChatGPT を用いる。

5 コンセプト文の自動生成

本章では、4.2節で述べた提案手法に基づいてコンセプト文の生成を行う。

本研究は、類似する複数のコミックのストーリー内容の共通点からコンセプト文の生成を試みる。そこで、コミックのストーリー内容の共通点の抽出のために、コンセプト文の素材データとしてコミックのストーリー内容がテキストベースで記述されるあらすじ文を使用する。コンセプト文の生成は主に (1) 形態素解析によるあらすじ文からの特定の品詞の単語の抽出, (2) 文章ベクトルの算出, (3) クラスタリング, (4) コンセプト文の作成に使用する単語の選定, (5) 大規模言語モデル (LLM) によるコンセプト文の作成の5つのステップを経る。(1) は各コミックに対する文章ベクトル算出のための前処理であり、あらすじ文から各コミックに登場するキャラクターの特徴を表す名詞と形容詞, キャラクターのストーリー上の主軸となる行動を表す動詞を抽出する。(2) では、類似するコミックごとに分類するために、抽出した単語から各コミックの文章ベクトルを算出する。(3) では、各コミックの文章ベクトルをもとにクラスタリングを行う。(4) では、クラスターごとにそのクラスターを特徴づける単語をコンセプト文の作成に使用する単語として選定する。(5) では、クラスターを特徴づける単語を変数として、LLM を用いてコンセプト文を生成する。

これより、それぞれのステップについて詳しく説明していく。

5.1 あらすじ文データの収集

本研究で用いるあらすじ文データは、漫画全巻ドットコム内の歴代発行部数ランキングに掲載されている全309作品(2024/12/12時点)それぞれの作品ページにある「作品概要」とし、人手で収集した。3.2節で述べた予備実験においてのあらすじ文の収集と同様に、作品概要の一部に商品説明や同名作品の他シリーズの作品説明が記述されている箇所は除いた。加えて、あらすじ文から物語作品のストーリーを端的に要約した文章であるコンセプト文を生成することから、ストーリーの内容とは関係のないWebサイト上で取り扱っている商品の説明やメタ情報が作品概要の大半を占めているものや、ライトノベル作品を除いた。この処理の結果、条件に当てはまった269作品のコミックのあらすじ文を本研究の対象とした。

5.2 あらすじ文の文章ベクトル算出とクラスタリング

コンセプト文は類似する複数の物語作品を説明可能な文章でもある。こういった観点からコンセプト文を自動生成するためには、複数の類似する物語作品にクラスタリングする必要がある。そのために、269作品のコミックのあらすじ文に対して前処理を施した後、各コミックのあらすじ文の文章ベクトルをそれぞれ算出する。

あらすじ文の文章ベクトル算出のために、あらすじ文に対して形態素解析などの前処理を行った。形態素解析器にはMeCab(ver. 0.996)を採用し、形態素解析用辞書にはipadic-NEologdを用いた。ストップワードは日本語ストップワード辞書であるSlothlibに含まれる単語と、1文字のひらがなとカタカナとした。加えて、コンセプト文は物語作品を特定可能な固有名詞を含めないことから、各コミックを特徴づける人物名や地名などの固有名詞

表 5.1: 漫画全巻ドットコムから収集した 269 作品のあらすじを 20 クラスターにクラスタリングした結果. 各クラスターに含まれる作品数と単語数, 異なり単語数, またそれぞれ分散と標準偏差を記す.

クラスター ID	作品数	単語数	異なり単語数
A1	8	143	125
A2	20	473	374
A3	24	525	432
A4	11	305	265
A5	23	654	510
A6	14	302	260
A7	17	299	256
A8	13	313	251
A9	12	339	273
A10	17	387	305
A11	8	171	148
A12	16	354	277
A13	12	344	302
A14	7	205	183
A15	7	169	141
A16	12	191	161
A17	16	445	324
A18	18	445	378
A19	7	131	113
A20	7	142	125
分散	27.15	19,463.23	11,398.13
標準偏差	5.21	139.51	106.76

や意味を成さない単語もストップワードとして設定し, あらすじ文から削除した. しかし, 本稿で除外した地名の固有名詞は, 都会や田舎, 国内, 国外などといったストーリーの主となるおおまかな舞台の判断材料となり, ストーリーの内容にも関わる可能性があるため, コンセプト文への地名の活用方法については今後の検討課題である. 形態素解析をした後, 名詞と動詞 (それぞれ非自立を除く), 形容詞を抽出した. その結果, 対象とする全 269 作品のあらすじ文から得られた単語の総数は 6,339 語, 総異なり数は 3,311 語となった. 269 作品中得られた単語数が最も多かったコミックの単語数は 60 語であり, 反対に得られた単語数が最も少なかったコミックの単語数は 5 語であった. 得られた単語数の平均は 23.57 語であった.

前処理を施した各コミックのあらすじ文の文章ベクトルの算出には, 3.2 節の予備実験 (2) と同様に sentenceBERT を用いた. モデルは, 日本語用の sentenceBERT モデルである

表 5.2: 漫画全巻ドットコムから収集した 269 作品のあらすじを 25 クラスターにクラスタリングした結果. 各クラスターに含まれる作品数と単語数, 異なり単語数, またそれぞれ分散と標準偏差を記す.

クラスター ID	作品数	単語数	異なり単語数
B1	6	125	109
B2	25	526	428
B3	21	573	466
B4	16	451	328
B5	12	295	257
B6	11	256	210
B7	20	583	462
B8	11	210	186
B9	11	260	222
B10	7	192	166
B11	10	247	216
B12	6	133	121
B13	17	362	311
B14	13	296	232
B15	3	53	51
B16	20	442	363
B17	8	195	164
B18	11	238	200
B19	9	189	148
B20	3	101	96
B21	5	70	66
B22	7	123	106
B23	3	64	58
B24	6	126	105
B25	8	227	182
分散	35.22	23,455.77	14,336.27
標準偏差	5.93	153.15	119.73

表 5.3: 漫画全巻ドットコムから収集した 269 作品のあらすじを 30 クラスターにクラスタリングした結果. 各クラスターに含まれる作品数と単語数, 異なり単語数, またそれぞれ分散と標準偏差を記す.

クラスター ID	作品数	単語数	異なり単語数
C1	3	70	64
C2	7	146	128
C3	8	153	137
C4	22	425	357
C5	16	404	333
C6	3	94	75
C7	5	116	95
C8	8	181	158
C9	12	302	250
C10	8	172	145
C11	9	199	166
C12	16	496	391
C13	3	36	31
C14	9	177	157
C15	14	410	324
C16	14	318	262
C17	8	157	147
C18	17	360	303
C19	10	164	144
C20	7	158	129
C21	6	126	111
C22	6	172	145
C23	4	102	88
C24	5	103	87
C25	13	382	316
C26	7	163	142
C27	4	105	94
C28	5	129	121
C29	5	106	88
C30	15	413	340
分散	23.30	15,729.61	10,014.51
標準偏差	4.83	125.42	100.07

sonoisa/sentence-bert-base-ja-mean-tokens-v2¹を採用した。その結果、768次元の文章ベクトルを得た。

得られたあらすじ文の文章ベクトルをもとにk-means法により、各コミックをクラスタリングした。4.2節で述べた通り、物語作品を表すコンセプト文は1つとは限らず、抽象的なものから具体的なものまでであるため、クラスター数を調整することで、コンセプト文の抽象化、具体化が起こると仮説を立てた。そのため、最適なクラスター数を求めることはせず、各クラスターに含まれるコミックの数が平均で15作品程度、10作品程度、10作品以下となるように、クラスター数をそれぞれ20クラスター、25クラスター、30クラスターに設定した。k-means法の実装にはPythonのオープンソースとして公開されている機械学習ライブラリのScikit-learn (ver. 1.3.2)²のクラスであるsklearn.clusterのKMeansを使用した。その結果を表5.1~5.3に示す。クラスターIDは区別のために便宜上振り分けたものであり、順序に意味はない。以下、個々のクラスターに対する結果、考察を述べる際にも、このクラスターIDを用いて説明を行う。また、各クラスター群に属するコミックの結果の一部を表5.4に示す。

20クラスター（表5.1参照）においては、クラスターに属する作品数が最大で24作品、最小で7作品となった。表5.4より、A2はアクション系やファンタジー系といった少年漫画の要素が強いコミックのクラスター、A7はラブコメディ系の要素が強いコミックのクラスター、A10はスポーツ系の要素が強いコミックのクラスター、A14は裏社会系の要素が強いコミックが中心であるものの、他のクラスターと比較すると全く別系統のコミックも多く含まれるクラスターとなった。

25クラスター（表5.2参照）においては、クラスターに属する作品数が最大で25作品、最小で3作品となった。表5.4より、B14はファンタジー系の要素が強いコミックのクラスター、B15は恋愛要素が強いコミックのクラスター、B18はラブコメディ系の要素が強いコミックのクラスター、B19はスポーツ系の要素が強いコミックのクラスターとなった。

30クラスター（表5.3参照）においては、クラスターに属する作品数が最大で22作品、最小で3作品となった。表5.4より、C1は歴史系の要素が強いコミックのクラスター、C4はラブコメディ系の要素が強いコミックのクラスター、C7はファンタジー系の要素が強いコミックのクラスター、C20はスポーツ系の要素が強いコミックのクラスターとなった。

また、B14とC7、B19とC20のようにクラスタリングするクラスターを増やすことにより、包含関係にあるクラスターも確認することができた。

5.3 各クラスターのクラスター代表語の抽出

コンセプト文の生成のために、前節でクラスタリングを行った各クラスターを特徴づける1つ以上のクラスター代表語の抽出を試みる。クラスター代表語の決定は以下の手順で行う。

- (1) 各クラスターにおいて、クラスターに含まれるコミックのあらすじ文を結合し、1つの文章とみなし、各単語のTF-IDF値を算出し、上位の最大20単語をクラスター代

¹<https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2> (2025/2/13 存在確認)

²<https://scikit-learn.org/stable/> (2025/2/13 存在確認)

表 5.4: k-means 法によりクラスタリングした各クラスター群に属するコミックの結果の一部。クラスター ID は表 5.1～表 5.3 を参照。

クラスター群	クラスター ID	コミック一覧
20	A2	ONE PIECE, 金田一少年の事件簿, キングダム, 七つの大罪, 聖闘士星矢, 王家の紋章 , 黒執事, ゴールデンカムイ, D.Gray-man, マギ , カードキャプターさくら, 葬送のフリーレン, 賭博堕天録カイジ 24 億脱出編, ツバサ , 銀牙伝説 WEED オリオン, CAT'S EYE, エア・ギア, 花の慶次 -雲のかなたに- , ブラッククローバー, 盾の勇者の成り上がり
	A7	H2, うる星やつら, ふたりエッチ, みゆき, かぐや様は告らせたい～天才たちの恋愛頭脳戦～ , きまぐれオレンジ・ロード, ToLOVE る-とらぶる-, ラフ, ニセコイ, 僕等がいた , ヲタクに恋は難しい, その着せ替え人形は恋をする, ママレード・ボーイ, 快感フレーズ , 天使なんかじゃない, 咲 -Saki-, あなたがしてくれなくても キャプテン翼, キン肉マン, ハイキュー !!, シュート!
	A10	, DEAR BOYS ACT4, BE-BOP-HIGHSCHOOL , ダイヤの A act2, ブルーロック, 黒子のバスケ, サラリーマン金太郎, アイシールド 21 , あひるの空, アオアシ, プレイボール, エリアの騎士, DAYS デイズ, アオハライド
	A14	こちら葛飾区亀有公園前派出所, ミナミの帝王, クッキングパパ, 闇金ウシジマくん, 神の雫 , CLAMP PREMIUM COLLECTION ××× HOLiC, 嘘喰い
	B14	NARUTO, 転生したらスライムだった件, 遊☆戯☆王, マギ, RAVE , テラフォーマーズ, ブラッククローバー, ARMS, 魔法陣グルグル, DAYS , まじかる☆タルるートくん, 無職転生～異世界行ったら本気だす～, マッシュル-MASHLE-
25	B15	BOYS BE…～young adult～, みゆき, 古見さんは、コミュ症です。 君に届け, ときめきトゥナイト, ANGEL HEART 1st シーズン, あさきゆめみし, ピーチガール
	B18	, ニセコイ, 僕等がいた, ヲタクに恋は難しい, その着せ替え人形は恋をする , 瓜を破る, 快感フレーズ
	B19	キャプテン翼, ハイキュー !!, テニスの王子様, MAJOR, DEAR BOYS ACT4 , ダイヤの A act2, WORST, あひるの空, おおきく振りかぶって
30	C1	蒼天航路, 花の慶次 -雲のかなたに-, 銀牙 -流れ星 銀- 美味しんぼ, H2, SPY × FAMILY, 宇宙兄弟, ときめきトゥナイト, ふたりエッチ , BOYS BE…～young adult～, みゆき, かぐや様は告らせたい～天才たちの恋愛頭脳戦～
	C4	, ToLOVE る-とらぶる-, バクマン。 , ラフ, 電影少女 Video girl Ai, ピーチガール , ニセコイ, ヲタクに恋は難しい, L・DK, からかい上手の高木さん, ママレード・ボーイ , 天使なんかじゃない, 咲 -Saki-, あなたがしてくれなくても
	C7	遊☆戯☆王, マギ, ブラッククローバー, 無職転生～異世界行ったら本気だす～ , マッシュル-MASHLE-
	C20	ハイキュー !!, テニスの王子様, メジャー MAJOR, DEAR BOYS ACT4 , ダイヤの A act2, WORST, おおきく振りかぶって

表語の候補単語とする。

- (2) 1作品のあらすじ文を属するクラスターから削除した後、計 268 作品のあらすじ文で 1 と同様に TF-IDF 値を算出する。これを全コミックに対して行う。
- (3) 任意のクラスターにおいて、手順 (1) から得た候補単語と、その同一クラスターに属するコミックを削除した手順 (2) から得た候補単語をそれぞれ比較し、共通する単語を抽出する。
- (4) 手順 (3) をクラスターに含まれるコミックの数だけ行い、どのコミックを削除しても抽出可能な共通単語をそのクラスターの代表語として決定する。

各クラスターの候補単語の決定は、c-TF-IDF (class-based TF-IDF) の考え方をもとに行った。各クラスターにおいて、クラスターに含まれるコミックのあらすじ文を結合し、1つの文書とみなすことで、各単語の TF-IDF 値を算出した。各クラスターに含まれるそれぞれの単語の TF-IDF 値の算出には Scikit-learn (ver. 1.3.2) のクラスである `sklearn.feature_extraction.text` の `TfidfVectorizer` を使用した。今回は漫画全巻ドットコム³のあらすじ文という単一のリソースのみでコンセプト文の作成を試みているため、単語の出現頻度に大きな差が見られないことから、TF-IDF 値も同値になる単語が多い。そのため、TF-IDF 値が同値の単語はランダムに並べられ表示されることから、20 位付近において、20 位以下と 20 位以上の単語の TF-IDF 値が同値である場合、同値の該当単語は含めないこととした。したがって、算出結果の TF-IDF 値が高い上位の最大 20 単語を候補単語とした。1 作品のあらすじ文を属するクラスターから削除して、同様に TF-IDF 値を算出する。これは、クラスターに属する 1 作品のみに登場する単語が TF-IDF 値に影響を及ぼすことがあるため、その 1 作品だけの特徴を排除するためである。これを全コミックに対して行い、削除したコミックのあらすじ文が属する候補単語と同数の TF-IDF 値上位単語を取得した。このとき取得した単語と候補単語を比較し、共通して含まれる単語を抽出する。同一クラスターに属するどのコミックのあらすじ文を削除しても共通して含まれた単語をクラスター代表語として決定する。

5.4 LLM を用いたコンセプト文の生成

5.3 節で得たクラスター代表語から、LLM を用いてコンセプト文の生成を試みる。本稿では LLM として、Open AI 社³が開発・提供する ChatGPT を使用した。ChatGPT のモデルは GPT-4o (GPT-4 omni) を使用した。

コンセプト文の生成において、ChatGPT には次の 6 つの指示を与えた。

- クラスター代表語を用いて、ストーリーの内容を端的に表現した短い文章を作成すること
- コミックが特定できるような固有名詞を含めないこと
- 文末を「物語」で締めくくること
- クラスター代表語を使用する順番はどんな順番でも構わないこと

³<https://openai.com/ja-JP/> (2025/2/14 存在確認)

表 5.5: 5.3 節で得られたクラスター代表語とそのクラスター代表語をプロンプトに入力したことで得られたコンセプト文の結果. クラスター ID とそれぞれのクラスターに属するコミック一覧は表 5.4 を参照.

クラスター ID	クラスター代表語	コンセプト文
A2	伝説, 時代, 少年, 勝負	伝説を追い求める少年が時代を超えて勝負に挑む物語
A7	ラブコメディ, 二人	二人の関係が揺れ動きながら繰り広げられるラブコメディの物語
A10	高校, 入学	高校に入学した少年少女が新たな出会いと成長を重ねる物語
A14	—	—
B14	魔法, 少年, 使える	魔法を使える少年が未知の運命に立ち向かう物語
B15	—	—
B18	少女	少女が運命に立ち向かい成長していく物語
B19	名門, 高校, 大会, 優勝, 天才	名門高校に集う天才たちが大会で優勝を目指す物語
C1	—	—
C4	恋愛, ラブストーリー, 二人	二人の想いが交錯しながら紡がれる恋愛ラブストーリーの物語
C7	魔法	魔法が織りなす奇跡と冒険の物語
C20	名門	名門の誇りをかけた挑戦が繰り広げられる物語

- コンセプト文には句読点を用いないこと
- コンセプト文のみを返すこと

1つ目はクラスター代表語をきちんと使うことを制約するためである. 2つ目と3つ目は, 3章の予備実験でコンセプト文を収集した際と同様の指示を与えた. 4つ目は, キーワードが入力された順番で使用され, 作成するコンセプト文の自由度が限られることを防ぐためである. 5つ目は, コンセプト文に句読点を用いられることで, 2文以上の文章がつながることを防ぎ, ChatGPT による回答が長くなることを防ぐためである. 最後の6つ目は, 回答に全く関係のない情報が入ることを防ぐためである. これら6つの指示を踏まえて, ChatGPT には以下のようなプロンプトを与えた. 以下の例は A2 のコンセプト文を生成するために, ChatGPT に与えたプロンプトである.

グループ化された複数のコミックから得られたそれらのコミックを特徴づけるキーワードを用いて, このグループ化されたコミックのストーリーの内容を端的に表現した短い文章であるコンセプトを作成してください. キーワードを使用する順番は問いません. コンセプトのみを返してください. コンセプトにはコミックが特定できるようなキャラクター名や具体的な地名などの固有名詞は入れないでください. コンセプトの文末は「～する物語」や「～の物語」などのように「物語」で終わらせてください. コンセプトには句読点をつけないでください.

グループ化されたコミックを特徴づけるキーワード: 伝説, 時代, 少年, 勝負

表 5.5 に 5.3 節で得たクラスター代表語とそのクラスター代表語をプロンプトに入力したことで得られたコンセプト文の結果を示す.

まず, A2, A7, B14, B19, C4, C7, C20 のクラスターでは各クラスターの要素が強いコミックの系統を代表するような単語を得ることができた. こういった単語を元に得られ

たコンセプト文からは、その系統を示すかのようなコンセプト文が得られた。A2のコンセプト文はファンタジー系の他にもアクション系といった少年漫画の要素が特徴的なクラスターであったことから、コンセプト文も少年漫画の説明文のようになっていることが伺える。A7, C4の二つのコンセプト文はラブコメディ系の要素が強いコミックのクラスターであったことから、コンセプト文にも“ラブコメディ”や“恋愛”という単語が使われており、説明文にも反映されていることが伺える。B14, C7の二つのコンセプト文はファンタジー系の要素が強いコミックのクラスターであったことから、コンセプト文にも“魔法”という単語が使われており、ファンタジーを彷彿とさせる説明文となっていることが伺える。B19, C20の二つのコンセプト文はスポーツ系の要素が強いコミックのクラスターであったことから、コンセプト文にも“名門”という単語が使われている。ここで、C7のコンセプト文はB14のクラスターから作られたコンセプト文とは異なり、魔法がある世界を舞台とした物語という抽象的な説明文となっている。通常、少ない物語作品に当てはまるような説明文は、多くの物語作品に当てはまる説明文よりも具体的になると想定される。しかし、B14よりもコミックの数が少なくなったにも関わらず、C7のコンセプト文の方が抽象的になってしまっている。これは、コンセプト文の作成に使用したクラスター代表語の数が少なかったことが原因であると考えられる。B19, C20にも同様のことが言える。

次に、A10, B18のクラスターでは各クラスターの要素が強いコミックの系統に関連する単語を得られているが、上記の6つのクラスターから得られたクラスター代表語と比較すると、関連性が弱い。加えて、生成されたコンセプト文も上記の6つのクラスターと比較すると、その系統を示しているとするには説明が曖昧である。A10は、スポーツ系のコミックの多くが中学や高校の部活動を描いた物語であることから、スポーツ系を学園ものという大きな枠組みで捉えれば説明はできていると考えられるが、クラスターの特徴を掴んだ説明文になっているとは言い難い。B18は単純にクラスター代表語が少ないことから、クラスター内の系統を反映しきれなかったと考えられる。

最後に、A14, B15, C1はクラスター代表語を得られなかったことから、コンセプト文の作成はできなかった。特にA14のクラスターに属するコミックの系統が複数あったことから、クラスター内で共通する単語がほとんどなかったと考えられる。また、それぞれクラスターに属するコミックの数が最小のクラスターであったことから、クラスター内に出現する単語数も他のクラスターに比べて少なくなったことが原因であると考えられる。

6 実験

本章では、5章で生成されたコンセプト文について、生成されたコンセプト文の物語作品のストーリー内容の説明妥当性の評価実験とその結果について述べる。

6.1 実験の狙い

5章では、物語作品のあらすじ文から抽出した単語をLLMを用いてコンセプト文の生成を行った。本研究では、物語作品のストーリー内容を表現するために、コンセプト文という物語作品のストーリー内容を短く簡潔に説明した文章を用いることで、他の物語作品との共通点からその物語作品のストーリー内容を把握できるのではないかと仮定を立てた。この仮説を検証するために、3.2節の予備実験(2)で得た類似する物語作品のあらすじ文には共通の特徴が出現することから、クラスタリングされたクラスターに属するコミックのあらすじ文から抽出したクラスター代表語をLLMを用いてコンセプト文の生成を試みた。本実験では、生成されたコンセプト文が物語作品のストーリー内容の説明として妥当かどうかをアンケートを用いて検証する。また、物語作品のストーリー内容を既知かどうか、コンセプト文の受け取り方にどう影響を与えるのかについても検証する。ここで得られた結果から、生成されたコンセプト文の有効性について検討していく。

6.2 コンセプト文の評価実験

本実験では、生成されたコンセプト文の評価実験を行うために、Yahoo!クラウドソーシングにおいて199名の回答者にアンケートを実施した。アンケート参加者の属性の内訳は、図6.1～図6.3の通りである。

本実験では、アンケート参加者に対してアンケートの内容を説明を行った。アンケートの内容は大きく分けて3種類の設問からなり、(1)アンケートに使用するコミックをそれぞれ知っているかどうかを尋ねる設問、(2)1つのコミックに対してそのコミックのコンセプト文が適切かどうかの評価をする設問、(3)1つのコンセプト文に対して該当クラスターに含まれるそれぞれのコミックが当てはまるかの評価をする設問とした。(1)での回答項目は、“どんな内容か知っている、説明できる”、“名前だけ知っている”、“知らない”とした。(2)、(3)での評価は、“当てはまる”、“部分的に当てはまる”、“あまり当てはまらない”、“当てはまらない”の4段階のリッカート尺度を用いた。このとき、アンケートに使用した全てのコミックのストーリー内容を把握している人はごく僅かであることから、コミックを知らない人でも回答ができるようにするために、設問ごとに本研究で使用した漫画全巻ドットコム各コミックのあらすじ文を載せることとした。また、(3)で評価するコミックは、コンセプト文に該当するクラスターから著者がランダムで選んだ。アンケートには、表5.5で記載している9文のコンセプト文を使用した。

6.3 コンセプト文の評価結果

本節では、6.2節で実施したコンセプト文に関するアンケート評価の結果について述べる。

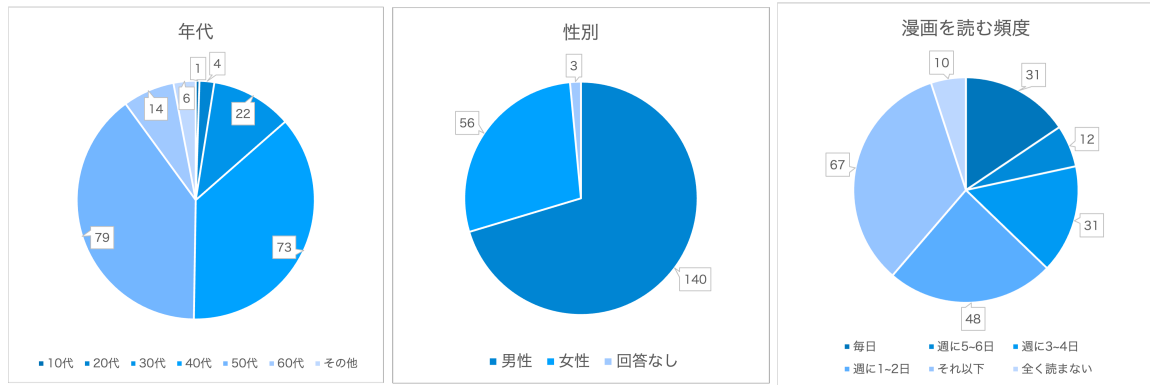


図 6.1: アンケート参加者の年代
図 6.2: アンケート参加者の性別
図 6.3: アンケート参加者の漫画を読む頻度

6.3.1 1つのコミックに対するコンセプト文の評価

図 6.4 は、6.2 節での実施したアンケート設問 (2) 1つのコミックに対してそのコミックのコンセプト文が適切かどうかの評価をする設問についての結果である。本設問では、「ヲタクに恋は難しい」に対して、当該コミックが属するクラスターから作成された3つのコンセプト文がストーリー内容を説明できているかどうかを尋ねた。その結果、A2 から作成された「二人の関係が揺れ動きながら繰り広げられるラブコメディの物語」と C4 から作成された「二人の想いが交錯しながら紡がれる恋愛ラブストーリーの物語」はともに、“当てはまる”，“部分的に当てはまる”と答えた人の割合が9割を超えた。これは、あらすじ文からメインのキャラクターが“二人”いると読み取れることや，“ラブコメディ”，“恋愛”，“ラブストーリー”といった恋愛を題材にしたコミックという特徴をコンセプト文に上手く反映することができたためであると考えられる。一方で、B18 から作成された「少女が運命に立ち向かい成長していく物語」は，“当てはまる”，“部分的に当てはまる”と答えた人の割合は3割を下回る結果となった。これは、上記の A2, C4 の二つのコンセプト文とは異なり、コンセプト文に含まれる恋愛に関連する単語が少ないことが原因であると考えられる。しかし、「ヲタクに恋は難しい」以外のコミックも複数同時に提示することで、同一クラスターに属するコミック全体に対してのコンセプト文の妥当性に関しては評価が上がる可能性がある。この検証は今後の検討課題とする。

6.3.2 同一クラスターに属する個々のコミックに対するコンセプト文の評価

本設問では、ファンタジー系の要素が見られたコミックからなるクラスターとスポーツ系の要素が見られたコミックからなるクラスターのそれぞれのコンセプト文が、コンセプト文の作成の元となったクラスターに属する個々のコミックに対して網羅的に妥当な説明となっているかの検証のために実施した。図 6.5~6.7 は、ファンタジー系の要素が強いコミックのクラスターから生成された各コンセプト文と、そのコンセプト文の作成の元となったクラスターに属する5つのコミックに対する評価である。図 6.8~6.10 は、スポーツ系の要素が強いコミックのクラスターから生成された各コンセプト文と、そのコンセプト文の作成の元となったクラスターに属する5つのコミックに対する評価である。

図 6.5 は A2 から作成されたコンセプト文と A2 中の5つのコミックそれぞれに対するコ

ンセプト文の評価の結果である。「ONE PIECE」に関しては“当てはまる”，“部分的に当てはまる”と回答した人の割合は6割を超えたが，それ以外のコミックは4～5割程度に留まっている。これは，多くのコミックが該当する抽象的な説明の文章となっていることから，コンセプト文の内容と個々のコミックのストーリー内容が合致するかどうかについては意見が二分したためだと考えられる。どの作品も“部分的に当てはまる”と回答した人が3割以上おり，大きな割合を占めていることから，完全に説明できているとは言えないがおおよその説明はできていると考えられる。この結果もまた，6.3節の結果と同様に，同一クラスターに属するコミック全体に対してのコンセプト文の妥当性に関しては評価が上がる可能性があると考えられる。

図6.6はB14から作成されたコンセプト文とB14中の5つのコミックそれぞれに対するコンセプト文の評価の結果である。「マギ」，「魔法陣グルグル」，「無職転生～異世界行ったら本気出す～」の3作品は，どのあらすじ文にもコンセプト文と同じ“魔法”という単語が入っていたことから，6割以上の回答者が“当てはまる”，“部分的に当てはまる”と回答をしている。一方で，「NARUTO」と「転生したらスライムだった件」のあらすじ文には“魔法”という単語が含まれておらず，特に「NARUTO」は忍者を題材としたコミックであるため，“当てはまる”，“部分的に当てはまる”と回答した回答者は2割程度となった。これはコンセプト文の内容がクラスターの大部分を占めるコミックに引っ張られてしまい，具体的な説明になってしまったためであると考えられる。

図6.7はC7から作成されたコンセプト文とC7中の5つのコミックそれぞれに対するコンセプト文の評価の結果である。「マギ」，「ブラッククローバー」，「無職転生～異世界行ったら本気出す～」，「マッシュル-MASHLE-」の4作品はどれも魔法を中心としたコミックであることがあらすじ文からも読み取れることから，6～8割の回答者が“当てはまる”，“部分的に当てはまる”と回答している。一方で，「遊☆戯☆王」にはあらすじ文から“魔法”に関連する単語がなかったことから，“当てはまる”，“部分的に当てはまる”と回答した回答者は2割程度に収まっている。

図6.8はA10から作成されたコンセプト文とA10中の5つのコミックそれぞれに対するコンセプト文の評価の結果である。A10のコンセプト文は学園ものを表すような説明文となっているからか，高校の部活動のスポーツがメインの題材である「ハイキュー!!」，「DEAR BOYS ACT4」，「アイシールド21」は“当てはまる”，“部分的に当てはまる”の回答が5割程度に留まった。特に「キャプテン翼」はあらすじ文から主人公が小学生のスポーツであると明記されていることから，“当てはまる”，“部分的に当てはまる”と回答した人の割合は2割程度となった。一方で，「アオハライド」に関しては，8割を超える回答者が“当てはまる”，“部分的に当てはまる”と回答した。

図6.9はB19から作成されたコンセプト文とB19中の5つのコミックそれぞれに対するコンセプト文の評価の結果である。B19のコンセプト文は全体的な評価として“当てはまる”，“部分的に当てはまる”と回答した人の割合が6～8割程度となっており，今回作成したコンセプト文の中で最も全体的に良い結果となっている。コンセプト文作成の際に使用することのできた単語数が最も多かったことから，それぞれのコミックに共通する特徴を的確に抽出でき，それをコンセプト文に反映できたと考えられる。

図6.10はC20から作成されたコンセプト文とC20中の5つのコミックそれぞれに対する

コンセプト文の評価の結果である。C20のコンセプト文もB19のコンセプト文と同様に“名門”というスポーツチームの強豪を表す単語が使用されていることから、スポーツを題材とする「ハイキュー!!」,「テニスの王子様」,「MAJOR」,「DEAR BOYS ACT4」では、“当てはまる”, “部分的に当てはまる”と回答した人の割合は5割を超えている。しかし, B19のコンセプト文よりも抽象的な説明となっているからか, どの作品も“当てはまる”, “部分的に当てはまる”と回答している人の割合は減少していることが観察できる。一方, 「WORST」は不良漫画であることから, “当てはまる”, “部分的に当てはまる”と回答した人は3割以下となっている。

以上の結果から, 今回生成されたコンセプト文はそれぞれの当該クラスターに属するコミックの多くをおおまかに説明できていると考えられる。

6.4 コミックの既知性がコンセプト文の評価に影響を与えるかについての検証

物語作品のストーリー内容を知っている人と知らない人でコンセプト文の受け取り方に差異が表れるのかについて分析を行った。知っているかどうかの回答は“どんな内容か知っている, 説明できる”, “名前だけ知っている”, “知らない”の3項目から回答をしてもらったが, 今回の分析では物語作品の内容を知っているかどうかに着目していることから, “どんな内容か知っている, 説明できる”と回答した人を“知っている”とし, “名前だけ知っている”, “知らない”とどちらかに回答した人を“知らない”とした。それぞれのコミックの既知性から回答に有意差が見られるかを検証するために, カイ二乗検定を行った。

6.4.1 1つのコミックの既知・未知によるコンセプト文の評価

図6.11は, 6.3節で使用したコミック「ヲタクに恋は難しい」の既知性によるそれぞれの受け取り方の差異の結果である。「二人の関係が揺れ動きながら繰り広げられるラブコメディの物語」では, 知っている人は全員が“当てはまる”, “部分的に当てはまる”と回答をしている。一方, 知らない人の中には, 少数の人が“あまり当てはまらない”と回答しており, この回答は有意に多いことが確認された。 $(p = 0.0006 < 0.05)$ これは, あらすじ文には記述されていない特徴がコミックのストーリー内で描かれており, 知らない人があらすじ文から読み取れなかったことが原因であると考えられる。あらすじ文とコンセプト文を詳しく比較してみたところ, 恐らく「二人の関係が揺れ動きながら」の説明が該当すると考えられる。関係が揺れ動くということは, 親密になる段階と疎遠になる段階がストーリー中に表れるため, それはストーリー内容を知っていないと判別が難しいためである。「少女が運命に立ち向かい成長していく物語」では, 知っている人と知らない人で受け取り方に差はあまり見られなかった。これは, あらすじ文ではラブコメディとストーリーの内容を具体的な明記がされているのに対し, コンセプト文はラブコメディ以外のジャンルにも当てはまるような抽象的な説明文となっていることから, 知っている知らないに関係なく大多数の回答者が当てはまる要素が少ないと考えたと推察できる。「二人の想いが交錯しながら紡がれる恋愛ラブストーリーの物語」では, 知っている人は9割程度の回答者が“当てはまる”, “部分的に当てはまる”と回答しており, 知らない人も同様に8割程度の回答者が“当てはまる”, “部分的に当てはまる”と回答した。このコンセプト文からは恋愛漫画であることが読み取れることができ, “二人の想いが交錯する”ことも容易に想像がつくことから,

コミックの既知性がコンセプト文に対する受け取り方に大して影響を与えなかったと考えられる。

6.4.2 同一クラスターに属する個々のコミックの既知・未知によるコンセプト文の評価

図 6.12～6.14 は、6.3.2 項で使用したファンタジー要素の強いコンセプト文の評価の際に回答してもらったそれぞれのコミックの既知性によるそれぞれの受け取り方の差異の結果である。

まず、図 6.12 の「伝説を追い求める少年が時代を超えて勝負に挑む物語」というコンセプト文では、どのコミックも既知であるかどうかはコンセプト文の受け取り方に影響を与えていないように伺える。これは、6.3.2 項でも述べた通り、コンセプト文の内容が王道的な少年漫画の説明文となっており、あらすじ文も同様に少年漫画を表すような説明文となっていることから、コミックの既知性がコンセプト文に対する受け取り方に影響を与えなかったと考えられる。

次に、図 6.13 の「魔法を使える少年が未知の運命に立ち向かう物語」というコンセプト文では、回答者の既知性によって受け取り方に差異があるコミックをいくつか確認することができた。「NARUTO」はあらすじ文からも魔法ではなく忍術を題材にしたコミックであることがわかるため、受け取り方は既知性に関係のない結果となった。「魔法陣グルグル」は魔法使いがコミック中に出てくるとあらすじ文で明記されていることから、受け取り方に差異が表れにくかったと考えられる。しかし、この魔法使いは女の子であり、主人公は勇者（盗賊）である¹ことはあらすじ文からは読み取りにくいいため、知っている人の方が“当てはまる”，“部分的に当てはまる”と回答した人の割合は小さかった。知っている人の方が“当てはまる”，“部分的に当てはまる”と回答した人の割合が大きくなった「転生したらスライムだった件」($P = 0.023 < 0.05$)，「マギ」($P = 2.7 \times 10^{-4} < 0.05$)，「無職転生～異世界行ったら本気出す～」($P = 7.39 \times 10^{-5} < 0.05$) の 3 作品において有意差が認められた。これは、あらすじ文には明記されていないが、実際にその物語作品を読むことでコンセプト文のような解釈が可能であるということを示唆していると考えられる。

最後に、図 6.13 の「魔法が織りなす奇跡と冒険の物語」というコンセプト文でも、回答者の既知性によって受け取り方に差異があるコミックをいくつか確認することができた。コンセプト文の説明する内容が魔法をテーマにしたものであるため、「遊☆戯☆王」のあらすじ文には魔法に関連する記述は見られずゲームがメインの話題であると読み取れることから、あらすじ文を読んだだけでも“当てはまる”，“当てはまらない”の判断はコミックの内容を知らずとも容易であったと考えられる。「マギ」，「ブラッククローバー」，「マッシュル-MASHLE-」に関しては、あらすじ文中にも魔法に関連する語彙が現れていることから、コンセプト文の魔法がある世界を舞台にしたコミックという説明に合致すると考えられるため、既知性による差異は見られなかったと推察する。しかし、「無職転生～異世界行ったら本気出す～」のあらすじ文にも魔法に関連する語彙は観察できるが、有意差が見られた($p = 3.25 \times 10^{-5} < 0.05$)。これは、コンセプト文中の“冒険”という単語から判断されていると考えられる。実際に、「ブラッククローバー」，「無職転生～異世界行ったら本気出す～」，「マッシュル-MASHLE-」の 3 作品のあらすじ文には冒険を示唆するような内容が記述

¹<http://guruguru-anime.jp/character/> (2025/2/13 存在確認)

されておらず、それら3作品に対する知らない人の“当てはまる”，“当てはまらない”の回答の割合は同程度のものとなっている。一方で、「マギ」に関してはあらすじ文中に“マジカルアドベンチャー”と明記されていることから，“当てはまる”，“部分的に当てはまる”と答えた人の割合はそれら3作品と比べて大きいことがわかる。このことから、「無職転生～異世界行ったら本気出す～」ではそのコミックを読むことで冒険があると理解できることから、既知性による差異が表れたものだと考えられる。

6.4.3 コミックの既知性に関するコンセプト文への評価結果のまとめ

6.4節では、物語作品のストーリー内容の既知、未知がコンセプト文の妥当性の評価に影響を与えるのかについて調査を行った。6.4.1項と6.4.2項の結果から、既知性によるコンセプト文の受け取り方にはいくつかのコミックにおいて有意差が見られるものがあった。この有意差が生まれた要因としては、あらすじ文中には明記されていない事柄がコンセプト文に表れ、その内容を示す描写がその評価対象のコミックで描かれていることだと考えられる。しかし、既知性の違いから得られた回答の“当てはまる”，“当てはまらない”の割合が逆転するといったことはなかった。このことから、コンセプト文は物語作品の既知性に関わらず、その物語作品の内容を説明可能であることが示唆された。

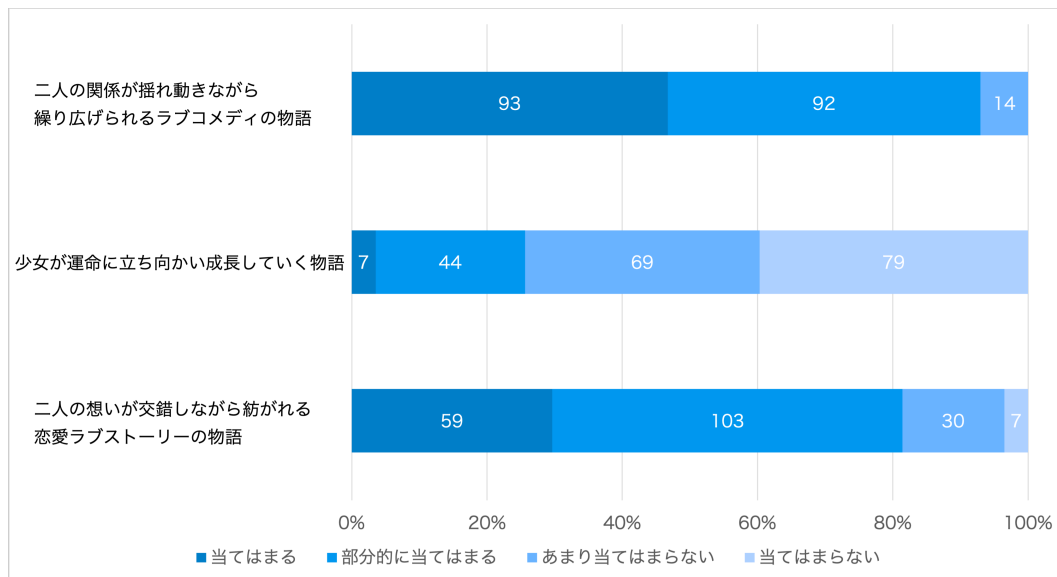


図 6.4: 「ヲタクに恋は難しい」が属するクラスターから作成された概念文の評価結果. 上から A7 (20 クラスター), B18 (25 クラスター), C4 (30 クラスター) の概念文である.

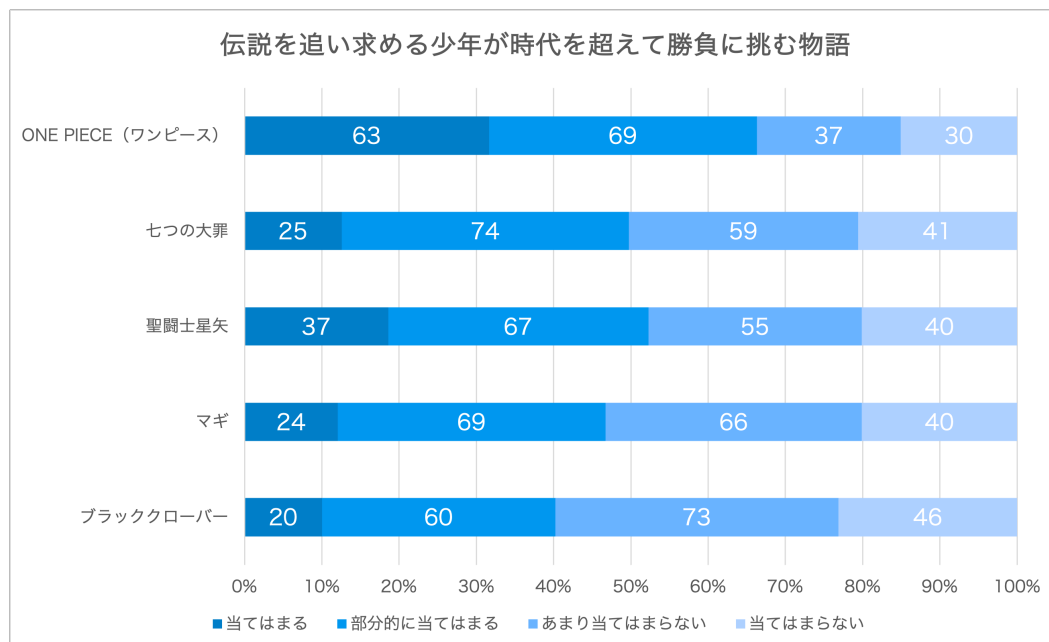


図 6.5: ファンタジー系の要素が見られたコミッククラスター A2 から作成した概念文と, A2 に属する 5 つのコミックそれぞれに対する概念文の評価.

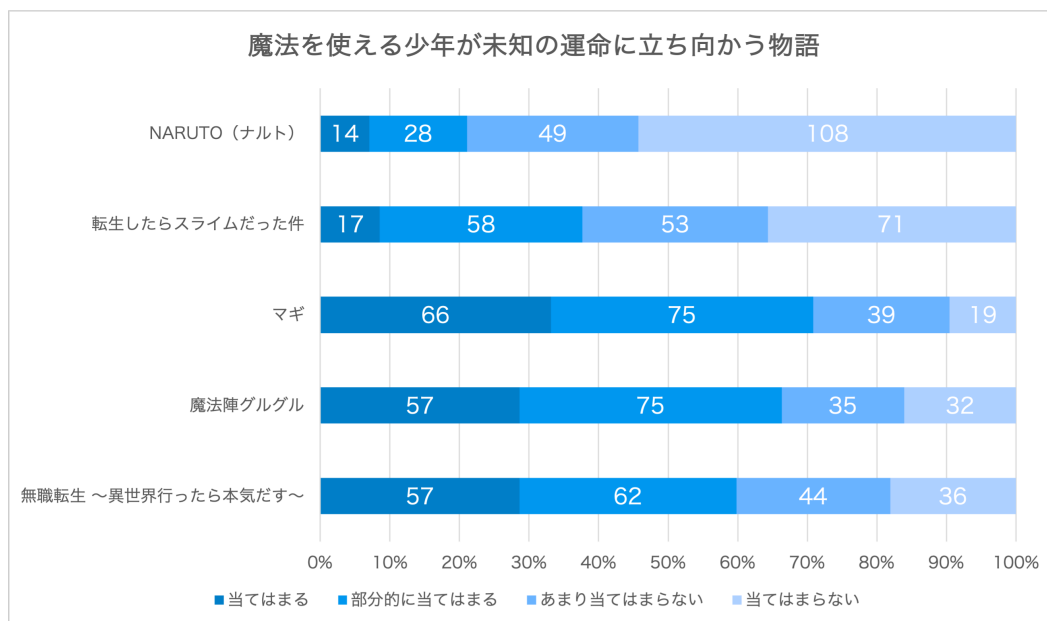


図 6.6: ファンタジー系の要素が見られたコミッククラスター B14 から作成したコンセプト文と、B14 に属する 5 つのコミックそれぞれに対するコンセプト文の評価。

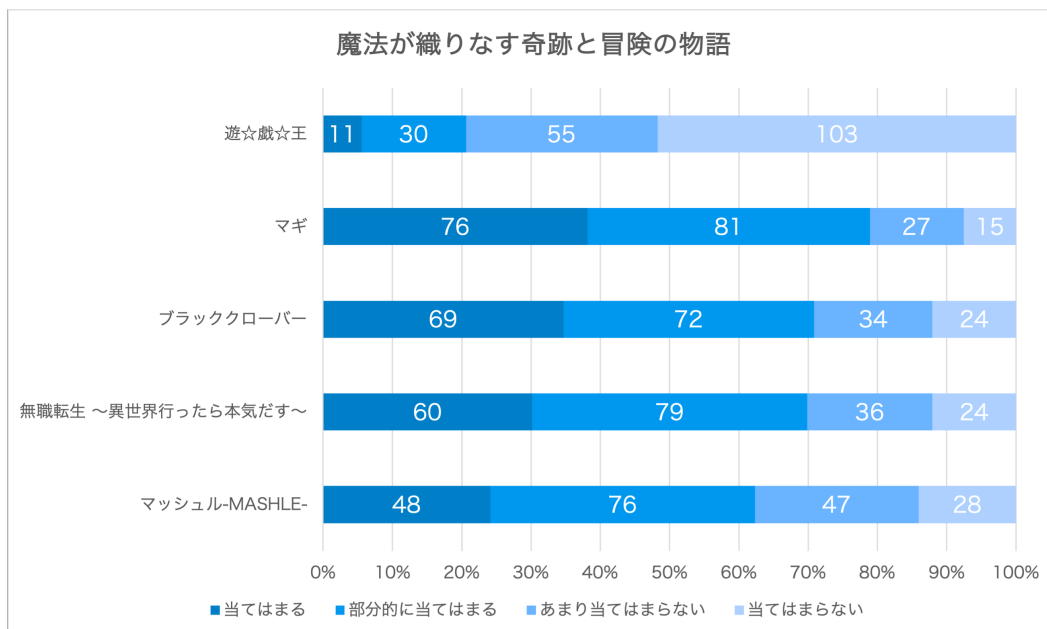


図 6.7: ファンタジー系の要素が見られたコミッククラスター C7 から作成したコンセプト文と、C7 に属する 5 つのコミックそれぞれに対するコンセプト文の評価。

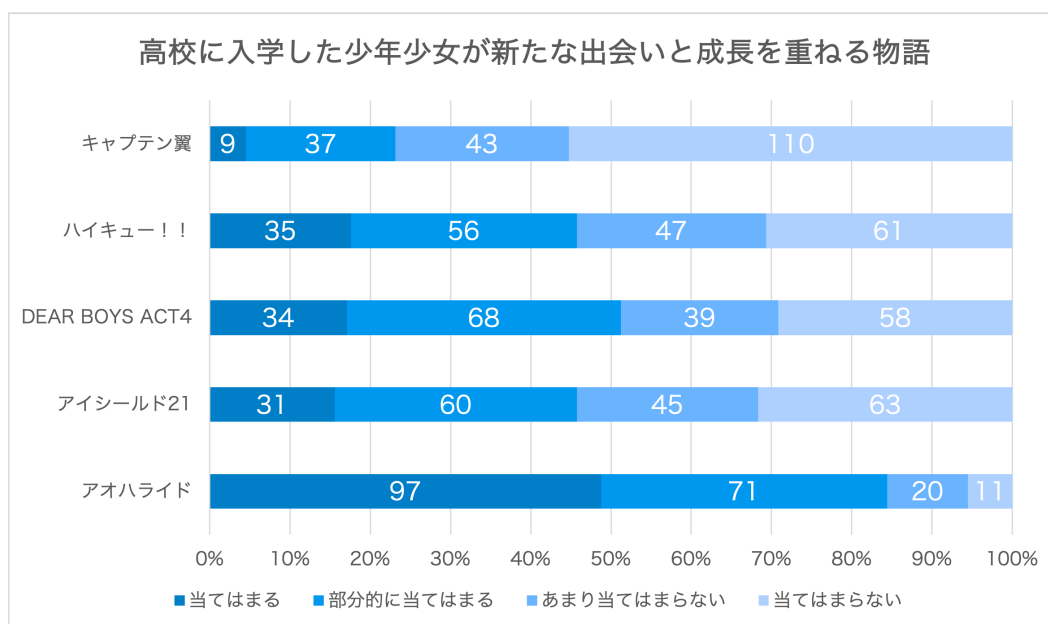


図 6.8: スポーツ系の要素が見られたコミッククラスター A10 から作成したコンセプト文と、A10 に属する 5 つのコミックそれぞれに対するコンセプト文の評価。

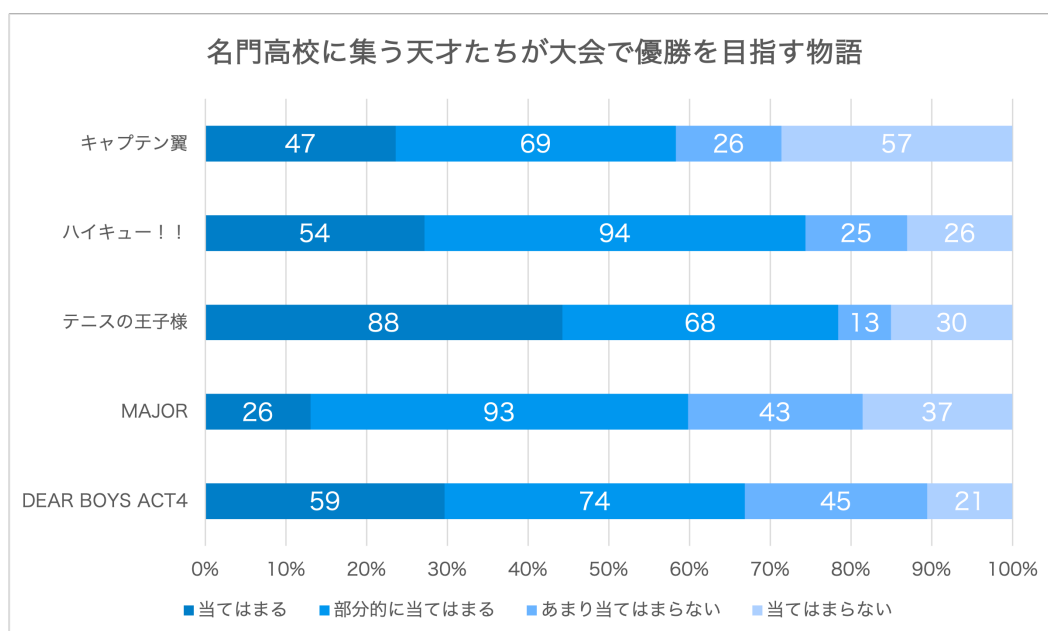


図 6.9: スポーツ系の要素が見られたコミッククラスター B19 から作成したコンセプト文と、B19 に属する 5 つのコミックそれぞれに対するコンセプト文の評価。

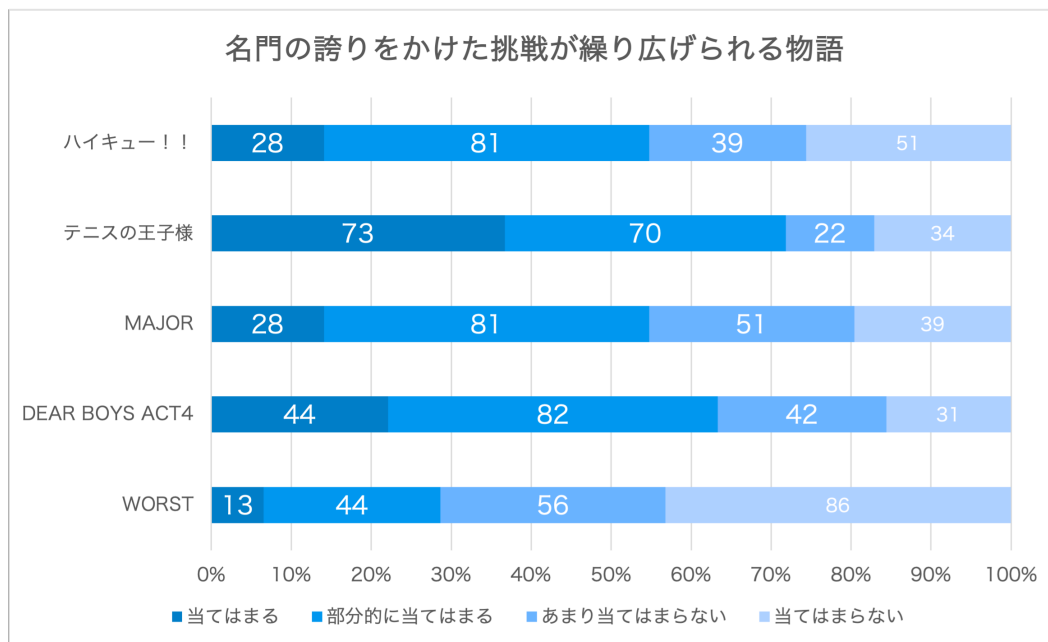


図 6.10: スポーツ系の要素が見られたコミッククラスター C20 から作成したコンセプト文と、C20 に属する 5 つのコミックそれぞれに対するコンセプト文の評価。

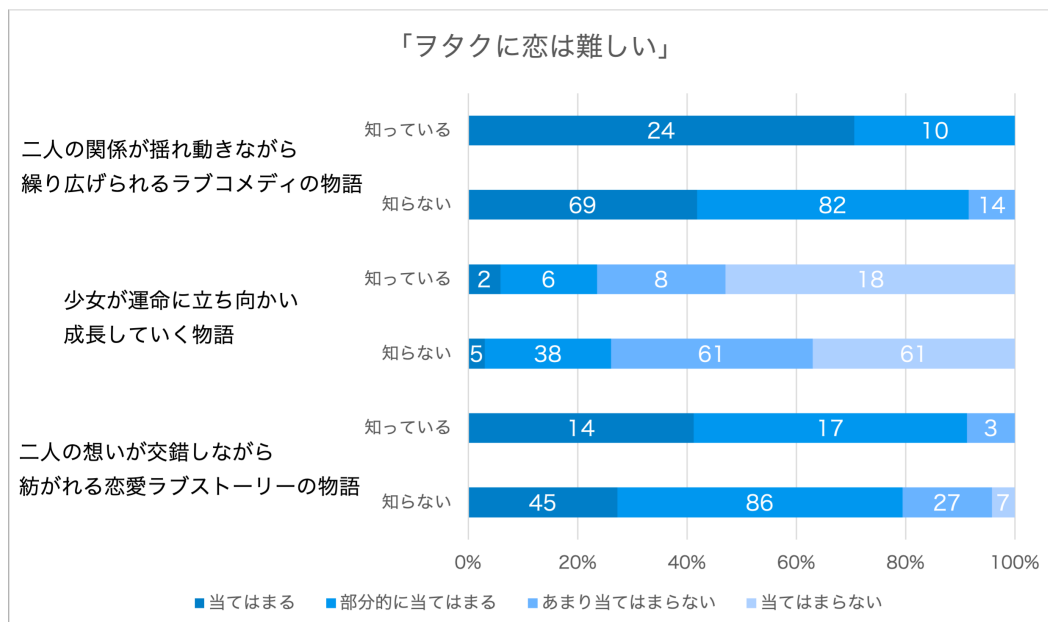


図 6.11: 「ヲタクに恋は難しい」の既知性の違いから見るコンセプト文の評価。上から A7 (20 クラスター), B18 (25 クラスター), C4 (30 クラスター) のコンセプト文である。

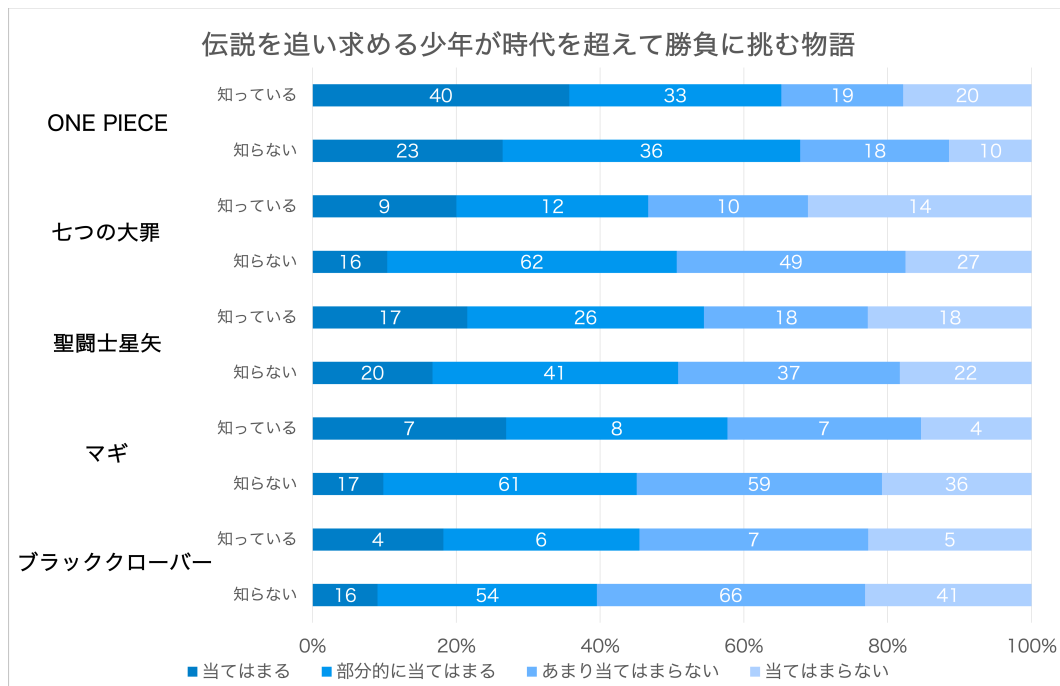


図 6.12: クラスタ A2 に属するそれぞれのコミックの既知性の違いから見るコンセプト文の評価.

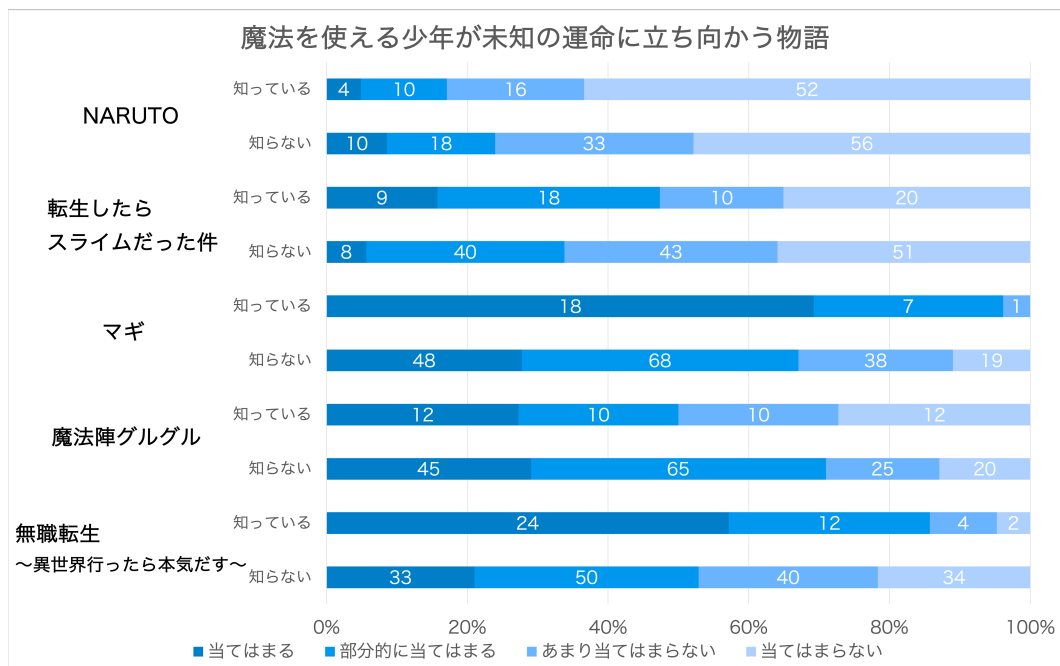


図 6.13: クラスタ B14 に属するそれぞれのコミックの既知性の違いから見るコンセプト文の評価.

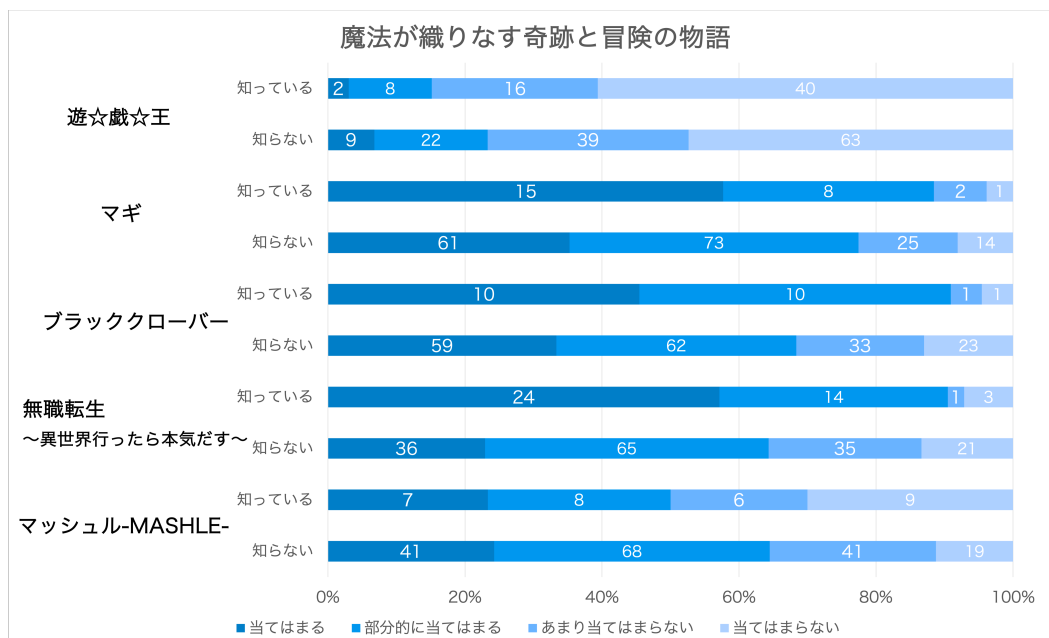


図 6.14: クラスタ C7 に属するそれぞれのコミックの既知性の違いから見るコンセプト文の評価。

7 考察

本章では、5章で生成されたコンセプト文および、6章での評価実験の結果から、生成されたコンセプト文とコンセプト文の活用について考察する。

生成されたコンセプト文に関する考察は、コンセプト文を生成する際の最初の重要なポイントであるクラスタリング、コンセプト文の生成に使用する単語の抽出方法、生成されたコンセプト文の妥当性の3点について議論する。類似する物語作品をそれぞれのクラスターに分類することができていないと、たとえコンセプト文が生成されてもそれぞれの物語作品の共通点からストーリー内容を説明しづらく、またその説明に納得がいかない点が生まれることから、コンセプト文の生成に際して適切にクラスタリングできたかについて議論する必要がある。続いて、コンセプト文の生成にはLLMを用いたことから、LLMにプロンプトとして渡す単語はコンセプト文の生成に非常に重要な役割を果たす。このことから、コンセプト文の生成に使用するクラスター代表語の抽出方法が上手く機能したのかについて議論する必要がある。最後に、生成されたコンセプト文の妥当性を考察することにより、本研究の提案手法が適切であったかを議論し、コンセプト文の生成に関する今後の展望を述べる。

また、本研究を通して得られた知見から、コンセプト文の活用方法についても議論を行う。

7.1 生成されたコンセプト文に関する考察

7.1.1 コンセプト文の生成の要となるあらすじ文のクラスタリングについて

コンセプト文を自動生成するために、類似する複数の物語作品群必要となることからクラスタリングを行った。まず、5.2節の表5.4の結果より、各クラスターには類似すると考えられるコミックがクラスタリングされていることが確認された。加えて、3章で行った実験の結果より、コンセプト文によるストーリー内容の説明は当該クラスターに属するおおよそのコミックを説明できていることがわかる。これらのことから、コミックのあらすじ文に対するクラスタリングは一定程度機能していたと考えられる。

つづいて、5.2節で述べた通り、クラスター数を増やしてクラスタリングを行うことにより、クラスター数の大小による包含関係が確認された。このことは、クラスター数の増減により、クラスターに含まれるコミックの数を増減させ、生成するコンセプト文の抽象度を調整できる可能性があると考えられる。しかし、包含関係にあるクラスター同士のコンセプト文の抽象度はクラスター数を増加させても具体的になるどころか、抽象的になることもあった。これは、あらすじ文から抽出できたコンセプト文に使用するクラスター代表語の数が少なかったためであると考えられる。

最後に、表5.4の結果より、B15とB18のようにそれぞれのクラスターは登場人物たちの恋愛模様を描いた物語作品の系統のクラスターと推察できるが、同一のクラスターには属さず、異なるクラスターとしてクラスタリングされていることがわかる。本研究では、B15からクラスター代表語の抽出ができずコンセプト文が作成できなかったため、類似する系統の物語作品のコンセプト文にどのような違いが表れるのかについて検証ができなかった。この検証はコンセプト文の生成に使用するクラスター代表語の抽出方法を再度検討し、今後の課題とする。

本研究では、クラスタリングを用いることにより、共通点を持つコミック同士に分類することを達成した。加えて、クラスター数を変更することにより、コンセプト文の抽象度について調整可能であることが示唆された。

7.1.2 コンセプト文の生成に使用するクラスター代表語の抽出方法について

5.3節において、本研究で提案するあらすじ文からコンセプト文に使用する単語の抽出方法を用いて、各クラスターからのクラスター代表語抽出を試みた。表5.5の結果から、多数のクラスターからクラスター代表語を抽出することができた。しかし、1語や2語しか抽出できていないクラスターが多く、そういった場合はLLMが関連しそうな語彙を補完してコンセプト文を作成していることが伺える。また、クラスター代表語が抽出できないクラスターもあった。こういったクラスターは大抵の場合が、そのクラスターに含まれるコミックの数が最小値のクラスターであった。この原因には大きく3つ考えられる。まず、コミックが最小値となるクラスターは、類似するコミックの少ない外れ値としてクラスタリングされたと考えられる。これは、本研究では269作品と限られた作品数で実験をしていることから、取り扱うコミックの数を増やすことで外れ値を減らせる可能性がある。つづいて、コミックのストーリー内容を表すあらすじ文のリソースの問題である。クラスターに含まれるコミックの数が最小値のクラスターということは、当然そのクラスターに含まれる単語の数も少なくなる。単語の数が少なくなることは、他の物語作品に共通する単語が少なくなる可能性が大きくなることに繋がる。これは、本研究で使用したのは単一のWebサイトから収集したあらすじ文のみを使用していることから、他のリソースのあらすじ文も使うことや、1つ目に挙げた原因と同じく取り扱うコミックの数を増やすことで解決できる可能性がある。最後にあらすじ文に出現する単語の表記揺れの問題である。あらすじ文には“二人”や“ふたり”、“バスケ部”や“バスケットボール部”、“全国制覇”や“全国優勝”などの表記揺れがいくつか見られた。あらすじ文での表現の違いも考慮したため、本研究では表記揺れを統一しなかったが、クラスター代表語をより抽出しやすくするために表記揺れは統一させる必要がある。以上のことから、本研究で提案したクラスター代表語の抽出方法は十分であるとは言い難い。

本研究では、クラスター代表語を決定する際、コミックのあらすじ文を1作品だけ削除し、同一クラスターに属するどのコミックのあらすじ文を削除しても共通して含まれるTF-IDF値の高い単語をクラスター代表語として決定した。ここで、コミックのあらすじ文を消した際に、その削除されたコミックのあらすじ文に含まれていたことによりTF-IDF値が大きく下がり、候補単語となる上位の単語から外れてしまった単語（以下、候補除外単語と記す）に着目する。候補除外単語は基本的に削除されたコミックに対しての特徴を強く表す単語であると考えられるが、同一クラスター内において複数のコミックで同一の候補除外単語が抽出されることがあった。同一の単語が候補除外単語として複数回抽出されるということは、その単語は同一クラスター内の複数のコミックのあらすじ文で1回以上出現していると考えられる。複数のコミックに同一の候補除外単語が出現しているのであれば、その単語もまたクラスター代表語として扱うこともできる可能性がある。ただし、候補除外単語はコミックのあらすじ文の削除によってTF-IDFが大きく下がることから、同一クラスター内でも一部のコミックにしか当てはまらない特徴であると考えられる。そのため、

候補除外単語の抽出回数が設定した基準回数を超えた場合のみクラスター代表語とすることや、クラスター代表語よりも重みづけを低くするなどの方法は模索する必要がある。

クラスター代表語の抽出方法の改善は今後の課題であり、この課題を解決することに妥当性の高いコンセプト文の作成やコンセプト文の抽象度の調整などにも繋がると考えられる。

本研究では、多くのクラスターでクラスター代表語を抽出することができ、コンセプト文の生成を達成した。提案したクラスター代表語の抽出方法により、類似する複数のコミック間に共通する話題を抽出することが可能であることが示唆された。

7.1.3 生成されたコンセプト文の妥当性について

6章において、生成されたコンセプト文が物語作品のストーリー内容を説明できているか検証するために、アンケート評価による実験を行った。その結果、本研究で生成されたコンセプト文はそれぞれの当該クラスターに属するコミックの多くを大まかに説明できていることが示唆された。また、それぞれのコミックに対する既知性に関しても、コンセプト文の受け取り方への影響は小さいことも示唆された。これらのことから、本研究で作成したコンセプト文は妥当であったと考えられる。しかし、本研究にて検証したことは、コンセプト文とその作成の元となったクラスターに属するコミックの1対1の評価であり、コンセプト文とその作成の元となったクラスターに属するコミック全体の1対多の評価は行っていない。今後はこの検証を行っていく必要がある。

物語作品のストーリー内容は1通りだけの表現ではなく、多数の解釈を行うことが可能である[25]。これは実際に、予備実験(1)の結果の1つである図3.1のC群からも観察できるように、物語作品の解釈は多岐にわたると考えられる。本研究では、コミックのあらすじ文を用いてコンセプト文の生成を試みたが、使用したあらすじ文は1種類のみであり、1種類の限られたリソースから多数の解釈を行うことは難しい。そこで、レビュー文を使用することが物語作品の多数の解釈を可能とすることに繋がるのではないかと考える。レビュー文は実際の読者の解釈が記述されることもあり、あらすじ文とは別の観点でその物語作品のストーリー内容の抽出ができることが期待される。レビュー文を活用することにより、1つの物語作品に対して多数のコンセプト文が付与可能になると考えられる。レビュー文を活用したコンセプト文の生成方法についても今後検討していく。

本研究で生成したコンセプト文は、そのコンセプト文が対象とするコミックのストーリー内容を既知・未知にかかわらず、説明可能であることが示唆された。

7.2 コンセプト文の活用に関する今後の展望

コンセプト文の活用は主に、本研究の目的である類似する物語作品への情報アクセスと物語作品の内容理解の2点が考えられる。

類似する物語作品への情報アクセスに関しては、コンセプト文から物語作品にアクセスする方法と、物語作品に付与されているコンセプト文から別の物語作品にアクセスする方法が考えられる。前者の方法は、ユーザーが自身が鑑賞したい物語作品があるときに、それをコンセプト文として表現することでそのコンセプト文と近しい意味となるコンセプト文に属する物語作品を提示するといった方法である。この方法では、ユーザーは自身が入力したコンセプト文と、コンピュータが計算した近しい意味となるコンセプト文を見比べるこ

とができ、自身が読みたい物語作品の探索ができるようになると考えられる。一方、後者の方法は、自身が鑑賞したことのある物語作品、もしくは興味がある物語作品に付与されているコンセプト文から、そのコンセプト文に属する別の物語作品にアクセスするといった方法である。この方法は、前者の方法とは異なり、鑑賞したい物語作品が漠然としていても物語作品にアクセスすることができると考えられる。両者の方法は、コンセプト文そのものを読むだけでもおおよその内容を理解できることから、各物語作品のあらすじ文を逐一確認しなくてもよくなるといった利点がある。こういったことから、コンセプト文は物語作品への情報アクセスとしての活用が期待できると考えた。

コンセプト文は類似する複数の物語作品に対しての共通点を切り取った短い説明文であることから、物語作品の内容理解について役立つと考えられる。そのため、コンセプト文を読むことでそれに該当する物語作品の内容をおおよそ把握することができる。しかし、コンセプト文の作成の元となった類似する物語作品は全く同一であることはなく、それぞれに突出する特徴があるはずである。コンセプト文を読むことでそれらの物語作品の大体のストーリー内容を理解すると同時に、それらの物語作品の間でどういった差異があるのかを示すことができれば、その物語作品の理解をより深めることができると考えられる。その差異を表す単語として、候補除外単語が利用できると考えられる。7.1.2節でも述べた通り、候補除外単語は基本的には削除された物語作品に対しての特徴を強く表す単語であることから、候補除外単語が有用である可能性がある。また、同一クラスターに属する物語作品のみで一般的なTF-IDF値を算出する方法もある。これをする場合は、より多くの物語作品の内容を示す他のあらすじ文やリソースが必要となる。コンセプト文を読むことで、それに該当する物語作品がどういったストーリーであるか、同一コンセプト文の他の物語作品とはどういった差異があるのかを把握することが期待される。

8 結論

本研究は、計算機によるコンセプト文の生成を目的とし、コンセプトが記述されたリソースが少なく、物語作品のコンセプト文を収集することが困難である問題を解決するために、あらすじ文に含まれるコミック間の共通点からコンセプト文の生成を試みた。

本研究では、コンセプト文と同じく物語作品のストーリー内容を説明するあらすじ文を用いて、コンセプト文が複数の類似する物語作品の共通するストーリー内容の説明文であることから、共通要素を抽出するために、あらすじ文の類似度によって物語作品のクラスタリングを行った。クラスタリングされた複数の物語作品のあらすじ文に共通して出現する特徴的な単語を抽出し、その単語を用いてLLMからコンセプト文の生成を行った。生成されたコンセプト文の多くは、そのコンセプト文の元となった複数の物語作品のストーリー内容の系統を表現したコンセプト文となった。また、生成されたコンセプト文のストーリー内容の説明が妥当なものとなっているか検証するためアンケートによる評価実験を実施した。アンケート評価実験の結果、本研究で生成されたコンセプト文はそれぞれの当該クラスターに属するコミックの多くを大まかに説明できていることが示唆された。しかし、コンセプト文の生成に使用する単語の数が少なくなってしまうことから、複数の物語作品のあらすじ文に共通して出現する特徴的な単語の抽出には改善の余地があることがわかった。

本研究では、あらすじ文という単一のリソースのみを使用したため、物語作品を複数の側面から表現することが困難であった。そこで今後は、コンセプト文に使用する単語の抽出方法を再考しつつ、多数のユーザーによる物語作品を複数の側面から評価されたレビュー文を用いることで、多様なコンセプト文の生成に取り組む。

謝辞

本研究を纏めるにあたり、関西大学総合情報学部総合情報学研究科の松下光範教授には、学部生からの4年間を通して数々の学会発表の機会をいただきました。手厚いご指導、ご鞭撻を賜りましたこと、厚くお礼申し上げます。

また、同じく関西大学総合情報学部総合情報学研究科の山西良典教授にもさまざまな場面でご意見、ご指摘を賜りましたこと、厚くお礼申し上げます。

松下研究室での日々を過ごすにあたって、11期、12期、14期、15期、16期の皆様には大変お世話になりました。感謝の意を表します。また、2024年の夏に東京に行った際、3泊も泊まらせていただいた敬愛なる大先輩、樋口亮太氏には多大なる感謝の意を表します。じゃんけんによるベッド争奪戦で3日間勝ってしまったことで、家主である樋口亮太氏を差し置いて3泊全てふかふかのベッドに寝させていただいたことは私は死んでも絶対に忘れません。

この松下研究室で共に4年間を過ごした同期である13期の皆様には、感謝してもしきれないくらいお世話になりました。櫛力輔氏とは、毎シーズン「今期は何のアニメ見る？」などとアニメの話で盛り上がったり、授業終わりに映画を一緒に見に行ったりと、たくさんお世話になりました。心より感謝いたします。岩本駿平氏は、ドッカンバトルの話で一緒に盛り上がったり、お酒に付き合ってくれたり、他愛もない話をして笑ったりと、たくさんお世話になりました。心より感謝いたします。新川晴紀氏は、同じ奈良県民として行きと帰りではばったり会って一緒に通学したり、今ハマっているVTuberの話で盛り上がったり、研究に疲れた時はキャッチボールをしたりと、たくさんお世話になりました。心より感謝いたします。杉原昂紀氏も、同じ奈良県民であるだけでなく、使ってる路線まで一緒だったことから、よく一緒に帰ったり、懐かしのアーケードカードゲームの話で盛り上がったりと、たくさんお世話になりました。心より感謝いたします。田中瑠慧氏には、しんどい時に何度も励ましの言葉をもったり、毎週のようにJR高槻駅周辺のラーメンを食べに行ったり、沖縄に行った時には一緒にホテルに泊まったりと、最高に楽しい時間を過ごしました。心より感謝いたします。畑玲音氏は、松下研究室の運営に積極的に関わってくれていたおかげで、先輩、後輩への架け橋となり、研究室を盛り上げてくれたりとたくさんお世話になりました。心より感謝いたします。林沙也加氏も、玲音氏と松下研究室の運営に積極的に関わってくれていたおかげで、先輩、後輩との仲を取り持って、研究室を盛り上げてくれたりとたくさんお世話になりました。心より感謝いたします。改めて、13期の皆様には心より感謝いたします。

また、小学校時代からの友達であり、今や大親友とも言える水本好祐氏にも、感謝の意を示します。共通の好きなものが多く、会える時間があるなら一緒に夜遅くまでお酒を飲みに行ったり、東京に行く用事があると毎回泊めてもらったりと大学生活には欠かせないくらい大変お世話になりました。心より感謝します。

一般財団法人吉村財団の皆様には、大学院の2年間において金銭的なご支援をいただきました。奨学生として様々なご支援をいただきましたこと、心より感謝申し上げます。

大学院の夏シーズンにおいて、2年とも楽しくバイトとして働かせていただき、サントリ屋上ビアガーデンと一緒に働いた皆様に、感謝の意を示します。

私のこの4年間の研究生生活のモチベーションとなった、数々の漫画の作者の方々と、それらに携わった出版社の方々、アニメ制作の方々に、感謝の意を示します。

最後に、大学院に通うための学費を支払ってくれたり、3食の食事を用意してくれたりなど、帰ることのできる暖かい家庭を守ってくれた両親と妹、その他親族の皆様に感謝の意を示すとともに謝辞と致します。

参考文献

- [1] Chaturvedi, S., Srivastava, S. and Roth, D.: Where Have I Heard This Story Before? Identifying Narrative Similarity in Movie Remakes, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 2, pp. 673–678 (2018).
- [2] Chu, W. T. and Li, W. W.: Manga FaceNet: Face Detection in Manga based on Deep Neural Network, *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, Vol. 2019, pp. 412–415 (2017).
- [3] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2019).
- [4] Friedland, L. and Allan, J.: Joke retrieval: recognizing the same joke told differently, *CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 883–892 (2008).
- [5] Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure, *arXiv preprint arXiv:2203.05794* (2022).
- [6] Hattori, S. and Takama, Y.: Recommender System Employing Personal-Value-Based User Model, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 18, No. 2, pp. 157–165 (2014).
- [7] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *NIPS2013*, pp. 3111–3119 (2013).
- [8] Nguyen, D., Trieschnigg, D. and Theune, M.: Using Crowdsourcing to Investigate Perception of Narrative Similarity, *CIKM '14: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 321–330 (2014).
- [9] Reimers, N. and Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks, *arXiv preprint arXiv:1908.10084* (2019).
- [10] ウラジーミル・プロップ (著), 北岡誠司, 福田美智子 (訳): 昔話の形態学, 水声社 (1987).
- [11] ブレイク・スナイダー (著), 廣木明子 (訳): 10のストーリー・タイプから学ぶ脚本術 SAVE THE CAT の法則を聞いたおす!, フィルムアート社 (2014).

- [12] 田中海斗, 齋藤吉平, 堀田政二, 澤野弘明: モーションコミック自動生成におけるコマ抽出に関する検討, 電子情報通信学会第8回コミック工学研究会予稿集, pp. 20–23 (2022).
- [13] 白石絵里奈, 田村亮介, 浅沼爽汰, 白井聡一, 藤田和成, 町田翔, 延澤志保: 概要と読者レビューに基づく漫画特徴抽出, 情報処理学会第81回全国大会講演論文集, Vol. 2019, No. 1, pp. 443–444 (2019).
- [14] 松下光範: コミック工学—漫画を対象としたコンテンツ指向研究, システム/制御/情報, Vol. 63, No. 3, pp. 113–117 (2019).
- [15] 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 (2010).
- [16] 澤野弘明, 大西諒太, 堀田政二: 漫画画像のモーションコミック化のためのコマ抽出および無地背景推定, 電気学会論文誌C (電子・情報・システム部門誌), Vol. 142, No. 12, pp. 1281–1287 (2022).
- [17] 涌井春那, 奥野拓: 漫画のあらすじから抽出した登場人物の関係を考慮したエピソード推薦, 情報処理学会第86回全国大会講演論文集, Vol. 2024, No. 1, pp. 423–424 (2024).
- [18] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与, 言語処理学会第22回年次大会, pp. 797–800 (2016).
- [19] 全国出版協会出版科学研究所: 2024年版 出版指標 年報, 全国出版協会出版科学研究所 (2024).
- [20] 元山直輝, 岡田真, 森直樹: 座標情報に基づく漫画のオブジェクト順位推定, 情報処理学会研究報告, Vol. 2019-NL-241, No. 26, pp. 1–6 (2019).
- [21] 安尾萌, 服部正嗣, 藤田早苗, 松下光範: 物語の類型に着目した絵本の類似探索手法に関する一検討, 電子情報通信学会技術研究報告, Vol. 116, No. 436, pp. 103–108 (2017).
- [22] 秋山明日香, 山西良典: BERTopic を用いた漫画のあらすじに対する特性表現の検討, Web インテリジェンスとインタラクション研究会, Vol. 2024, No. 21, pp. 125–128 (2024).
- [23] 藤川雄翔, 松下光範, 山西良典: 物語作品の短文要約によるストーリーの特徴分析, 第10回コミック工学研究会, pp. 34–36 (2023).
- [24] 藤川雄翔, 畑玲音, 松下光範: 物語のコンセプトに基づく情報アクセス手法の基礎検討: 短文要約文とあらすじ文との比較, 2024年度人工知能学会全国大会 (第38回) 論文集 (2024).
- [25] 橋本陽介: ナラトロジー入門 プロップからジュネットまでの物語論, 水声社 (2014).
- [26] 山西良典, 西原陽子, 松下光範: 語彙の標本化と量子化によるあらすじの特性表現に関する基礎検討, 第7回コミック工学研究会, pp. 31–38 (2022).

- [27] 山下諒, 朴炳宣, 松下光範: コミックの内容情報に基づいた探索的な情報アクセスの支援, 人工知能学会論文誌, Vol. 32, No. 1, pp. WII-D-1-11 (2017).
- [28] 迎山和司, 野寺由規, 蛭名紬: MMPose を用いたマンガキャラクターにおける姿勢推定の検討, 電子情報通信学会第7回コミック工学研究会予稿集, pp. 27-30 (2022).
- [29] 朴炳宣, 松下光範: 深層学習を用いたコミックの内容推定, 日本画像学会誌, Vol. 58, No. 5, pp. 528-537 (2019).
- [30] 朴炳宣, 松下光範, 服部正嗣: トピック分類を用いた絵本の類似検索に関する検討, 電子情報学会技術研究報告, Vol. 117, No. 420, pp. 113-118 (2018).