The University of Osaka
Institutional Knowledge Archive

Title	ビッグデータとしてのウェブ小説ー言語特徴およびトレンド解析ー
Author(s)	黄, 晨雯
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/89502
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

2022 年博士学位申請論文

ビッグデータとしてのウェブ小説

―言語特徴およびトレンド解析―

大阪大学大学院言語文化研究科

言語文化専攻

執筆者： 黄 晨雯

目次

1．研究背景 1
1.1 先行研究 2
1.2 研究目的 5
1.3 本論文の構成 7

2．分析データ 9
2.1 「小説家になろう」 9
2.2 「アルファポリス」 13

3．基本情報 15
3.1 「小説家になろう」 15
3.2 人気度 16
3.3 「アルファポリス」 22

4．小説のタイトル 25
4.1 タイトルの長さ 25
4.1.1 全体 25
4.1.2 「小説家になろう」 27
4.1.3 「アルファポリス」 29
4.2 タイトルにおける語彙 30
4.2.1 単語分割 30
4.2.2 語彙多様性 32
4.2.3 コレスポンデンス分析 35
4.2.3.1 全体 35
4.2.3.2 人気作品 53

5．小説のあらすじ 56
5.1 あらすじの長さ 56
5.2 トピックモデル 57
5.2.1 今までのトピックモデル 57
5.2.2 Top2vec 59
5.3 モデルの実行 60
5.4 年度別のトピック 64

5.5 人気作品のトピック 75
5.6 ジャンル別のトピック 80
5.6.1 ファンタジー 82
5.6.2 恋愛 87
5.6.3 SF 94
5.6.4 推理（文芸）／ミステリー 95
5.6.5 歴史物語 97
5.7 異世界転生・転移トピック 98

6. 本文の冒頭部分 104
6.1 婚約破棄 106
6.2 ハイファンタジー 108
6.3 異世界転生・転移 110
6.4 キャラクター紹介トピック 115

7. 感想コメント 118

8. おわりに 126
8.1 まとめ 126
8.2 今後の課題と展望 127

参考文献 128

付録 1 「小説家になろう」の投稿ジャンル 133

付録 2 あらすじのトピックを構成する主要単語表 134

付録 3 コメントのトピックを構成する主要単語表 137

謝辞 140

1. 研究背景

インターネットとデジタル化の迅速な発展により、ウェブ小説が大きなブームとなっている。ウェブ小説とはウェブ上のプラットフォームに投稿された小説、もしくは高い人気と大きな注目を獲得した上で、書籍化されて紙媒体で発売された小説を指す。飯田一史作『ウェブ小説の衝撃』（2016）はウェブ小説投稿プラットフォームが支持される理由、ウェブ小説の流行作品に対する質的な内容分析、および書店とウェブコンテンツと読者との複雑な関係などのウェブ小説に関する様々な内容を扱った本である。ウェブ小説投稿プラットフォームに莫大な数の書き手が集まる理由としては、長さに縛られない自由さ、流行りものを題材に遊べること、タグと検索機能により読者のマッチングがしやすいこと、実際の読者から直接評価をもらえること、読者の反応を即時に掴めることなどが考えられると飯田は述べている。つまり、ウェブ小説プラットフォームにおいては様々な題材を集めた上に、現在の流行りを素早く把握し、読者との距離も非常に短いという特徴がある。ウェブ小説が日本の代表的なサブカルチャーになったと言っても過言ではないであろう。

ウェブ小説の投稿および閲覧プラットフォーム「小説家になろう」[1]は、2004 年から運営を始め、今では日本最大級の小説サイトとされている。2021 年 9 月 1 日の時点では、小説掲載数が 90 万を突破し、登録ユーザー数が 210 万人以上を超えている。小説を自由に掲載でき、掲載された小説を無料で読める特徴がある。

飯田（2016）では、「小説家になろう」の人気上位作品、書籍化作品に多いのはファンタジーであり、特に RPG を想起させる中世風世界ファンタジーだと述べている。流行作品の内容からすれば、人生をループする、即ち 2 度目の人生を送るパターンが人気を博し、主人公が概ねある種のスキルに傑出していると指摘した。ただし、ウェブ小説は流行の興廃変化が急速に生じるため、飯田（2016）が述べた流行作品内容の特徴が現在主流となっている作品群には必ずしも当てはまるとは限らない。現在に至るまでのウェブ小説の実態を解明するにはビッグデータ解析が必要だと思われる。

Doug Laney が 2001 年にビッグデータのキーワードを 3 つの V、即ち Volume（データ量）、Velocity（高頻度で新たなデータが発生し高速で流れていくこと）と Variety（データ種類の多様性）で定義している。近年ではさらに Veracity（データの真実性）、Variability（データが表す意味の変動性）、Visualization（データの可視化）、Value（価値、データによって達

[1] https://syosetu.com/

成できること）が加わり、7 つの V によってビッグデータが解釈されている[2]。ウェブ小説におけるビッグデータの応用は、大量で手作業における分類や分析が不可能なものを解析でき、多種類のデータを総合的に要約することもできる。また、デジタル社会において、データの更新が著しく早いため、リアルタイムで新たなものを追っていくことはできないが、ビッグデータにより風潮の変わり目などを機械的に分析することができる。意図的に抽出したデータより、ビッグデータによって得られた結果の方がより信頼性があると考えられる。さらに、数値データは論ずるまでもなく、テキストデータまで可視化することが可能なため、文字より直感的に変化と相違を感じられる特徴がある。研究価値として、デジタルヒューマニティーズ[3]（Digital Humanities）への貢献も期待できる。上述した点を踏まえ、本研究ではビッグデータとしてのウェブ小説に焦点を当て、数値データのみならず、テキストデータを大量に扱い、言語特徴を中心とした傾向分析を行う。

1.1 先行研究

日本のウェブ小説に対する言語特徴およびトレンド変化の分析に関しては、主に「小説家になろう」を分析対象とした先行研究がいくつか行われている。

浦川（2013）は「小説家になろう」を対象に、作者が小説へ付与するキーワードの時系列傾向分析を行い、流行の興廃変化を解析することを目的とした。具体的には、時系列順の出現頻度による単語の流行の視覚化および、与えられた検索語に対して類似度の高い言葉を計算し関連語として出力を試みた。その結果、いくつかの単語に関して、特徴的な出現頻度の傾向及びそれに連動する傾向を示す関連語の出現を確認した。

本田（2017）は現在 Consumer Generated Media（CGM）サイト[4]への投稿コンテンツの画一化が懸念され、既に見たことのあるようなコンテンツや派生コンテンツの増加が感じられると指摘し、「小説家になろう」を対象に、小説に付与されたキーワードを分析し、小説キーワードの多様性動向を調査した。多様性動向を定量的に分析するために提案されたのは

[2] https://datasciencedegree.wisconsin.edu/data-science/what-is-big-data/
[3] コンピューターによる情報科学の手法を、広く人文科学の研究に応用する学際的な学問分野。古典籍のデジタル化、史料や芸術作品のデジタルアーカイブの構築と利用、ビッグデータを用いたテキスト解析などを指す。（https://dictionary.goo.ne.jp/）
[4] Web サイトのユーザーが投稿したコンテンツによって形成されるメディアのことである。

コサイン類似度[5]を用いる方法である。毎月投稿される小説群における類似性や相関を調査するために，コサイン類似度の総和である SumCos 値を算出し，キーワードの類似性を時系列的に分析した。通時的に見れば SumCos 値の平均が徐々に増加していることが確認できる。つまり、キーワードが類似する小説が増えているということは、ウェブ小説のテーマにおける多様性が減少しているということを表している。

飯田（2018）は文書の分散表現を生成する Doc2Vec 手法を用いて小説のあらすじをベクトル化し、ベクトル値によるクラスタリングの結果をもとに偏りのあるジャンルを特定する方法を提案した。この研究は提案にとどまり、具体的な分析と考察は行われていないが、アイディアとしては非常に優れており、あらすじに対する類似度推定によってジャンルの特性を考察することに研究の価値があると思われる。

実崎（2019）は数値データに注目している。将来人気になる作品の発見を目指し、2018 年 11 月時点で収集した 521095 件の小説のメタデータを対象に、1 日ごとの平均ブックマーク増加数の対数値を人気度（目的変数）として回帰分析による人気度の推定を試みた。残念ながら、決定係数も誤差率も高い数値を得ることができず推定には不十分という結果になっているが、試みとしては非常に有益であり、過去の数値データを人気度推定に用いることで興味深い結果が得られるのではないかという発展の可能性を垣間見せている。

ウェブ小説以外にも、テキストデータを対象とする特徴分析に関わる先行研究は多数行われている。

『*Corpus Stylistics: Theory and practice*』（McIntyre and Walker, 2019）はテキストコーパスの構築方法を紹介し、様々な統計手法を用いてテキストコーパスに対する文体分析を試みた本である。第 4 章と第 5 章では語彙の多様性について説明し、Charles Dickens の散文と 19 世紀のフィクションコーパスを研究対象として、STTR という手法を用いて語彙の多様性を比較した。その結果、2 つのコーパスにおいて統計的有意差がないと認められた。統計検定の結果より、STTR によってコーパス間の語彙の多様性を評価できるという点は、特筆に値すると考えられる。

田畑（2004a）は、テキストデータの多変量解析に目を向け、プログラミング言語 R によるコレスポンデンス分析、主成分分析、クラスター分析の実行方法と考察の仕方を解説し、

[5] 機械学習におけるコサイン類似度（Cosine Similarity）とは、2 つのベクトルが「どのくらい似ているか」という類似性を表す尺度で、具体的には（ベクトル空間における）2 つのベクトルがなす角のコサイン値のことである。この値は、2 つのベクトルの内積（＝向きと大きさを持つベクトル同士の掛け算）を、2 つのベクトルの大きさで割ることで計算される。（https://atmarkit.itmedia.co.jp/ait/articles/2112/08/news020.html）

異なる手法を比較しながら、テキストマイニングの多様なやり方を示した。田畑（2004b）は、Dickens と Smollett の作品における–ly 副詞を対象として、主成分分析とコレスポンデンス分析を行い、結果に対する比較考察を試みた。テキストデータを扱う研究ではコレスポンデンス分析の方がより適切だということが示された。水本（2009）は主成分分析、コレスポンデンス分析、行列データを転置して行う主成分分析の 3 つの多変量解析手法を比較した。その結果、変数間の差異を調べるにはコレスポンデンス分析がより適切だということが確認され、目的に応じて使い分けすることも重要だと指摘された。

また、大規模な文書を構成する各々の文書に内在するトピックを推定するために、トピックモデルという研究手法が使用されている。文学作品もしくはコーパスデータを対象として LDA という代表的なトピックモデルを用いた応用研究は少なくない。

Jockers and Mimno（2013）は 19 世紀のイギリス、アイランドとアメリカのフィクション小説計 3346 点を研究対象として、小説のテーマの選択に作家の性別が影響するかどうかに焦点を当てトピック解析を行った結果、25 個のトピックにおいて作家のジェンダーによる相違が見られることを解明した。例えば「Female Fashion」「Flowers And Natural Beauty」「Tears and sorrow」といったトピックは女性作家に好まれ、女性作家の作品に多く出現する一方、男性作家の作品にはほぼ出現していないことを明らかにした。Jockers の研究により、フィクション創作における潜在的なトピックの選択に関しては、19 世紀の作家では男女別に好みがあることが証明された。

田畑（2017）は機械学習による自然言語処理ツールキット MALLET を用いて LDA を実行することにより生成されたトピックモデルが、FLOB コーパスのサブコーパスとなっている 15 のレジスター及び下位の標本テキストに内在する意味構造をどの程度適切に抽出しうるかを検討した。具体的には、ワードクラウドによる個々のトピックの解釈を試み、ネットワークグラフ、ヒートマップなどによる可視化を通して、各トピックと FLOB コーパスの言語使用域の標本テキストとの関係性を明らかにしている。構築されたトピックモデルはコーパスに潜在している意味構造を分類記述するのに有効であることが確認された。

黒田（2017）は、コナンドイルのミステリー小説と歴史小説を対象とし、どのような語がトピックを構成しているのか、両者がそれぞれどのようなトピックを持っているのか、ワードクラウドとネットワークグラフなどの視覚化したデータをもとに考察し、ジャンル間の相違を反映する特徴を明らかにした。

Jockers and Mimno（2013）、田畑（2017）と黒田（2017）は英語のテキストデータを扱った研究である。一方で、筆者は中国のミステリー作家 4 人の作品を対象として、LDA に基づくトピックモデルを用いて解析を行い、時代やモチーフの設定によるトピックの相違を明らかにした（黄 2020）。さらに、黄（2021）は新たな Top2Vec というトピックモデルを用い、中国の有名な推理作家である程小青の作品を考察した結果、冒頭部分において社会問題に関する論述が導入的な話として多く使われることを解明した。また、トピック間の関係性を可視化し、室内を中心として事件を展開する形が好まれることも考察できた。このように、テキストデータにおける潜在的な意味解析を行うには、トピックモデルが非常に有効な手法だということが確証された。

本研究で取り扱う語彙多様性分析、コレスポンデンス分析、およびトピックモデルに関する説明は、それぞれ第 4 章の 4.2.2、第 4 章の 4.2.3、および第 5 章の 5.2 において詳細に述べる。

最後に、本研究は Karsdorp, F.の著作『*Humanities Data Analysis: Case Studies with Python*』（2021）を主な参考として、プログラミング言語 Python によるデータ分析を行う。Karsdorp（2021）はテキストデータを中心に、Python によって実現できる様々な統計解析の実行の仕方を説明し、データに関する多種多様な操作方法も詳述した。Web ページからデータを取得する方法も紹介され、実用性が非常に高い。

1.2 研究目的

前節で取り上げた「小説家になろう」に対する先行研究はウェブ小説の特徴にさまざまな観点からアプローチし一定の成果を上げてはいるものの、不足している部分も多数ある。

まず、「小説家になろう」のキーワードとあらすじを対象として傾向分析が行われているが、タイトルと本文については言及されていない。あらすじに対する研究に関しては、研究手法の提案にとどまり、具体的な内容分析には触れていない。また、人気度推定に関しても、ブックマークという変数しか取り扱っていないため、不十分だという問題点もある。さらに、先行研究ではウェブ小説のデータをもっぱら「小説家になろう」のみに依存しているため、ウェブ小説の特徴として論じられた項目や統計量は、「小説家になろう」という特定の集積拠点の特徴を反映しているに過ぎないのではないかという疑義を免れないと考えられる。

ウェブ小説以外のテキストデータに関わる先行研究では、様々なテキストデータ解析の手法が挙げられており、扱うデータによって適切な手法も異なることが分かる。例えば、目視で確認できないものは多変量解析とトピックモデルのような統計手法およびワードクラウドとヒートマップのような視覚化したものを使った方が質的分析で掴むことのできない特徴を考察することができる。本研究では先行研究を踏まえ、異なるデータ項目に対して最適な解析手法を考えた上で分析を進める。

また、本研究では「小説家になろう」のみならず、もう一つの利用者型小説投稿サイト「アルファポリス」[6]からも大量の小説情報を取り上げ分析対象に入れて、全般的にウェブ小説の特徴を掴むことを目指す[7]。大規模データを用いるため、テキストマイニングや機械学習の手法によって統計的に分析を進め、結果に対して質的考察を加える方針をとる。主に統計情報と言語情報に目を向け、作品数の傾向変動などを明らかにし、さらにタイトルの言語特徴、あらすじと本文のトピック構成に焦点を当て、時系列的に流行要素がどのように変化するのか、また人気作品を代表するポイントが存在するのかを検討する。

本研究では、以下のような具体的なリサーチクエスチョンを事前に設置し、それを解明することを試みる。

RQ1. ウェブ小説の投稿数は 2021 年までにどのように変化しているのか。またそれぞれのジャンル間にはどのような差異があるのか。（第 3 章の 3.1 節と 3.3 節で解明する）

RQ2. 小説サイトにおいて高い人気度を得ているのはどのような作品なのか。（第 3 章の 3.2 節と 3.3 節で解明する）

RQ3. 小説のタイトルは年とともに明らかな変化を見せているのか。またジャンル間にどのような相違が見られるのか。語彙の使用のみならず、文字数など量的情報に関してはどのような傾向があるのか。（文字数は第 4 章の 4.1 節、語彙項目は 4.2.3 節で解明する）

RQ4. 小説のあらすじに関してはどのようなトピックが潜在的に含まれているのか。年とともに鮮明な変化を見せているのか。（第 5 章の 5.4 節～5.7 節で解明する）

---

[6] https://www.alphapolis.co.jp/

[7] 他にもカクヨム、アルファポリス、エブリスタ、pixiv などの大きな利用者型小説投稿サイトが存在するが、データの取りやすさも含めて検討した結果、「アルファポリス」にした。

RQ5.　人気作品に限定し、本文の冒頭部分に関してどのような特徴的なトピックがあるのか。特定のパターンが観察されるのか。（第 6 章で解明する）

RQ6.　読者はどのような感想コメントを書きやすいのか。感想コメントの特徴は年々変化するのか。（第 7 章で解明する）

以上のようなリサーチクエスションに基づき、章ごとに解析を進めていく。

本研究は人文学におけるビッグデータの可能性を発掘する試みである。機械学習は大量のデータの中から有用な情報を取り出すことができるため、研究者が遠読[8]の視点から入った上でミクロな分析を結合し、より客観性を踏まえた解釈ができると考えられる。デジタルヒューマニティズの分野におけるサブカルチャー研究へ新視点をもたらすことも期待できる。

また、本研究を通して近年の日本のウェブ小説を概観的に見ることができる。新人作家に対しては、少しでも創作の示唆を与えることができれば幸いである。

中国においてもウェブ小説が非常に流行ったサブカルチャーに発展してきている。本研究で行われた分析結果は今後の中日のウェブ小説の比較にもつながるのではないかと思われる。文化の異同を考察するための一助となることを期待する。

1.3 本論文の構成

第 2 章では、 分析対象としたデータを詳細に説明する。本論文では小説投稿サイト「小説家になろう」と「アルファポリス」のメタ情報、人気作品の本文第 1 話のテキストファイル、および感想欄の書き込みを分析対象とする。第 3 章ではウェブ小説投稿サイトの実態を明らかにするために小説投稿数のトレンド、および人気ジャンルについて解析を行う。第 4 章では小説のタイトルを分析対象として考察する。まず、タイトルの長さ（文字数）の時系列的な変化、およびジャンル間の相違を明らかにする。次に、タイトルの語彙項目に注目し、コレスポンデンス分析手法を用いて、ジャンル間、年度別の語彙使用頻度の違いを解明する。第 5 章ではあらすじの分量の変化を明らかにし、Top2Vec という新たなトピックモデルを

[8] Moretti, F. (2013) *Distant Reading*, Verso において紹介された、文学データに計算手法を適用する文学研究のアプローチのことを指す。

用いてあらすじの内容を中心にトピック解析を行っていく。それにより、トピックのトレンド変化や人気トピックに共通する要素を解明する。続いて第 6 章においても第 5 章と同じ研究手法を使用し、人気作品の本文第一話の部分に対してトピック考察を試みる。特に、人気テーマの書き出しのパターンを探り出す。第 7 章では「小説家になろう」サイトの感想欄に書かれるコメントを取り上げ、第 5 章と第 6 章と同じ手法でトピック解析を行う。この解析を通して、ビッグデータからするウェブ小説の読者コメントの特徴を明らかにする。最後に第 8 章では、本論文のまとめ、および今後の課題と展望について述べる。

2. 分析データ

本研究では第 3 章から第 5 章まではウェブ小説のメタ情報を用いて分析を進める。第 6 章は小説の本文データを抽出して解析を行う。第 7 章ではウェブ小説の感想欄に書かれるコメントを分析対象として扱う。第 6 章と第 7 章のデータに関してはそれぞれの章の中において詳細に述べる。

この章ではメタ情報に関して、「小説家になろう」と「アルファポリス」からそれぞれデータを取得する方法について説明する。

2.1「小説家になろう」

小説のメタ情報とは何かといえば、図 2.1 が表示しているように、小説のタイトル、作者の ID、あらすじ、ジャンル、キーワード、更新日、ブックマークといったような本文以外の小説情報である。

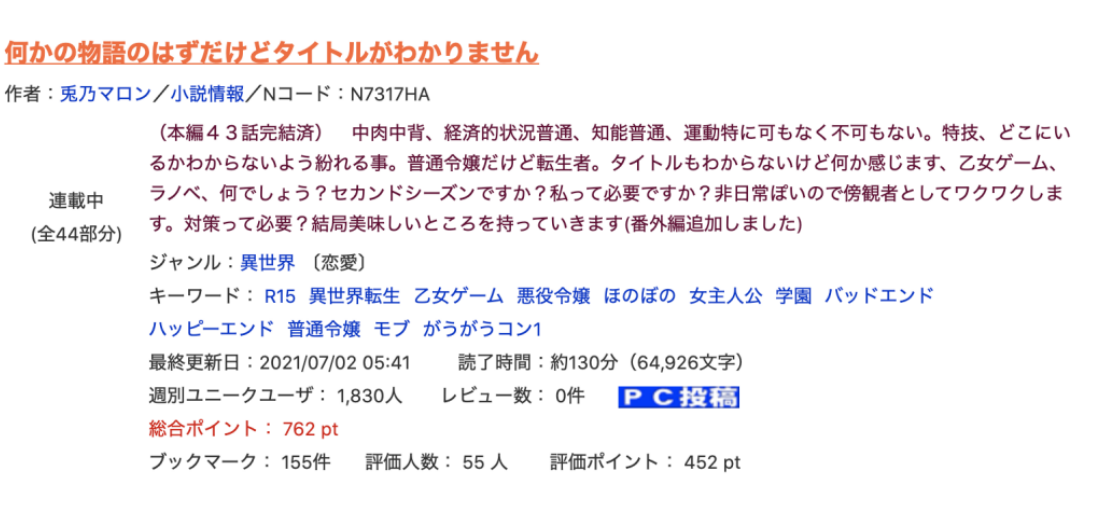


図 2.1　「小説家になろう」における作品のメタ情報

2021 年 9 月 1 日の時点で、「小説家になろう」には 90 万近くの作品が掲載されている。「小説家になろう」に関しては、掲載されている小説のメタ情報が全て「なろう小説 API」[9]に保存される。なろう API は HTTP でのリクエストに対して「小説家になろう」の小説情

[9] https://dev.syosetu.com/man/api/

報を JSON[10]形式または YAML[11]形式で応答する。図 2.2 で「なろう小説 API」に保存された小説情報の一例を挙げている。


```yaml
title: デビルエンジェルAYA
ncode: N7449HE
userid: 738795
writer: 八上 みつか
story: |
  関東某県の山中にある新興都市。
  そこは企業が『ＪＯＹ使い』という能力者を秘密裏に開発するための施設だった。
  昼間は普通の学園都市であるが、夜は制限を解かれた能力者たちが争い合う危険な街へと変貌する。
  何も知らずに高校からこの街で一人暮らしを始めた主人公、星野空人。
  彼は同じ新入生でありながら最強クラスのＪＯＹ能力を持ちヒーローに憧れる少女、赤坂綺と出会い一目惚れをする。
  それぞれ特徴のある三つの学園、夜間に勢力争いをする少年能力者グループ、そして街を支配する企業の闇。
  さまざまな陰謀渦巻く街で、空人は時に争いに巻き込まれながらも、綺に並び立つ男になるため日夜奮闘していく。
biggenre: 2
genre: 202
gensaku:
keyword: >
  R15 残酷な描写あり 青春
  異能力バトル 学園 現代
  男主人公 群像劇 超能力
general_firstup: 2021-09-07 00:00:00
general_lastup: 2021-09-22 02:00:00
novel_type: 1
end: 1
general_all_no: 138
length: 795817
time: 1592
isstop: 0
isr15: 1
isbl: 0
isgl: 0
iszankoku: 1
istensei: 0
istenni: 0
pc_or_k: 2
global_point: 30
daily_point: 12
weekly_point: 4
monthly_point: 18
quarter_point: 18
yearly_point: 18
fav_novel_cnt: 5
impression_cnt: 1
review_cnt: 0
all_point: 20
all_hyoka_cnt: 2
sasie_cnt: 0
kaiwaritu: 29
novelupdated_at: 2021-09-22 02:00:01
updated_at: 2021-09-22 02:02:04
```


図 2.2　なろう API に保存されている作品の一例（YAML 形式）

図 2.2 に示されているように、小説情報に関わる様々な項目が保存されており、それぞれが意味していることについては、「なろう小説 API」のホームページにも書いてあるが、ここでは表 2.1 に項目の説明をまとめている。後述する考察の部分においては、表 2.1 における多数の項目を用いて情報をカテゴライズし解析を行う。

[10] JavaScript のオブジェクト記法を用いたデータ交換フォーマット

[11] 何らかの構造を持ちデータ集合を簡素な文字列の並びとして表記するための記法を定めたデータ形式

表 2.1　なろう API のメタ情報の各項目の説明[12]

項目	説明
title	小説名
ncode	作品の ID
userid	作者のユーザーID
writer	作者名
story	小説のあらすじ
biggenre	大ジャンル (1:恋愛 2:ファンタジー 3:文芸 4:SF 99:その他 98:ノンジャンル)
genre	ジャンル（具体的な分類は付録 1 で説明する）
gensaku	未使用項目
keyword	キーワード
general_firstup	初回掲載日
general_lastup	最終掲載日
novel_type	連載は 1、短編は 2
end	短編と完結は 0、連載中は 1
general_all_no	全掲載部分数。短編の場合は１
length	小説文字数
time	読了時間
isstop	長期連載停止中は 1、それ以外は 0
isr15	登録必須キーワードに「R15」が含まれる場合は 1、それ以外は 0
isbl	登録必須キーワードに「ボーイズラブ」が含まれる場合は 1、それ以外は 0
isgl	登録必須キーワードに「ガールズラブ」が含まれる場合は 1、それ以外は 0
iszankoku	登録必須キーワードに「残酷な描写あり」が含まれる場合は 1、それ以外は 0
istensei	登録必須キーワードに「異世界転生」が含まれる場合は 1、それ以外は 0
istenni	登録必須キーワードに「異世界転移」が含まれる場合は 1、それ以外は 0
pc_or_k	1 はケータイのみ、2 は PC のみ、3 は PC とケータイで投稿された作品
global_point	総合評価ポイント＝((ブックマーク数×2)+評価ポイント)
daily_point	日間ポイント
weekly_point	週間ポイント
monthly_point	月間ポイント
quarter_point	四半期ポイント
yearly_point	年間ポイント
fav_novel_cnt	ブックマーク数
impression_cnt	感想数
review_cnt	レビュー数
all_point	評価ポイント

[12] 表 2.1 は https://dev.syosetu.com/man/api/ より引用し一部改変

all_hyoka_cnt	評価者数
sasie_cnt	挿絵の数
kaiwaritu	会話率
novelupdated_at	小説の更新日時
updated_at	最終更新日時

本研究では、Python コードで書かれた情報取得用のスクリプト[13]を用いて「なろう小説API」で保存されているデータセットにアクセスし、現時点における全作品のメタ情報をExcel 表として出力する。

パラメータの指定で抽出するときに不要な情報を取り上げないことも指定できるが、細かな前処理は後でするように決め、全ての情報を出力するようプログラムを動かし、合わせて 835,099 件取得し、小説情報のコーパスを構築した。

表 2.2　「小説家になろう」メタ情報コーパスの一部抜粋

title	ncode	userid	writer	story	biggenre	genre
WHITE・HEART	N4212GT	1382972	新規四季	少女は夏休みに、田舎に遊びに	2	202
幽棲博物誌	N6583HE	2158549	john doe	この世にある理解の範疇を大き	4	403
世界の治し方	N5534HD	793850	士田 松次	―汚れた校舎、ダサい制服、破	1	102
レベルってなんですか？	N5084HA	2194462	Nombre	大学の入学式にバイクで向かう	2	201
「もうお前は用済みだ」と	N4412HE	761640	みまさと　みか	「もうお前は用済みなんだよ」Sラ	2	201
『赫怒の魔』～E級成り上	N3940HE	2249457	チョトツー	あの日、人類最悪の日を四日連	2	201
悠久のクシナダヒメ「日	N5963HD	1744944	Hiroko	奈良時代の日本から戻った和也	2	202
勇者は転職して魔王にな	N9131GY	2147304	nauji	勇者が仲間達と共に魔王を倒し	2	201
妹のドアマットから解放さ	N8451HC	1428410	春乃紅葉	家族から使用人のように扱われ	1	101
絶望を生き抜いた男は、	N0362HD	1878921	乙夜　然	突然の転移により異世界へと舞	2	201

表 2.2 では構築されたコーパスの一部を取り上げたものである。行ごとに作品が分かれ、タイトルや作者などの情報が示されている。表 2.2 の列項目では 7 項目しか表示されていないが、他にも例えば初めての投稿年月日、最新の更新日、総合評価ポイント、ブックマーク数など表 2.1 で示されたように 40 項目も整理されている。不要な情報の削除、必要な情報の抽出、もしくはカテゴリーを分けて特徴をみるには、様々な項目を活用することが肝心である。

「小説家になろう」の投稿作品は幅から言えば短編（セクションが 1 つしかない作品）と連載（短編以外）の 2 種類に分かれているが、本研究では短編を除き、連載中もしくは完結

[13] 参照 Github URL: https://github.com/karupoimou/NarouDataAllDownLoad

した連載長編作品のみを研究対象とする。（処理としては”novel_type”==1 の行のみを取り出す。合わせて 448839 件。）

2.2「アルファポリス」

「アルファポリス」では、「小説家になろう」のように小説のメタデータを保存する API は設置されていない。そのため、筆者はスクレイピングするためのコードを書き、それを使用して小説に関わる諸々な情報を取得する。

図 2.3　「アルファポリス」の作品のメタ情報

図 2.3 は、累計ポイント 1 位[14]の作品に関わる諸情報をスクリーンショットしたものである。タイトル、作者、ジャンル、連載状態、長さ、あらすじ、累計ポイント、お気に入り数、文字数、投稿日、最新更新日などが使える情報として考えられるため、全て抽出し表 2.3 のようにコーパスを構築した（2021 年 7 月 6 日時点で、10 万件ほど取得した）。

[14] アクセス日： 2021 年 10 月 18 日

表 2.3　「アルファポリス」メタ情報コーパスの一部抜粋

Year	BookName	Author	Category	Status	Length	Intro	Point	Bookmark	Tokens	Created	Updated
2016	とあるおっさんの	椎名ほわほわ	ファンタジー	連載中	長編	ＶＲＭＭＯＲＰＧが普及	95530675	22747	4212299	2016.08.16	2021.07.06
2016	異世界ゆるり紀行	水無月 静琉	ファンタジー	連載中	長編	神様のミスによって命を	95302387	30102	1339339	2016.06.02	2021.07.03
2016	月が導く異世界道	あずみ　圭	ファンタジー	連載中	長編	月読尊とある女神の手	85991656	31901	3501413	2016.07.11	2021.06.23
2016	Bグループの少年	櫻井春輝	青春	連載中	長編	クラスや校内で目立つ	61128267	7617	1314082	2016.08.25	2021.01.23
2016	素材採取家の異世	木乃子増緒	ファンタジー	連載中	長編	２８歳会社員、ある日突	55948499	30627	1431055	2016.07.07	2021.07.02
2018	装備製作系チート	tera	ファンタジー	連載中	長編	※まだまだまだまだ更新	53868314	27967	2445084	2018.09.09	2021.06.23
2016	異世界に飛ばされ	シ・ガレット	ファンタジー	連載中	長編	おっさんが異世界を満喫	50810052	17266	1720182	2016.12.22	2021.07.06
2016	異世界転生騒動記	高見　梁川	ファンタジー	連載中	長編	とある貴族の少年は前世	41723706	13778	1887364	2016.08.30	2021.05.31
2019	大自然の魔法師ア	さとう	ファンタジー	連載中	長編	書籍１～５巻好評発売中	39080501	22514	1720827	2019.03.09	2021.07.05
2017	落ちこぼれ☆１魔	右薙光介	ファンタジー	連載中	長編	領主お抱えの立派な冒険	38994442	22968	1561297	2017.10.08	2020.12.21

「アルファポリス」において、小説の長さは長編、短編、ショートショート（短編より短い物語）の 3 つの類に作品が分かれている。「小説家になろう」と同じく、短編以下は取り除き、長編のみ（合計 49224 件）を研究対象とする。

「小説家になろう」と「アルファポリス」で合計 50 万近くのメタ情報を用いて、第 3 章では投稿作品の数を中心に統計し、第 4 章ではタイトルの分析を行い、第 5 章ではあらすじの特徴を考察する。

3. 基本情報

3.1「小説家になろう」

まずは全体的なトレンドを把握するために、投稿年ごとに作品を分類し、それぞれの年度における投稿作品の数を図 3.1 の棒グラフをもって表示している。図 3.1 から明らかなように、作品数が年々上昇する傾向にある。2021 年に関しては 9 月 1 日までのデータしか取っていないため、前の年度と比較することができないが、2020 年までに投稿作品数が年々増加していることから、ウェブ小説の産出が次第に多くなり、注目されつつあることが推察できる。

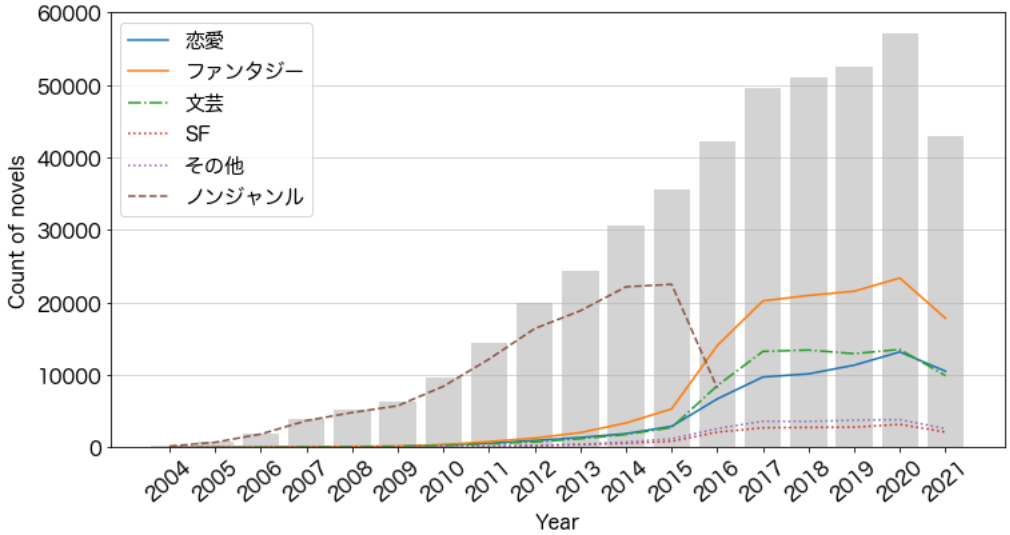
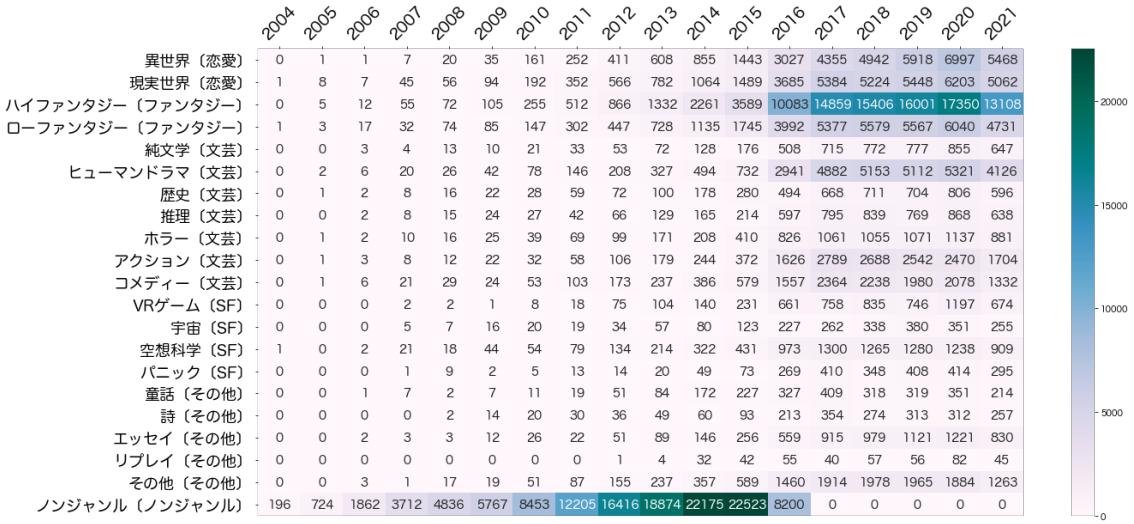


図 3.1　投稿年と小説数の時系列関係図（全作品）

また、ジャンルごとの作品数の傾向も検討したい。図 3.1 では全体の傾向を表す灰色の棒グラフ以外に、それぞれの年における各上位ジャンルの作品数も折線グラフで描画し、各ジャンルは色で分け、左側の凡例と対照しながら見ればそれぞれのトレンドが分かる。ノンジャンルの割合は 2015 年以前では圧倒的であったが、2017 年からは全てゼロとなっている。ノンジャンルは付録 1 のジャンル分類表には現れなかったが、実際、2016 年にジャンル分類が大幅に変わり、その時に一部の作品を除き、全作品のジャンルが「ノンジャンル」になり、その後のジャンル変更は作者自身がしなければならないということである。つまり「ノ

ンジャンル」は様々なジャンルが入っていると考えられ、実際のジャンル分布は不明である。ノンジャンルがなくなった 2017 年以降の傾向に注目すると、「文芸」「SF」「その他」に関してはほぼ毎年数がそれほど変わらない。「ファンタジー」と「恋愛」については上昇していることが見られ、特にファンタジーの作品数が圧倒的に多く、つまり「ファンタジー」が最もトレンドに乗っているといっても過言ではない。

細かい下位ジャンルの分布を図 3.2 のようなヒートマップ[15]の形で提示する。項目が非常に多くて、折れ線グラフでみるには重なり合いすぎて色合いの分別が難しいということもあり、ヒートマップのほうがより見やすいと判断した。図 3.2 では、横軸に年、縦軸に各ジャンルを設置し、各セルに色と数値を同時に表している。色については、薄ピンク色に近づくほど割合が低く、それに対して濃い青色に近づくほど割合が高い。



	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
異世界（恋愛）	0	1	1	7	20	35	161	252	411	608	855	1443	3027	4355	4942	5918	6997	5468
現実世界（恋愛）	1	8	7	45	56	94	192	352	566	782	1064	1489	3685	5384	5224	5448	6203	5062
ハイファンタジー（ファンタジー）	0	5	12	55	72	105	255	512	866	1332	2261	3589	10083	14859	15406	16001	17350	13108
ローファンタジー（ファンタジー）	1	3	17	32	74	85	147	302	447	728	1135	1745	3992	5377	5579	5567	6040	4731
純文学（文芸）	0	0	3	4	13	10	21	33	53	72	128	176	508	715	772	777	855	647
ヒューマンドラマ（文芸）	0	2	6	20	26	42	78	146	208	327	494	732	2941	4882	5153	5112	5321	4126
歴史（文芸）	0	1	2	8	16	22	28	59	72	100	178	280	494	668	711	704	806	596
推理（文芸）	0	0	2	8	15	24	27	42	66	129	165	214	597	795	839	769	868	638
ホラー（文芸）	0	1	2	10	16	25	39	69	99	171	208	410	826	1061	1055	1071	1137	881
アクション（文芸）	0	1	3	8	12	22	32	58	106	179	244	372	1626	2789	2688	2542	2470	1704
コメディー（文芸）	0	1	6	21	29	24	53	103	173	237	386	579	1557	2364	2238	1980	2078	1332
VRゲーム（SF）	0	0	0	2	2	1	8	18	75	104	140	231	661	758	835	746	1197	674
宇宙（SF）	0	0	0	5	7	16	20	19	34	57	80	123	227	262	338	380	351	255
空想科学（SF）	1	0	2	21	18	44	54	79	134	214	322	431	973	1300	1265	1280	1238	909
パニック（SF）	0	0	0	1	9	2	5	13	14	20	49	73	269	410	348	408	414	295
童話（その他）	0	0	1	7	2	7	11	19	51	84	172	227	327	409	318	319	351	214
詩（その他）	0	0	0	0	2	14	20	30	36	49	60	93	213	354	274	313	312	257
エッセイ（その他）	0	0	2	3	3	12	26	22	51	89	146	256	559	915	979	1121	1221	830
リプレイ（その他）	0	0	0	0	0	0	0	0	1	4	32	42	55	40	57	56	82	45
その他（その他）	0	0	3	1	17	19	51	87	155	237	357	589	1460	1914	1978	1965	1884	1263
ノンジャンル（ノンジャンル）	196	724	1862	3712	4836	5767	8453	12205	16416	18874	22175	22523	8200	0	0	0	0	0

図 3.2　投稿年とジャンルの作品数ヒートマップ図（全作品）

図 3.2 から明らかなように、3 行目のハイファンタジー（現実世界とは異なる世界を主な舞台とした小説）の作品数がどの投稿年においても極端に多く、異世界におけるファンタジー物語が最も流行しているジャンルだと言える。2015 年以前はノンジャンルが圧倒的であり、深掘りすることは困難であるが、実際は 2016 年以降の傾向と類似しているのではないかとも推測される。

3.2 人気度

[15] 2 次元データの個々の値を色や濃淡として表現した可視化グラフである。

図 3.2 によって作者側のトレンドは大まかに把握できているが、必ずしも読者の好みや傾向を反映しているとは限らない。読者にどれほど好かれているかを計る方法として、ブックマーク数や評価ポイント、週間ポイントなどが挙げられる。表 2.1 の”global point”で説明したように、総合評価ポイントはブックマーク数と評価ポイントの両方が考慮されており、より多くの情報量が含まれ、読者の好みを最大限に反映しているのではないかと考えられる。

図 3.3 では、ブックマーク数と総合評価ポイントの相関関係を示した散布図である。横軸ではブックマーク数、縦軸では総合評価ポイントの点数を表し、青色の点はそれぞれの作品がこの 2 次元散布図における位置を表示している（上と右の縦棒はそれぞれブックマーク数と総合評価ポイントのヒストグラムを表しているが、ヒストグラムについては図 3.4 と図 3.5 で詳細な説明を行い、図 3.3 ではブックマーク数と総合評価ポイントの相関関係に焦点を当てる）。図 3.3 から分かるように、ブックマーク数が多いほど、総合評価ポイントが高いという傾向があり、回帰直線からも明らかなように両者は基本的に正比例していると結論付けられる。もちろん、ブックマーク数が高いのに総合評価ポイントがゼロという例外も観察されたが、実際こうした外れ値のページをチェックすると、ポイントが未公開なため 0 として統一されているという事情があることが明らかになり、全体の傾向に影響を及ぼすことはないと思われる。要するに、人気度を測る指標としては、実はどちらを取り扱っても変わらないことが分かった。ただし、読者の評価ポイントも含まれている点から、総合評価ポイントがよりふさわしいのではないかと判断できる。

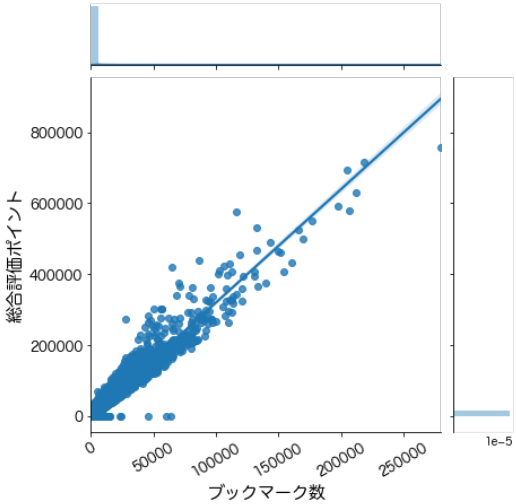


図 3.3　総合評価ポイント（縦軸）とブックマーク（横軸）の関係図

ここでは総合評価ポイントの全体像を把握するために、全作品の総合評価ポイントをヒストグラム[16]で描画したのは図 3.4 である。図 3.4 を見ると、0 に近い小説の数が格段に多いため、その 1 本の縦棒しか見ることができない。図 3.5 では対数スケール[17]を設定した上で描画した総合評価ポイントのヒストグラムである。図 3.5 から明らかなように、ポイントが上がるとともに数が減少し、何十万ものポイントを獲得している作品も存在するが、非常に稀である。

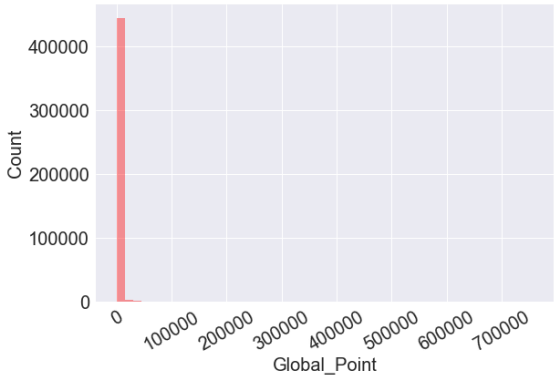


図 3.4　総合評価ポイントのヒストグラム

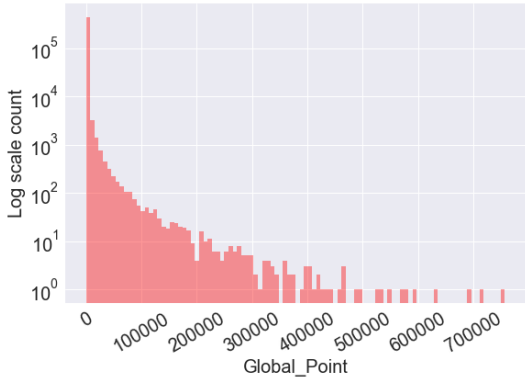


図 3.5　対数スケール使用のヒストグラム

人気作品とそれ以外の作品との間に明確な一線を画すことができるかどうかは議論が分かれる問題であるが、0 に近い作品が多いということを前提に、ここでは試しに総合評価ポイントを「0」、「1 以上 99 以下」、「100 以上 999 以下」、「1000 以上 9999 以下」、および「10000 以上」の 5 つのカテゴリーに分けてそれぞれの割合を出してみた。図 3.6 の円グラフが示すように、100 ポイント以下の作品が 7 割以上である一方、総合評価ポイント 10000 以上の作品は全体の 1.36%で、全体の割合からすれば非常に上位だと言えるのではないか。もちろん 10000 以上でさらに細分化することもできるが、一旦 10000 というところで一線を置き、10000 以上の作品を人気作品と定義し、人気作品の特徴を探っていく。

[16] ヒストグラムは、量的データの分布の様子を見るのに用いられ、データをいくつかの階級に分け、度数分布表を元にグラフ化したものである。横軸にデータの階級を、縦軸にその階級に含まれるデータの数をとる。

[17] 対数スケールは、基底値 10 をある数でべき乗した値を示す。対数スケールは、表示しているデータが残りのデータより極端に少ないか多い場合、あるいは値の間の割合または差異が重要である場合に便利である。

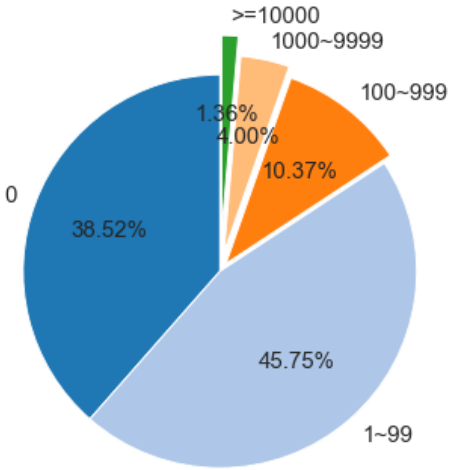


図 3.6　総合評価ポイントの分布図

そこで、人気作品の内訳をさらに明らかにすることを試みる。総合評価ポイント 10000 以上の作品総計 6104 件に対して投稿年ごとの分布を図 3.7 の棒グラフで表している。2004 年から 2007 年までは総合評価ポイント 10000 の作品がないため、2008 からの表示となっている。2021 を除き、2008 年から 2020 年までは明らかに年々人気作品の数が上がっていくことになっている。投稿年が古いほど読まれるチャンスが多くポイントがもっと積めるのではないかと思われるが、実際 2021 年 9 月の時点で 2020 年の総合評価ポイント 10000 以上の作品が最も多いということが観察され、やはり近年の作品がより注目され、作者と読者の相互交流が多いということが推察できる。

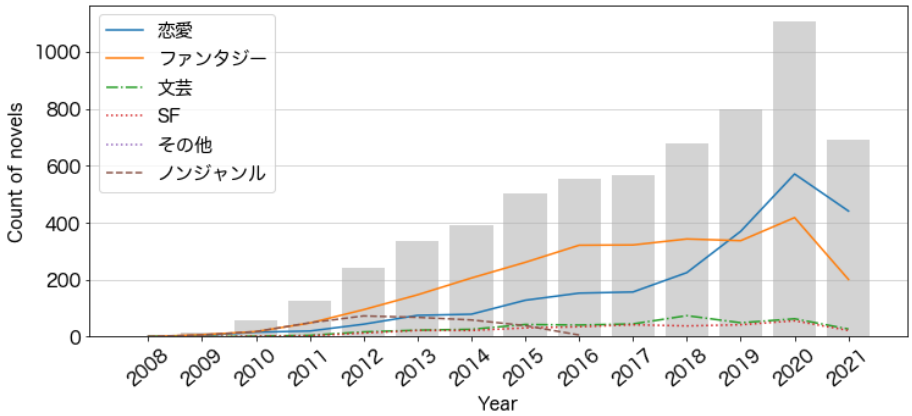


図 3.7　投稿年と小説数の時系列関係図（人気作品）

また、人気作品はほとんどファンタジーと恋愛ジャンルに属していることは明らかである。2019 年から恋愛作品が逆転し、ファンタジーを超えてさらに伸びることも非常に興味深い。その下位ジャンルの分布状況を図 3.8 のように表示すると、1 行目の異世界恋愛と 3 行目のハイファンタジーが非常に目立っている。この二つのジャンルの共通点と言えば、物語の背景が異世界にあることであり、やはり異世界を舞台とする恋愛ストーリーとファンタジー話が流行しているのであろう。

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
異世界（恋愛）	1	4	14	15	35	55	65	101	138	131	187	305	461	380
現実世界（恋愛）	0	0	2	5	9	20	14	27	15	26	38	65	111	64
ハイファンタジー（ファンタジー）	0	5	14	39	84	128	183	233	279	282	299	264	342	158
ローファンタジー（ファンタジー）	0	2	4	9	11	19	23	28	42	40	45	74	78	45
純文学（文芸）	0	0	0	0	2	0	0	0	0	1	0	1	1	1
ヒューマンドラマ（文芸）	1	0	0	2	2	8	5	13	12	6	17	12	19	11
歴史（文芸）	0	0	1	1	6	5	3	11	9	18	27	18	26	12
推理（文芸）	0	0	0	1	1	1	2	0	1	1	0	3	0	1
ホラー（文芸）	0	0	0	0	2	0	1	2	0	1	1	2	0	0
アクション（文芸）	0	0	0	1	0	3	2	5	4	6	13	5	11	1
コメディー（文芸）	0	0	1	0	4	6	13	12	15	12	16	8	6	1
VRゲーム（SF）	0	0	1	3	11	17	17	22	31	33	34	28	44	16
宇宙（SF）	0	0	0	0	2	3	0	3	1	1	3	9	4	3
空想科学（SF）	0	0	1	0	0	2	2	4	2	5	1	6	5	2
パニック（SF）	0	0	0	0	0	0	3	2	1	3	0	0	3	1
童話（その他）	0	0	0	0	0	0	0	0	0	0	0	0	0	0
詩（その他）	0	0	0	0	0	0	0	0	0	0	0	0	0	0
エッセイ（その他）	0	0	1	0	2	0	0	0	0	1	0	0	0	0
リプレイ（その他）	0	0	0	0	0	0	0	0	0	0	0	0	0	0
その他（その他）	0	0	1	1	2	1	1	2	2	1	2	3	1	0
ノンジャンル（ノンジャンル）	1	6	18	49	73	69	59	39	6	0	0	0	0	0

図 3.8　投稿年とジャンルの作品数ヒートマップ図（人気作品）

図 3.2 で明らかなように、ハイファンタジーの作品数自体が抜群に多いため、割合からすればハイファンタジーにおいて人気作品が多いことは自然だと考えられる。異世界恋愛、現実世界恋愛、ローファンタジー、ヒューマンドラマの 4 つのジャンルの作品数がそれほど変わらないことが図 3.9 から見られたので、異世界恋愛に人気作品数が格段に多いことが非常に興味深く思える。特にここ 3 年では逆に異世界恋愛の人気作品数がハイファンタジーを超える趨勢になっていることが観察され、異世界恋愛が近年の大ブームだとも考えられる。ここでは頻度ではなく、元の作品数も考慮に入れて、各カテゴリーにおける人気作品の割合を計算してさらなる検討を試みる。

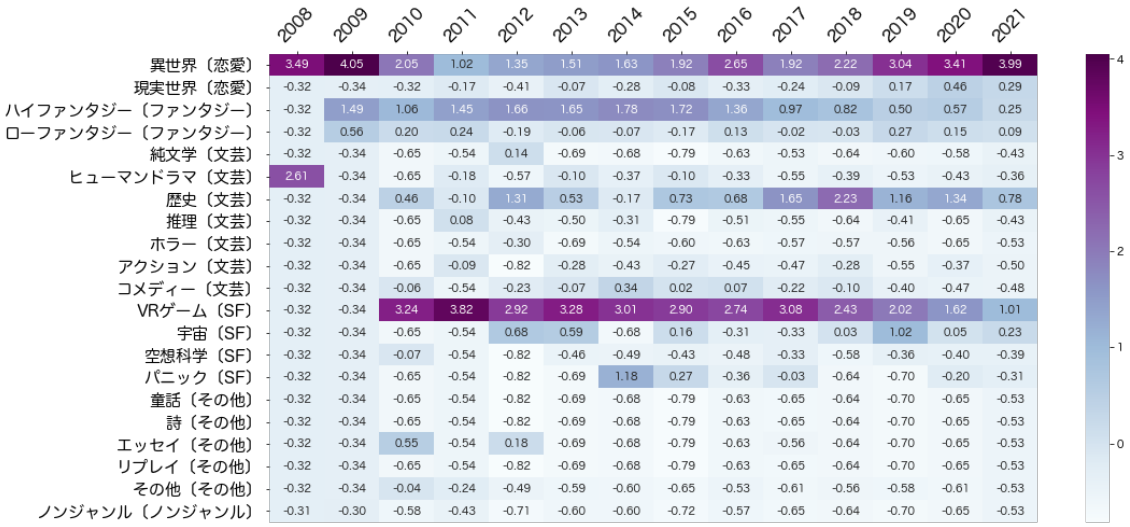


図 3.9　投稿年とジャンルの作品割合ヒートマップ図（人気作品）

図 3.9 はヒートマップを通して、投稿年とジャンルでグループ化した各カテゴリーの人気作品の重みを提示している。それぞれのセルが表す数値は、そのカテゴリーにおいて総合評価ポイント 10000 以上の小説が占める割合をさらに標準化[18]したものである。図 3.9 を吟味すると、投稿の年とほぼ関わらず、異世界恋愛と VR ゲームを題材とした小説が非常に人気を得やすいことが分かる。異世界恋愛については再度その人気性が確証されたと言えようが、意外なことに、VR ゲームに関しては数自体がそれほど多いくないが、高い評価ポイントを獲得できる確率が格段に高いことが明らかになった。つまり、VR ゲームに強い興味を抱くファンダムが存在することが証明された。

ハイファンタジーも依然として特徴的だと言えるが、異世界恋愛や VR ゲームほどではない。文芸ジャンルにおいては歴史的な物語が読者に好かれると言えるのではないか。「小説家になろう」の人気作品を扱った研究（ゆうゆう@なろうデータ収集卿, 2018）[19]によれば, 2018 年 10 月 13 日までに同サイトに掲載された全作品を対象に、10000 ポイントに到達した作品が属するジャンルをヒストグラムで可視化した結果、VR ゲームの割合が最も高く、続いて異世界恋愛、ハイファンタジー、歴史の順であることが明らかになっている。ただし、この研究では通時的あるいは時系列的な分析は行っていない。本論文が拠った研究では、その特徴は年と関わらず、全体的にそういった傾向があることが分かり、さらに 2019 年から

[18] 統計学における標準化は、与えられたデータを平均 0 、標準偏差 1 のデータに変換する操作である。ここでは、列ごとに標準化が行われた。

[19] https://ncode.syosetu.com/n3630fb/30/

2021 年にかけての 3 年間で投稿し始めた作品に関しては、異世界を舞台とする恋愛物語がとりわけ好まれているという興味深い現象も観察できた。

3.3「アルファポリス」

次に、小説投稿サイト「アルファポリス」の集計情報を検討する。図 3.10 は投稿年ごとに分かれて投稿数を棒グラフで表示したものである。結果から明らかなように、2021 年の不完全な情報を除き、2014 年設立以来投稿数が年々上昇しており、「小説家になろう」と類似したトレンドが見られている。全般的にいえば、日本においてウェブ小説への関心が年々高まっているのであろう。

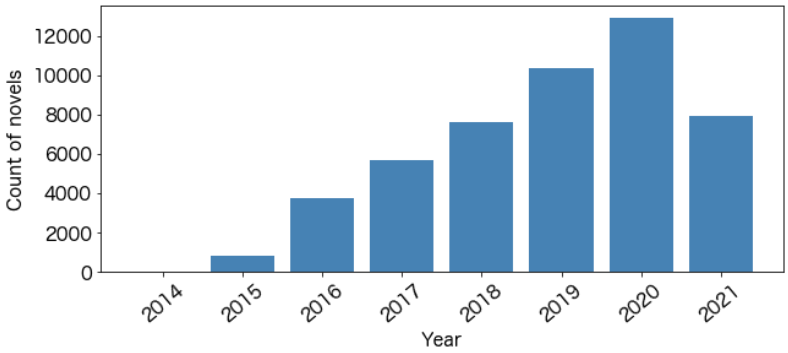
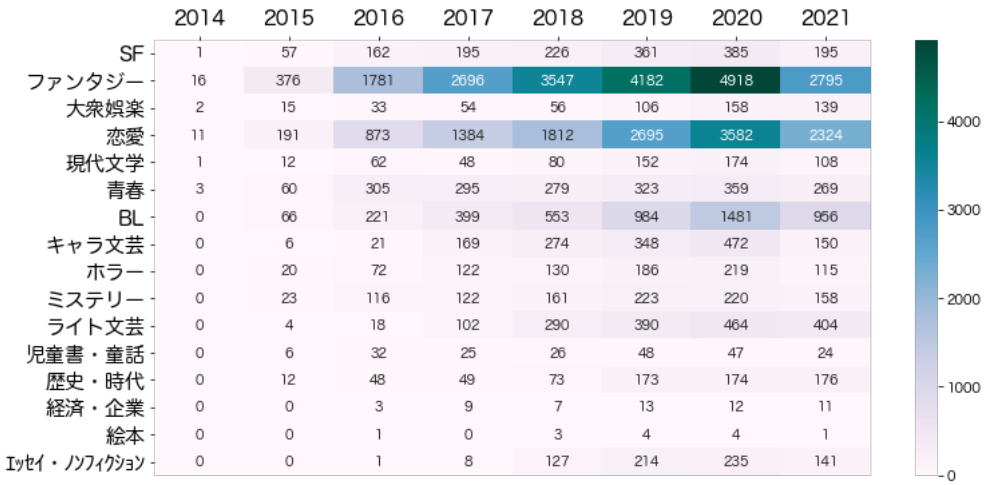


図 3.10　投稿年と小説数の時系列関係図（全作品）

投稿ジャンルの状況を把握するために、図 3.11 のようにヒートマップで各年のジャンルごとの投稿数を示している。



	2014	2015	2016	2017	2018	2019	2020	2021
SF	1	57	162	195	226	361	385	195
ファンタジー	16	376	1781	2696	3547	4182	4918	2795
大衆娯楽	2	15	33	54	56	106	158	139
恋愛	11	191	873	1384	1812	2695	3582	2324
現代文学	1	12	62	48	80	152	174	108
青春	3	60	305	295	279	323	359	269
BL	0	66	221	399	553	984	1481	956
キャラ文芸	0	6	21	169	274	348	472	150
ホラー	0	20	72	122	130	186	219	115
ミステリー	0	23	116	122	161	223	220	158
ライト文芸	0	4	18	102	290	390	464	404
児童書・童話	0	6	32	25	26	48	47	24
歴史・時代	0	12	48	49	73	173	174	176
経済・企業	0	0	3	9	7	13	12	11
絵本	0	0	1	0	3	4	4	1
ｴｯｾｲ・ﾉﾝﾌｨｸｼｮﾝ	0	0	1	8	127	214	235	141

図 3.11　投稿年とジャンルの作品数ヒートマップ図（全作品）

各投稿サイトには独自のジャンル分けの基準があるため、単純な比較はできないが、王道のジャンル、例えば、SF、ファンタジーと恋愛ジャンルはどのサイトにも大きなジャンルとして存在している。図 3.11 からわかるように、ファンタジーと恋愛を題材とする作品数が圧倒的に多く、年々増加している。この結果は図 3.1 で表示されている「小説家になろう」のジャンル別トレンドともマッチしている。面白いことに、図 3.11 ではファンタジーと恋愛に続き、BL（ボーイズラブ）作品の投稿数も際立っている。

続いて人気作品の分布状況を考察する。各投稿サイトにおいて小説に対する読者の評価基準が異なるため、「小説家になろう」の方で用いた総合評価ポイント 10000 を共通の境界線にすることはできないが、割合を参照することができる。ここでは、「小説家になろう」で定義した 1.36%に近い割合、つまり長編連載作品の上位 1~2%を基準に、境界線を置くことにする。図 3.12 は、ポイント 200 万以上（上位 2%）の作品を人気作品と定義し、合わせて 986 編の小説を取り出して、年度別の投稿数を表したものである。これもまた図 3.7 と類似した結果が得られ、やはり投稿年が近い作品ほど注目されやすいのではないかと考えられる。

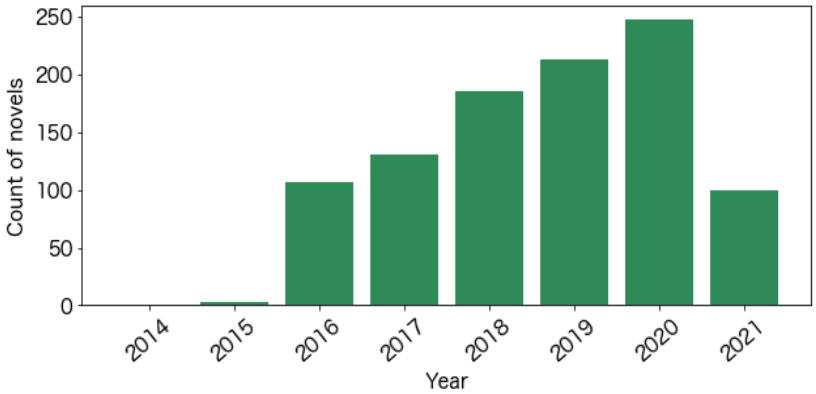


図 3.12　投稿年と小説数の時系列関係図（人気作品）

ジャンル別に人気作品の数を見れば、予想通り総投稿数がトップであるファンタジーと恋愛ジャンルが人気作品の大半を占めている。興味深いことは、ファンタジーに代わり、2019 年から 2021 年までは恋愛作品の人気度がファンタジーを上回っている。これもまた図 3.8 で示された「小説家になろう」の傾向とは類似しており、共通性が見られた。

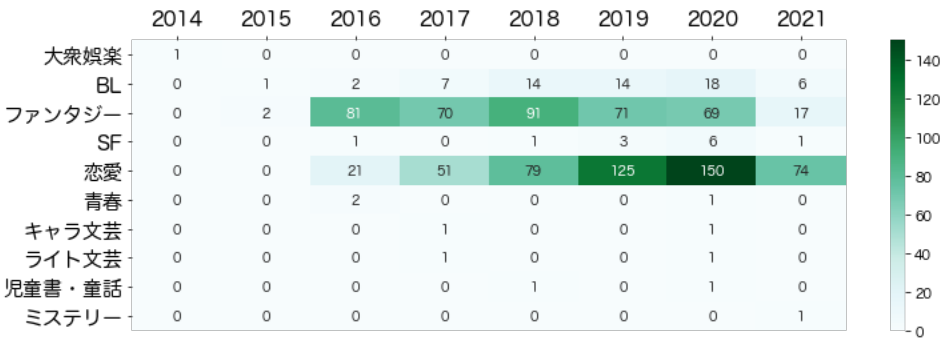


図 3.13　投稿年とジャンルの作品数ヒートマップ図（人気作品）

また、投稿年ごとの人気作品の割合を描画した図 3.14 が示したように、割合で見ると、2017 年と 2018 年では、恋愛ジャンルがファンタジーを超えている。実際 2017 年から 2021 にかけて「アルファポリス」で最も人気を博しているのは恋愛ジャンルである。

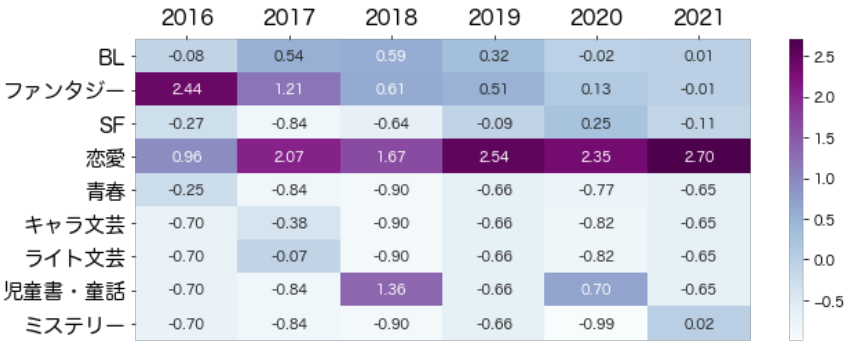


図 3.14　投稿年とジャンルの作品割合ヒートマップ図（人気作品）

「小説家になろう」では、VR ゲームを題材とする SF 作品が非常に目立ったが、「アルファポリス」ではあまり重要視されていないようである。これもまたサイトそれぞれの特徴と関係しており、SF 愛好者は「小説家になろう」で投稿した方が良いということも明らかである。

4. 小説のタイトル

前章では基本情報に対して様々な集計を行った。この章では主に言語情報に注目して、タイトルにおける各種傾向を掴むことを試みる。4.1 ではタイトルの長さに変化が見られるかどうかを検証する。4.2 ではタイトルに使われる単語情報をもとに語彙の多様性を検証し、また高頻度語の特徴を検討する。

4.1 タイトルの長さ

4.1.1 全体

ここでは小説のタイトルは次第に長くなっているのではないかという点について検証する。Jiang(2013)において分かち書きの複雑さと精度が問題だと指摘され、中国語の文の長さを数えるには単語数より文字数を用いるべきだと述べている。日本語でも類似した問題が存在するので、文字数の方が良いと判断される。従って、タイトルの長さを計測するには、単語数ではなく、文字数をそのまま使用する。

そこで、それぞれの小説のタイトルを文字数でカウントし、図 4.1 のような boxenplot によって可視化して考察を行う。

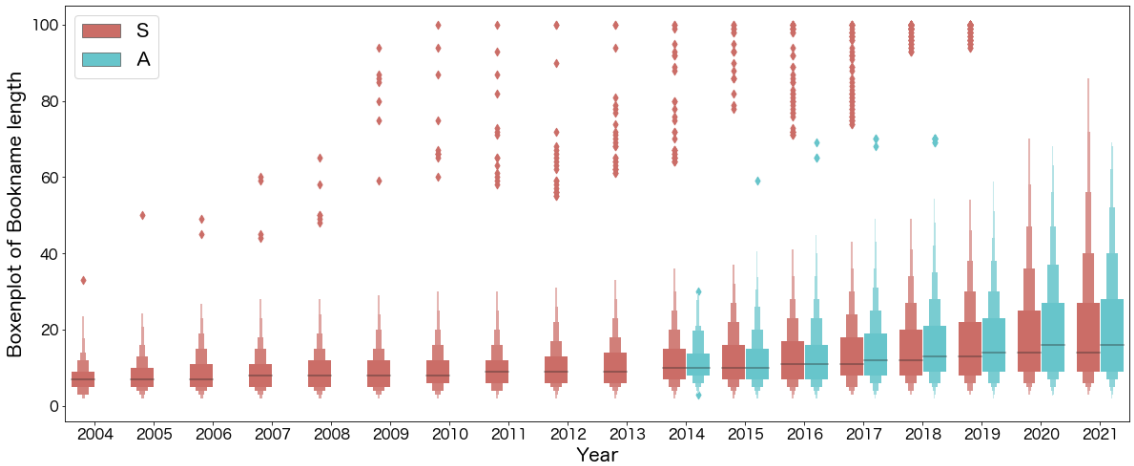


図 4.1　サイトごとのタイトルの文字数の boxenplot

図 4.1 は箱ひげ図を拡張させた boxenplot というものである。箱ひげ図はデータのばらつきをわかりやすく表現するための統計図であり、最小値、第 1 四分位点（25%）、中央値

（50%）、第 3 四分位点（75%）、最大値といった 5 つの要約統計量から成り立っている。図 4.1 で言えば、それぞれの箱の中央あたりに描いてある黒い線が中央値であり、中央値に最も近い上下の線はそれぞれ第 1 四分位点と第 3 四分位点である。Boxenplot は最大値と最小値の間にさらにいくつか小さな箱を設けている。箱と少し離れた上に位置する多数の点は、外れ値を意味している。今回扱うデータの量が多いため、普通の箱ひげ図より boxenplot を用いた方が見やすいと判断された。

図 4.1 では、「小説家になろう」（S, 赤色で表示）と「アルファポリス」（A, 青色で表示）の両サイトに分けて、投稿年ごとに文字数の状況を描いている。「小説家になろう」の方はタイトル制限が 100 文字であるのに対して、「アルファポリス」の方は 70 文字である。上限が異なるとしても、図 4.1 の結果から見るとサイトに関係なく、タイトル文字数の中央値が年々緩やかに上昇しているのではないかと思われる。中央値の上昇がそれほど明らかではないかもしれないが、2021 年に近づくほど、箱の最大値が両方とも伸びていることは明白である。つまり、中央値を遥かに超える長いタイトルが年々増加している。要するに、この 2 つのサイトを見る限り、ウェブ連載小説に関しては年々タイトルが長くなる傾向がある。

次に、人気作品の特徴も考察する。図 4.2 の boxenplot では、両サイトのデータを前章で定義された人気作品とそれ以外の作品のカテゴリーにさらに分けて、ラベルに「1」がついたものを人気作品、「0」がついたものをそれ以外の作品と表示している。S0 と A0 はほぼ図 4.1 の全体傾向と一致している。「小説家になろう」の人気作品を表す S1 は、2011 年以前はあまり目立たず S0 の傾向と変わらないが、2012 年からは徐々に差を開き、ここ数年は中央値自体の差がほぼ倍になっていることが読み取れる。一方で、「アルファポリス」の A1 では、作品数があまりにも少ない 2014 年と 2015 年を除き、2016 からの数値を見れば、A0 の作品群と次第にかけ離れていくことが見てとれる。タイトルというのは投稿後でも作者自身によって変えられるものであるため、長いから人気になりやすいのか、それとも人気になったから情報の追加や変更を行ったのかは現段階では解明できないが、人気作品全般において、タイトルを長くつけることが趨勢になっているのは確かなことである。興味深いことに、2020 年からは両サイトともほぼ文字数の制限までに人気作品の箱が伸びている。極端に長いタイトルをつけられた作品が格段に増えているためだと考えられ、2020 年からはウェブ小説の新時代になったと言っても過言ではない。

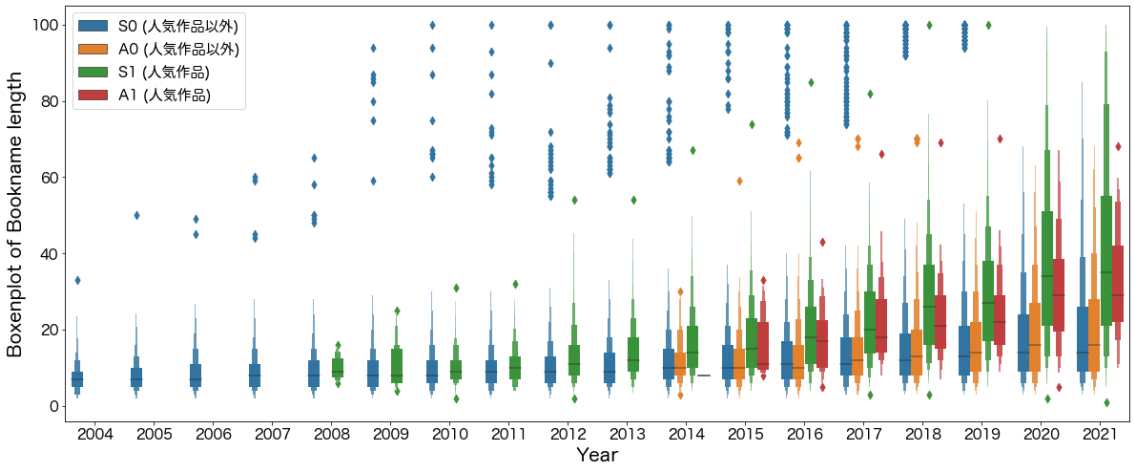


図 4.2　サイトごとの人気作品・それ以外の作品のタイトルの文字数の boxenplot

4.1.2「小説家になろう」

4.1.1 では、全体のトレンドを分析し考察を試みた。この節からは、各ジャンルに焦点を当て、さらなる細かい対比や分析を進めていく。前章でも述べたように、両方のサイトではジャンルの分け方が異なり、そのまま比較することは難しく思われるため、4.1.2 ではまず「小説家になろう」を対象に、ジャンルごとの文字数変化を描画し解析を試みる。次の節では、「アルファポリス」における人気のジャンル「ファンタジー」と「恋愛」を取り上げ、「小説家になろう」の結果と比較する。

図 4.3 では、6 つの図を提示している。上の 2 行 2 列はそれぞれファンタジー、恋愛、SF と文芸ジャンルにおけるタイトル文字数の時系列箱ひげ図である。同じように総合評価ポイント 10000 以下の作品をラベル「S0」、総合評価ポイント 10000 以上の人気作品をラベル「S1」で表す。また、それぞれのジャンルにおいて、色の濃い方が人気作品、つまり「S1」の状況を反映している一方で、色が浅い箱が人気作品以外のもの、つまり「S0」を表している。

また、3 行目における 2 つのグラフに関しては、左側が総合評価ポイント 10000 以下の作品（S0）を中心に、投稿年ごとのタイトルの文字数の中央値を折線で表示している。ジャンルによって異なる色とマークで示している。それに対して、右側が総合評価ポイント 10000 以上の作品（S1）の状況を表している。

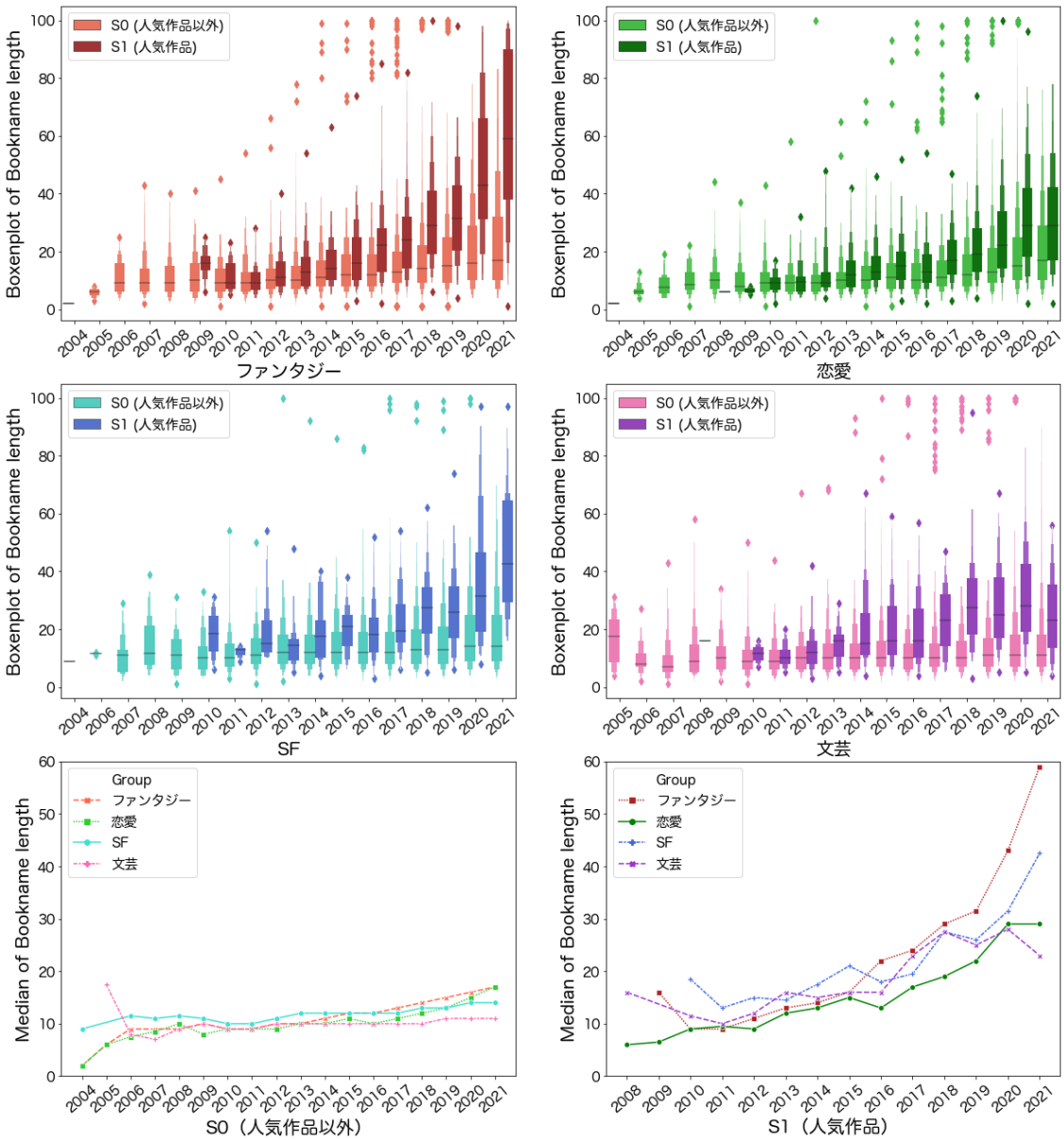


図 4.3　ジャンルごとのタイトルの文字数の boxenplot（「小説家になろう」）

図 4.3 からわかるように、どのジャンルにおいても基本的には年々伸びる傾向があり、人気作品の方が一般に長いタイトルが多い。特にファンタジーの方ではその傾向が極めて強く、2020 年からは一気に上がった様子も見られ、2021 年では人気作品の中央値が 60 にも上り、箱も大幅に伸びており、制限まで長くつけるタイトルが大幅に増加していると考えられる。SF も類似した傾向を成し遂げているが、ファンタジーほどではない。恋愛と文芸の方は伸び率が比較的に緩い。さらに文芸ジャンルはここ数年において下降が見られ、安定し

た成長を続けているとは言えない。ただし、激しい落下でもないため、今後観察し続ける必要があると思われる。

4.1.3「アルファポリス」

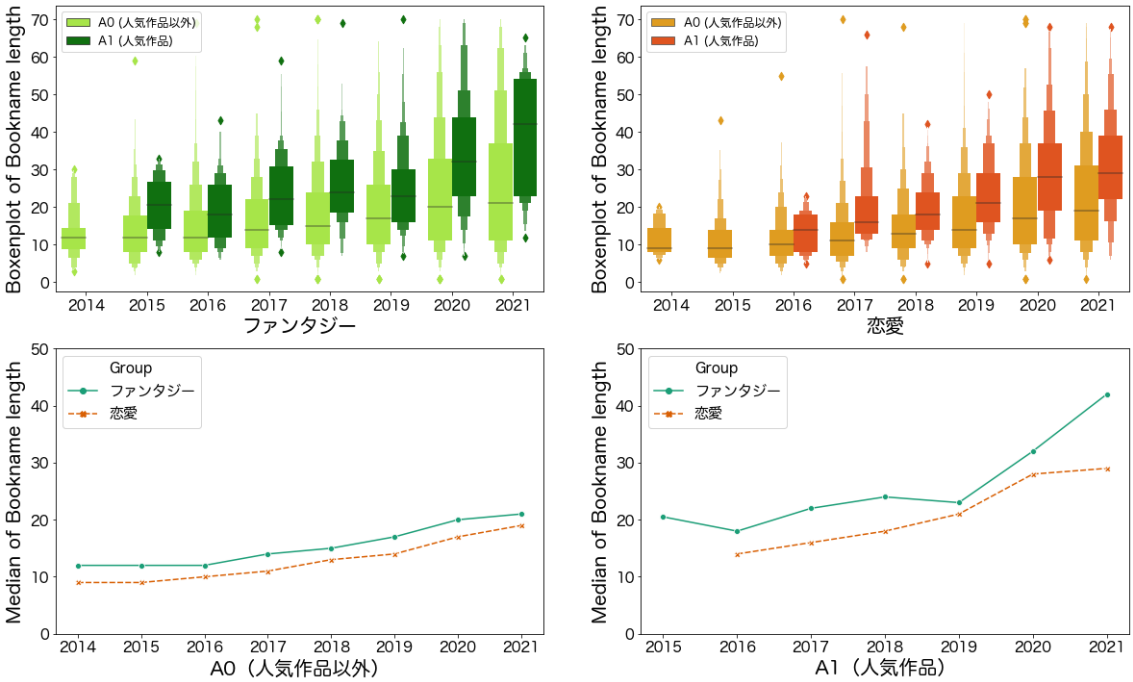


図 4.4　ジャンルごとのタイトルの文字数の boxenplot（「アルファポリス」）

一方で、図 4.4 は「アルファポリス」の方のファンタジーと恋愛ジャンルを取り上げて同じ手法で比較を行った。文芸と SF などは人気作品の数が少ないため比較しづらいと思われ、ここでは扱わないことにした。

図 4.4 から分かるように、どちらのジャンルにおいても人気作品の方がより中央値が高い。ファンタジーでは急速に伸びているのに対して、恋愛ジャンルでは緩やかに成長している。この趨勢は「小説家になろう」で得られた結果とほぼ同じである。要するに、投稿サイトは関係なく、ウェブ小説におけるタイトルの特徴に関しては長く書かれる傾向にあり、さらにファンタジーがとりわけ目立つと結論づけられる。その原因として考えられるのは、現在各サイトにおいて投稿作品数が爆発的に増加しているため、作品が読者に気づかれずに埋もれる可能性が非常に高い。クリックさせる手段、もしくは読者の注意を引くための工夫

なのではないかと思われる。ファンタジーが他のジャンルよりもその傾向が強いのは、競争が最も激しいからだと言えなくもない。

4.2 タイトルにおける語彙

この節では、タイトルに使われる言葉に焦点を当て、語彙の多様性検証と特徴語に対する考察を試みる。まずは前処理として必要な分かち書きについて説明する。

4.2.1 単語分割

日本語のテキストを機械に学習させるには前処理として単語分割、つまり分かち書きが必要である。つまり、英語のように単語の間にスペースを入れて区切る作業である。

この処理によって、単語頻度の計算や様々な機械学習手法を適応することができる。日本語の分かち書きは一般的に形態素解析といわれ、たくさんのツールも存在しているが、最も主流であるのは MeCab だとされている。

MeCab は京都大学情報学研究科—日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジンである。特徴としては、辞書やコーパスに依存しない汎用的な設計を基本方針とし、パラメータの推定に条件付き確率場（CRF）を用いており、高い解析精度を実現させている。同程度の解析精度を持つ Chasen と比べ、処理速度が 3、4 ほど速い。また近年では、Sudachipy のような専門家による語彙の追加や調整がなされた高品質な辞書を搭載した形態素解析ツールも次々と登場してきたが、大規模なデータを処理する場合、MeCab の方が抜群的に速いという点で評価される。

MeCab でデフォルトとして使われる標準的な形態素解析辞書は IPAdic であるが、短単位と階層的見出し構造に基づき開発された UniDic を辞書にすることがより推奨されるようになっている。小木曾（2010）では、新聞・文学作品・ブログの 3 ジャンルを対象として、適合率[20]、再現率[21]と F 値[22]で評価をした結果、いずれもよほど良い精度を見せているが、全て

[20] 正しく分割できた単語の数を c とする。誤って分割された単語を e とする。正解数を N とする。適合率とは、c/N の値であり、高いほど良いとされている。
[21] 再現率は、c/(c+e)の値であり、高いほど良いとされている。
[22] F 値は(2*適合率*再現率)/(適合率+再現率)の値であり、高いほど良いとされている

のタスクにおいて UniDic の解析精度がほかの辞書に勝っているということを結論付けられた。そのため、本研究では日本語の分かち書きツールとして UniDic 辞書を搭載した MeCab を導入し、Python で単語分割作業を実行する。

ここではタイトルの一例を示す：

「**異世界に召喚されわがまま言ったらガチャのスキルをもらった**」

単語のみの分かち書きを行えば、このように語彙の間にスペースが入った文になる。

「**異 世界 に 召喚 さ れ わがまま 言っ たら ガチャ の スキル を もらっ た**」

品詞タグづけを付与したまま出力することもできる。表 4.1 ではそれぞれの単語に付与された品詞情報の一部を表示している。

表 4.1　MeCab による品詞タグ付分ち書き[23]

語彙	大分類	中分類	小分類	活用型	活用形	読み	レンマ
異	接頭辞					イ	異
世界	名詞	普通名詞	一般			セカイ	世界
に	助詞	格助詞				ニ	に
召喚	名詞	普通名詞	サ変可能			ショウカ	召喚
さ	動詞	非自立可能		サ行変格	未然形-サ	スル	為る
れ	助動詞			助動詞-レ	連用形-一般	レル	れる
わ が ま	名詞	普通名詞	形状詞可能			ワガママ	我が儘
言っ	動詞	一般		五段-ワア	連用形-促音便	イウ	言う
たら	助動詞			助動詞-タ	仮定形-一般	タ	た
ガチャ	副詞					ガチャ	がちゃ
の	助詞	格助詞				ノ	の
スキル	名詞	普通名詞	一般			スキル	ス キ ル -
を	助詞	格助詞				ヲ	を
もらっ	動詞	非自立可能		五段-ワア	連用形-促音便	モラウ	貰う
た	助動詞			助動詞-タ	連体形-一般	タ	た

分かち書きにより単語の計算ができるようになる。さらに、特定の品詞を抽出することや取り除くこともできるため、必要な情報のみ確保することも可能である。

[23] https://ccd.ninjal.ac.jp/unidic/UNIDIC_manual.pdf を参照

4.2.2 語彙多様性

発話や文章の質を分析する際、語彙の多様性という項目がある。これは言い換えれば語彙の豊富性、つまりどれほど多様な語が使われているかを計算するものであり、延べ語数における異なり語数の割合（英語では Type Token Ratio、略して TTR と呼ぶ）を使用するのが一般的である。McIntyre and Walker（2019）[24]によると、長文では機能語がたくさん使われるため、TTR の値が非常に小さくなる恐れがある。よって、STTR（Standardised Type Token Ratio）の方がより適切な評価指標になる。STTR は、文書を 1000 語（数は自分で設定可能）ごとに分けてそれぞれの TTR を計算し、平均値を取る方法である。STTR の結果をもとに、語彙の多様性を判定することが可能である。

この節では、カテゴリーごとの語彙の多様性に変化があるかどうかを検討したい。それぞれ単一の小説のタイトルを対象とするのではなく、カテゴリー間の相違を考察するため、作品のタイトルを結合した上で、上述した STTR の手法を使用し、語彙の多様性を数値で比較する。

表 4.2 は、投稿年ごとのタイトルにおける語彙の豊富性を見るために集計した延べ語数と異なり語数の表である。できる限り多くのカテゴリーが比較できるように、1000 語のような大きい数ではなく、100 語ごとの TTR の平均をとって考察する（以下全ての語彙多様性分析に適用する）。人気作品では 2008 と 2009 年の延べ語数が 100 語以下であるため、分析対象から外すことになるが、作品数自体が少なく、個体差に左右されやすいと考えられるため、分析対象から除外することは妥当であると考えられる。

表 4.2　延べ語数（Token）と異なり語数（Type）の集計表

	全作品			人気作品		
Year	**作品数**	**Token**	**Type**	**作品数**	**Token**	**Type**
2004	199	778	506			
2005	747	3,286	1,481			
2006	1,925	9,028	3,170			
2007	3,959	18,775	5,361			
2008	5,221	26,262	6,769	3	21	20
2009	6,318	31,577	7,892	17	93	75
2010	9,573	50,080	10,554	56	322	232
2011	14,262	78,193	13,838	125	747	436

[24] McIntyre, D. and Walker, B. (2019) *Corpus Stylistics: Theory and practice*, Edinburgh University. p.126

2012	19,740	113,680	17,147	242	1,696	795
2013	23,934	147,222	19,655	336	2,742	1,052
2014	29,884	197,360	22,970	392	3,775	1,297
2015	34,410	241,443	25,016	505	5,203	1,584
2016	39,666	302,121	27,252	663	7,716	2,016
2017	45,979	376,741	29,322	696	9,478	2,249
2018	47,393	430,050	30,153	866	14,212	2,846
2019	48,703	483,562	31,504	1,013	17,307	3,136
2020	53,325	621,878	34,076	1,357	31,350	4,148
2021	40,426	522,779	30,193	795	20,059	3,200

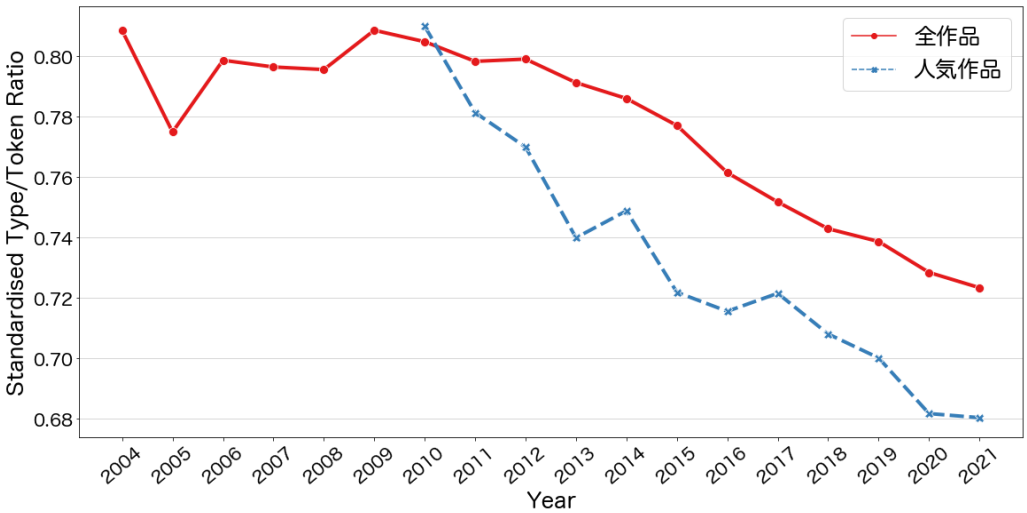


図 4.5　語彙多様性の時系列関係図[25]

図 4.5 の折れ線グラフでは全ての作品と人気作品に分けて、投稿年ごとの STTR の数値結果を表している。全体として多少の上下変動はありつつも、年々数値が下がっていくことが明らかであり、語彙の多様性が低下してきていると結論づけられる。その上、人気作品の方では、全作品と比べて数値がさらに低いことが見てとれる。つまり、人気作品はより共通性が高く、特定のトピックや言葉遣いが非常に好まれることが考えられる。

[25] McIntyre, D. and Walker, B. (2019) *Corpus Stylistics: Theory and practice*, Edinburgh University. p.146 では Dickens の散文コーパスと 19 世紀のフィクションコーパスとの STTR の比較研究が行われたが、どちらも 0.43 に近い値であり、ここで得られた最低値 0.68 より遥に低いが、区切りの仕方（McIntyre 2019 では 1000 語あたりの平均 TTR を計算している）が異なるためこれらの結果をもとにウェブ小説のタイトルは散文とフィクションより語彙多様性が低いと論ずることは不適切である。

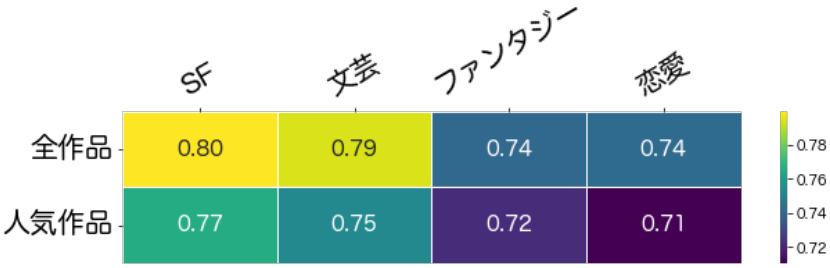


図 4.6　ジャンルと語彙多様性のヒートマップ関係図

図 4.6 ではジャンルに分けて、各ジャンルにおける語彙の使用状況を反映している。人気作品の数値が全作品より全般的に低いのは予想通りである。ジャンル間の相違では、やはり SF と文芸作品と比べると、ファンタジーと恋愛ジャンルの語彙多様性が非常に低い。ファンタジーと恋愛ジャンルのタイトルにおいて重複して使われる単語が極めて多いことが推測できる。

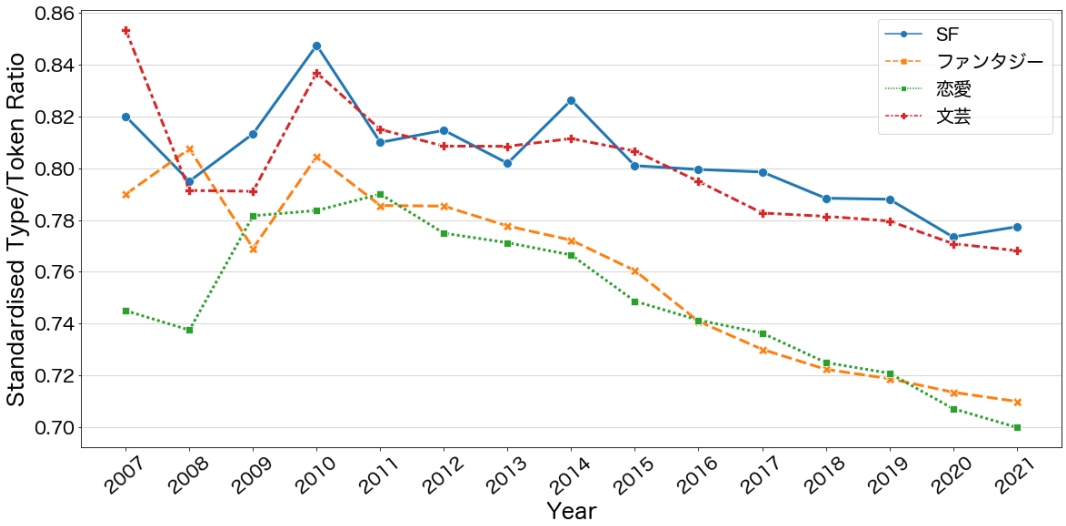


図 4.7　ジャンル別の語彙多様性の時系列関係図（「小説家になろう」）

図 4.7 と図 4.8[26]ではさらに投稿サイトに分けて総合的に語彙の多様性数値をグラフで可視化している。図 4.7 は「小説家になろう」を対象に、それぞれのジャンルにおいて、年ごとの語彙多様性を表示している。2007 年から 2010 年まで数値は上がり下がりしているが、

[26] 図 4.8 の文芸に含まれるのは「アルファポリス」のキャラ文芸、ホラー、ミステリー、ライト文芸、歴史・時代、現代文学、青春の 7 つのジャンルである。図 4.11 でこの 7 つが「小説家になろう」の文芸ジャンルに近いため、1 つの大ジャンルとして見なす。

2011 年以降は SF にいくつか上昇が見られるものの、年々全般的に減少していることが明らかである。特に、ファンタジーと恋愛ジャンルの低下が著しい。

図 4.8 では、「アルファポリス」側の語彙多様性状況を示している。恋愛ジャンルに顕著な低下が観察されたのは図 4.7 との共通点である。ファンタジーも減少傾向にあるが、図 4.7 の 2015 年以降の時系列の変化と比べると、減少幅が非常に緩やかである。SF と文芸ジャンルでは、一定した増加もしくは減少が観察されず、上下していることが明らかである。図 4.7 と合わせて見ると、SF の語彙多様性は年との関係が比較的薄いように考えられる。

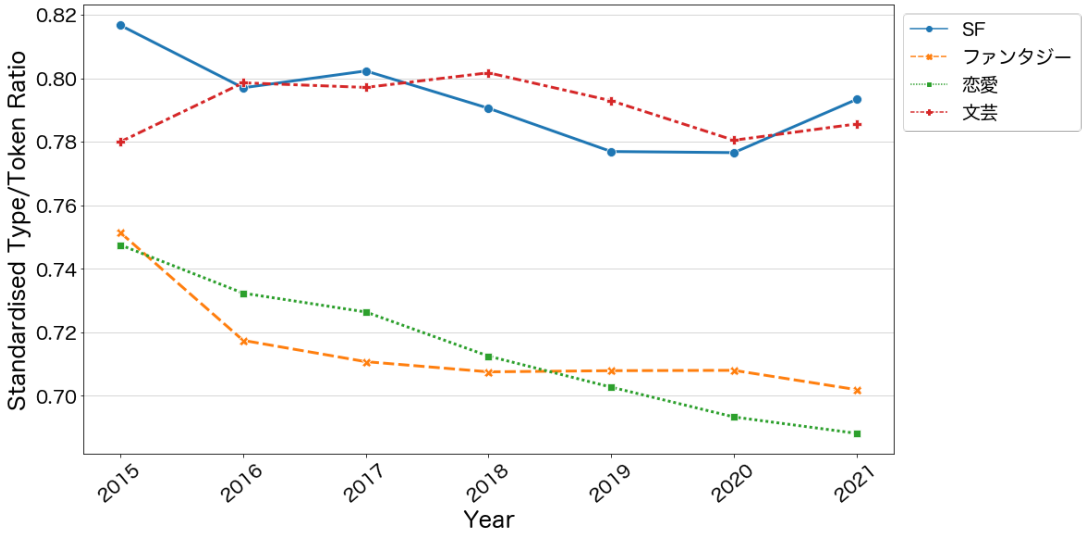


図 4.8　ジャンル別の語彙多様性の時系列関係図（「アルファポリス」）

ウェブ小説のタイトルに使われる語彙の使用傾向をまとめると、語彙の多様性が全体として減少しつつあり、ファンタジーと恋愛においてはその傾向が顕著である。また、ポイントが高い人気作品群ではさらにタイトルの使用語彙の共通性が高いことが言える。

4.2.3 コレスポンデンス分析

4.2.3.1 全体

この節では、タイトルで使われる語彙の使用頻度に重点を置き、カテゴリーごとの言語特徴を比較することを狙う。まずは、作品ごとの単語の出現頻度を統計する。表 4.3 はそれぞれの小説タイトルにおける単語の生起度を表す語彙頻度プロファイルの一部である。行ごとに作品を並べ、最初の 3 列では作品の情報、つまり id、ジャンルと投稿年を表示している。4 列目からは語彙情報を表し、使われる頻度を埋め込んでいる。実際行項目は 44 万以

上、語彙の列項目も 4 万以上あるが、ここでは説明のために一部のみ抽出している。表中の数字はそれぞれの作品における当該単語の出現頻度を表している。タイトルは短い文であるため、標準化は行わずそのまま粗頻度[27]を使用する。

表 4.3　テキストファイルごとの語彙頻度

ID	Group	Year	『	』	【	】	世界	令嬢	俺	冒険	勇者	君	日
1	ファンタジー	2021	0	0	0	0	0	0	0	0	0	0	0
2	SF	2021	0	0	0	0	0	0	0	0	0	0	0
3	恋愛	2021	0	0	0	0	1	0	0	0	0	0	0
4	ファンタジー	2021	0	0	0	0	0	0	0	0	0	0	0
5	ファンタジー	2021	0	0	0	0	0	0	1	1	0	0	0
6	ファンタジー	2021	1	1	0	0	0	0	0	0	0	0	0
7	ファンタジー	2021	0	0	0	0	1	0	0	0	0	0	0
8	ファンタジー	2021	0	0	0	0	0	0	0	0	1	0	0
9	恋愛	2021	0	0	0	0	0	0	0	0	0	0	0
10	ファンタジー	2021	0	0	0	0	0	0	0	1	0	0	0
11	恋愛	2021	0	0	0	0	0	1	0	0	0	0	0
12	恋愛	2021	0	0	0	0	0	0	0	0	0	1	1
13	ファンタジー	2021	0	0	1	1	0	0	0	1	0	0	0
14	恋愛	2021	0	0	0	0	0	0	0	0	0	0	0
15	文芸	2019	0	0	0	0	0	0	0	0	0	0	0

もちろん目的としては個々の作品ではなく、年度とジャンルに分けて、同じカテゴリーに当たる全ての作品に対して語彙の使用頻度の平均を計算し、カテゴリー間の類似性や差異を明らかにすることである。列項目に関しては、全体における使用頻度の上位 100 位までの語彙項目のみを分析対象とする。

図 4.9　上位単語のワードクラウド（機能語あり）

[27] 粗頻度とはコーパスから取得した生の頻度のことで、何の調整も加えていない頻度を指す。

図 4.9 では出現頻度上位 100 の単語をワードクラウド[28]で表したものであり、単語の割合が高いほど大きく表示される特徴がある。「の」「は」「に」「と」「を」「で」「が」「た」「て」などの助詞や助動詞が主要単語として最も使われることも明らかである。ただし、単語のみの分析であれば、文の構成に欠かせない機能語より内容語に注目したいため、助詞と助動詞、さらに句読点や波線を含む記号も取り除くことが妥当だと考えられる。よって、助詞、助動詞と記号が含まれない上位 100 語に限定し、図 4.10 が示したように内容語を中心に分析を進める方針である。異世界転生というモチーフが流行しているということは既に明らかである。

図 4.10　上位単語のワードクラウド（機能語なし）

一方で、行項目に関して、カテゴリーとして特徴を捉えたいため、数が少なすぎるとカテゴリーの特徴を反映しているとは言いづらいのではないかと思われる。よって、前処理として作品数の下限を 10 にし、10 件もないカテゴリーであれば取り除くことにする。最終的には 90（カテゴリー）*100（語彙項目）の表ができた。その一部を取り上げて表 4.4 にまとめた。最初の 3 列はカテゴリーを表し、例えば 1 行目は 2014 年の「小説家になろう」の SF ジャンルにおける各語彙項目の平均使用状況を意味している。表の情報を目視で分析することは容易ではないため、視覚化したデータをもとに分析を進めることとする。ここでは、コレスポンデンス分析を取り上げる。

表 4.4　カテゴリーごとの語彙割合

Site	Group	Year	ゲーム	スキル	ダンジョン	チート	ライフ	世界	中	人	人生	令嬢
S	SF	2014	0.027	0.000	0.003	0.005	0.002	0.085	0.008	0.014	0.003	0.000
S	SF	2015	0.036	0.005	0.003	0.006	0.003	0.082	0.012	0.013	0.002	0.000

[28] ワードクラウドとは単語の頻度を可視化する方法である。頻度が高いほど大きく、低いほど小さく表示される特性がある。

S	SF	2016	0.035	0.002	0.004	0.008	0.005	0.094	0.011	0.014	0.001	0.001
S	SF	2017	0.038	0.005	0.002	0.005	0.008	0.097	0.011	0.014	0.002	0.000
S	SF	2018	0.044	0.006	0.003	0.009	0.009	0.116	0.012	0.015	0.003	0.002
S	SF	2019	0.041	0.004	0.002	0.008	0.005	0.113	0.011	0.023	0.003	0.002
S	SF	2020	0.065	0.015	0.002	0.010	0.008	0.125	0.013	0.023	0.005	0.003
S	SF	2021	0.061	0.009	0.002	0.007	0.009	0.112	0.010	0.027	0.006	0.003
A	恋愛	2014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	恋愛	2015	0.016	0.000	0.000	0.000	0.000	0.026	0.010	0.026	0.005	0.010
A	恋愛	2016	0.013	0.000	0.000	0.000	0.005	0.039	0.002	0.016	0.003	0.026
A	恋愛	2017	0.012	0.001	0.000	0.008	0.001	0.048	0.010	0.016	0.004	0.055
A	恋愛	2018	0.018	0.001	0.000	0.003	0.004	0.050	0.013	0.023	0.006	0.098
A	恋愛	2019	0.023	0.001	0.001	0.003	0.004	0.066	0.013	0.021	0.007	0.141
A	恋愛	2020	0.022	0.003	0.001	0.003	0.005	0.060	0.017	0.024	0.011	0.163

コレスポンデンス分析、いわゆる対応分析は、林知己夫により提唱された「数量化法 3 類」とほぼ同じ原理による解析手法である。コレスポンデンス分析の基本の考え方は、クロス集計表を構成する行項目と列項目の相関が最大になるように計算されて、近い項目が隣り合うように並べ替えるということである。

目的は、分析対象としている多数の変数の持つ情報を、できるだけ少ない次元に圧縮し、2 次元か 3 次元の散布図にプロットし、位置関係を観察することによって、各カテゴリー間の関係と類似性を視覚的に把握することである。

類似した多変量解析手法として主成分分析も挙げられるが、大きな違いは主成分分析が量的データ（間隔尺度、比率尺度）を扱うことが多く、質的データ（名義尺度、順序尺度）の場合であればコレスポンデンス分析の方がよりふさわしいとされている。田畑（2004a）、田畑（2004b）と水本（2009）では、カテゴリーデータ、つまり類似関係や差異を見るにはコレスポンデンス分析が最も有効であることを証明された。そこで、コレスポンデンス分析の手法を用いて表 4.4 で示されたデータセットを可視化することを試みる。コレスポンデンス分析を Python で行うためには、mca ライブラリをインポートする必要がある。

まずは、「小説家になろう」と「アルファポリス」のそれぞれのジャンル間の関係性を考察しようと思う。mca の実行および散布図の作画を行った結果、最も寄与率[29]の高い Factor1

[29] データ全体の変化に対して、その構成要素ではるデータがどれくらい影響したかを表す指標である。

（60.08%）と Factor2（30.59%）[30]の得点を二次元散布図に配置したものが図 4.11 である。合わせて 90%ほどの寄与率があり、カテゴリー間の差異をほぼ解釈できていると思われる。

図 4.11 では、それぞれのラベルをより見やすくするために、ジャンルを基準に色分けしてマークをつけている。「S」と「A」はそれぞれ「小説家になろう」と「アルファポリス」のことを指している。変数の頻度パターンの近いカテゴリー（ラベル）が近く布置され、頻度パターンの異なるカテゴリー（ラベル）同士は遠く布置される特徴がある。特に原点から反対の方向に配置されるラベルは相違があり、離れるほど差異が大きいとされている。

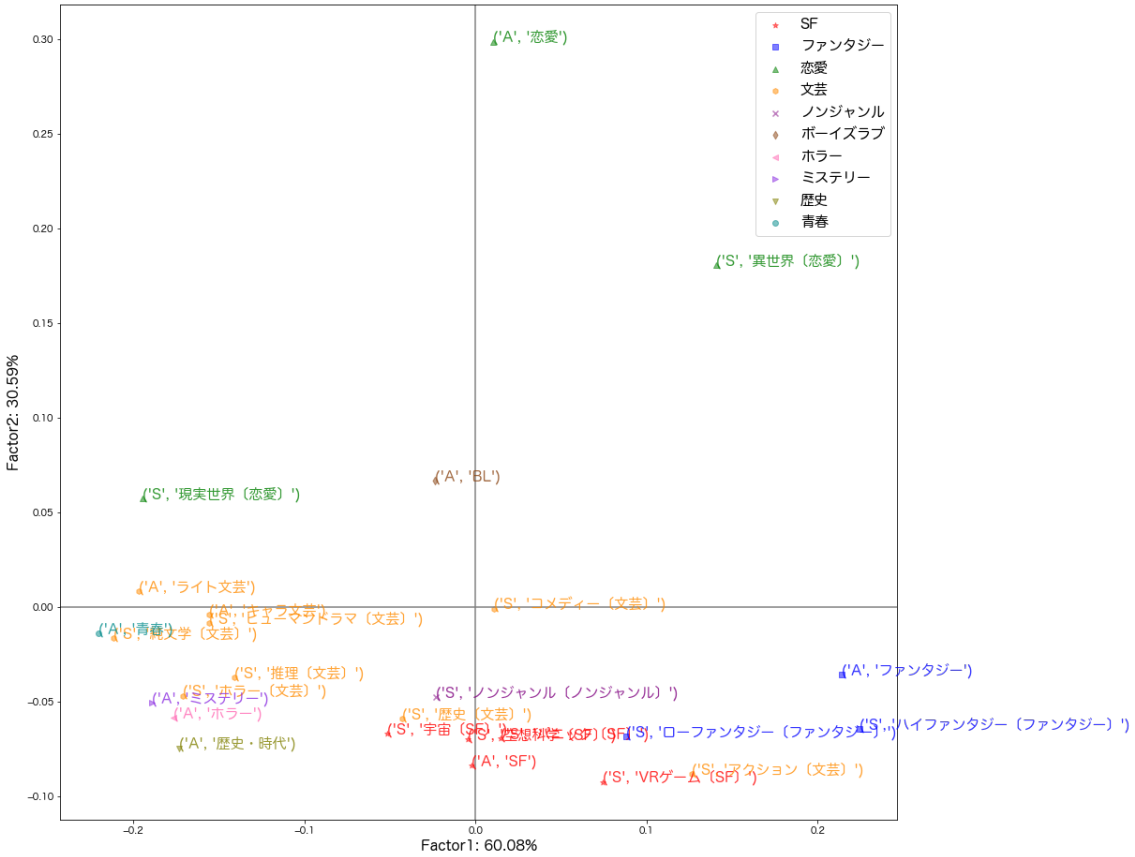


図 4.11　内容語に基づく対応関係図（ジャンルによる色分け）

図 4.11 から明らかなように、ファンタジーと SF に属する下位ジャンル（赤と青色のラベル）は右下に集中している。恋愛ジャンル（緑色のラベル）に関しては、異世界と現実世界によって左右に綺麗に分かれている。同じく恋愛ジャンルではあるが、はっきりと相違が観

[30] Factor1 は表 4.3 の行項目と列項目の相関が最大になるように行われた 1 回目の並び替えである。ただし、1 回の並び替えだけでは全てのパターンを表しきれないため、2 回以上が必要となる。Factor2 は 2 回目の並び替えである。Factor1 と Factor2 のパターンをそれぞれ横軸と縦軸で表したのは図 4.11 である。

察できる。「アルファポリス」の恋愛ジャンルは（'S','異世界[恋愛]'）[31]に近いため、どちらかといえば異世界の成分が高いこともわかった。BL（ボーイズラブ）は比較的に現実世界の恋愛に近寄っている。

文芸ジャンルも概ね左下に偏っており、「アルファポリス」の「青春」「ミステリー」「ホラー」「歴史・時代」も文芸とみなして差し支えないと判断できる。ただし、例外として「小説家になろう」の「アクション」ジャンルは右下のファンタジーの近くに配置されていることがわかる。つまり、「アクション」ジャンルは文芸というより内容的にはファンタジーに近いことが言えるであろう。

続いては、さらに年代に分けて具体的に考察する。

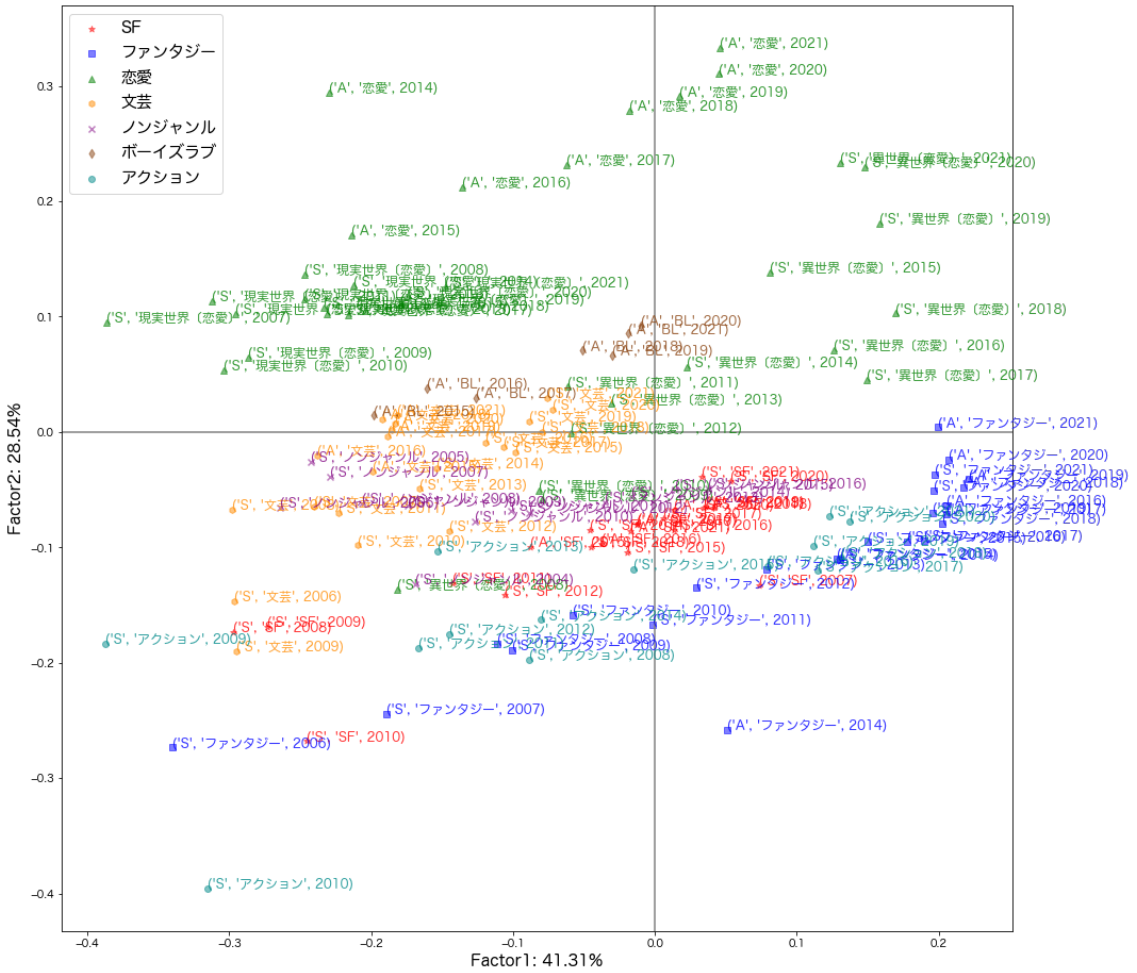


図 4.12　内容語に基づく対応関係図（ジャンルによる色分け）

[31] 「小説家になろう」の異世界恋愛ジャンル

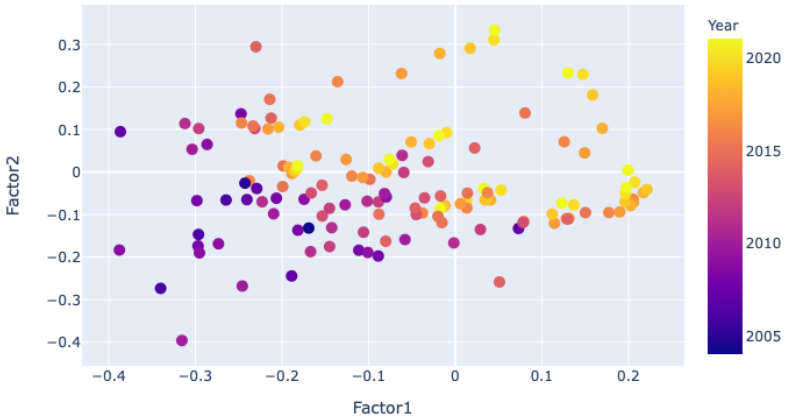


図 4.13　内容語に基づく対応関係図（投稿年による色分け）

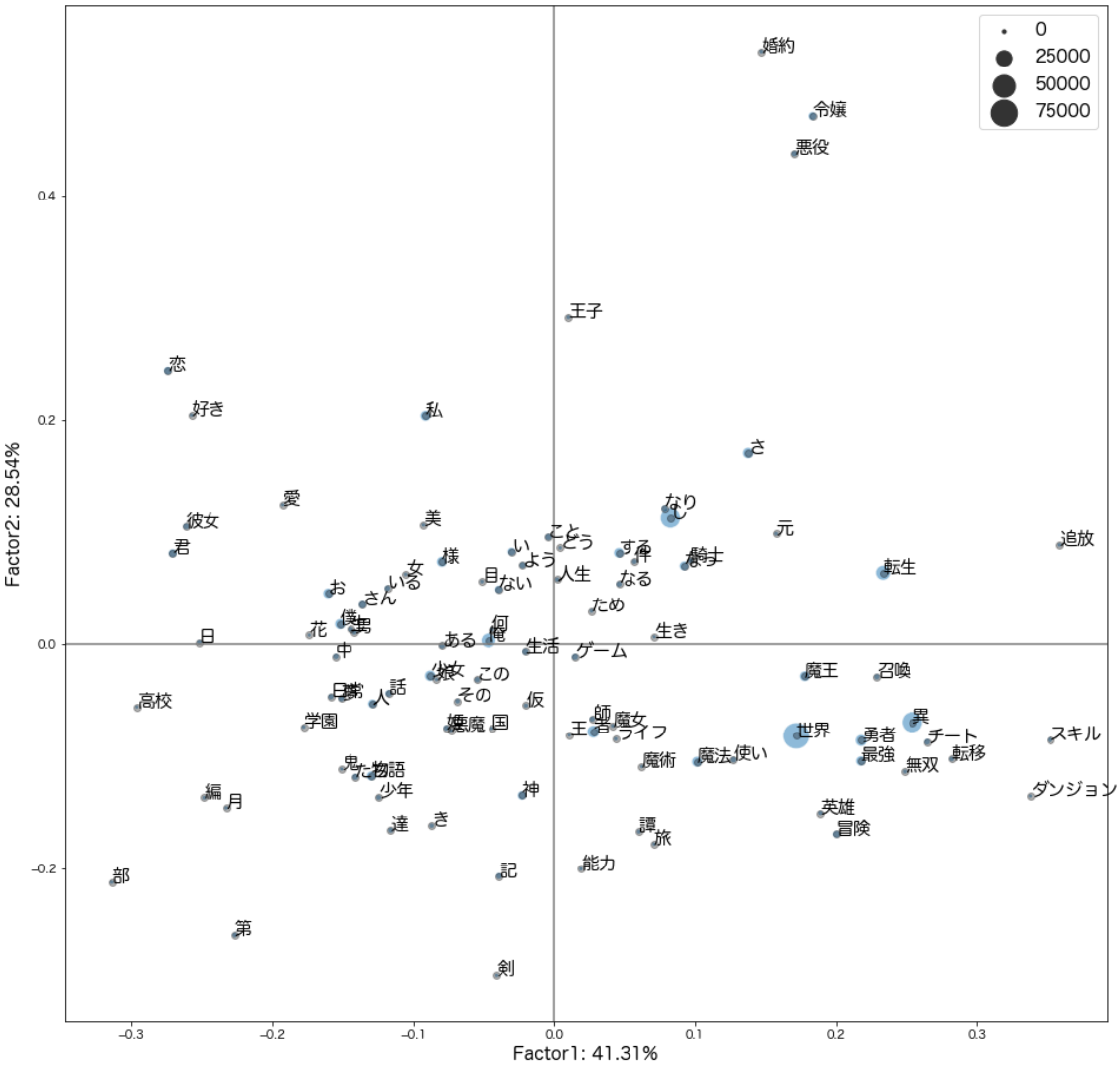


図 4.14　内容語に基づく対応関係図（語彙のみ）

図 4.12 は、カテゴリー間（サイト、ジャンル、投稿年でグループ化したもの）の関係を可視化した図であり、ジャンルによってラベルを色分けしている。「青春」「ミステリー」「ホラー」「歴史・時代」は全て文芸としてカテゴライズしたが、アクションは単独のジャンルとして取り出している。図 4.13 は年代関係を考察するためにそれぞれのラベルを年代順に濃い紫色から薄い黄色まで分けて表示している。

図 4.14 では、それぞれの変数、つまり単語間の関係を視覚化し、点の大きさでそれぞれの頻度をウェイトとして具現化している。図 4.12、4.13 と 4.14 を照らし合わせて見れば、それぞれのカテゴリーを代表する語彙項目も一目瞭然である。

図 4.14 では、右に最頻語である「異」「世界」が位置づけられ、Factor1(41.31%)を解釈する横軸は物語の背景、つまり舞台との関連性が強いと思われる。左には現実恋愛と文芸のような現実を舞台とするジャンルが多数配置されているのに対して、右には異世界恋愛、SF とファンタジーがほとんどであり、異世界という設定に大きく影響されていると考えられる。また、それぞれのジャンルの投稿年からすれば左に配置されているのは昔の作品が多いが、右にいくほど投稿年が近いということも言えなくはない。ノンジャンルが投稿年の変化とともに右へと移動することからもその傾向が推察できるのではないか。一方で、Factor2（28.54%）を表す縦軸には、上がほぼ恋愛ジャンルのみであり、つまり恋愛が主題であるかどうかに大きく左右されているという解釈ができる。このように、タイトルの語彙の使用状況でそれぞれのジャンルの特性を表している。

また、内容語に注目すると：

(1). ファンタジーと異世界恋愛ジャンルに関しては投稿年によって大きく変化していることが見てとれる。ファンタジーの方では英雄や勇者など冒険物語からチート能力を獲得して強者になるような転生ファンタジーへと変化し、また最近では追放ファンタジーへと流行が推移していることが分かる。異世界恋愛の方では婚約をめぐる展開、また悪役令嬢キャラが近年非常に好まれるようになったと言える。

(2). 主要単語の中に、「君」「僕」「私」「俺」のような代名詞が見られ、恋愛作品や文芸作品に特徴的に使われることがわかる。「君」「私」の位置からすれば、恋愛ジャンルにおいて多く使われることが推測できる。特に初期の現実恋愛作品のタイトルには「君」という二人称が愛用されていることも言える。

(3). ウェブ小説初期の作品が集まったと思われる左下側には、「第」「部」「編」のような接頭辞と接尾辞が特徴的である。表 4.5 における「部」の使い方の頻度統計から分かるように、「部」に関してはいくつか部活を指している「部」も存在するが、連続シリーズのことを指す「部」が格段に多く、「第」との関連性が見てとれる。

表 4.5　「部」の n-gram

Rank	Bigram	Frequency	Trigram	Frequency
1	一 部	245	第 一 部	245
2	二 部	182	第 二 部	172
3	研究 部	110	第 三 部	83
4	文芸 部	100	第 1 部	83
5	三 部	97	第 2 部	58
6	1 部	94	吹奏 楽 部	43
7	帰宅 部	89	学園 高等 部	39
8	野球 部	80	第 四 部	33
9	高等 部	54	の 文芸 部	18
10	演劇 部	50	会 執行 部	17

「編」に関しては、実例から見れば物語の追加説明としてタイトルの末尾で使われているのがほとんどである。

「編」の実例：

1. アンリーズナブル：異世界編～ダークでシリアスな異世界で超辛いけど、頑張って生き延びます～
2. 六感　～辺境の編～
3. DeadMan'sHand―臥龍転生編―
4. サイコパス一家が遊園地に行ったら（番外編)
5. 能面顔の悪役令嬢　IF　～あり得る１つの未来　青石兄編～
6. れっつ世界を救おう！　挿話・番外編
7. すれ違う恋の行方〈大学編〉
8. 赤城山の月(純白の S と共に・ココロノツバサ番外編)
9. 氷の彩光、砂の紅-Septarche1 リリーシア編
10. なぜ僕はこんな怖いパーティの一員になっているのでしょうか？　（後編）（継続用）

表 4.6 では「編」の頻出パターンを列挙している。全体的に言えば「番外編」が最も使われていることがわかる。「第」「部」「編」の使い方により、連続シリーズに分けて書くことが初期作品の特徴なのではないかと考えられる。

表 4.6　「編」の n-gram（¥t はスペースを意味する）

Rank	Bigram	Frequency	Trigram	Frequency
1	番外 編	940	¥t 番外 編	235
2	生 編	57	異 世界 編	23
3	特別 編	50	年 生 編	23
4	世界 編	34	の 番外 編	21
5	過去 編	33	¥t 特別 編	19
6	完結 編	33	シリーズ 中学 編	19
7	中学 編	26	神 番外 編	13
8	者 編	20	小学 生 編	11
9	学園 編	20	シリーズ 番外 編	11
10	総集 編	17	高校 生 編	9

また、図 4.14 の中央下側には「記」「譚」のような接尾辞が見られ、実際の意味と使い方は左に位置している「物語」とほぼ同様であるが、「記」と「譚」がよりファンタジーを題材とする作品のタイトル（近年ではない）に頻出していることが明らかである。実例からも分かるように、冒険や戦争など事件の締めくくりとして使われることが多い。

「記」の実例：

1. もう当て馬になるのはごめんです！～巻き戻り令嬢のやり直し奮闘記～
2. トルコ旅行記　兄と妹のちょっとだけクレージーなジャーニー
3. モブの奮闘記
4. 聖輪奇聞・転生記！～神を導きし救世主～
5. 異世界冒険記『ストレージ・ドミニオン』
6. 大戦乱記
7. 長宗我部転生記戦国オタクは２度転生する
8. 斬月記
9. ダンジョン討伐記
10. キモオタニートの異世界開拓記！

「譚」の実例 ：

1. エボルブルスの瞳―特殊事案対策課特命係傀異譚―
2. 白蓮伝「マナ無し冒険者と八百屋の娘の冒険譚」
3. 【全話挿絵あり】彼女に魅入られた『最強の魔王／最強の勇者』である転生者の英雄譚
4. 盾持ち令嬢の英雄譚
5. チーム・イリスの事件譚
6. 異世界満喫冒険譚
7. 元伯爵令嬢の下克上と恋愛譚
8. 蒼ノ英雄譚　～最強旅団の魔滅記録～
9. 婚約破棄から始まるとんでも異世界冒険譚～黒猫の騎士とポンコツ姫～
10. キコリの異世界譚

表 4.7　「記」の n-gram

Rank	Bigram	Frequency	Trigram	Frequency	Four-gram	Frequency
1	奮闘 記	687	の 奮闘 記	194	異 世界 冒険 記	175
2	冒険 記	566	世界 冒険 記	186	異 世界 放浪 記	125
3	放浪 記	322	の 冒険 記	143	異 世界 奮闘 記	112
4	転生 記	302	世界 放浪 記	133	異 世界 転生 記	102
5	旅行 記	257	世界 奮闘 記	115	異 世界 旅行 記	88
6	漫遊 記	89	世界 転生 記	104	異 世界 漫遊 記	44
7	創世 記	84	世界 旅行 記	99	の 異 世界 記	36
8	建国 記	83	異 世界 記	63	異 世界 漂流 記	33
9	漂流 記	82	の 放浪 記	54	異 世界 生活 記	25
10	世界 記	72	世界 漫遊 記	44	異 世界 探検 記	18

表 4.8　「譚」の n-gram

Rank	Bigram	Frequency	Trigram	Frequency	Four-gram	Frequency
1	冒険 譚	945	の 英雄 譚	429	異 世界 冒険 譚	226
2	英雄 譚	826	の 冒険 譚	372	の 異 世界 譚	99
3	世界 譚	151	世界 冒険 譚	232	異 世界 英雄 譚	62
4	転生 譚	96	異 世界 譚	145	勇者 の 英雄 譚	44
5	復讐 譚	86	の 復讐 譚	65	勇者 の 冒険 譚	34
6	怪奇 譚	54	世界 英雄 譚	64	異 世界 転生 譚	28

7	怪異 譚	53	の 転生 譚	32	者 の 英雄 譚	25
8	前日 譚	36	世界 転生 譚	28	から 始まる 英雄 譚	25
9	後日 譚	33	始まる 英雄 譚	26	の 成り上がり 英雄	17
10	浪漫 譚	31	な 冒険 譚	24	少年 の 英雄 譚	17

表 4.7 と表 4.8 では「記」と「譚」のそれぞれの bigram, trigram と four-gram の頻出パターンを挙げている。表 4.7 と表 4.8 から明らかなように、「記」と「譚」の使い方は非常に類似しており、冒険に関しては両方とも使われている。英雄や勇者の物語に関しては「譚」が比較的に使われ、放浪、旅行、探検などは「記」が好まれる。異世界背景との関わりが深いことは明白である。

「物語」、および「物語」の少し上に位置している「話」は文芸作品のラベルに近いため、文芸作品に使われることが多いと考えられる。「話」に関しても、実例から分かるように対話を表すはなしというより、「物語」のように文の語尾に使われることが多い。

「物語」の実例：

1. 悠久のクシナダヒメ　「日本最古の異世界物語」　第二部
2. ハーフフリート ～わけあって半分人間やめました。これは私が世界を滅ぼすまでの物語です～
3. 夢物語
4. あの日、死んだ君へ届ける物語。
5. 悪女と蔑まれ宣告された婚約破棄から始まる逢瀬物語！　～運命の王子様は全員同性！？百合が咲いたり咲かなかったりするハーレムライフ！！～
6. 最強魔女が大魔王を名乗るまでの物語
7. シルベ=テルミチのチートなき、異世界ライフの物語
8. 親王様は元大魔法師～明治の宮様に転生した男の物語～戦は避けられるのか？
9. あなたとの愛をもう一度　～不惑女の恋物語～
10. ラールノダ国～人と神を紡ぐ王女と仲間の物語～

「話」の実例：

1. 【ストック１００話超え】軍師アルベルトは生き残りたい！　～とある転生者の異世界成り上がり戦記～

2. 迷宮攻略と領土防衛と特産品作りを一手に担っていた魔術師を追放した伯爵家の話

3. 転生して俺 TUEE 無双出来るはずが周りが強すぎるって話。

4. 吸血鬼な悪役令嬢に転生した私が SSS 級国際指名手配犯となり最強チートスキルで婚約破棄王子や元クラスメイトの勇者などを倒していくお話

5. 学園のアイドルな幼馴染の姉ちゃんと同居する話

6. 魔剣に転生したおっさんと剣が使えない僕　～スローライフに憧れる僕が魔王になるまでの話～

7. カラスとすずらん～意識低い系冒険者パーティの台所を預かっています～平凡なわたしと、平凡に生きられず幸せが分からない傷だらけの彼が出会って恋をして変わっていく、人生と心の話～

8. 魔王軍の書庫の主の僕が、転生して面食いアイドルと付き合う話。

9. 私の周りに現れるキャラたちの話

10. オバ魔女アジャルタ　～おばさん魔女アンジェリカ・アジャルタが英雄？になるお話し～

表 4.9 と表 4.10 は「物語」と「話」の上位パターンを示している。共通点としては「の物語」「の話」のように「の」と共起するパターンが最も多い。「物語」の方は異世界や恋と共起する頻度が高く、関連性が強いと認められる。一方で、「話」の方は何かしらの内容語との関連性がほぼ見られないが、Four-gram から明らかなように、動詞と同時に出現するパターンが頻用される。

表 4.9　「物語」の n-gram

Rank	Bigram	Frequency	Trigram	Frequency	Four-gram	Frequency
1	の 物語	3,144	異 世界 物語	268	の 異 世界 物語	122
2	恋 物語	355	の 恋 物語	170	異 世界 転生 物語	61
3	世界 物語	291	少女 の 物語	104	な 恋 の 物語	49
4	転生 物語	206	恋 の 物語	101	異 世界 転移 物語	38
5	な 物語	158	世界 の 物語	88	なる まで の 物語	34
6	ない 物語	140	まで の 物語	81	一 つ の 物語	32
7	夢 物語	127	勇者 の 物語	79	の 国 の 物語	26
8	た 物語	123	少年 の 物語	78	た 男 の 物語	22
9	学園 物語	92	から 始まる 物語	73	異 世界 冒険 物語	21
10	始まる 物語	82	たち の 物語	64	から 始まる 恋 物語	21

表 4.10　「話」の n-gram（¥t はスペースを意味する）

Rank	Bigram	Frequency	Trigram	Frequency	Four-gram	Frequency
1	の 話	1,672	の お 話	488	に なっ た 話	100
2	お 話	895	第 1 話	139	¥t 第 1 話	52
3	た 話	633	第 一 話	112	て しまっ た 話	41
4	する 話	370	なっ た 話	109	¥t 第 一 話	37
5	1 話	251	に なる 話	107	まで の お 話	32
6	一 話	165	し た 話	99	なる まで の 話	31
7	なる 話	124	まで の 話	86	転生 し た 話	29
8	な 話	120	する お 話	65	た 俺 の 話	28
9	れる 話	94	だけ の 話	62	て い た 話	27
10	ない 話	93	れ た 話	61	さ れ た 話	26

近年の作品において語尾の使い方も変化しているのではないかと考えられる。実際、図 4.14 の右上に位置した「件」は非常に注目に値する。「件」の具体例からすれば、助数詞としての使い方が見られず、意味としては事柄を指す「話」や「物語」と一致している。図 4.13 と図 4.14 から、「件」が近年の作品に頻繁に使われていると判断できる。実例からも見られるように、タイトルが全般的に長くなっている。

「件」の実例：

1. 空から女性が落ちてきた件～自殺するところだったのを助けたら学年のアイドルでした、でもどうやら生きるのが嫌のようです～

2. 【WEB版】うちの弟子がいつのまにか人類最強になっていて、なんの才能もない師匠の俺が、それを超える宇宙最強に誤認定されている件について

3. 俺の姉妹が鬼デレ過ぎていつまでも彼女が出来ない件。

4. 【9 月書籍・コミカライズ発売！】植物魔法チートでのんびり領主生活始めます～前世の知識を駆使して農業したら、逆転人生始まった件～

5. 時之螺旋階段　～王立魔道学院の卒業試験で洞窟に魔術剣を抜きに行ったら、初対面のはずが 33 日前に会ったと主張する少年が現れたが何も覚えていない件～

6. 転生したら才能があった件　～異世界行っても努力する～

7. 神秘斬滅【ルナイレイズ】の少女　～落ちこぼれな【僕】の封印を断ち切ったら、異世界最強覇王の【俺】が復活してしまった件について～

8. 異世界転生でチートを授かった俺、最弱劣等職なのに実は最強だけど目立ちたくないのでまったりスローライフをめざす～奴隷を買って魔法学園で底辺生活を送ってたらいつのまにか英雄視され美少女に囲まれてた件～

9. 隣に住む変態なお姉さんに愛されたいけど、年下の女友達のせいで全然うまくいかない件について

10. 「聖少年様（仮）」は「聖女」のようです！～勘違いで呼ばれたらしい私の力が本物だった件について～

表 4.11 の「件」の使い方から明らかなように、「件」はほぼ過去形の動詞と関連づけられている。これも近年のタイトルが長い文になっていることと関わりがあると推測できる。「記」「譚」「物語」などと比べ、長文との相性がいい「件」が近年愛用されていると思われる。「話」の使い方は多少「件」に類似しているが、日常物語の場合は「話」を使う傾向がある。

表 4.11　「件」の n-gram

Rank	Bigram	Frequency	Trigram	Frequency	Four-gram	Frequency
1	た 件	2,721	だっ た 件	749	に なっ た 件	358
2	な 件	240	なっ た 件	384	て い た 件	262
3	ない 件	215	い た 件	291	て しまっ た 件	158
4	いる 件	189	れ た 件	252	さ れ た 件	122
5	すぎる 件	127	し た 件	192	転生 し た 件	73
6	てる 件	63	て いる 件	177	て き た 件	51
7	する 件	35	しまっ た 件	164	れ て いる 件	45
8	の 件	35	て た 件	100	なっ て た 件	44
9	くる 件	34	き た 件	56	世界 だっ た 件	38
10	れる 件	30	られ た 件	50	し て いる 件	38

ここまで内容語を中心に分析を行ったが、記号の使い方にも変化があるのではないかと仮設し、記号を中心にコレスポンデンス分析を試みる。

図 4.15 と図 4.16 は出現頻度トップ 20 の記号を取り上げて描画した結果である。図 4.15 ではジャンルによって異なる色のマーカーを付けており、図 4.16 では投稿年の相違を紫色から黄色まで年代順に反映している。各ジャンルが混雑していることが明らかであり、ジャンルより、投稿年との関係が深いことが図 4.15 と図 4.16 から分かる。

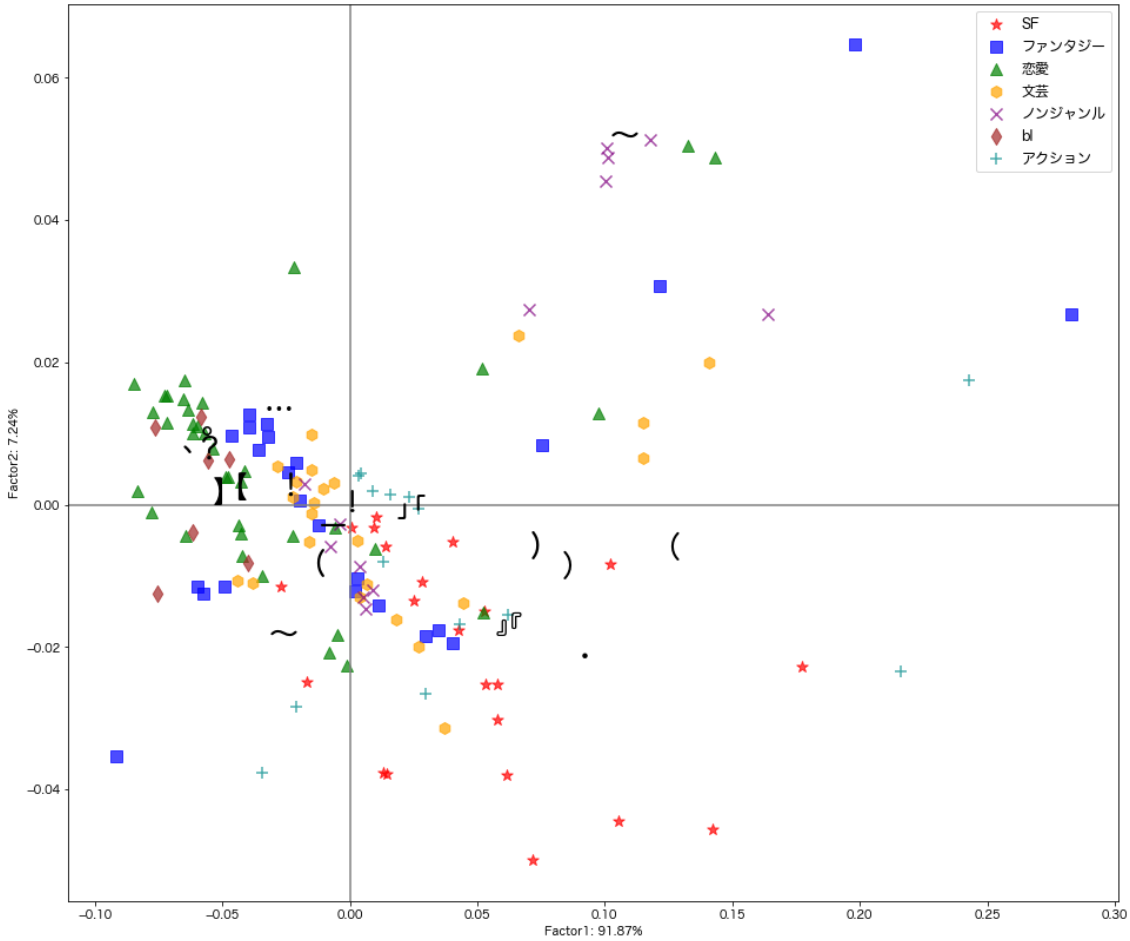


図 4.15　記号の対応関係図（ジャンルによる色分け）

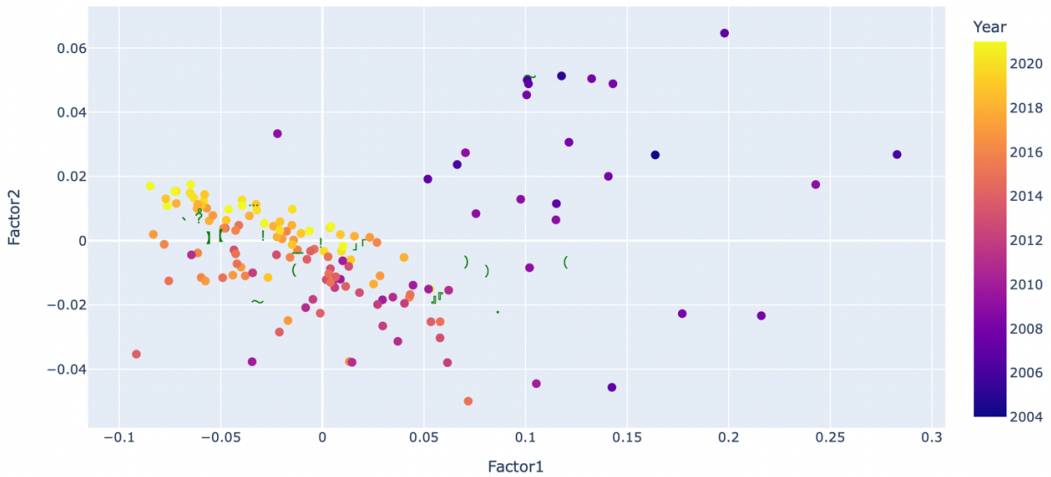


図 4.16　記号の対応関係図（投稿年による色分け）

SF のタイトルでは中点の「・」が比較的に使われていることが図 4.15 から分かる。図 4.16 から見れば近年において、「、」「。」のような句読点と、「！」「？」のような文末の感嘆符が特徴的である。長文の特徴と関連しているのであろうと考えられる。興味深いのは【】の使用である。表 4.12 では【】の使い方を集計してまとめたものである。ほとんど情報の追加として使われているということがわかる。近年におけるウェブ小説のタイトルはできる限り多くの情報を取り入れるようにしていると思われる。

表 4.12　【 】の使い方（トップ 10）

Rank	Words	Frequency
1	【 完結 】	165
2	【 連載 版 】	96
3	【 書籍 化 】	47
4	【 Web 版 】	27
5	【 書籍 化 決定 】	21
6	【 WEB 版 】	17
7	【 本編 完結 】	15
8	【 番外 編 】	12
9	【 連載 】	11
10	【 書籍 化 ＆ コミカライズ 】	10

また興味深いことに、図 4.15 に同じ「～」が 2 つ見られるが、使われる傾向が異なることが分かる。UTF-8 の文字コードで確認したら、図 4.15 の右上に位置するのは波ダッシュであり、左下に位置するのは全角チルダである。

図 4.17 は図 4.16 の情報をまとめ、投稿年と記号の対応関係をクラスタリングが行われたヒートマップ[32]で表している。波ダッシュと全角チルダの使い方はほぼ対立していることが分かる。波ダッシュは 2004 年から 2009 年にかけて、また 2018 年から 2021 年までの頻用が見てとれる。それに対して、全角チルダは 2009 から使われるようになり、2008 年以前はほぼ使われなかった。この 2 つの記法は類似した表現なので、意識しながら使い分けしているとは思えないが、端末により文字コードの表記が異なるのではないかと予想される。

[32] 階層クラスターはユークリッド距離・ウォード法を使用した。本論文におけるデンドログラム付きヒートマップ図は全てこの手法と距離尺度が適用される。

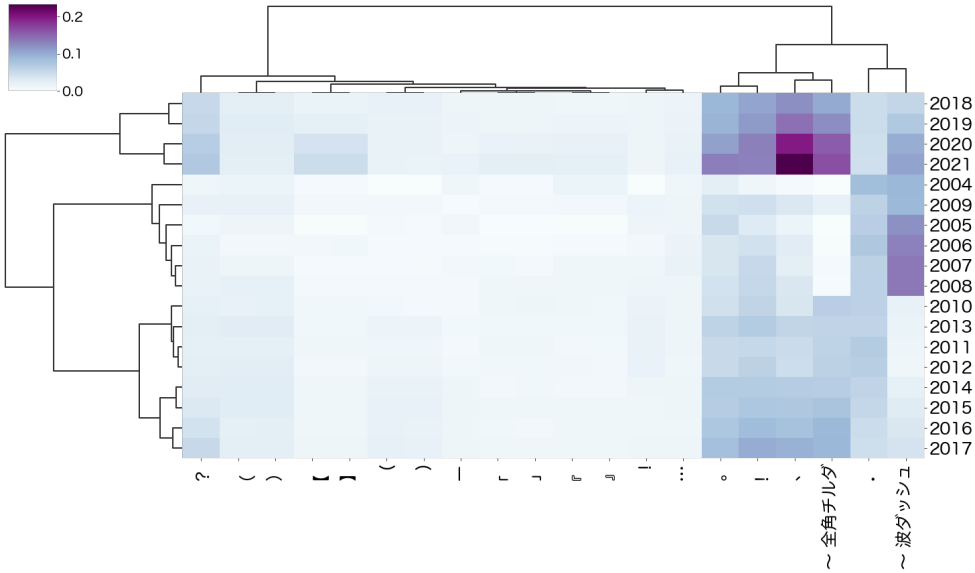


図 4.17　投稿年と記号項目のヒートマップ図

表 2.1 では「pc_or_k」という項目の説明を示した。「pc_or_k」を用いて、どの端末から小説が投稿されたのかを解明することができる。ケータイの場合は「1」、パソコンの場合は「2」、両方を使う場合は「3」で表示される。投稿年と端末情報を結合して波ダッシュと全角チルダの分布を可視化したのは図 4.18 である。図 4.18 から分かるように、大きな分かれ目は 2009 年以前と 2010 年以降であり、端末に相違があるとは言えない。もちろん、投稿に使われる端末が表示されるだけで、どの端末で執筆されたかを表すものではない[33]ため、深掘りすることは容易ではない。

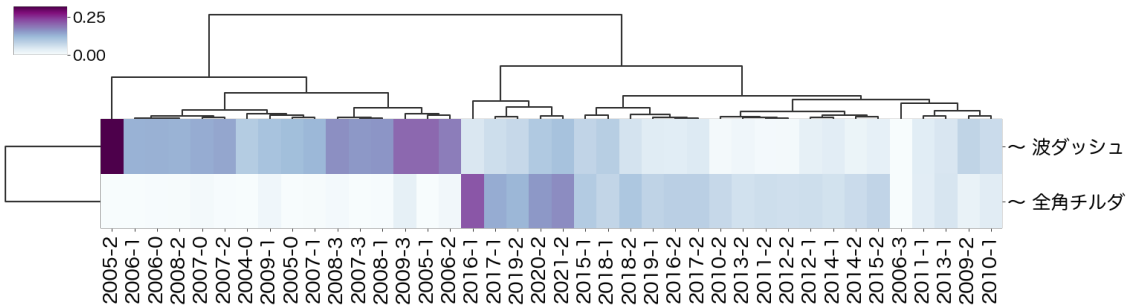


図 4.18　投稿年と投稿端末を結合したヒットマップ図

[33] https://dev.syosetu.com/man/api/

投稿の端末による表記違いという予想を覆したが、日本語の文字コードとの関わりがあることは考えられる。日本語の CP932 文字コードから UTF-8 に変換すれば、波ダッシュが全角チルダになると言われる[34]。2009 年から Windows7 が発売されるようになり、Windows に実装される文字コードとの関連があるように思われる。近年では波ダッシュの表記が多くなってきたのはアップル製のパソコンが用いられるためではないかと考えなくもない。波ダッシュであれ全角チルダであれ、「～」が好まれて使われる傾向があるのは結論づけられる。

4.2.3.2 人気作品

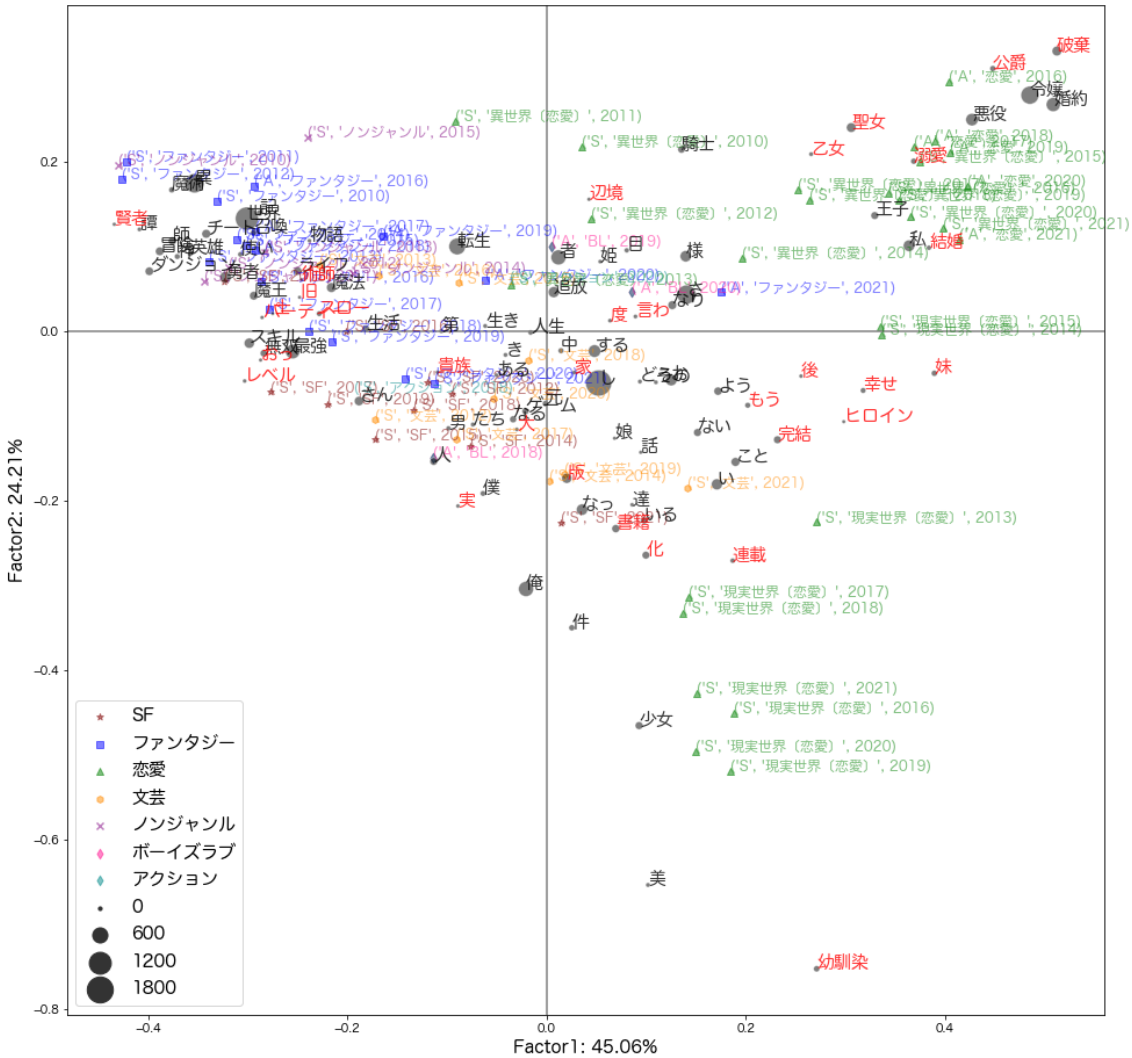


図 4.19　人気作品内容語の対応関係図（Factor1 と Factor2）

[34] https://qiita.com/kasei-san/items/3ce2249f0a1c1af1cbd2

この節では、人気作品のタイトルにおける語彙特徴を解明することを試みる。

図 4.19 と図 4.20 は人気作品のみを対象に、カテゴリーごとの語彙の分布状況を描いた図である。図 4.12 と同じように、上位 100 語までの内容語を語彙項目として取り扱っている。図 4.19 は Factor1(45.06%)と Factor2(24.21%)、図 4.20 は Factor1(45.06%)と Factor3(8.16%)の対応関係図を示している。Factor1 は恋愛ジャンルであるかどうかで左右に分かれ、右側にはほぼ恋愛作品しか配置されていない。Factor2 は物語の背景と関わり、異世界という設定によるところが大きい。また Factor3 は年度の差異を解釈し、下にいくほど 2021 年に近づいていると判断できる。赤い字は図 4.14 で現れなかった語彙を表示している。

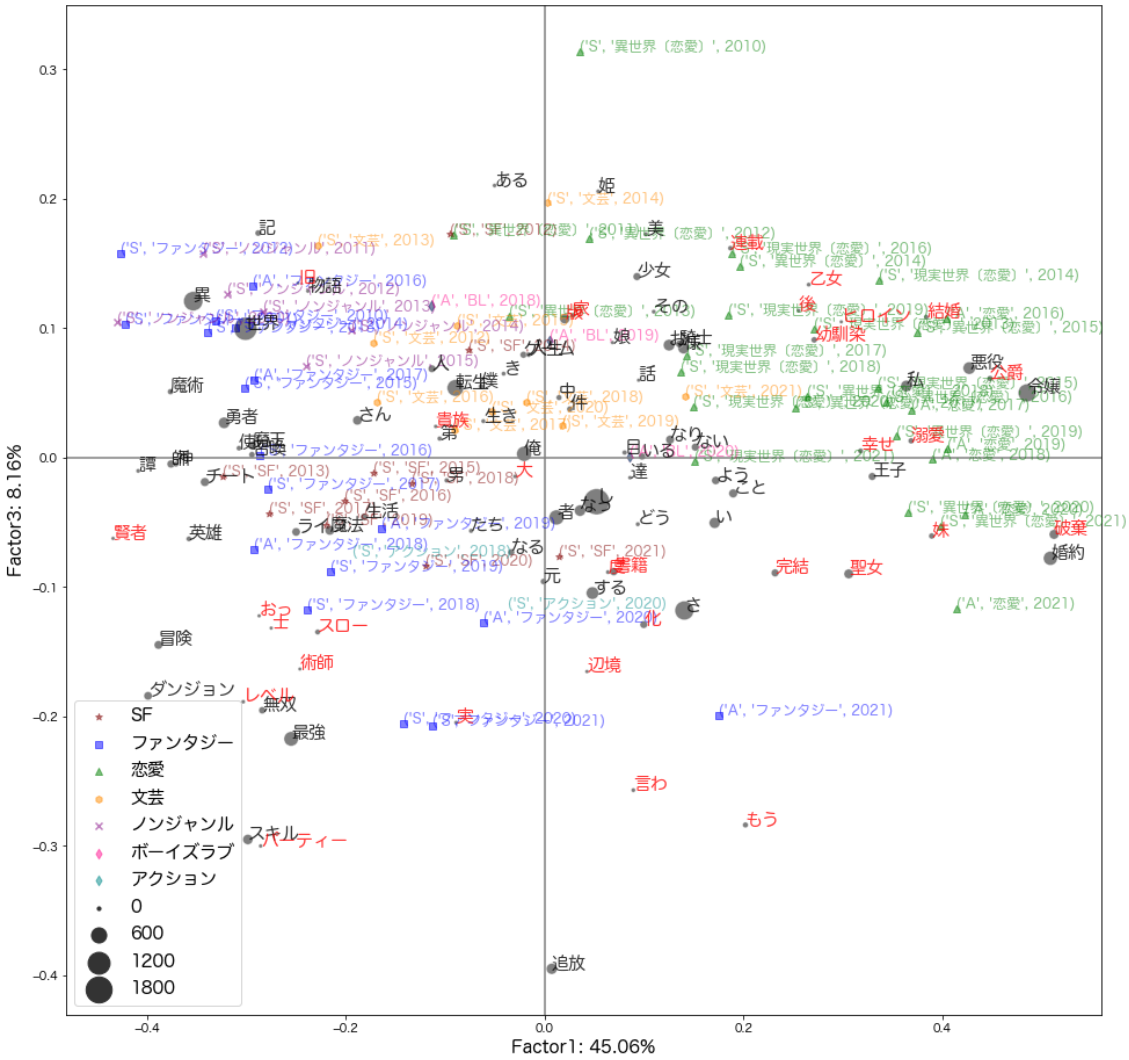


図 4.20：人気作品内容語の対応関係図（Factor1 と Factor3）

人気作品では内容自体と関係性がない「完結」「書籍」「化」が特徴的に出現しており、タイトルで宣伝効果をあげるための追加説明を付与する傾向があると考えられる。これも前述した【】との関連性が見られ、人気作品のタイトルの文字数の増加に寄与していると考えられる。こうした語句が 2020 年、2021 年の作品群に近く位置していることから、ここ 2 年では書籍化が盛んになっている、もしくはタイトルに宣伝情報を挿入することがトレンドになったと推測できる。また、近年の現実恋愛作品では「幼馴染」という設定が非常に好かれていることが読み取れ、興味深い現象だと思われる。

5. 小説のあらすじ

第 5 章では、各作品のあらすじの各種傾向を掴むことを方針とする。まずはあらすじにおいても文字数に特徴が見られるかどうかを検討する。また、トピックモデルを用いてトピックという視点からあらすじの特徴、特に年度別の相違や人気作品を特徴づける話題などを考察する。

5.1 あらすじの長さ

第 4 章ではタイトルの長さが年々上がることが証明された。そこで、あらすじにおいてもそのような傾向があるのではないかと考えられる。それを検証するため、投稿年ごとに両サイトの文字数の集計情報を図 5.1 の boxenplot によって可視化した。

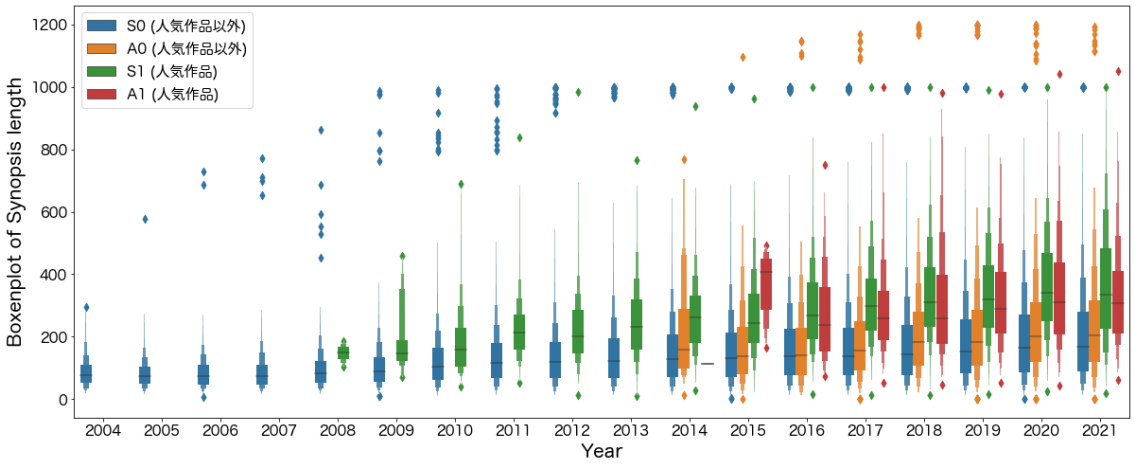


図 5.1　サイトごとの人気作品・それ以外の作品のあらすじの文字数の boxenplot

予想された通り、タイトルの長さの傾向と同様に、あらすじもほぼ年々長くなりつつあり、人気作品の中央値はさらにそれ以外の作品を大きく上回ることが観察できる。長いというのも情報量が多く含まれていると考られるのであろう。つまり、近年の作者は大量の情報をあらすじで提示する傾向があり、人気作品において紹介文がより長くなるのも、読者が長いあらすじを好むだと考えられ、あらすじにおいて情報量が豊富であることがトレンドになっている。この章では、あらすじを対象として、トピックという視点から解析し、内容に焦点を絞る。まずはトピックモデルについて紹介する。

5.2 トピックモデル

トピックモデルとは何かといえば、田畑（2017）で指摘したように、文書中で共起語句は同じ話題を構成しているという前提に立脚し、確率論的アルゴリズムによって共起する語群、即ちトピックを抽出するのがトピックモデルである[35]。つまり、文書データ群があった際に、それぞれの文書は何についての話なのか、各文書のトピック構成を判断するために、トピックモデルが構築される。

5.2.1 今までのトピックモデル

トピックモデルの先駆として位置付けられているのは Deerwester（1990）が発表した潜在的意味索引（Latent Semantic Indexing, 略称 LSI）である。LSI は高次元の文書の行列を特異値分解（SVD）[36]に基づき低次元に縮約してトピックを推定する手法であり、圧縮されたベクトル空間内では, 関連の強い項目が近くに配置され、文書とそれに対応する特徴語をマッピングすることができる。

確率的統計手法で LSI を改良して提案したのは Hofmann が 1999 年に発表した確率論的潜在意味解析モデル（Probabilistic Latent Semantic Indexing, 略称 PLSI）である。図 5.2[37]は Hofmann が作成した PLSI モデルの構成図を示している。P(d)は文書 d が選択される確率、P(z|d)は文書 d におけるトピック z の出現率、P(w|z)はトピック z における単語 w の出現率である。EM アルゴリズムを用いることによって、それぞれの P (z|d)と P (w|z)を算出することができると Hofmann が指摘した[38]。

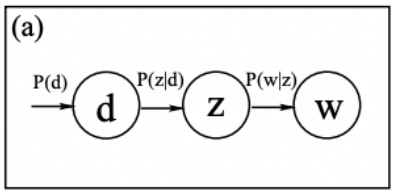


図 5.2　PLSI モデル（Hofmann, 1999, p290）

[35] 田畑智司（2017）「FLOB コーパスの意味構造：確率論的トピックモデルによる言語使用域の特徴づけ」『言語文化共同研究プロジェクト 2016』p.6

[36] 任意の実行列が二つの直交行列と特異値からなる対角行列の内積に分解できることである。

[37] 図 5.2 は Hofmann, T. (1999) Probabilistic Latent Semantic Analysis, *UAI'99 Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, p.290 より引用

[38] Hofmann, T. (1999) Probabilistic Latent Semantic Analysis, *UAI'99 Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp.290–291

PLSI をさらに汎化したのは Blei（2003）によって提案された LDA（Latent Dirichlet Allocation）、つまり潜在的ディリクレ配分法である。LDA では、個々の文書におけるトピックの確率分布、単語の確率分布が、ディリクレ分布という確率分布を仮定して生成させる。Blei（2003）において使用された LDA のグラフィカルモデル（図 5.3[39]）は、トピックの確率分布 θ がディリクレ分布 α に従い、単語の確率分布がディリクレ分布 β に従うことを示している[40]。つまり、PLSI ではトピックの確率分布パラメーターが確定されているのに対して、LDA では変数となり、ディリクレ分布に従うのである。PLSI に関しては、コーパスのサイズに比例してパラメーター数が線型的に増加し、過剰適合に関わる重大な問題をもたらすが、LDA ではこの問題が発生しないと Blei が指摘した[41]。

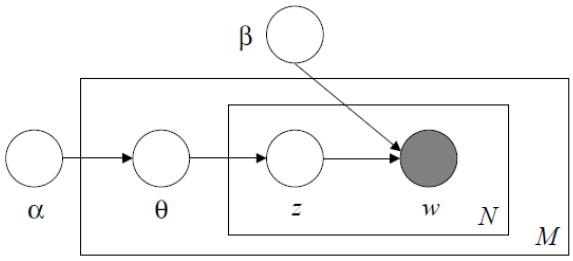


図 5.3　LDA モデル（Blei, 2003, p997）

MALLET[42]というツールキットを通して LDA を実行することができる。MALLET の実行によって、入力した文書集合にどのようなトピックが存在しており、また、それらのトピックが各文書にどの程度の割合で出現しているのか、単語が各トピックにどの程度の割合で生成されているのかを知ることができる。

モデルの実行により生成されたトピック、つまり語群をどのように解釈するかは分析者に委ねられる。例えば、「アップル」、「ジョブス」、「スマホ」というキーワードからこのトピックは「iPhone」に関わると判断できる。一方で、「アップル」、「ジュース」、「イチゴ」であれば、「果物」か「果汁」のトピックになる。同じ「アップル」が含まれているにもかかわらず、トピックは大きく異なる。分析にあたり、「キーワード」を十分に考慮し、

[39] 図 5.3 は Blei, D.M., Ng,A. and Jordan, M. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3:993-1022, p.997 より引用
[40] Blei, D.M., Ng,A. and Jordan, M. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3:993–1022, p.996
[41] Blei, D.M., Ng,A. and Jordan, M. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3:993–1022, p.1001
[42] https://mimno.github.io/Mallet/topics

総合的にトピックの解釈をする必要がある。トピックに対する適切な解釈は考察において非常に重要な部分である。

LDA モデルの応用は様々な領域において盛んに行われている。人文学における LDA モデルの応用研究は、第 1 章の 1.1 先行研究において既に詳述した。小説の言語特徴と内容性に関する分析にあたり、トピックモデルを用いる価値があると考えられる。

しかし、LDA を実行する際には、トピック数を研究者自身で設定しなければならないが、最適な数を決めることが難しく思われる。また、統計的なアルゴリズムによるモデルがゆえに、単語の出現頻度によるところも非常に大きい。助詞や助動詞のように頻繁に使われるが意味を含めない機能語が書くトピックの中心単語になる可能性が高いことが実証でも分かった。つまり、よりまとまったトピックにするには前もって機能語のようなストップワーズを取り除く作業が必要になる。ただし、ストップワーズが比較的曖昧なことであり、ストップワーズの選定は主観性が入りやすいものだと思われる。また、LDA や PLSI はテキストの bag-of-words 表現に依存し、つまり、文章に単語が含まれるかどうかのみ考え、単語の順番と意味的関係性を無視する。これらの問題点を踏まえて、分散表現を取り入れた Top2Vec モデルを紹介する。

5.2.2 Top2vec

Top2Vec を紹介する前に、まず分散表現について説明する。分散表現とは、単語やドキュメントのような文字的な表現を高次元の実数ベクトルで表すことである。Mikolov（2013a）では単語をベクトルで表現する言語モデル、すなわち Word2Vec を提案し、単語間の類似度を計算することができると指摘した。また Le（2014）では Word2Vec のモデルを拡張させ、ドキュメントのベクトル化を考え、Doc2Vec という文章の類似度を計算できる手法を提案した。このように、分散表現の手法は単語やテキストの意味的関係性を掴むことができ、類似語検索や文書分類などに有効に使われている。

Angelov（2020）で発表した Top2Vec は、ドキュメントの埋め込みを利用してトピックベクトルを検索するモデルである。Angelov によると、Top2Vec はまず Doc2Vec（もしくは他の事前学習モデル）を介して各文書をベクトルに変換する。次に UMAP で次元削減を行い、HDBSCAN で文書ベクトルを分類し、同じグループに分類された文書ベクトルの平均を取ったものをトピックベクトルにした上で、トピックベクトルから近い順にそれぞれのトピックの構成単語を決めるという仕組みになっている。トピックベクトルとドキュメントベ

クトルとのコサイン類似度を計算することによって、各ドキュメントを最も類似度の高いトピックに割り当てる。[43]

Top2Vec の特徴としては意味的関係性を考慮した上に、トピックの数をあらかじめ指定する必要がない。またストップワーズが基本的にどのトピックベクトルからも均等な距離を取るため、キーワードとしてトピックに配属されることはほぼない。よって、ストップワーズの前処理を行う必要がないと Angelov が指摘している。黄（2021）では、Top2Vec を用い、中国の有名な推理作家である程小青の作品を考察した。出力されたトピックには、代名詞の「她」（彼女）に関わる異質なトピックが発見され、そのトピックの分布を考察すると、異なる作品における代名詞の使い方の相違を見ることができた。このように、機能語など LDA でストップワーズとして扱いされやすい語でも、実際はドキュメントの特徴を反映できる可能性があるため、ストップワーズを除去しない Top2Vec の方がより多くの情報を得られると考えられる。

また、Angelov（2020）は、トピックの Information gain を評価したところ[44]、訓練データ[45]において LDA や PLSA など従来のトピックモデルより Top2Vec の方が遥かに高い得点を得たということを証明した。そこで、この新たな Top2Vec を使用してあらすじの解読を試みる。

5.3 モデルの実行

Angelov（2020）は個人の Github アカウント[46]で Top2Vec モデルのコードと実行の仕方をアップロードし詳細に説明を加えている。それを参照しながら、モデルの実行を試みる。実行環境に関しては、Jupyter Notebook[47]を用いて、Python 言語によってプログラムを書き込み、モデルの実行やデータの整理、可視化などの作業を行う。

実行に必要なのは、リスト化されたファイルコーパス（合わせて 473323 件）である。それぞれのファイル形式としては、分割済みかつスペースが入ったものを使用する（実際使用するリストを一部表示する）。

---

[43] Angelov, D.（2020）TOP2VEC:Distributed Representations of Topics, *Computation and Language*, arXiv:2008.09470, pp.4–9

[44] トピックがドキュメントをどれほど適切に記述しているかを評価する方法。Angelov, D.（2020）TOP2VEC:Distributed Representations of Topics, *Computation and Language*, arXiv:2008.09470, pp.9–10

[45] Scikit-learn の fetch_20newsgroups というサンプルデータセット

[46] https://github.com/ddangelov/Top2Vec

[47] Jupyter Notebook は、Python などを web ブラウザ上で記述・実行できる統合開発環境である。対話型の開発環境であるため、前の実行結果に応じて、次に実行するプログラムや作業を選択できる。

実行リストの一部表示：

```
corpus[:3]
```

```
['少女 は 夏休み に 、 田舎 に 遊び に 来 まし た 。 その 田舎 に は 山 が あり まし た 。 不思議 な 逸話 が 沢山 あり ます 。 そ
んな 森 に 少女 は 足 を 踏み入れ て しまい まし た ！ さて 、 皆 さん 。 少女 は どう なる ん でしょう か ？ 何 を 見る の でしょ
う か ？ 何 と 出会う の でしょう か ？ フフフ 、 楽しみ です ね 。',
 'この世 に ある 理解 の 範疇 を 大きく 逸脱 し た 物 。 多く の 人 は 見 て 見 ぬ 振り し て 放っ て おく 不可思議 。 これ は そ
う いっ た 不 可能 、 不可解 、 不 条理 を 好む 物好き 達 の 物語 で ある 。',
 '－－ 汚れ た 校舎 、 ダサい 制服 、 破け た カバン 、 他愛 ない 会話 が 世界 の 全て だっ た －－ タバコ を ふかし て は バイク を
 乗り回し 、 喧嘩 に 明け暮れる 日々 。 なん の 面白み も ない 教科 書 に 退屈 な 授業 。 つまら ない 日々 の 中 で 学校 を 抜け出
し た 紀雄 が 出会っ た の は 、 変わっ た 風景 画 を 描い て いる 女の子 だっ た 。 佐々 原 凪 －－ 世界 を 壊し て い た 彼女 と
の 出会い が 、 さらに 新た な 出会い を 紡い で いく 。 紀雄 の 淡く 騒々しい 日々 が 始まっ た 。 彼女 、 友達 、 親 、 教師 。
恋愛 、 夢 、 喧嘩 、 勉強 、 いじめ －－ 。 必死 で 駆け抜け た 、 十 六 の 季節 。 教科 書 以上 に いろんな こと を 教え て くれ
た 、 青い 季節 。 ※ タバコ は 二十 歳 に なっ て から で お 願い し ます ！ ※ この 作品 は カクヨム さん の ほう に も 掲載 し て
い ます 。']
```

データセットが 47 万にのぼるため、実行後のトピック数は 2306 に達している。数があまりにも多いため、Top2vec の Hierarchical topic reduction[48]機能を使用し、トピックの数を 100 と設定し、類似したトピックを結合させる。最終的には、この 100 個のトピックを対象に、解析を進めていく。付録 2 では各トピックのサイズ（各トピックに含まれるドキュメント数）と各トピックを構成する主要単語を表示している。トピックサイズの大きい順にトピックが出力されたので、非常に見やすいと思われる。

例えば、最も大きいトピック Topic0 を構成する主要単語のワードクラウド図 5.4 を検討すれば、「王太子」「殿下」「侯爵」「公爵」「伯爵」などから背景は皇室や宮廷にあると考えられる。「婚約」「破棄」「政略」等は中心的な話を指し、つまり Topic0 は皇室や宮廷における婚約破棄や政略結婚の話題だと解釈できる。

図 5.4　Topic0 のワードクラウド

こうしたトピックに対する説明は研究者によるものであり、唯一の正解があるわけではないが、このトピックに配属された最も類似度の高い具体例を 10 件あげて吟味すると、やはり Top2Vec で出力されたキーワードがトピックの特徴を良くまとめており、貴族の婚約、特に婚約破棄の展開が多いことが観察できた。こうして具体例も参照しながら良い解釈が

48 この機能を True に設定すれば、トピック間のコサイン類似度を計算し、類似したトピックを同じクラスターに配属し、研究者が指定した数まで減らすことができる。

できると思われる。やはり、婚約破棄系の作品がウェブ小説において最もトレンドに乗っていることが結論づけられる。

Topic0 の具体例：

1. 【小説家になろう | 『嫌われ令息と悪役令嬢』 | 2021_異世界〔恋愛〕】

王太子と隣国王女との政略結婚を進めていたら、王太子が勝手に恋人を作り、その仲を邪魔したとして、断罪されそうな伯爵令息の話し

---

2. 【小説家になろう | 『ある不器用な恋の物語』 | 2021_異世界〔恋愛〕】

仕方ない…彼は私を嫌っているのだから…。　無愛想で無表情で目も合わせてくれない人…。それがスペリア伯爵令嬢アイリスの婚約者公爵家嫡男ディリク余りにも不器用な公爵令息と伯爵令嬢の物語

---

3. 【小説家になろう | 『公爵令嬢は心の中で毒を吐く ～まさか王位継承権第 1 位王子と婚約するなんて～』 | 2019_異世界〔恋愛〕】

王位継承権第四位のアホ殿下に断罪されると思っていたら、まさかの次期王妃！？

---

4. 【小説家になろう | 『あなたに呼んでほしいから』 | 2019_異世界〔恋愛〕】

伯爵令嬢のクレアは婚約者の不貞により婚約を破棄するが、家の事情ですでに結婚が遅れていたため新たな婚約者を探すことは諦めていた。ところが、王宮の夜会で６年ぶりに再会した幼馴染のセドリックに求婚される。彼の望みは昔のようにクレアから「セディ」と呼ばれること？しっかり者の令嬢と、歳下のちょっと訳あり貴公子の結婚までのお話です。

---

5. 【小説家になろう | 『薔薇乙・悪役令息の悩み』 | 2017_ローファンタジー】

公爵家次男、ロアナは第一王子の婚約者。「私の婚約者であり、公爵令息ともあろうお前が嫌がらせとは……」　婚約破棄、追放され平民身分にされた後で、商家の娘（美人）の婿として嫁ぐ。　そんなシナリオを思い描いていたら、予想外の方向に。「罰として婚約者から第八妾妃候補へと落とす事とする！」

---

6. 【小説家になろう | 『婚約破棄と言われたのですが既に婚約解消済みです』 | 2021_コメディー〔文芸〕】

「真実の愛を見つけた！だから婚約破棄してくれ！」幼馴染の王太子から婚約破棄をお願いされた公爵令嬢のエクレール。「殿下、お言葉ですが私達の婚約は既に解消されています」呆れたように返すエクレールに王太子はショックを受けるのでした。かなり短い話です。アルファポリス、カクヨムにも投稿しています。

---

7. 【アルファポリス｜『婚約破棄、喜んで！』｜2016_恋愛】

豊かな王国で、侯爵令嬢エレーヌは、公爵家子息の婚約者から浮気のあげく、一方的に婚約破棄を言い渡された。それに対して、エレーヌは喜んでそれを受け入れたのは公爵家に関わる秘密のためであった。小説家になろう様でも掲載しております。

---

8. 【アルファポリス｜『訳あり公爵令嬢と癖あり皇太子と巻き込まれた訳あり伯爵嫡男のあれこれ』｜2021_恋愛】

ミリアは公爵令嬢しかも、帝国では 4 番目に身分が高い女性しかも聖女の証の銀色を持って生まれたならば同じ銀色を持って生まれた皇太子の婚約者に!が皇太子の本性を知ってるミリアは絶対に皇太子の婚約者になりたくないと訳ありの伯爵の嫡男と婚約するが訳あり公爵令嬢(聖女)と癖あり皇太子(聖人)と巻き込まれた訳あり伯爵嫡男 3 人のそれぞれの思いが夜会から動き出す

---

9. 【小説家になろう｜『政略結婚ですけれど、あなたを愛したいと思うのです。（愛するとはどんなものか分かりませんけれど）』｜2020_異世界〔恋愛〕】

レヴァン・オックスブラッドは最低な気分であった。なぜなら突如辺境伯の爵位と領地を与えられ更には婚約を言い渡されたからだ。婚約者である令嬢はアイビー・ロータス、公爵家のご令嬢。母親は国王陛下の姉であり、つまりは陛下の姪であるというまさに令嬢の中の令嬢である。このご令嬢が婚約解消を言い出すであろうと思っていたが、令嬢の口から出た言葉は思いがけないものだった。「私、あなたを愛することにしましたの」

---

10. 【アルファポリス｜『猫に花束を』｜2020_恋愛】

モンテナス国の第三王女、ユーアリティは大国のルデンの王太子妃となった……のはいいが、肝心の王太子は女嫌いらしい。そんな折、つい最近婚約破棄騒動があった他国の王太子夫妻が来国する。ユーアリティは他国の婚約破棄騒動に大きく巻き込まれる。R15 は保険

Topic0 の構成単語と実際のあらすじと合わせて見れば、トピックの分類が適切に行われていると考えられる。5.4 節からはトピックの具体的な考察を行う。まずは投稿年を比較対象として、それぞれのトピックの分布状況を確認する。

5.4 年度別のトピック

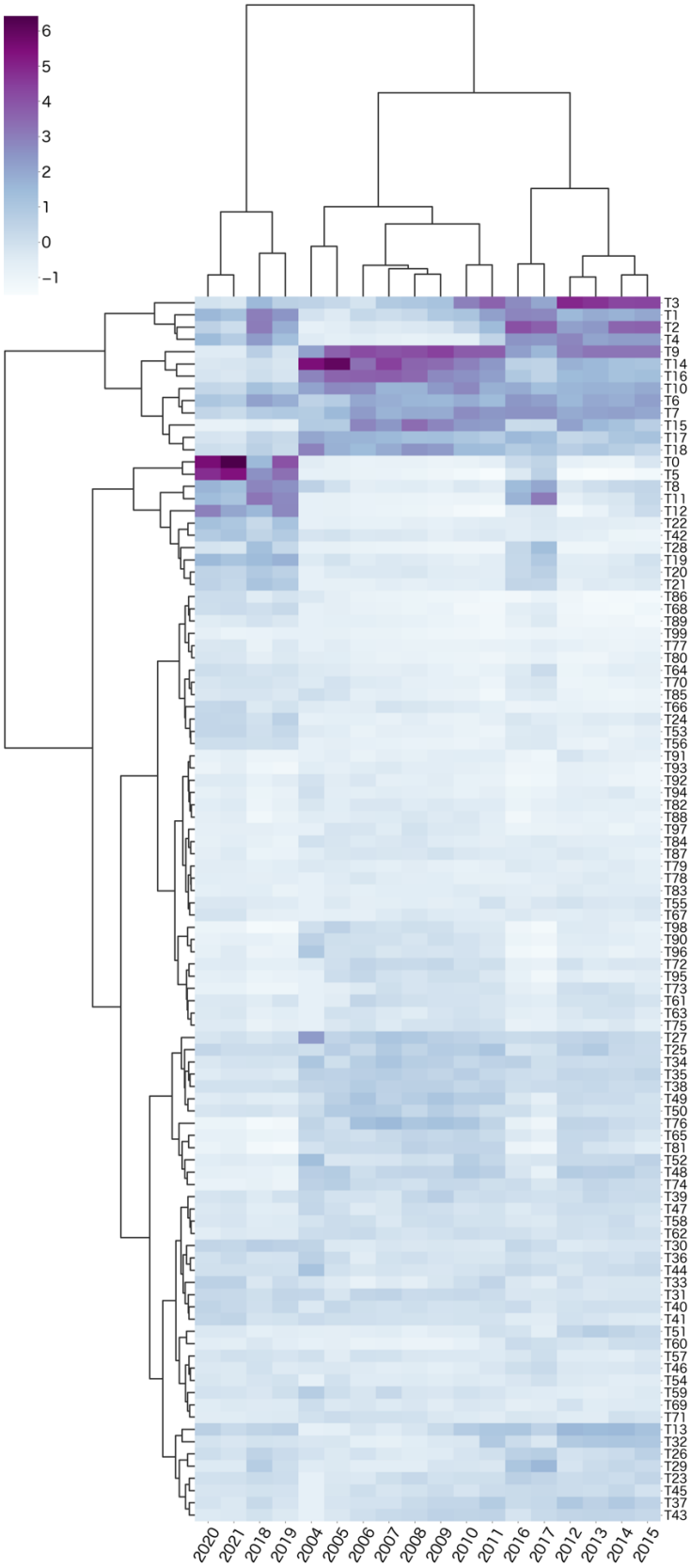


図 5.5　投稿年ごとの各トピックの割合ヒートマップ図

図 5.5 はトピックと投稿年の関係を表すヒートマップである。縦軸にトピックの番号、横軸に投稿年のラベルを配置し、トピック間、年度間の関係性をクラスタリングの結果によって可視化している。ヒートマップはセルの濃淡の違いでトピック密度の差を表している。その年におけるトピックの割合が高いほど、当該セルの色が濃くなる。左上のカラーバーを参照すればわかるように、割合が低いほど薄いブルーになり、高いほど紫色に近寄る。縦軸ではトピック間の関係性を表し、同一のクラスターに属するトピックは生起するテクストが類似している傾向があることを示している。横軸では投稿年の関係性を表し、似た特徴を備えた年度は近接し、類似しない年度は遠く配置される。

横の投稿年クラスターの分岐から明らかなように、トピックの構成によって 2004 年～2017 年、2018 年～2021 年の 2 グループにはっきりと分けることができる。前者はさらに 2011 年を境界線とし、2004 年~2011 年と 2012 年~2017 年との明確な違いが見られる。縦のトピックの分岐から言えば、上側のグループには比較的に大きいトピックが配置され、鮮明な特徴を持っているように見える。こうしたトピックを中心に考察を行う。

2004 年～2017 年のグループを特徴づけるのは図 5.5 の Topic9～Topic18 である。これらのトピックの分布に注目し、年ごとの類似度スコアの平均割合を可視化したのは図 5.6 である。それぞれのトレンドからすれば概ね年々下がる傾向にあると考えられる。

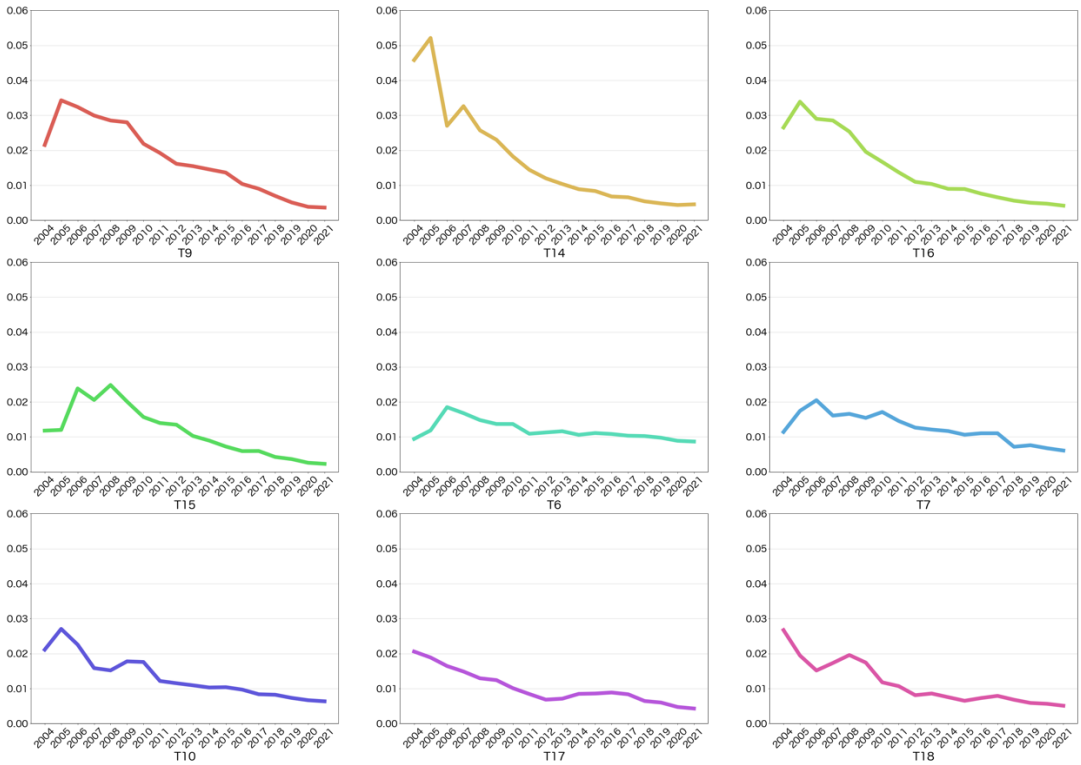


図 5.6　トピックの割合変化図

Topic9、Topic14 と Topic16 に関しては比較的激しい下降傾向がみられ、その上 2004~2011 年において割合が高いため、初期のウェブ小説の特徴を表しているであろうと言える。この 3 つのトピックを中心に考察を行う。図 5.7 は Topic9 の構成単語の割合を可視化したものである。Topic9 の主要単語を吟味すると、空虚や儚さが中心に描かれ、古風で運命に関わる言い回しが主題になっているのではないかと考えられる。

図 5.7　Topic9 のワードクラウド

実際 Topic19 に属するあらすじから、最も類似度スコアの高いものを 10 件取り出し、以下のように示す。

1.　【小説家になろう | 『RUUCAMI　ADVENT』 | 2006_ノンジャンル】

平凡な高校生「辰児」は、時の止まった村「お伽村」に迷い込む。不思議な村人や、村に隠された数々の謎は、彼の内に秘めた力を目覚めさせてゆく。「神殺し」の名を背負った辰児を待つものは・・・

2.　【小説家になろう | 『幻想の壁を越えて』 | 2010_ノンジャンル】

あの日、ある好奇心から僕の人生の歯車が狂ってしまった僕は奇妙で不思議な出会いと体験を通して狂ってしまった人生の歯車を元に戻すことができるのだろうか

3.　【小説家になろう | 『神の力~回りだす運命~』 | 2015_ローファンタジー】

全ての始まりの力を持つ少年少女の物語。いろんな感情が交差する時、君はどんな答えをだすのかな?

4.　【小説家になろう | 『止まない雨』 | 2010_ノンジャンル】

一人の少年は、ただ過ぎていく日々に諦観していた。一人の少女は、辛苦の毎日に、救いの手を望んでいた。交わるはずのない二人の運命は、ほんの小さな出来事によって変わる。人は人を救えるのか。止まない雨は、降り続く。

---

5. 【小説家になろう | 『堕天の騎士と聖なる王女』 | 2015_ノンジャンル】
かけがえのない大切な人と回り続ける世界の歯車を守るため、少年は立ち上がる。

---

6. 【小説家になろう | 『僕の人生を狂わせたもの。』 | 2014_ノンジャンル】
ごくごく一般的に、生活していた僕。ある日の朝に起きた出来事によって人生の歯車が壊れだす。壊れだした歯車を、戻すことが出来るのか？それとも壊れたままの人生を、歩み続けるのか？この物語は、1 人の僕という男の壊れた人生を取り返していくという物語です。

---

7. 【小説家になろう | 『ALICE』 | 2010_ノンジャンル】
「忘れられた”物語”はやがて壊れてしまう」不思議の国に迷い込んだアリスは、忘れてしまった‘物語”を取り戻すため、自らの記憶に立ち向かうために歩き出す。

---

8. 【小説家になろう | 『あの日誓ったこと』 | 2012_ハイファンタジー】
齢十六歳にして家族、村の人々、自分にあったすべてを失った少女。彼女の瞳に映るのは幸か不幸かあるいは希望？彼女は今、大切な人との誓いを果たすため、前に進む。

---

9. 【小説家になろう | 『ダークレイズキングダム　レイズサイド』 | 2012_ハイファンタジー】
ある日、炎が運命を狂わせた。　少女は憧れた人の影を追い、聖なる騎士としての道を歩み始める。　過去に狂った歯車が、死者に支配された世界を大きく変えていく。　——今、運命が動き出す。

---

10. 【小説家になろう | 『咆吼』 | 2006_ノンジャンル】
己の過去と罪悪感に縛られた青年と、束縛された少女。二人の出会いが、止まっていた、互いの時間に息を吹き込んだ。二人の、運命という名の歯車は、今動き出した……。

これらの具体例を吟味すると、やはり人生の大きな変わり目を迎えたというような運命の転機を述べることがほとんどである。初期のウェブ小説ではそのような書き方が多いことが明らかになった。

続いて、Topic14 と Topic16 に対して考察を行うが、Topic14 と Topic16 のキーワードに高い関連性があることは図 5.8 と図 5.9 のワードクラウドから観察することができる。

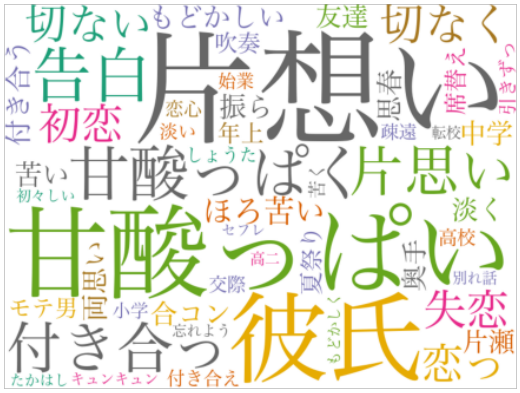


図 5.8　Topic14 のワードクラウド

図 5.9　Topic16 のワードクラウド

構成単語から分かるように、Topic14 は恋愛的な話が中心で、特に片思いや、付き合う中における甘酸っぱい気持ちに関わるトピックだと判断できる。隣合わせで似たような分布をしている Topic16 の構成を見ると、特に悲しみ、苦しみや切なさなどをめぐる心理的な描写が主題となっており、恋との関連も見られるのではないか。Topic14 が「小説家になろう」サイトが運営し始めた最初の 2 年、つまり 2004 年と 2005 年における割合が格段に高いことは図 5.6 から明らかである。甘酸っぱい恋物語がウェブ小説の最初の流行なのではないかと考えられる。

次に、2012~2017 という中期グループにおいて特徴的なのは、図 5.5 からすれば Topic1、Topic2、Topic3 および Topic4 だと読み取れる。図 5.10 では Topic1 から Topic4 までの 4 つのトピックのトレンドをそれぞれ折線によって表している。

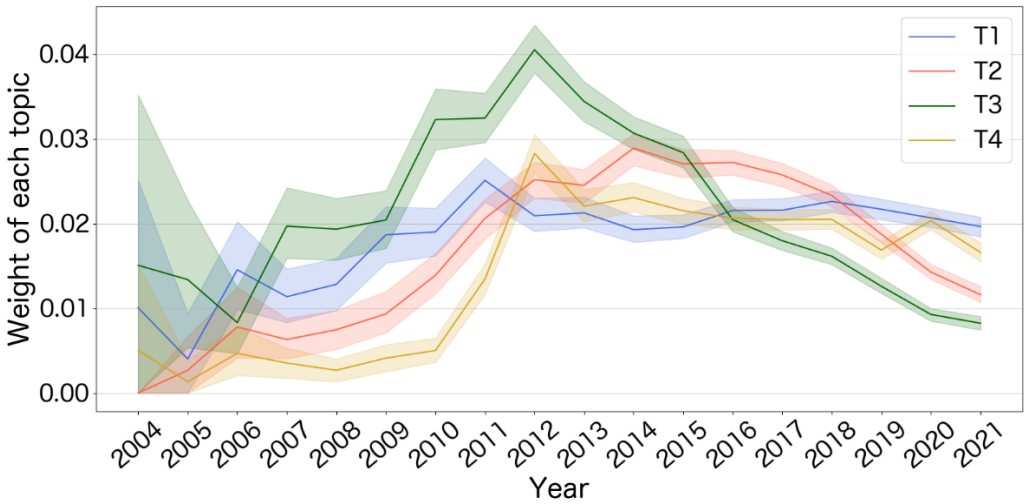


図 5.10　トピックの割合変化図

年ごとの分布からすれば、Topic3 は激しい上昇と下降が見られ、検討する価値が高いと思われる。図 5.12 の Topic3 の構成単語から、作者の思い、特に思いつきで書き始めることや暇つぶしという意図に関する説明を中心としている。ウェブ小説は年々投稿作品が多くなることが既にわかり、新人作者も年々増加していると推測できる。よって、まともな紹介よりこうした軽い説明が好まれたと思われる。2013 年以降下がりつつあることから、より推敲されたあらすじ、もしくは情報量が多く含まれるような文にすることが定番になっているのではないかと考えられる。

図 5.11　Topic2 のワードクラウド

図 5.12　Topic3 のワードクラウド

もう 1 つ明らかな上昇と下落が見られるのは Topic2 である。図 5.11 に示されたように、Topic2 は脱字、誤字や文法などの指摘願いに関わるトピックである。これも「はじめまして」や「処女」などのキーワードから新人作者との関連性が見られる。Topic2 も Topic3 も内容を詳細に説明するようなものではなく、むしろ読者を意識しながら、対話しているようなトピックである。特に Topic2 は「くださる」「ありがたい」などから対話感が濃厚に感じられる。つまりウェブ小説が発展しているとともに、読者とコミュニケーションしようとする作者の意識も高まってきたのではないか。2017 年から徐々に下がることも明らかであるが、あらすじの文字数がとりわけ長くなっていることからも、あらすじに文章自体の情報が大量に書かれることが流行になっていると考えられる。

図 5.10 から分かるように、黄色で描かれた Topic4 に関しては 2010 年から急上昇を見せており、2012 年で頂点に至った。2013 年では少し下落したが、その後は一定の割合で推移していることが見られる。図 5.13 で示された Topic4 の内容と合わせて見れば、VR ゲーム

や RPG ゲームのトピックだと言え、このトピックの時系列的な傾向からも、現実における
オンラインゲームの爆発的人気沸騰や科学技術の発達との関わりが深いと考えられる。

図 5.13　Topic4 のワードクラウド

Topic1 はそれほどトピックらしいトピックではないように思える。図 5.14 の主要単語からすると、話し言葉がほとんどであることが分かる。

図 5.14　Topic1 のワードクラウド

Topic1 は 2011 までに割合が徐々に上がってきたことは図 5.10 から明らかである。正式な書き言葉を用いる従来のあらすじと違い、軽い話し言葉を使うあらすじの割合が高くなってきたと考えられる。また、実際 Topic1 に属した具体例を吟味すると、会話文や主人公の心の声がそのままあらすじになっており、それらを通して文中の見どころなどを先に読者に送るような形をとっていることが分かる。特に不本意や怪訝な気持ちなどが含まれるものが多くみられ、読者の興味をそそる工夫だと考えられる。こうした心境や対話が含まれる内容が大きなトピックとして使われることが興味深い特徴である。

Topic1 の具体例：

1. 【小説家になろう｜『酔いつぶれてるおっさん助けたら美少女になった。』｜2020_コメディー〔文芸〕】

いや、マジで意味のわかんねー話するから酒のつまみ食いながら聞いてくれや…私さ、実は男だったんだわ。

---

2. 【アルファポリス | 『タネツケ世界統一　下品な名前で呼ばれてるけど、俺、世界を救うみたいです』 | 2018_ファンタジー】

だから俺の名前はタネツケじゃねえっつてんだろ！

---

3. 【小説家になろう | 『異世界転生日記』 | 2020_ハイファンタジー】

さてさて、疲れて帰って寝たら死んで異世界に！？いやいやどうゆうこっちゃねん。まあ、家族もいなかったし、未練もないしいいけど…。なんも説明もないし？勝手にいろいろやってみようかなぁ…？なんだかわからないうちに異世界転生していたのでいろいろなことをやりはじめます。

---

4. 【小説家になろう | 『最後だと思ったら生まれ変わってましたw』 | 2020_異世界〔恋愛〕】

なんか、こんな所来ちゃった。よくわからないけど、転生しちゃったのかな.....？ここだと、なんか不安だなぁ

---

5. 【小説家になろう | 『くっころ神に任命された俺は、今日も高飛車女子を恥ずかし…めるわけないじゃん！！怖いじゃん！！みんな仲良くすれば、いいじゃん！！！』 | 2021_コメディー〔文芸〕】

異世界で神様やることになっちまった！……ちょ、はい？！くっころ神ってなんなんだよ！俺は絶対にやんねぇぞ？！

最後に図 5.5 から読み取れた 2017 年以降の特徴トピックである Topic0、Topic5、Topic8、Topic11 および Topic12 を中心に考察を加える。図 5.15 ではこの 5 つのトピックのそれぞれの時系列変化を示した。

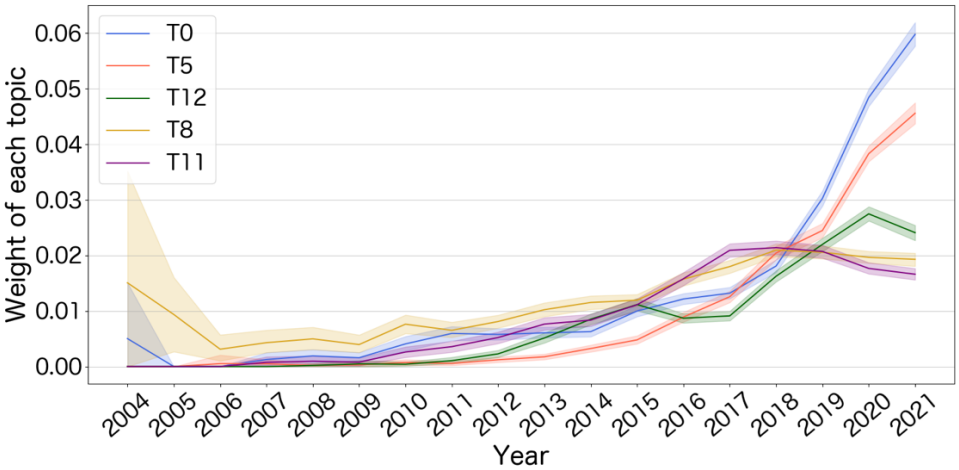


図 5.15　トピックの割合変化図

2017 年から激しく上がるのは Topic 0 と Topic5 である。Topic0 は前述した婚約破棄の大トピックである。それが近年さらに好まれていることがわかる。また、Topic5 で示されたように、能力が低いためパーティー（一行／集まり）から追放されるストーリーがここ数年において非常に流行っていることが明らかである。また、Topic0 と Topic5 に続くトレンドとして、Topic12、つまり乙女ゲームにおける悪役令嬢の話題も上昇傾向にあり、トレンドに乗っていることが分かる。近年における貴族をめぐる婚約破棄、乙女ゲームの悪役キャラに転生する物語、パーティーから追放されるストーリーの流行は、タイトルにおけるコレスポンデンス分析の結果も側面的に裏付けしているのではないか。

図 5.16　Topic5 のワードクラウド

図 5.17　Topic12 のワードクラウド

次に Topic8 と Topic11 について考察する。Topic8 は魔族討伐に関わるトピックだと図 5.18 から判断できる。王道的なファンタジートピックとして安定した成長を見せているが、正義の仇を取るような熱血冒険物語より、Topic5 のような逆襲系がより好まれることも今時の趨勢である。Topic11 は主人公の背景を表すトピックだと考えられ、社畜、サラリーマン、ニート、ブサメンなどの負け組人物設定が好まれていると思われる。具体的なあらすじを参照すれば、やはりこうした世間でよく言われる負け組が（転生して）スキルを手に入れ、勝ち組になるような展開プロットが多く書かれている

図 5.18　Topic8 のワードクラウド

図 5.19　Topic11 のワードクラウド

Topic11 の具体例：

---

1. 【小説家になろう｜『３２歳ニートの異世界譚』｜2019_ハイファンタジー】

３２歳ニートが異世界で大活躍！？スキル？チート？ハーレム？異世界に来たからには好きにさせてもらうぜ！

---

2. 【小説家になろう｜『裸の戦士』｜2016_ハイファンタジー】

裸戦　勇引きこもりニート独身が、あるスキルで世界を救うお話。

---

3. 【小説家になろう｜『ただイケ（忠志イケメンに宿る）』｜2012_ハイファンタジー】

平凡な高校生が念願（？）の異世界イケメンライフを？　ファンタジー世界を舞台に中身も外も残念な憑依系イケメンが頑張るかもしれない話。チート、知識チート無し・・・って何が出来るんだ？

---

4. 【小説家になろう｜『チート無しで異世界転生した結果ｗｗｗｗｗｗ』｜2018_ハイファンタジー】

脱糞した勢いで自決したら異世界転生したったｗｗｗｗｗｗｗチート能力も特別なアイテムも何も貰えんかったけどワイが主人公なことには変わりないんや！無双してハーレム築くやで！果たして彼はいわゆるチート主人公のようにこの世界で活躍できるのか、乞うご期待！

---

5. 【小説家になろう | 『チートが無いけど異世界で頑張ります(´・ω・｀)』 |2019_ハイファンタジー】

人生負け組のおじさんが、異世界へでも、チートスキルなし！若返りなし！知識チートも微妙？気付けば、美少女エルフに養われてる異世界で、リスタート(第 2 部分から、ヒロインが出てきます)

Topic8 と Topic11 は 2017 年からあまり上昇が見られないが、大きなトピックとして安定していることが結論づけられる。

5.5 人気作品のトピック

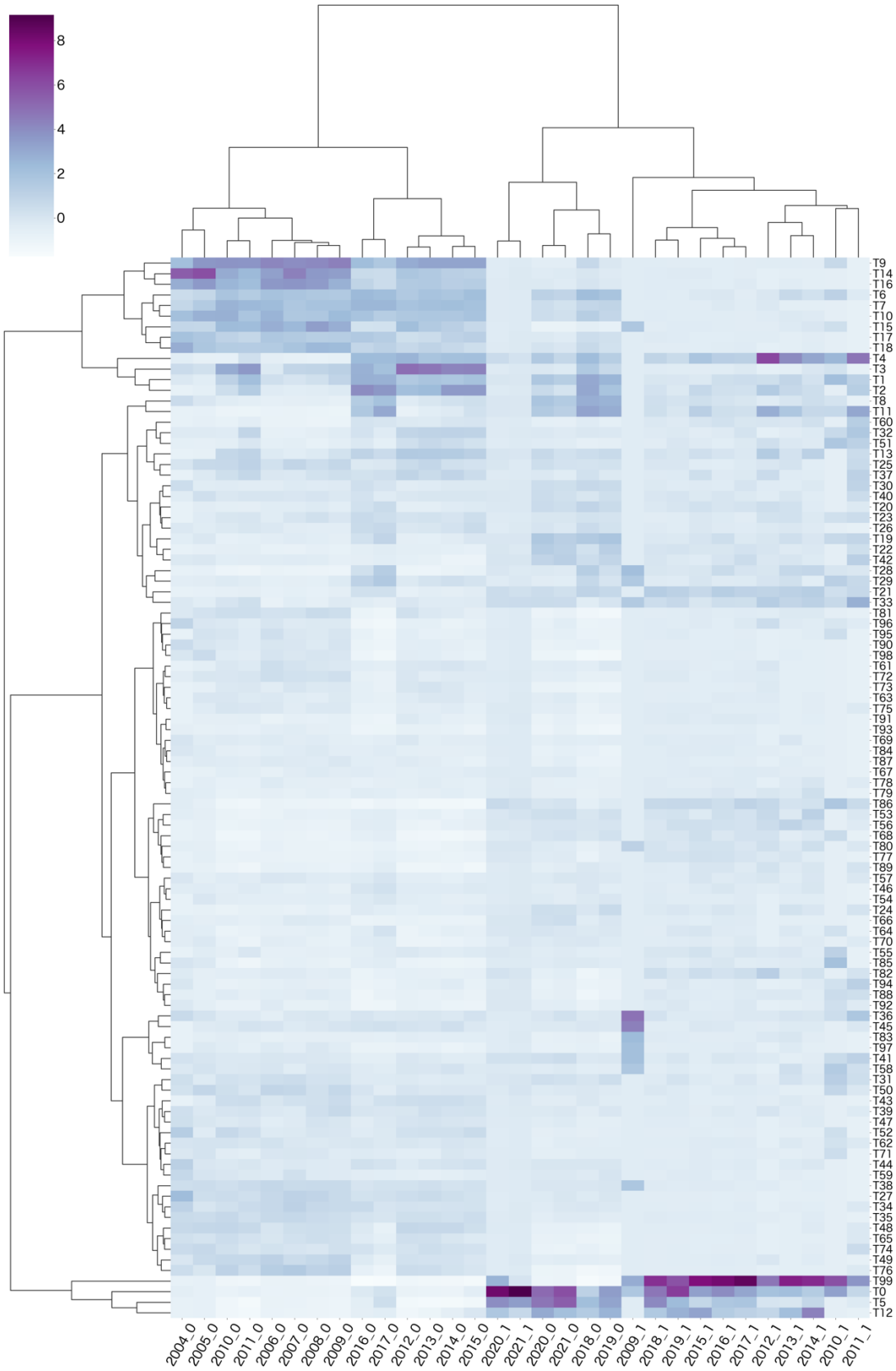


図 5.20　カテゴリーごとの各トピックの割合ヒートマップ図

続いてさらに人気作品とそれ以外の作品に分けて時系列変化を検討する。図 5.20 では年度に加え、人気作品のカテゴリーを「1」、それ以外のカテゴリーを「0」に分けてトピックの割合を可視化したものである。例えば、「2020_1」は 2020 年に投稿された人気作品であり、「2020_0」は 2020 年に投稿された人気作品以外のものである。クラスターの分岐から明らかなように、基本的には、人気作品とそれ以外の作品に明確な境界線が見られる。人気作品の方では、下側に位置付けられている Topic99、Topic0、Topic5、Topic12 の 4 つが非常に目立つことが分かる。ただし、2018 年から 2021 年までの人気作品とそれ以外の作品が非常に近く配置されていることも観測できる。その原因としては、人気度に関わらず、Topic0、Topic5、Topic12 の占めた重みが高いことだと考えられる。

表 5.1　年ごとの人気作品トピック（トップ 5）

2010			2011			2012			2013		
Rank	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**
1	**T99**	6	1	**T4**	9	1	**T4**	27	1	**T99**	50
2	**T0**	4	2	**T99**	8	2	**T99**	23	2	**T4**	28
3	**T1**	3	3	**T11**	6	3	**T11**	14	3	**T12**	18
4	**T4**	3	4	**T33**	6	4	**T0**	11	4	**T0**	18
5	**T55**	2	5	**T36**	5	5	**T2**	9	5	**T11**	13
2014			**2015**			**2016**			**2017**		
Rank	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**
1	**T99**	55	1	**T99**	94	1	**T99**	132	1	**T99**	162
2	**T12**	33	2	**T0**	43	2	**T0**	74	2	**T0**	64
3	**T4**	27	3	**T12**	40	3	**T5**	32	3	**T5**	37
4	**T0**	16	4	**T4**	24	4	**T12**	28	4	**T12**	28
5	**T21**	14	5	**T11**	17	5	**T4**	27	5	**T4**	28
2018			**2019**			**2020**			**2021**		
Rank	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**	**Rank**	**Topic**	**Size**
1	**T99**	161	1	**T0**	182	1	**T0**	375	1	**T0**	321
2	**T0**	112	2	**T99**	171	2	**T5**	170	2	**T5**	112
3	**T5**	91	3	**T12**	80	3	**T99**	131	3	**T12**	31
4	**T12**	42	4	**T5**	67	4	**T12**	82	4	**T33**	27
5	**T21**	39	5	**T21**	40	5	**T4**	38	5	**T21**	18

表 5.1 では投稿年ごとの人気作品における上位トピックを示している。表 5.1 からすれば、最初のブームは Topic4 の VR ゲームと Topic11 の負け組転生に関する物語であった。2013 年から Topic0 婚約破棄と Topic12 悪役令嬢が徐々に流行となり、2016 年からは Topic0、Topic5、Topic12 が定着してトップ 4 から外れることがない（Topic99 に関しては次の段落で述べる）。前述したように、2018 年から人気度に関わらずこの 3 つのトピックが高い割合を占めたことから、近年において題材や物語の設定に関しては、昔より全般的に多様性が低下していることが言えるのではないか。

図 5.21　Topic99 のワードクラウド

Topic99 はコミック化、書籍発売に関わるトピックであり、人気作品独特の話題だと図 5.20 のクラスターの分岐からわかる。内容からうかがえるように、おそらく書籍化やコミカライズが決定した後に宣伝効果を果たすため、作者があらすじを書き換える傾向があると推定できる。よって、Topic99 から書籍発売やコミカライズ、受賞などの実際の傾向情報も推測できるのではないかと考えられる。

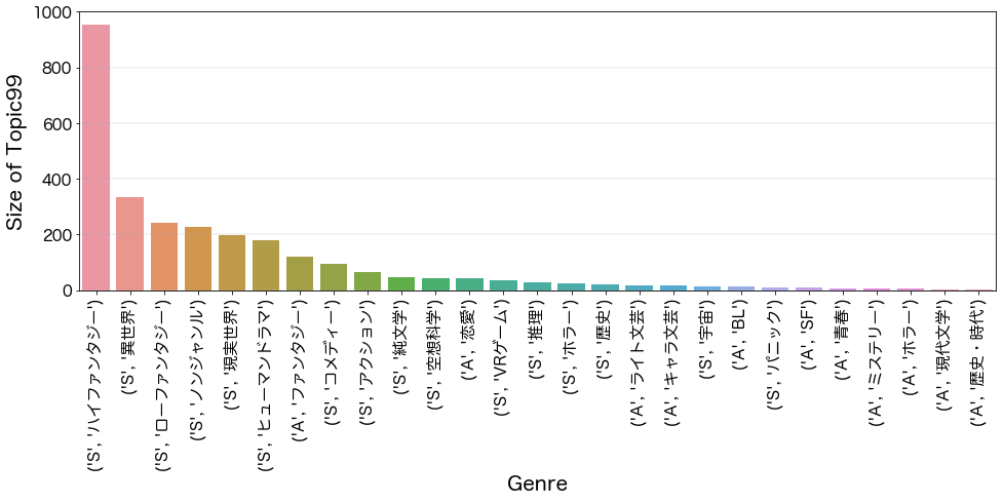


図 5.22　Topic99 のジャンルごとの数

図 5.22 ではあらすじが Topic99 に属する作品の数をそれぞれのジャンルごとに表示している。ハイファンタジーが極めて数が多く、最も書籍化やコミカライズの対象として好まれるのだと考えられる。続いては、異世界恋愛、現実世界、ローファンタジーやヒューマンドラマなど、基本的には人気ジャンルが出版されやすい印象である。ただし、そもそもハイファンタジーの作品数が著しく多いため、出版される作品の数が多いのも予想通りのことだと思われる。

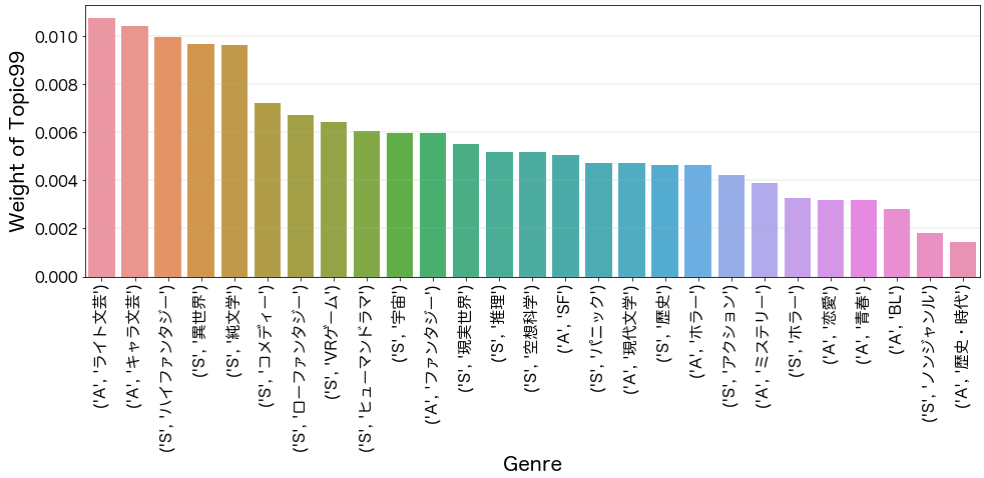


図 5.23　Topic99 のジャンルごとの割合図

図 5.23 では作品数ではなく、各ジャンルにおいて Topic99 に属する作品数の割合を表示した。割合からすれば、「アルファポリス」の方では、ライト文芸とキャラ文芸のようなライトノベルに近い作品が比較的に書籍化のチャンスを得やすいことが分かる。「小説家になろう」の方では、相変わらずハイファンタジーと異世界恋愛ジャンルが飛び出ているが、純文学が数の割に比率が高いことが興味深い現象である。

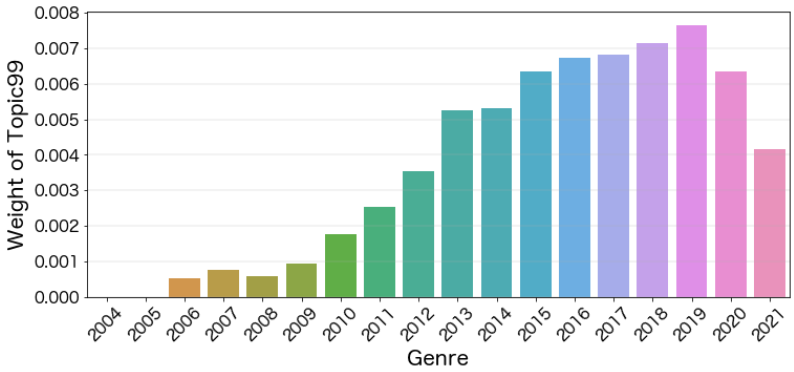


図 5.24　Topic99 の投稿年ごとの割合図

また、投稿年ごとの Topic99 の割合情報を図 5.24 で表している。時系列的に Topic99 の傾向を見れば、2019 年で頂点になり、2020 年からは落下していくことが分かる。ただし、書籍化やコミカライズ自体が没落したとは思えない。なぜなら、第 4 章のタイトル分析で既に述べたように、特に 2020 年と 2021 年に書籍化の情報がタイトルに特徴的に出現している。つまり、宣伝形式が転移していることが考えられる。あらすじより、タイトルでお知らせすることが 2020 年より好まれるようになったのではないか。また、2021 年は投稿年が近すぎて書籍化が進まない可能性もあると考えられ、引き続き発展的な課題として追跡する必要があるように思われる。

5.6 ジャンル別のトピック

5.6 ではジャンルに分けてそれぞれのトピックの分布状況を考察する。まずはヒートマップを通して、全体像を掴むことにする。図 5.25 のクラスタリングの結果を吟味すると、潜在的な意味をもつトピックからすれば、歴史小説、推理（ミステリー）小説、異世界恋愛と SF 系の作品に関してはトピックの多様性が少なく、特に歴史小説、異世界恋愛、VR ゲームと宇宙題材が非常に類型化されていることが分かる。ファンタジーもいくつかのトピックに偏っていることが見てとれる。それに対して、現実性と人間性を重視する文芸作品の方ではトピック選択が多い。

人気作品の方では、やはり多くのジャンルにおいて Topic99 が中心的なトピックになっていることが言うまでもない。また、異世界恋愛以外にも、多数のジャンルにおいて貴族婚を背景とする Topic0 が浸透しており、ジャンルに関わらずこのトピックを読者が受け入れやすいのではないかと考えられる。

さらにミクロな視点から考察するべく、特にファンタジー、恋愛、SF、推理小説および歴史小説を中心に、各ジャンルの具体的なトレンドや変化などを検討する方針で分析を進めていく。

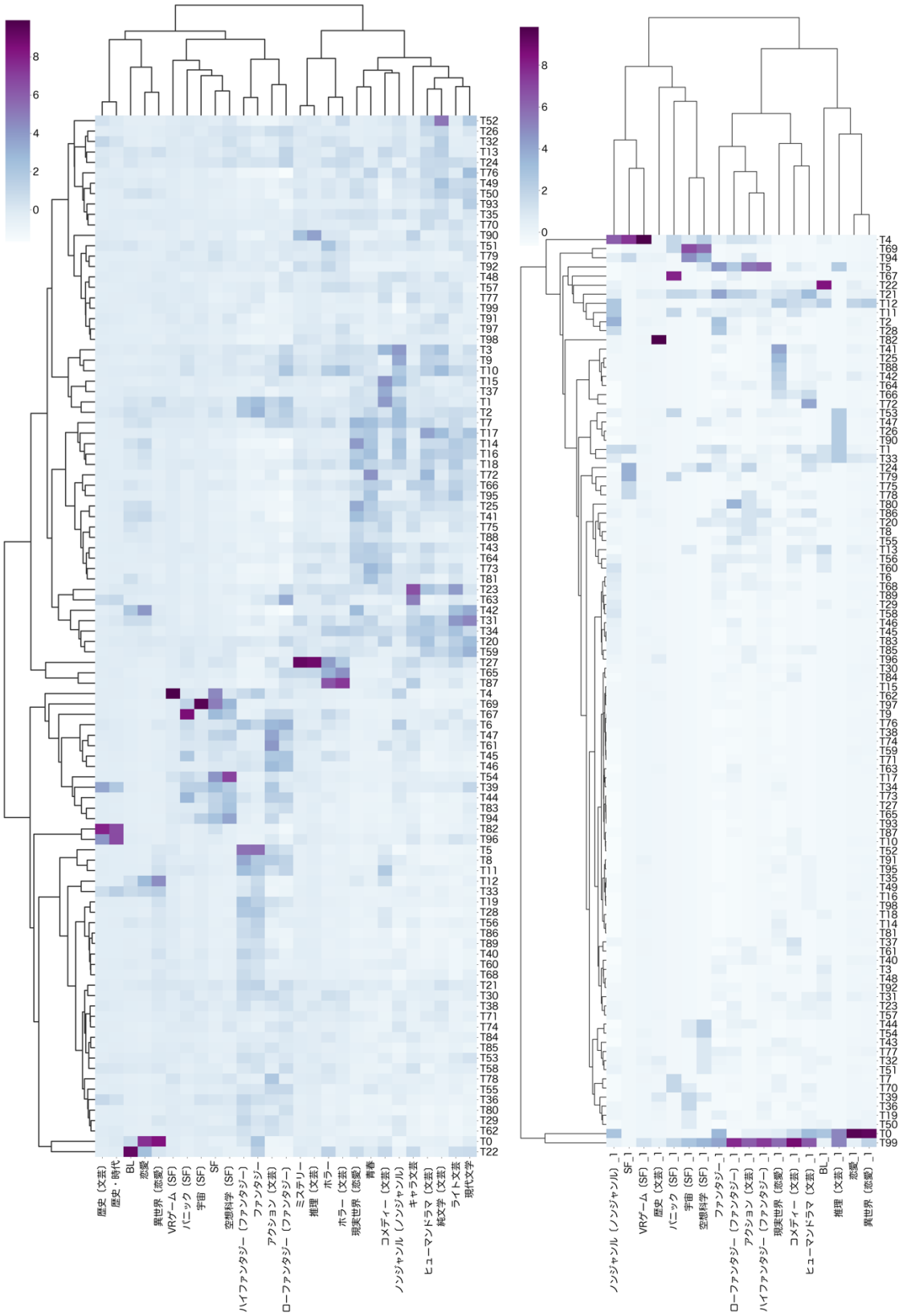


図 5.25　ジャンルとトピックのヒートマップ関係図（左：全作品　右：人気作品）

5.6.1 ファンタジー

この節からはジャンルに分けてトピックの分布状況を考察し、ジャンルごとの実態を明らかにする。5.6.1 ではファンタジーの全体状況を検討する。まずは、「小説家になろう」における異世界を舞台とするハイファンタジーを中心に検討する。

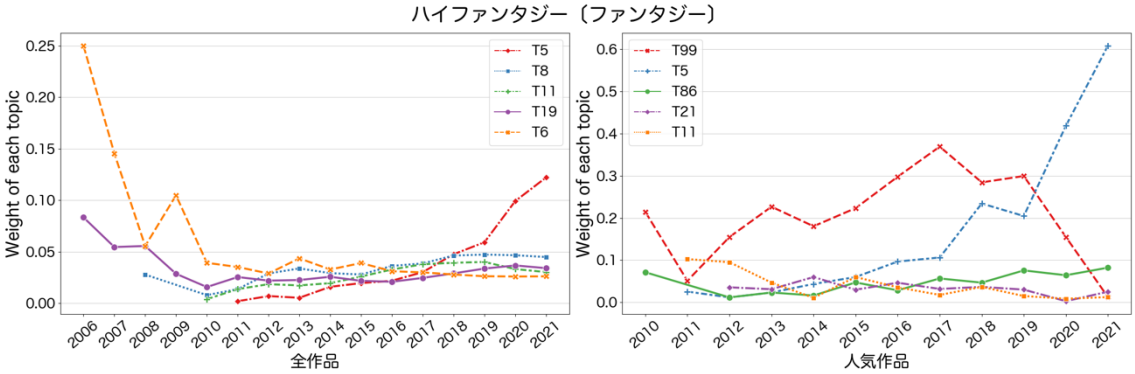


図 5.26　「小説家になろう」（ハイファンタジー）

図 5.26 では 2 つのグラフを提示している。左の図はハイファンタジーに属する全ての作品を対象に、全体の割合がトップ 5 に入るトピックの年ごとの平均重みを示している。右の図では対象を人気作品のみに絞っている。それぞれの時系列図の凡例では、色と形でトピックを見分け、さらに、サイズの高い順に上から下まで並べている。

人気度に関わらず、Topic5（パーティーから追放される物語）はファンタジーにおける超人気トピックであることは判断できる。特に人気作品の方では Topic5 の上昇度合いが一層激しい。2018 年から急上昇が見られ、2021 年で言えば 60%にも達し、ファンタジーを描く人気作品のうち半分以上が追放系のストーリーだという顕著な傾向が見てとれる。

全体の方では、Topic8 魔族討伐と Topic11 負け組の逆転がファンタジーの大きなトピックとして安定した成長を見せている。王都や農村を背景にして、悲惨な出身を持つ捨て子の成長などをめぐる Topic19 も重要なテーマとして存続していることが分かる。

図 5.27　Topic19 のワードクラウド

Topic6 は時系列的に減少傾向を見せてはいるが、2006 から 2009 年までのデータ数自体が少ないため、急下降しているとは言い切れない。ただし、割合が年々次第に減少していることは確かなことであり、2015 年までは最も使われていたが、2016 年からは徐々に他のトピックに追い抜かれるようになったことが見られる。Topic6 は戦乱や災厄に関わる話題であり、初期のファンタジー小説では多く書かれていたが、現在では魔族と人類の陣営戦をめぐる物語が定番になっているのではないか。

図 5.28　Topic6 のワードクラウド

Topic8 魔族討伐に関して、人気作品の方で上位に入っていないことは意外なことである。多く書かれているにもかかわらず、それほど読者の中では人気を稼いでいないのではないかと思われる。もちろん、Topic99（書籍化）と Topic21（作者のお詫び）は数が多いのにも関わらず物語の内容に触れておらず、実際のストーリー構成がわからないままなので、魔王と戦うような王道系が必ずしも評価されていないとも限らない。ここで言えるのはやはり人気作品としては、注目されるほど読者との関係が深まり、あらすじにおいて物語自体と関係のないことが書かれるのもありふれたことになっている。ここでもう一つ注目すべきなのは数自体が少ないが高評価作品が増加してきた Topic86 である。薬草師や錬金術師のような知恵や技術を運用する非戦闘系の職業をめぐる物語が好まれるようになっていることも面白い現象である。

図 5.29　Topic86 のワードクラウド

次に、「小説家になろう」におけるローファンタジーを考察する。

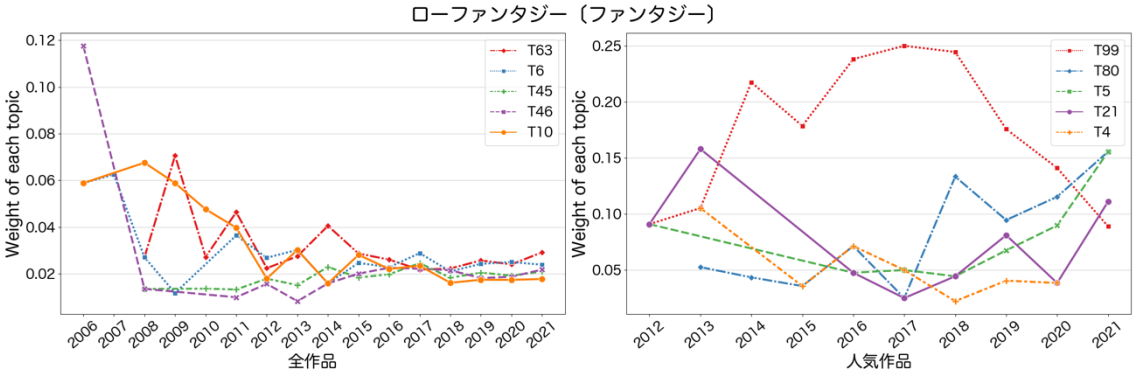


図 5.30　「小説家になろう」（ローファンタジー）

現実に近い世界を舞台とされるローファンタジーで、上位トレンドにおいて全作品と人気作品のトピックに重なりがないことが面白い現象である。Topic63（陰陽師）、Topic45（化け物）、Topic46（超能力者）などが多く描かれているのにも関わらず、高評価もしくは読者に人気なトピックには偏りがあり、作者が書きがちなものとの間にズレが生じていると考えられる。内容性が含まれない Topic99、Topic21 を除き、迷宮探索をめぐる Topic80 が人気作品の最上位に来ているとともに、2018 年からのトレンドにもなっていることが明らかである。

図 5.31　Topic80 のワードクラウド

また、ハイファンタジーにおいても高人気を得た Topic5 に関しては、ローファンタジーの 2018 年からの上昇トピックにもなっていることが図 5.30 から分かる。実際の具体例を見ると、ダンジョン即ち迷宮との関わりが深く、同時に弱者からスキルアップする追放系の重要な要素が含まれることが分かる。ローファンタジーにおける Topic5 のあらすじ全 26 件のうち「ダンジョン」というキーワードが含まれるものが 16 件見つかり、ヒット率が 62%にも至る。Topic80 と合わせて見れば、ダンジョンというモチーフがローファンタジーの趨勢であることに違いない。

ローファンタジーにおける Topic5 の具体例：

【小説家になろう ｜ 『最弱冒険者が【完全ドロップ】で現代最強　自分だけのレアスキルとカスタムアビリティを駆使して他の誰より強くなる！』 ｜2021_ローファンタジー】

天水ソラは、『底辺Ｆランク』と呼ばれる最弱の冒険者だった。Ｅランクの冒険者に暴行を受け、自前の装備を奪われた日、ソラはダンジョンで黒い玉を拾った。その玉の力が解放されたとき、ソラの目の前にはステータスボードが浮かんでいた。ランクの限界を突破して、レベルアップでステータスを分配する。アビリティをカスタムし、他の冒険者には出来ないビルドを行って行く。また超レアアイテムも、完全ドロップで確定入手。ダンジョン武具を次々と精錬して、ソラは強くなっていくのであった。――これは冒険者最強と呼ばれる者の、最弱から始まる物語である。毎日１８：００頃更新※本作はカクヨムにも同時投稿しております。

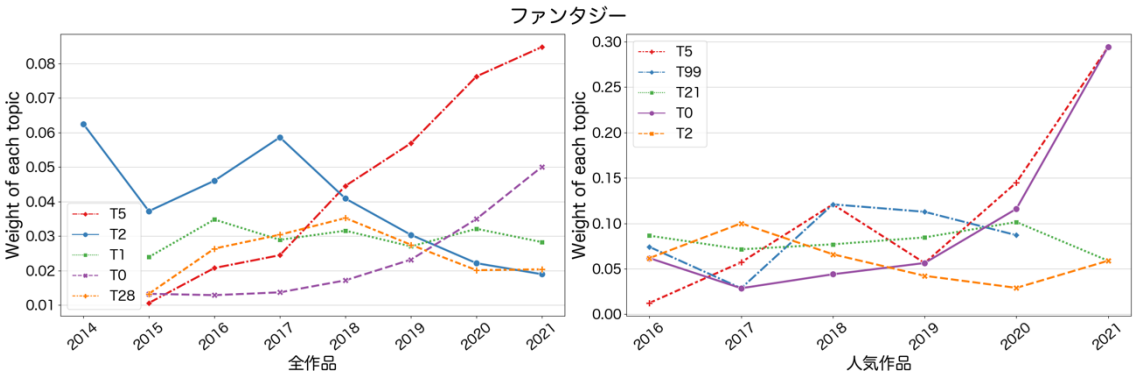


図 5.32　「アルファポリス」（ファンタジー）

一方で、「アルファポリス」においては「小説家になろう」の傾向と類似したように、やはり Topic5 追放系の超人気が見られる。これは投稿サイトや背景と関係なく近年のファンタジー小説に共通した大ヒットトピックだと言えるのであろう。また、図 5.32 から分かるように、異世界恋愛の定番だと思われる Topic0 婚約破棄が「アルファポリス」のファンタジージャンルにおいて人気トピックになっているのが非常に興味深いことである。実例を見れば、婚約要素を取り入れたファンタジー題材が多い。つまり恋愛とファンタジーが両立して進むストーリーが新たなトレンドになることも言えなくはない。

ファンタジーにおける Topic0 の具体例：

1.　【アルファポリス｜『(完)普通令嬢でも救世主になれるのか』｜2021_ファンタジー】

アイリス・アンダーソンは、婚約解消をされた男爵令嬢。１９歳なので後が無い。貧乏な貴族の令嬢でも自分の運命を良くする為に頑張ります！ゆるい、世を救うヒロインの物語。

2.　【アルファポリス｜『愛されない王妃は王宮生活を謳歌する』｜2021_ファンタジー】

小国メンデエル王国の第 2 王女リンスターは、病弱な第 1 王女の代わりに大国ルーマデュカ王国の王太子に嫁いできた。政略結婚でしかも歴史だけはあるものの吹けば飛ぶような小国の王女などには見向きもせず、愛人と堂々と王宮で暮らしている王太子と王妃のようにふるまう愛人。まあ、別にあなたには用はないんですよわたくし。私は私で楽しく過ごすんで、あなたもお好きにどうぞ♡恋愛要素の強いファンタジーです。初投稿です。世界観を楽しんでくだされば幸いに存じます。

最後に、ローファンタジーの上位トピックに出現しているTopic10 に関して深掘りしていく。Topic10 は全 100 個のトピックのうち、唯一の環境を描くトピックだと観察されている。図 5.33 のキーワードからすれば、明かりがつく夜、または夕方や夜になりかける頃の景色についての描写が多いと判断できる。天気状況に関わる雨の描写もキーワードに含まれている。要するに、あらすじに環境描写を中心とする場合は暗い要素が含まれることが多いと考えられる。

図 5.33　Topic10 のワードクラウド

それぞれのジャンルにおけるTopic10 の割合分布図 5.34 を描画したら、実はホラー系の作品に最もよく使われることが分かる。ローファンタジーも上位にはきているが、特徴トピックとまでは言えない。ホラーの作品で恐怖の雰囲気を醸し出せるように環境描写を用いることが多いのではなかと考えられる。また、純文学にも特徴的に出現していることは興味深い。やはり芸術性を重視するジャンルでは環境描写が大いに用いられるのであろう。

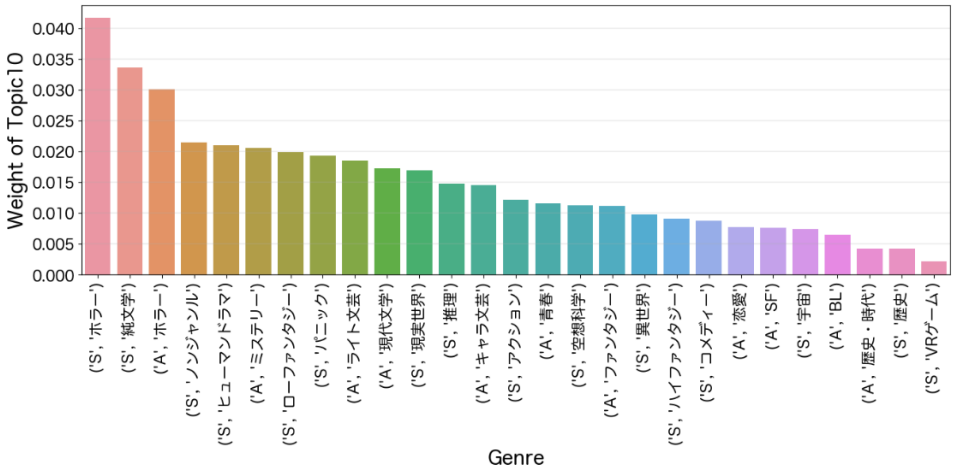


図 5.34　Topic10 のジャンルごとの割合図

5.6.2 恋愛

続いて、恋愛ジャンルについて考察を進める。

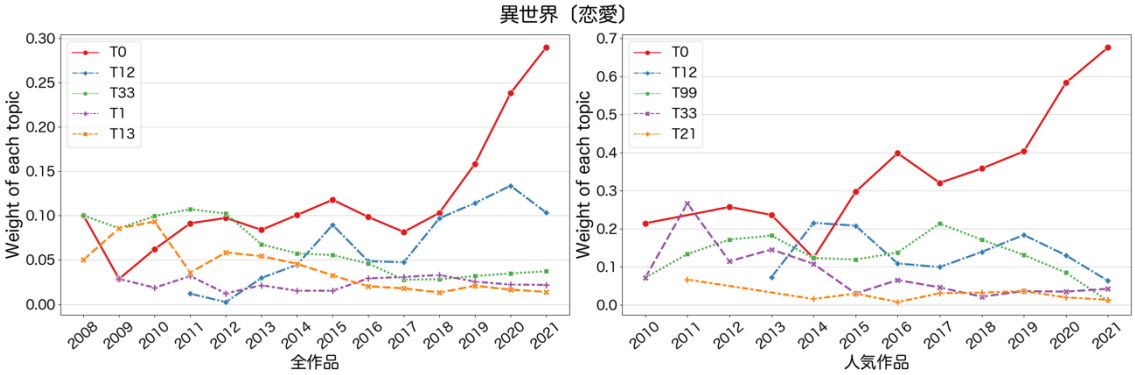


図 5.35　「小説家になろう」（異世界恋愛）

異世界を舞台とする恋愛ジャンルでは、Topic0 皇室や貴族をめぐる婚約破棄が最も人気であることがわかる。また、乙女ゲームにおける悪役キャラが近年好まれるようになっていることも Topic12 から見られる。2018 年からはこの 2 つのトピックから展開する物語がほとんどである。特に婚約破棄や政略結婚をめぐる話題が急上昇していき、人気作品の方を見ればさらに明らかなように、婚約者物語が 2015 年から既に制覇しており、そのトレンドが次第に進み、2021 年で言えば高評価作品のうち 70%がこの話題を中心としたストーリーである。恋愛トピックの多様性の低下も確証される。

2012 年までは Topic33 王位争奪や後宮話が人気であったが、近年ではこうした伝統的な異世界物語の割合が次第に減少していることも図 5.35 から明らかである。

図 5.36　Topic33 のワードクラウド

続いては現実世界を舞台とする恋愛作品のトピック傾向解析を行う。

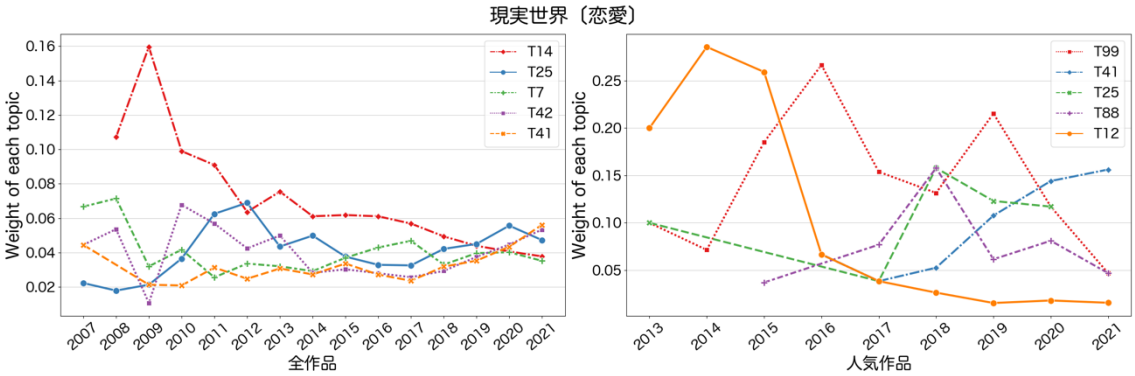


図 5.37　「小説家になろう」（現実世界恋愛）

現実世界をめぐる恋愛ジャンルにおいて、やはり片思いが最も大きな主題だということは図 5.37 の Topic14 と Topic25 から分かる。Topic14 は前述した甘酸っぱく切ない気持ちが中心的な話であり、Topic25 は内気、引っ込み、人見知りなどの性格特性に重点が置かれている。Topic14 が年々減少していることから、悲恋をめぐるストーリーが次第に書かれなくなってきたことが言えるのではないか。それと対比的に、Topic41 可愛い系彼氏をめぐるトピックが近年徐々に上がってきている。これは人気作品の方でも上昇傾向が見られ、甘い恋愛が近年のトレンドになっていると言える。キャラ設定から言えば、Topic25 の不器用系と Topic41 の可愛い系が人気だと思われる。

図 5.38　Topic25 のワードクラウド

図 5.39　Topic41 のワードクラウド

また、人気作品のトレンドから観察できるのは、Topic88 が 2015 年から高評価作品に選ばれやすいことである。Topic88 に関しては、完璧キャラという要素が入っていることが図 5.40 から読み取れる。もちろん、実例を吟味すると、人前では完璧な人物が人に知られない一面も持っているような設定が多いようであるが、完璧キャラという表設定があることは間違いない。こうした設定が近年において人気を得やすいことも明らかである。

図 5.40　Topic88 のワードクラウド

Topic88 の具体例：

1.　【アルファポリス｜『悪役令嬢は腐女子である』｜2021_恋愛】

容姿端麗。頭脳明晰。成績優秀。品行方正。運動神経抜群。　全て完璧なお嬢様である悪役令嬢の秘密。それは腐女子であること。　名もないモブが偶然その秘密を知ったことから始まるラブコメ。

2.　【小説家になろう｜『王子君は私にだけ冷たい』｜2016_現実世界〔恋愛〕】

容姿端麗、成績優秀、運動神経抜群なイケメン『王子君』は、何故か私にだけ冷たい。何で？

3.　【小説家になろう｜『如月先輩』｜2014_現実世界〔恋愛〕】

俺の通う学校には、学園のアイドルと呼ばれる女生徒がいる。如月姫紗良（きさら）。容姿端麗、成績優秀、運動神経抜群、品行方正……。そんな完璧過ぎる彼女にも、やはりというか、欠点や弱点は存在し……。

4.　【小説家になろう｜『キレイな花には棘がある』｜2016_現実世界〔恋愛〕】

至って普通の男子高校生佐藤の学校、桜木高校には高嶺の花がいた。浅倉まゆ。運動神経抜群、頭脳明晰、家柄もいい。そんな彼女には決定的にダメなところがあった。それは…毒舌だったのだ。

5.　【小説家になろう｜『僕の恋愛事情は複雑です』｜2021_現実世界〔恋愛〕】

学年１のイケメン男子成績優秀でスポーツ万能だけど彼女無し彼女居ない歴＝年齢！！女子達に言えない秘密とは・・・

図 5.37 に見られる Topic12 は実際背景から言えば異世界のはずであるが、2016 年以前は作者による誤分類の可能性が高い。ただし、2013 年から 2015 年までに見られた高い割合に関しては間違いとは言えない。

セフレや不倫浮気（Topic42）に関わる背徳的な恋要素を含む作品も近年においてまた増加していることは図 5.37 から分かる。もちろん、必ずしも物語の主人公が背徳したわけではないが、事情のある大人の恋が語られていることは確かである。このトピックが近年上がってきているというのは、純愛的な学園ストーリーが主流であるにもかかわらず、会社を舞台とする複雑な大人の恋を描く人も増えてきたことを表しているのであろう。

図 5.41　Topic42 のワードクラウド

Topic42 は「アルファポリス」の恋愛ジャンルの大きなトピックでもある。「アルファポリス」の恋愛トピックのトレンドと合わせて言えるのはまず婚約破棄トピックが主導的な地位にあることである。このトレンドがさらに続くのではないかと思われる。また、悪役令

嬢や大人の恋に関わるトピックが人気であることからも、やはり近年では純愛的なものより、もっと刺激的な恋愛ストーリーが好まれるようになっている。

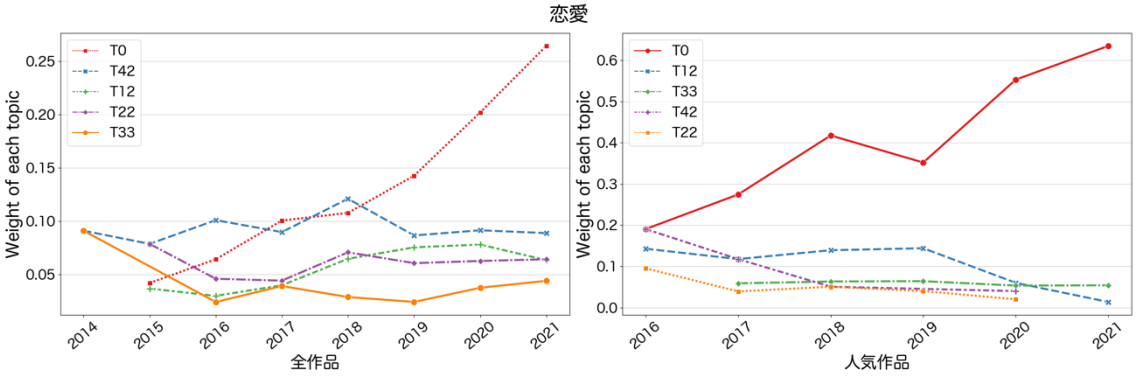


図 5.42　「アルファポリス」（恋愛）

最後に、1 つ特殊なトピックを検討する。現実世界の恋愛ジャンルに多く出現している Topic7 に関しては全てひらがなから構成され、非常に興味深くて単語のみで解釈することが難しい。実際の具体例を見れば、名前の振り仮名が多く出現していることが明らかになった。その上、複数の名前が同時に出現するパターンも多い。

図 5.43　Topic7 のワードクラウド

Topic7 の具体例：

1. 【小説家になろう | 『きっとそこには愛がある』 | 2015_ノンジャンル】

女子が苦手、見た目は不良の霧宮紳二《きりみやしんじ》が繰り広げる学園青春ラブコメ。

---

2. 【小説家になろう | 『僕の敵は〇〇だ』 | 2010_ノンジャンル】

主人公の龍旛　襲哉（たつば　しゅうや）は、親友の緩崎　隊徒（かんざき　たいと）と逢真津　憂揮（あいまつ　ゆうき）と弾那賀　冥人（たまなか　めいと）。クラスメイトの劉願磁　砂爺（りゅうがんじさや）と懐山　御久（なつやま　みく）たちと繰り広げるバトルコメディー。

---

3. 【小説家になろう | 『girl or girl』 | 2011_ノンジャンル】

空（そら）円巻優姫（えんまきゆき）滝沢菜夜日（たきざわなやか）三人が描く三角関係ラブストーリー。

---

4. 【小説家になろう | 『この恋を叶えるために』 | 2016_現実世界〔恋愛〕】

中学の卒業式、加藤蓮〈かとうれん〉は幼馴染の琴原春香〈ことはらはるか〉に告白し二人は結ばれるが、ある日、春香が事故に遭い記憶を失う。

---

5. 【小説家になろう | 『彼女の失われた青春』 | 2016_現実世界〔恋愛〕】

豊平高等学校《とよひらこうとうがっこう》に通う高 2 になったばかりの桐原翔太。翔太は彼女が欲しいと幼馴染みの花澤姫香《はなざわひめか》に相談すると星の降る夜にお願い事をしに学校へ。そこで、出会った不思議な少女、結城奈々《ゆうきなな》と一緒に青春を謳歌することとなる。

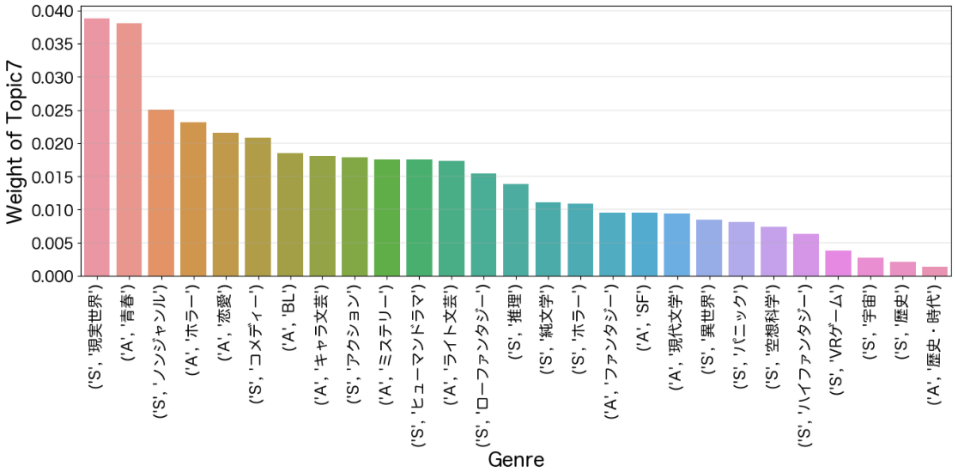


図 5.44　Topic7 のジャンルごとの割合図

Topic7 のジャンル分布図 5.44 を観察すれば、特に現実恋愛と青春において非常に高い割合が見られる一方、ファンタジーや SF における割合は非常に低い。Topic7 と類似したトピックとしてあげられるのは Topic18 である。Topic18 は漢字の名前から構成されたトピックだと思われ、現実世界を舞台とする作品に非常に多く出現していることは Topic7 と同様である。つまり、文芸作品や純愛作品において、あらすじに人物の名前と人物間の関係が重点的に書かれる特徴があるのではないか。特に図 5.45 のワードクラウドにおける「同級」「始

業」「終業」などから学園物語を描く場合は具体的な名前の出現頻度が高いと言えるのではないか。興味深いと思われるのは、「美咲」という名前の相関性が非常に高くて、頻用されることが Topic18 から分かる。

図 5.45　Topic18 のワードクラウド

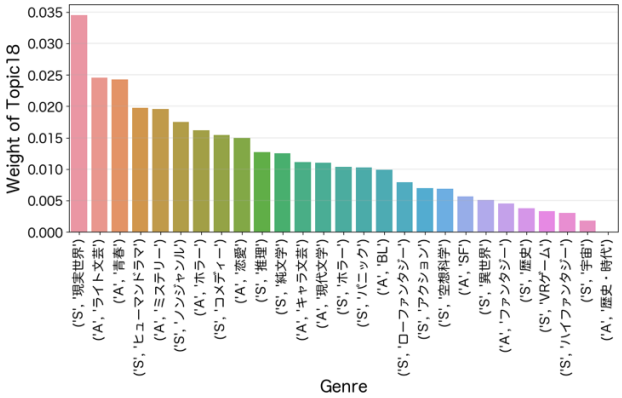


図 5.46　Topic18 のジャンルごとの割合図

明治安田生命の人名調査[49]の結果からすれば、「美咲」は 1991 年から 2014 年までの間において、ほぼ毎年女の子の名前ランキングのベスト 10 にランクインしている。やはり小説人物の名づけに関しても現実世界に影響されるのではないかと推測できる。

ここでは、現実世界を舞台とする恋愛ジャンルに限定し、ウェブ小説の人名調査を行い、その特徴を明らかにすることを試みる。表 5.2 では名前の上位 20 位の項目と頻度をまとめている。「ランクイン」という項目は、明治安田生命によって行なわれた生まれ年別ベスト 10 の平成以降のランキング情報[50]を参照し、表中の名前が当たれば○が付けられる（人名の分かち書きは分割ツールの精度にもよるが、名前自体が分かれて識別されることもあるため、ここでは字が名前に含まれれば、当たると判定する）。比率で言えば、現実恋愛小説の人名トップ 20 のうち、4 分の 3 はリアル世界のトレンド名前と重なり合うことがわかる。やはり現実世界の影響があると考えられる。

表 5.2　名前頻度トップ 20（現実世界恋愛ジャンル）

Rank	名前	頻度	ランクイン	Rank	名前	頻度	ランクイン
1	優	305	○	**11**	希	110	○
2	翔	230	○	**12**	遥	109	○
3	蓮	218	○	**13**	透	109	
4	悠	163	○	**14**	楓	107	○

[49] https://www.meijiyasuda.co.jp/enjoy/ranking/year_men/girl.html
[50] https://www.meijiyasuda.co.jp/enjoy/ranking/year_men/

5	葵	137	○	**15**	華	105	
6	咲	133	○	**16**	太	99	○
7	桜	113	○	**17**	進	94	
8	奏	111		**18**	玲	94	○
9	真	110	○	**19**	輝	88	○
10	ハル	110		**20**	結衣	86	○

また、苗字についても調査を行い、表 5.3 では頻度上位 20 の苗字と「名字由来 net」[51]において調べた当該苗字の全国順位を示している。「佐藤」「鈴木」「田中」「高橋」「山田」「佐々木」以外は、基本的に稀なものに偏っている。「柊」「橘・立花」のような貴族を想起させる古めかしい苗字が頻繁に使われることは面白い現象である。また、「一ノ瀬」「雨宮」「佐倉」「白石」などは芸能人、声優、二次元から影響を受けたのではないかとも考えられる。やはり身近なトレンドを取り入れやすいことが言えなくもない。

表 5.3　苗字頻度トップ 20（現実世界恋愛ジャンル）

Rank	苗字	頻度	全国順位	Rank	苗字	頻度	全国順位
1	佐藤	239	1	**11**	一ノ瀬	81	1408
2	柊	120	12603	**12**	雨宮	81	774
3	橘	119	558	**13**	佐倉	79	3843
4	鈴木	117	2	**14**	原	75	62
5	立花	111	636	**15**	七瀬	64	15507
6	田中	108	4	**16**	斎藤	63	233
7	高橋	105	3	**17**	小鳥遊	61	58615
8	神崎	99	787	**18**	白石	61	220
9	佐々木	98	13	**19**	天野	61	24
10	山田	84	12	**20**	成瀬	60	643

5.6.3 SF

SF（サイエンス フィクション）ジャンルの傾向に関しては、「小説になろう」の方では 2012 年より vrmmorpg に関わる Topic4 の圧倒的な人気が見受けられる。人気作品の方ではほぼ VR ゲームに支配されていることも明白であり、そのトレンドは 10 年間ずっと継続している。

[51] https://myoji-yurai.net/prefectureRanking.htm

図 5.47　「小説家になろう」（SF）

「アルファポリス」の方では明らかに VR 中心とはいえないが、Topic 4 の上昇傾向が明らかである。やはりゲーム題材の人気はこれからも続くのではないかと思われる。これは発展的な課題として、何年後にまた調査を続ければさらなる発見ができると思われる。（「アルファポリス」の方では SF 人気作品の数が非常に少ないため時系列図にはできず図 5.48 では全作品の図のみ表示している。ただし Topic4 の数がそのうち最も多いことが統計された）

他のトピックに関してもやはり両方の投稿サイトで重複することが多く、宇宙惑星をめぐる Topic69、人工知能に関わる Topic54 と感染ウィルスのパニックトピックが SF の定番トピックであることは確実である。

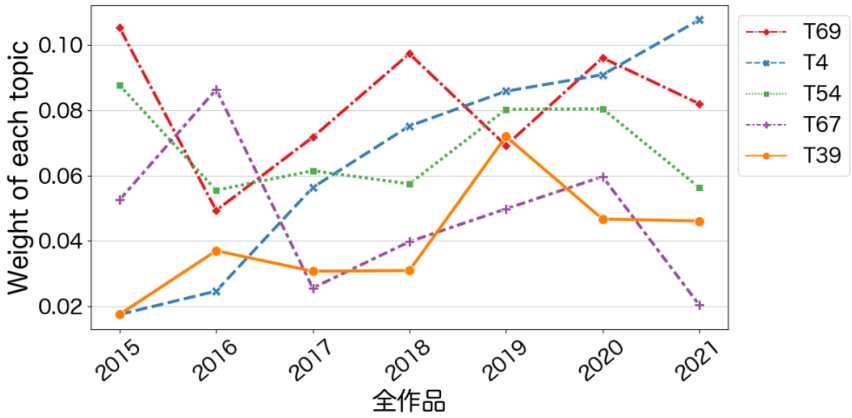


図 5.48　「アルファポリス」（SF）

5.6.4 推理（文芸）／ミステリー

推理小説に関しては、本格派、社会派、ハードボイルド、警察小説などさまざまな分類が存在し、主人公のキャラ設定も多種多様であり、長い歴史の中で色々と派生や変化が生じてきた。ではウェブ小説としての推理小説はどのようなトピックが最も書かれているのであろうか。

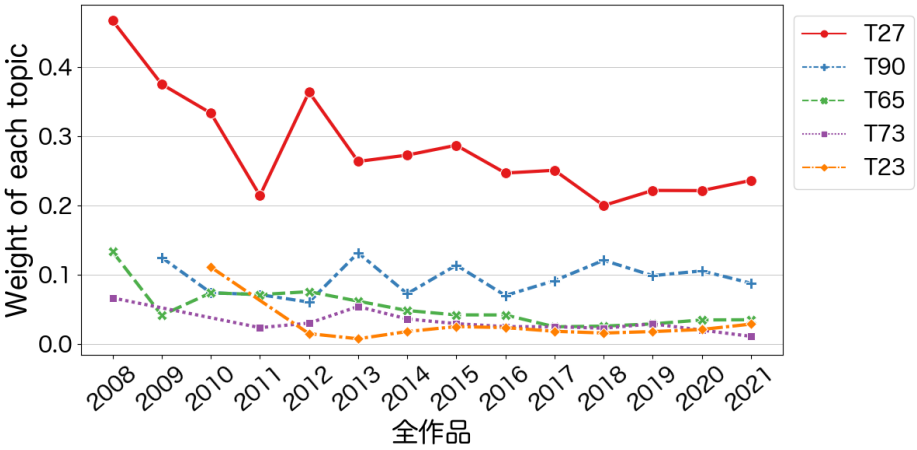


図 5.49　推理（文芸）／ミステリー

図 5.49 から明らかなように、Topic27 が最も大きなトピックとなっており、ここ 10 数年ではこのトレンドが変わることがない。図 5.50 の主要単語から推定できるように、Topic27 は刑事が犯人を探すような話題であり、つまり刑事物語が最も書かれていると結論づけられる。

また、Topic90 は「探偵」「助手」を中心とするトピックで本格推理小説によく出現する要素だと判断できる。「ホームズ」「ワソトン」との関連からシャーロックシリーズの二次創作が多いのではないかと思われる。こうした探偵が主人公となるパターンが依然として推理ジャンルにおいて一定の割合を持つが、刑事小説が既に主流になっていることは言うまでもない。現実では法的な事情に基づき、人命に関わる事件であれば探偵ではなく刑事が出動するため、刑事が主人公になるのも読者にとって受け入れやすい設定だと思われる。VR ゲーム題材の流行と同じように、小説は時代に影響されやすいと考えられる。

図 5.50　Topic27 のワードクラウド

図 5.51　Topic90 のワードクラウド

5.6.5 歴史物語

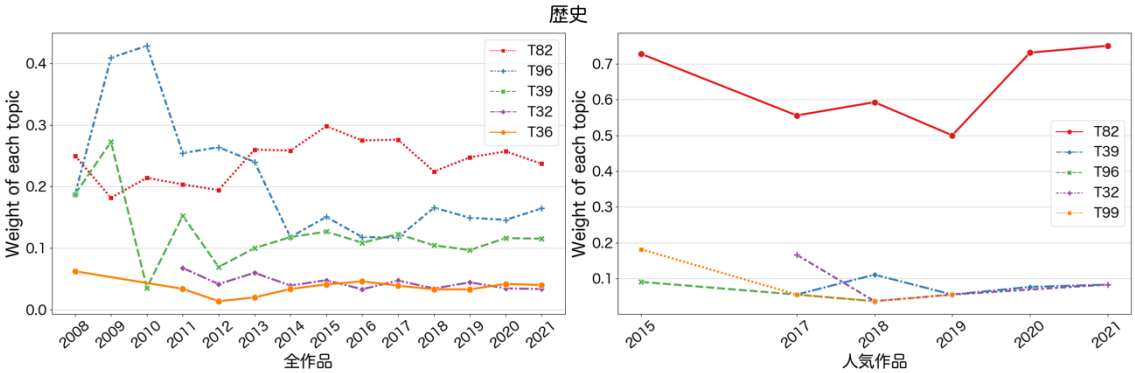


図 5.52　歴史（文芸）/ 歴史・時代

図 5.52 から分かるように、歴史の物語において、Topic82 戦国時代が圧倒的に人気である。特に図 5.53 のキーワードから、織田信長に関わる物語が好まれることが読み取れる。Topic82 に続き、Topic96 も大きなテーマとして書かれている。図 5.54 から明らかなように、Topic96 は江戸川幕府の末期と浪士をめぐるトピックだと判断できる。Topic96 のトレンドから言えば、2009 年と 2010 年で頂点に至っており、つまり初期のウェブ歴史小説ではこの幕末時代を描くものが最も多かったと思われるが、2011 年よりその割合が次第に減少し、戦国物語に追い抜かれた。ただし Topic96 は依然として大きなトピックとして存在することに相違ない。

図 5.53　Topic82 のワードクラウド

図 5.54　Topic96 のワードクラウド

Topic82 と Topic96 はどちらも比較的に遠い古代を舞台とするトピックである。この 2 つに及ばないが、Topic39 も比較的に好まれて書かれる歴史トピックだということは図 5.52 か

ら読み取れる。Topic39 は「陸軍」「海軍」「戦隊」など近代戦争に必要不可欠な要素から、近代軍備を備えた戦争物語だと言える。要するに、歴史物語では、近代戦争より大河的な物語がより好まれ、またいくつかの歴史時代のうち、日本の戦国時代が最も人気であり、とくに織田信長と関連のあるストーリーが好かれ、最も評価されるのもこの題材とした小説である。

図 5.55　Topic39 のワードクラウド

5.7 異世界転生・転移トピック

5.6 では各ジャンルにおける肝心なトピックを検討してきたが、異世界転生[52]と異世界転移[53]の要素が含まれる作品に関してはトピックがどのような構成比率を成しているのかについても考察に値すると思われる。

表 2.1 でも示したように、「小説家になろう」の方では異世界転生か異世界転移であるかどうかを示すラベルが存在する。もし「istensei」が「1」である場合、その作品が転生作品であり、「0」の場合は転生ではない。「istenni」も同じパターンで転移かどうかを表している。「アルファポリス」の方ではそのような正式な辞書が作られていないため、ここでは「小説家になろう」に絞って検討を進める。

表 5.4 では「小説家になろう」全体、下位ジャンルの異世界ファンタジーと異世界恋愛の 3 つに分けて、それぞれ転生要素が入ったトピックを表示している（上位 10 つまで）。Size が示すのはそれぞれのトピックに属する転生作品の数であり、Weight は当該トピックがどれほどの割合を占めているかを表す。

---

[52] 死んでから異世界にいく。
[53] そのまま異世界にいく。

表 5.4　異世界転生トピック(トップ 10)

全体				異世界ファンタジー				異世界恋愛			
Rank	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**
1	**T12**	3,660	0.08	1	**T11**	1,645	0.061	1	**T12**	2,615	0.250
2	**T11**	2,399	0.053	2	**T28**	1,643	0.061	2	**T0**	985	0.094
3	**T28**	2,203	0.048	3	**T5**	1,178	0.044	3	**T1**	354	0.034
4	**T1**	1,742	0.038	4	**T68**	1,046	0.039	4	**T11**	342	0.033
5	**T68**	1,542	0.034	5	**T2**	1,034	0.039	5	**T89**	262	0.025
6	**T2**	1,454	0.032	6	**T1**	1,033	0.039	6	**T28**	260	0.025
7	**T5**	1,440	0.032	7	**T89**	961	0.036	7	**T68**	251	0.024
8	**T89**	1,424	0.031	8	**T8**	852	0.032	8	**T2**	214	0.020
9	**T0**	1,321	0.029	9	**T21**	707	0.026	9	**T13**	147	0.014
10	**T8**	1,174	0.026	10	**T12**	633	0.024	10	**T21**	144	0.014

全体としては、Topic12 即ち乙女ゲームに転生する物語が最も多くて、このトピックは異世界恋愛における最上位でもある。Topic11 は前述したように、現実世界における負け組が新たな人生を始めるといったような話題である。これも転生系ファンタジーのあらすじにおいて最も好まれるパターンだと言える。Topic28 は Topic11 と互角するように重要なトピックだと割合の大きさから読み取れる。

図 5.56　Topic28 のワードクラウド

Topic28 のキーワードからすれば、お詫びや手違いなどに関わるトピックであり、実際のあらすじを吟味すれば、神様のミスで主人公が死んでしまい、異世界に転生するとともに、チートも手に入れるようなパターンが多い。こうしたシーンから始まる物語が異世界ファンタジーの典型的なプロット展開でもあると言えるのではないか。

Topic28 の具体例:

1. 【アルファポリス | 『遊星君、力技で何でも解決(仮名)』 | 2018_ファンタジー】

神様のミスで死んでしまった遊星君、お詫びとして特典を貰って、剣と魔法の世界へと転生する事になった。最強の体と魔力を貰うハズだったのだが…

---

2. 【小説家になろう | 『（仮）』 | 2018_アクション〔文芸〕】

神の手違いという不運な事故に巻き込まれた男の話

---

3. 【小説家になろう |『神をも超えるスキルを持って異世界を完全攻略する!!!』| 2018_異世界〔恋愛〕】

ある日、トラックにしゃがれそうになっている幼女を助けたら。その人が管理神であってお詫びとして、チート能力を一つもって異世界に転生させてくれるというから、スキル創造をもらって異世界最強を目指す

---

4. 【小説家になろう | 『異世界で俺なりに頑張る！』 | 2018_ハイファンタジー】

主人公は神様の手違いで死んでしまった、なので主人公は神様から 3 つスキルを貰い、異世界転生することを決意、はたして異世界で主人公はどうするのか。不定期投稿ですが、これから頑張っていくのでよろしくお願いします。

---

5. 【小説家になろう | 『俺は何度でもやり直せる』 | 2015_ハイファンタジー】

平凡なサラリーマンの俺は、ある日天使のミスで異世界に転移してしまう。お詫びに 3 つの願いを叶えてくれるというので、俺は願った。「人生を何度でもやり直せるようにしてくれ」

また、転生の原因と見られるトピックとして、Topic68 と Topic89 が挙げられる。Topic68 は「天寿」「来世」が中心語で、今世が終わってからの生まれ変わりというパターンになる。一方で、Topic89 は車に撥ねられて事故死するトピックだと判断できる。要するに、転生する大きなチャンスとして、神様のミス、人生が終わった後の生まれ変わり、また交通事故に遭うという 3 つのパターンが多く用いられると考えられる。3 つのうち、神様のミスでチートをもらい転生するパターンが最も好まれる。もちろん、ストーリーの内容性が含まれるトピックからすると、Topic5 と Topic8 をめぐるファンタジーの世界観が好まれることに変わりはない。

図 5.57　Topic68 のワードクラウド

図 5.58　Topic89 のワードクラウド

異世界恋愛の方で乙女ゲームと悪役令嬢をめぐる Topic12 の割合が非常に高い。つまり、女の子が転生の主人公となる傾向があり、悲惨なシナリオを破るというような展開が多いと思われる。それに対して、ファンタジーの方では特に失敗者の性質を持つ男性が主要人物になり、成功する新たな人生を送るパターンが多いのではないか。これは現実社会のストレスを反映し、それに対抗したい願望が小説を通して表されているのであろう。

次に、転移要素が含まれる作品のトピック構成を表 5.5 に表示している。

表 5.5　異世界転移トピック(トップ 10)

全体				異世界ファンタジー				異世界恋愛			
Rank	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**
1	**T11**	2,352	0.044	1	**T11**	1,571	0.049	1	**T1**	351	0.04
2	**T29**	2,029	0.038	2	**T5**	1,318	0.041	2	**T12**	327	0.037
3	**T1**	1,913	0.036	3	**T29**	1,306	0.041	3	**T29**	320	0.037
4	**T4**	1,697	0.032	4	**T8**	1,197	0.038	4	**T11**	309	0.035
5	**T8**	1,694	0.032	5	**T28**	1,153	0.036	5	**T0**	235	0.027
6	**T5**	1,634	0.03	6	**T1**	1,133	0.036	6	**T20**	196	0.022
7	**T2**	1,586	0.029	7	**T2**	1,127	0.035	7	**T2**	182	0.021
8	**T28**	1,570	0.029	8	**T4**	1,078	0.034	8	**T8**	181	0.021
9	**T21**	1,114	0.021	9	**T21**	784	0.025	9	**T57**	176	0.02
10	**T20**	1,057	0.02	10	**T60**	683	0.021	10	**T28**	175	0.02

表 5.5 から明らかなように、異世界転生にしても異世界転移にしても、Topic11 のキャラ設定が共通して重要な要素になっていることが分かる。また注目すべきトピックは Topic29 である。「マンホール」「足元」「吸い込まれる」「覚める」などのキーワードより、転移の仕方などに関わるトピックだと判断できる。実例を見れば、やはり吸い込まれて異世界に転送されたというような仕組みで、目覚めたら異世界にいたというような設定が多い。これは転移作品のあらすじによく用いられるパターンだと考えられる。

図 5.59　Topic29 のワードクラウド

Topic29 の具体例：

1. 【小説家になろう｜『帰宅冒険記』｜2019_ローファンタジー】

目が覚めたら異世界に来てしまった青年。彼の目的はただ 1 つ。元の世界に戻って家に帰るのみ。右も左もわからない世界で、彼は家に帰れるのか？

---

2. 【小説家になろう｜『高校卒業前に異世界に飛んで死にかけた件』｜2016_異世界〔恋愛〕】

目が覚めると見知らぬ世界にいたどこなのかもわからない元の世界には帰れるのか？

---

3. 【小説家になろう｜『とある者の異世界転移』｜2017_ローファンタジー】

赤山田久は、ヒーローだ ある日買い物をしようとして店に着いたら足元に穴が開いて落ちた

---

4. 【小説家になろう｜『繰り返す夢は儚く、脆く』｜2018_ローファンタジー】

少女が目を覚ましたのは、見知らぬ真っ白な空間でした

---

5. 【小説家になろう｜『3 人揃ってゲームの世界に来てしまったのだが』｜2019_異世界〔恋愛〕】

買い物に行こうとした主人公舞がゲームの世界に吸い込まれる！？一緒に吸い込まれたリア友朱理、紫織と一緒に元の世界に戻ろうとする物語

もう一つ注目に値するのは VR ゲームの象徴としての Topic4 である。異世界転生のファンタジーにおいて Topic4 が上位にきており、つまり vrmmorpg のゲームという舞台を転移先に設定する物語は少なくないと言えるのであろう。それに対して、異世界恋愛ではゲームと関連する背景であれば、乙女ゲームが主な舞台となっている。

6. 本文の冒頭部分

黄(2021)では中国の推理小説作家である程小青の作品の冒頭部分を中心にトピック解析を試みた。その結果、社会問題など導入的な話をするパターンが見られた。ウェブ小説は作家のような特定した書き方があるとは思えないが、類型化が進んでいることから冒頭部分にも確定したパターンがあることを期待される。また、あらすじのトピックを分析した上で気がついたことといえば、物語がどのように始まるのかをあらすじで書くことは少なくない。そこで、実際の冒頭部分と比較して新たな発見ができることが期待される。

この章では、「小説家になろう」の人気作品（6104 点）と「アルファポリス」の人気作品（941 点）に絞り、それぞれ第一話の内容を抽出して Top2Vec によりトピック解析をしてみた。人気のウェブ小説の始まりとしてどのような話題が使われやすいのかを解明する。特に婚約破棄や追放ファンタジーなど人気なトピックによく出現する書き出しのパターンに注目する。モデルを実行した結果、総じて 24 個のトピックが出力された。トピックを構成する主要単語を表 6.1 のようにまとめた。この出力結果をもとに考察を行う。

表 6.1：本文第一話のトピック構成

Rank	トピック	サイズ	キーワード
1	Topic0	1,231	殿下 婚約 王妃 王太子 社交 破棄 政略 解消 わたくし 婚姻
2	Topic1	506	女子 部活 高校 中学 教室 告白 彼氏 登校 放課 下校
3	Topic2	475	会社 コンビニ 残業 駐車 終電 職場 入社 アパート 電車 運転
4	Topic3	411	死体 一撃 咆哮 地面 敵兵 胴体 全身 獲物 切り裂か 振り下ろす
5	Topic4	384	パーティ パーティー クビ 役立た 冒険 メンバー 無能 ランク リーダー 雑用
6	Topic5	356	魔王 戦い 魔族 滅ぼし 強大 英雄 人類 魔神 滅ぼす 勇者
7	Topic6	351	転生 お主 チート じゃろ 天国 じゃっ 世界 輪廻 生まれ変わる のう
8	Topic7	339	プレイヤー プレイ VR システム VRMMO ゲーム ログ World ユーザー
9	Topic8	338	性格 外見 出身 得意 所属 無口 主人 人称 身長 茶髪
10	Topic9	337	愛し 幸せ ただただ 願っ 愛する 苦しみ 叶わ 温か 温もり 憎しみ
11	Topic10	293	ヒロイン 悪役 乙女 エンディング 攻略 ルート ライバル 断罪 令嬢 エンド
12	Topic11	247	鬱蒼 木々 辺り みる とりあえず 生え 岩肌 状況 見当たら 感覚
13	Topic12	198	授かっ 授かる 儀式 ギフト 神官 神託 神殿 神父 ハズレ 教会
14	Topic13	188	赤ちゃん 赤ん坊 泣き声 覚め 動かそう ぼやけ 赤子 思い出せ 抱き上げ 記
15	Topic14	171	召喚 帰し ルノ クラスメイト 勇者 ラノベ ローブ テンプレ 転移 状況
16	Topic15	168	作品 いただけれ 掲載 http イラスト 連載 加筆 書籍 追記 読み
17	Topic16	164	ダンジョン 探索 モンスター 階層 迷宮 出現 レベル ドロップ アイテム 潜る

18	Topic17	162	宇宙 艦隊 航空 惑星 搭載 兵器 植民 戦艦 機動 駆逐
19	Topic18	148	バター スープ 食べ 料理 美味しい 美味い ソース パン 美味 食べる
20	Topic19	142	柔らか 感触 体温 布団 シーツ 毛並み 首筋 じっと 寝台 温もり
21	Topic20	129	銅貨 銀貨 金貨 革袋 硬貨 貨幣 報酬 リゼル 買い取っ 依頼
22	Topic21	125	尾張 信長 織田 家臣 大名 戦国 天下 城主 今川 秀吉
23	Topic22	98	病院 看護 病室 入院 手術 救急 医者 患者 リハビリ 見舞い
24	Topic23	51	選手 野球 サッカー 試合 大会 優勝 チーム エース 出場 全国

まずは第一話のトピックと各ジャンルとの対応関係を把握する。

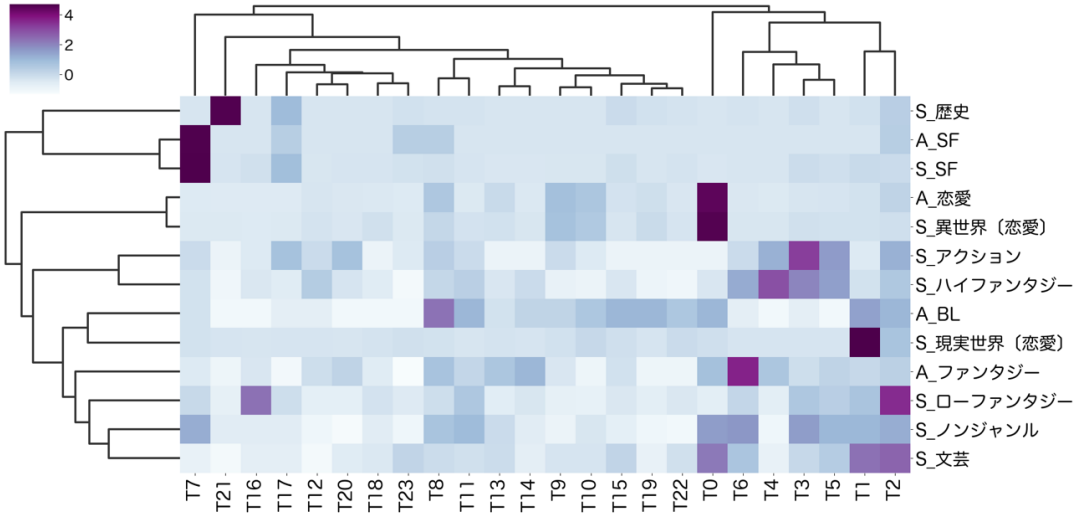


図 6.1　ジャンルとトピックのヒートマップ関係図

図 6.1 で示されたように、歴史、SF と異世界恋愛ジャンルは比較的に多様性が低くて、限られたトピックに縮約されているが、ファンタジーと文芸作品の方は多様性が割に高い。これはあらすじのトピック分布とも類似しているが、現実世界の恋愛ジャンルに関しては、あらすじのトピックがばらついているものの、冒頭文に統一性があることが見てとれる。その原因としては学園物語が多く描写されるため、書き出しの部分ではほぼ学校を舞台として、部活や教室などで行われる出来事が描かれ、Topic1 が通用するからだと思われる。

図 6.2　Topic1 のワードクラウド

6.1 婚約破棄

あらすじにおいて最も人気なトピックである婚約破棄をテーマとする作品 （第 5 章で出力された Topic0 の人気作品 1206 件）を対象として、書き出しの部分においてどのようなトピックが特徴的に使用されているのかを考察する。図 6.3 では各トピックの数の分布を棒グラフで最大から最小まで表示している。

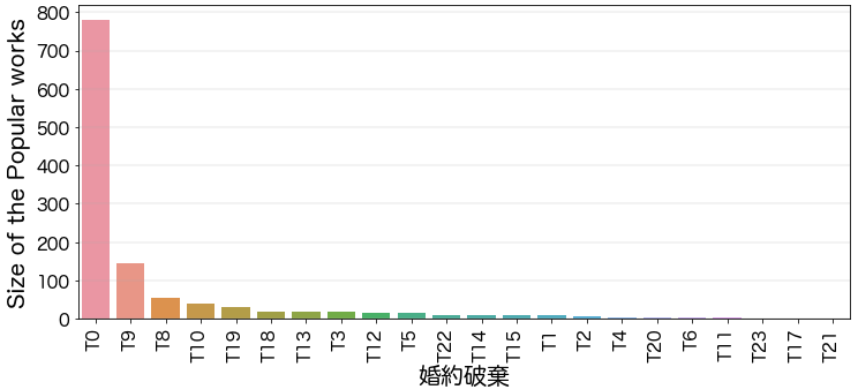


図 6.3　婚約破棄の第一話トピック分布

図 6.3 のトピック分布から分かるように、Topic0 は最も高い割合を占めている。図 6.4 のワードクラウドから、Topic0 があらすじの婚約破棄のトピック自体と非常に類似していることが分かる。具体的な冒頭部分を読むと、ここで挙げた 1 例のように書き出しで単刀直入に婚約破棄のことを告げるパターンが多いようである。

図 6.4　Topic0 のワードクラウド

Topic0 の冒頭例：

【小説家になろう｜『婚約破棄をされたら騎士様に彼女のフリをして欲しいと言われました』｜2021_異世界〔恋愛〕】

「婚約を破棄して欲しい」そう告げたのは、婚約者のハロルド様だ。ハロルド様はハーヴィ伯爵家の嫡男だ。私の婚約者のはずがどうやら妹と結婚したいらしい。いつも人のものを欲しがる妹はわざわざ私の婚約者まで欲しかったようだ。「ラケルが俺のことが好きなのはわかるが、妹のメイベルを好きになってしまったんだ」「お姉様、ごめんなさい」いやいや、好きだったことはないですよ。ハロルド様と私は政略結婚ですよね？目の前のソファーに座り寄り添っている二人は悲劇の主人公のように酔っていた。「では、婚約破棄ということでよろしいですね？」「ああ、婚約破棄のサインをしてくれ」用意周到に婚約破棄の書面を出し、私はサインをして、では失礼します。と部屋から出ていった。両親はどうせ反対はしないだろう。ハーヴィ伯爵家に娘が嫁ぐのだから。しかし、婚約破棄はきっと噂になる。しかも妹に婚約者を取られたのだから。しばらくは縁談の話もないだろう。そして、やはり両親からはしばらく夜会には行かないようにと言われた。婚約破棄されるようなことは恥さらしだと。そうなったのは妹のせいだが両親は妹を甘やかし、今回のことも私のせいだと決めつけていた。

Topic0 のような内容が婚約破棄をテーマとする作品の冒頭部に現れる典型的な書き出しではあるが、感情と心境をめぐる Topic9 も相当の数を持っているため見逃すことができない。愛情と幸せが苦しみにつながることがキーワードから見れとれる。実例で挙げたように、愛が報われないようなことから物語が展開されることが多い。

図 6.5　Topic9 のワードクラウド

Topic9 の冒頭例：

【小説家になろう｜『あなたのことは愛していますが、今回は辞退しようかと思います』|2019_異世界〔恋愛〕】

一目で恋に落ちました。一瞬で永遠の愛をささげようと思いました。陽に透けるような金色の髪、深い深いブルーの少し憂いを含んだ瞳。そのあたたかな眼差し。優しくて落ち着いた声音で私の名を呼ばれると身も心も溶けてしまいそうなほど嬉しく思っていました。心のそこから愛していました。見つめられる

と時が止まるのではないかと心配になるほどに。愚かな私は、あなたの憂いや哀しみ、苦しみをすべて取り除けると信じていました。あなたの楯になり癒しになれると疑うこともなく、生きてきました。それほど愛したあなたに疎まれ、私はどうすればよいのでしょう、どうすればよかったのでしょう。だから、もう二度と愛しません。遠くから、ただ、あなたの幸せをお祈りしています。あなたを愛する資格などないはずのなのにそれでもなおあなたをお慕いしています。

6.2 ハイファンタジー

次に、追放要素が含まれるファンタジー作品(534 件)をめぐって、冒頭部分の始まり方を解析する。このテーマについて、最も使われる始まりのパターンはあらすじのトピック自体と関係性が非常に深い Topic4 だということが図 6.6 および図 6.7 から明らかである。無能なためクビになることが直接導入的な話として描かれやすいと思われる。

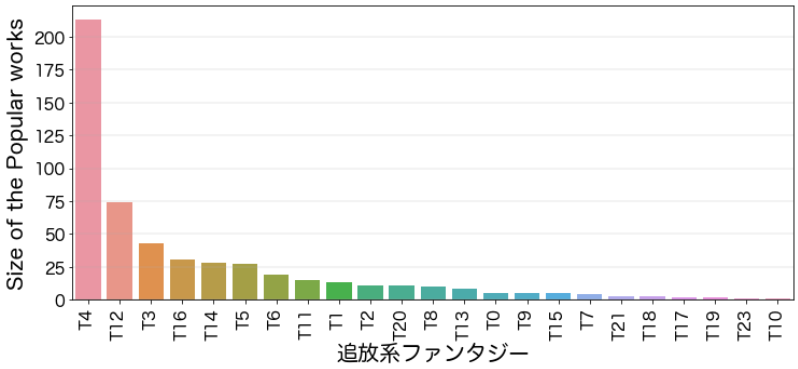


図 6.6　追放系ファンタジーの第一話トピック分布

図 6.7　Topic4 のワードクラウド

そのほか、2 位の Topic12 の作品数も少なくない。これは教会の儀式でギフトを授かることに関わるトピックだとキーワードから判断できる。つまり、鑑定を受け、職業やスキルが決められるシーンが最初に描かれる。主人公は概ね役に立たないスキルの持ち主であり、教

会に必要とされず追い出される展開につながることが多い。これも追放系作品の冒頭部にある典型的な場面設定だと言える。

図 6.8　Topic12 のワードクラウド

3 位の Topic3 に関しては、激しい戦闘シーンとゆかりの深いトピックだと考えられる。このトピックは追放系ではそれほど数が多くないが、図 6.10 から明らかなように、2012 年から 2017 年にかけて Topic3 が特徴的に多く出現している。つまり、追放系の小説が流行する前は、戦闘アクションが描かれる衝突式の書き出しが多く使われていた。

図 6.9　Topic3 のワードクラウド

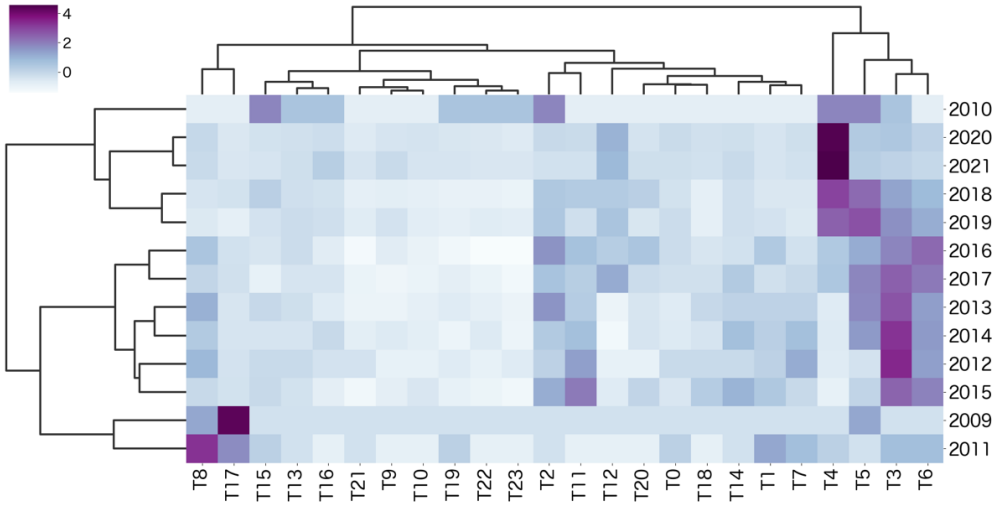


図 6.10　投稿年とトピックのヒートマップ関係図（ハイファンタジー）

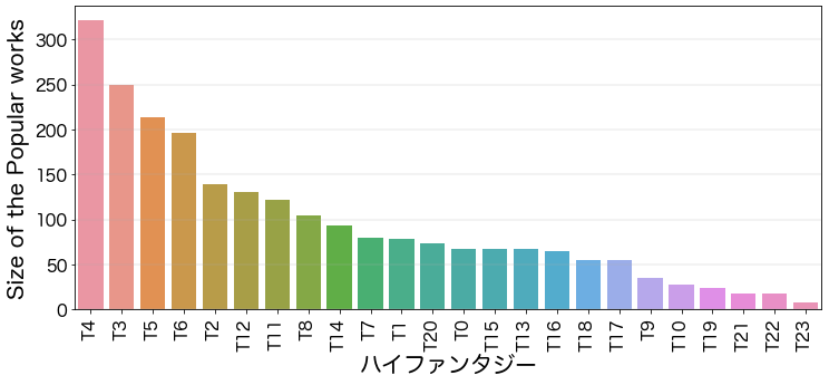


図 6.11　ハイファンタジーの第一話トピック分布

また、ハイファンタジー全体の第一話のトピック分布図 6.11 を見れば、Topic4 追放トピックおよび Topic3 激闘トピックに続き、魔物と戦うことが語られる Topic5 は 3 位に上がっている。5.6 では魔族討伐のトピックが人気作品において割合が高くないと述べたが、やはりあらすじが全てを反映しておらず、実際の本文内容の傾向から魔族討伐は追放系以外の人気題材であることに相違ないと考えられる。加えて、このトピックに関しては Topic3（激闘トピック）とのつながりが強く、図 6.13 から分かるように魔族討伐をあらすじとする作品の冒頭では、Topic3 が占める割合も相当高い。

Topic6 は転生に関わるトピックのため、次の 6.3 で恋愛ジャンルと合わせながらまとめて検討する。

図 6.12　Topic5 のワードクラウド

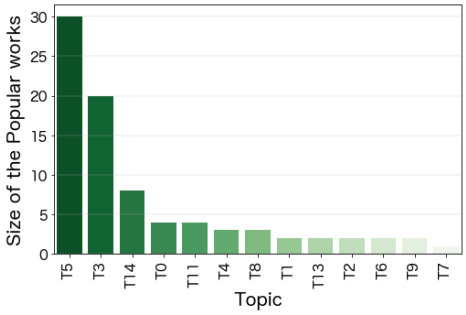


図 6.13　魔族討伐の第一話トピック分布

6.3 異世界転生・転移

この節では、異世界に転生する物語に限定して、異なるジャンル間の違いを考察する。前章でも述べたように、「アルファポリス」の方では転生と転移などに関わる情報が書かれていないため、表 6.2 でまとめたのは全て「小説家になろう」の情報である。

表 6.2　異世界転生トピック（トップ 5）

異世界ファンタジー				異世界恋愛			
Rank	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**
1	**T6**	118	0.173	1	**T10**	151	0.327
2	**T2**	70	0.103	2	**T0**	90	0.195
3	**T11**	54	0.079	3	**T8**	46	0.100
4	**T13**	51	0.075	4	**T9**	40	0.087
5	**T3**	48	0.070	5	**T13**	35	0.076

ファンタジーと恋愛では冒頭部分において異世界転生に関する共通トピックが非常に少なく、上位に位置するトピックを見る限りでは Topic13 のみである。

Topic13 は赤ちゃんのトピックで、つまり幼児に転生するというパターンはどちらもある程度使われていると言える。

図 6.14　Topic13 のワードクラウド

ファンタジーにおける異世界転生では Topic6 が最も特徴的である。図 6.15 のキーワードからすれば Topic6 は死んだ後の事情説明に関わるトピックだと考えられる。「お主」や「じゃろ」などの主要単語から対話が肝心な要素になっているのではないかと思われる。

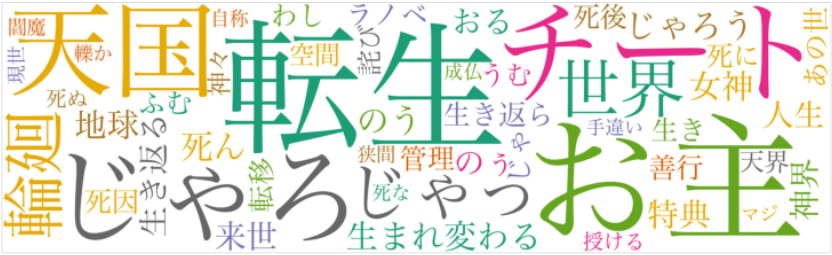


図 6.15　Topic6 のワードクラウド

実際の冒頭部分を 1 例取り出して吟味すると、主人公が死んだ後に、神様と対話されているシーンが描かれていることが分かった。こうしたパターンが異世界転生のファンタジーに最も使われていると思われる。

Topic6 の冒頭例：

【アルファポリス ｜ 『スキルで快適！異世界ライフ(笑)』 ｜ 2018_ファンタジー 】

とある地方に住み、親の庇護の下、嫌々仕事をしながら趣味のサブカルチャーにほぼ全ての収入を注ぎ込む日々の生活。　そんなダメ人間である彼、【工藤くどう蓮れん】(23)は若くしてその生涯を終えた。ある日突然彼の家に押し入った強盗犯の手によって。　そんな彼を見かねた１人の神が、自分の元に彼の魂を呼び寄せた。そして神の第一声が…。「お主なぁ、もうちぃ～とばかしマシな人生を送ろうとは思わなんだか？」「は？全然マシだし？自分が楽しけりゃよくね？ま、死んだみたいだけどなー。」　神は呆れた。「そんなんじゃから、惨たらしい死に様になるのじゃ。どうやらお主には人生を１からやり直させた方が良さそうじゃな。」「え、何？もしか異世界転生？チート ktkr」「チートなぞやるかバカタレ。己が身１つで逝け。良い人生を送っているとワシが判断したら、その時は何か１つワシから贈ろう。ではな。」　神はそれだけ言って消えた。「ちょ、待てよぉぉぉぉぉぉ～っ今チートよこ……うぉぁぁぁぁぁぁっ……」　謎空間で霊体化していた蓮の身体は、ストーンと暗闇へ向け垂直落下していくのであった。

また、Topic2 もファンタジー転生の方で多用されるトピックである。Topic2 は場面的なトピックだと思われ、通勤や路上に関連するようなものが頻繁に出現している。出来事が異なるかもしれないが、コンビニか会社帰りのシーンが描かれるのではないかと考えられる。これも前章で述べた出世できない主人公のキャラクター設定やトラックなどの事故線と関連しているのではないか。

図 6.16　Topic2 のワードクラウド

Topic6 と Topic2 に続き、3 位である Topic11 は草原の景色が描かれているとキーワードから推測することができる。また、「落ち着こう」や「覚める」などの主要単語からも、知らない場所にいる可能性が高い。つまり、Topic11 は転生後の最初の居場所を描くトピックだと推定することができる。その上、木々が生える草原が行き先として設定されることが多いのであろう。

図 6.17　Topic11 のワードクラウド

Topic6 と Topic2 はどちらも転生する前の場面が描かれているように思われる。つまり、前世のことについて書くのが冒頭部で最も使われるパターンだと結論づけられる。また、赤ちゃんとして転生する、または野外で目を覚めるというような転生後のシーンも書き出しで描かれることが多い。

それに対して、恋愛の方ではどのように転生したのかについてはそれほど説明を加えていないと思われる。最上位の Topic10 は、乙女ゲームのキャラクターに転生する物語と対応している。ヒロインと悪役令嬢の前世および既定のエンディングなどについて述べることが Topic10 の中心的な話だと判断できる。また、異世界転生の要素を取り入れた婚約破棄の話に関しては、婚約破棄の内容性が優先され、前述した Topic0 で総括されていると考えられる。

図 6.18　Topic10 のワードクラウド

要するに、ファンタジー作品に関しては冒頭部分において異世界に至る原因や元の世界で起きたハプニングなどについて詳細に書かれることが多い。一方で、恋愛ジャンルの方ではあまり転生の理由や経緯が重要ではなく、主人公がシナリオを読めているパターンが多くて人物の絡みに関する説明が冒頭で描写される仕組みとなっている。もちろん赤ちゃんとして転生するパターンはどちらも使われるが、割合で言えばどちらもそれほど高くはない。

次は異世界転移に関わるトピックを検討する。

表 6.3　異世界転移トピック（トップ 5）

異世界ファンタジー				異世界恋愛			
Rank	**Topic**	**Size**	**Weight**	**Rank**	**Topic**	**Size**	**Weight**
1	**T6**	82	0.123	1	**T2**	24	0.14
2	**T14**	78	0.117	2	**T0**	21	0.122
3	**T2**	67	0.101	3	**T19**	18	0.105
4	**T11**	59	0.089	4	**T14**	17	0.099
5	**T3**	52	0.078	5	**T1**	12	0.07

異世界転移に関して、ファンタジーの方では異世界転生の代表トピックとよほど重なっていることが表 6.3 の集計結果からわかる。つまり、物語の始まりシーンは類似したものであり、使い分けが曖昧である。しかし、ファンタジーと恋愛ジャンルの両方において高い割合を占めた Topic14（召喚による転移）は異世界転移作品の典型的な書き出しであることは判明した。その上、キーワードにおけるクラスメイトとの共起、および学校と関わる Topic1 も上位にあることから、クラスメイトで同時に異世界に転移するというのも典型的な設定であろう。また、乙女ゲームが支配する恋愛転生系と違って、転移系であれば、Topic2 と Topic14 のように元の世界からの転移について具体的に描写されるトピックが多いことが分かる。

図 6.19　Topic14 のワードクラウド

なお、「アルファポリス」のファンタジー作品の冒頭部分のトピック分布を図 6.20 で可視化してみると、転生トピックである Topic6 が最上位であり、転移を代表する Topic14 もそれに続き多いということから、「アルファポリス」では転生・転移要素が含まれるストーリーが現段階のトレンドだと考えられる。恋愛ジャンルに関しては図 6.1 からも分かるように両方の小説投稿サイトでは大差が見られず、冒頭部の形式がほぼ一致しているということが明らかである。

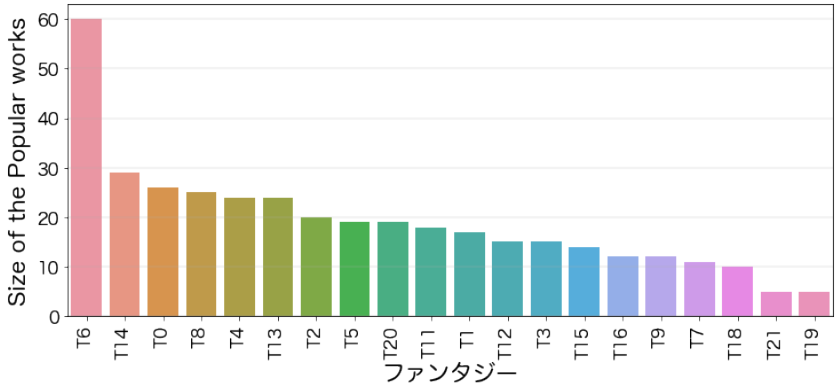


図 6.20　ファンタジー（「アルファポリス」）の第一話トピック分布

6.4 キャラクター紹介トピック

性格や外見、出身などキャラクターの設定に関する Topic8 も第一話でよく使われる書き方である。

図 6.21　Topic8 のワードクラウド

第 5 章の 5.6.2 で述べたように、現実世界を舞台とする恋愛作品のあらすじにおいて、性格の特性を示す人物設定のトピックが好まれるが、冒頭部に登場人物のトピックが書かれ

ることはほぼないということは表 6.4 の現実恋愛ジャンルの低い割合から判明した（Size はそれぞれのジャンルの作品数、Count はジャンルにおける Topic8 の作品数、Weight はジャンルにおける Topic8 の割合）。現実世界を背景とするジャンルと比べると、異世界の物語では人物説明がより重要であることが明らかとなっている。また「アルファポリス」のジャンルが概ね上位に位置することは、小説投稿サイトの違いを表しているのではないか。

表 6.4　各ジャンルにおける Topic8 の分布

Site	Group	Count	Size	Weight
S	推理〔文芸〕	2	11	0.182
A	BL(ボーイズラブ)	8	61	0.131
A	恋愛	44	475	0.093
A	SF	1	11	0.091
S	ノンジャンル〔ノンジャンル〕	20	319	0.063
A	ファンタジー	25	394	0.063
S	アクション〔文芸〕	3	51	0.059
S	異世界〔恋愛〕	96	1,888	0.051
S	ハイファンタジー〔ファンタジー〕	104	2,305	0.045
S	コメディー〔文芸〕	4	94	0.043
S	ローファンタジー〔ファンタジー〕	15	419	0.036
S	ヒューマンドラマ〔文芸〕	3	108	0.028
S	VR ゲーム〔SF〕	7	256	0.027
S	現実世界〔恋愛〕	5	396	0.013
S	歴史〔文芸〕	1	137	0.007

ここでは、第 4 章で紹介したコレスポンデンス分析の手法を通して、冒頭文のキャラクター紹介によく出現する言葉に関して、ジャンル間にどのような使い分けが見られるのかを解明する。キャラクターの特徴に関わる分析なので、Topic8 の構成単語の品詞を参照し、名詞を扱うことにする（上位 100 まで）。数詞の割合が高いため、数詞は分析対象から外す。ジャンルの数が少なければ外れ値にもなりやすいため、Topic8 の作品数が 10 未満のジャンルは除外し、灰色塗りの 6 つのジャンルに限定してコレスポンデンス分析を行い、図 6.22 のようにジャンルラベルと語彙項目を 2 次元空間に配置し描画する。

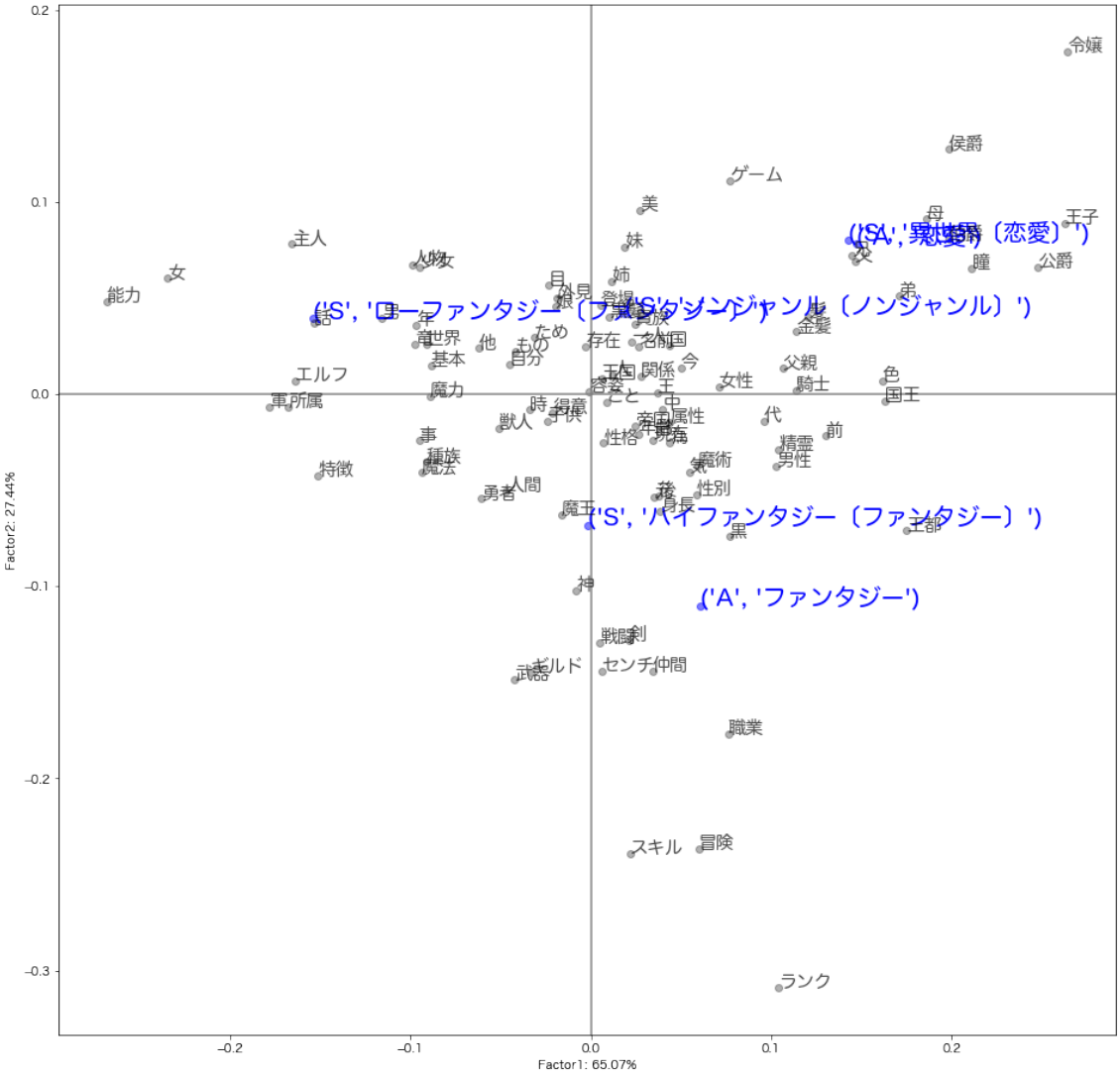


図 6.22　名詞とジャンルの対応関係図

図 6.22 からはジャンル間の語彙の使用頻度の差異がはっきりとわかる。右上の異世界恋愛ジャンルでは、瞳や髪に関わる外見、地位（王子、令嬢、伯爵、公爵など）および家族関係（兄、弟、父、母など）に関して重点的に描写されている。それに対して、ファンタジーの方では身長の高さがより重視されている。もちろん、スキル、武器、職業、種族、所属などのファンタジー定番の属性が特徴的に描写されていることが明らかである。

7. 感想コメント

5 章で述べたように、ウェブ小説は書き手と読み手のコミュニケーションが重視され、人気な作品ほどそのつながりが強い。この章では、読者と作者のやり取りに目を向け、小説のコメント欄の書き込みを中心に検討する。

團（2020）ではオンライン小説投稿サイトの感想欄に注目して、「小説家になろう」における代表的な作品である『無職転生 異世界に行ったら本気出す』の感想欄を分析対象として、2012 年 11 月 26 日から 12 月 30 日までの約 1 ヶ月間に感想欄に投稿されたコメントを事例として扱い、書き込みの特性を考察した。

感想欄におけるやりとりで目立つのは、書き手への労いや激励だと指摘されている。その上、誤字脱字の指摘にかかわる書き込みもたくさんあり、それに基づいて本文が修正されることが多いようである。本文の内容、表現について読み手による指摘が作者に採用されることがあり、その一例を團は取り上げて検討した。團によると、読み手が修正案を出すときに、自らの意見を格下げして伝えるのが見られる。つまり、書き手の作品に対する権威性を踏まえた上にコメントをデザインし、書き手はコメントを受けて本文を修正する相互行為の特性が明らかである。

しかし、筆者が問題として考えたのは事例が少なさすぎて、果たして丁寧な指摘が普遍現象なのかどうかは数を増やして考察しなければ言い切れないのではないか。そのため、1 つの作品にとどまらず、2004 年から 2021 年までの作品全般に対してコメントを大量に取り上げ、トピックモデルによる集約をもとに、指摘の丁寧さを考察することをこの節の 1 つの目的とする。ビッグデータによりウェブ小説全体に対するコメントの特徴などを解明することも期待される。

ここでは「小説家になろう」のコメントに限定して調査を行う。コメントの取り方に関しては、できるだけ偏りがないようにサンプルを取ろうとするため、それぞれの投稿年において、各ジャンルの総合評価ポイント順に上位 10 位までの作品を対象に、それぞれ 100 件までの投稿コメントをスクレイピングする。（100 件以上ある場合は最新の 50 件と最古の 50 件を取る）

総じて 756 作品に対する合計 79390 件のコメントを取り出した。感想欄は「良い点」「一言」「気になる点」という項目に分けられており、コメントの投稿年を含めた統計情報を表 7.1 にまとめた。

表 7.1　コメントの情報まとめ

コメント投稿年	コメント総数	良い点	一言	気になる点
2006	468	0	468	0
2007	788	0	788	0
2008	1,824	0	1,824	0
2009	2,539	203	2,200	136
2010	3,601	813	2,344	444
2011	4,569	1,007	2,937	625
2012	5,270	1,138	3,471	661
2013	5,058	1,207	3,251	600
2014	4,586	1,039	3,003	544
2015	5,103	1,095	3,371	637
2016	5,259	1,092	3,530	637
2017	4,671	1,010	2,888	773
2018	4,738	1,172	2,648	918
2019	5,564	1,442	2,984	1,138
2020	7,410	1,940	3,851	1,619
2021	17,942	4,514	9,753	3,675

この 79390 件のコメント（全て読者の書き込み）を分析対象とし、前章でも使った Top2Vec を用いてトピックモデルを実行し、トピック数を 100 個に絞り、出力された結果をもとに考察を行う。最も大きいトピック Topic0 から明らかなように、「面白い」という評価が最上位の定番コメントだということが読み取れる。（Topic0 の実例はほぼ「面白いです」「楽しみにしています」の 2 種類だけ）

図 7.1　Topic0 のワードクラウド

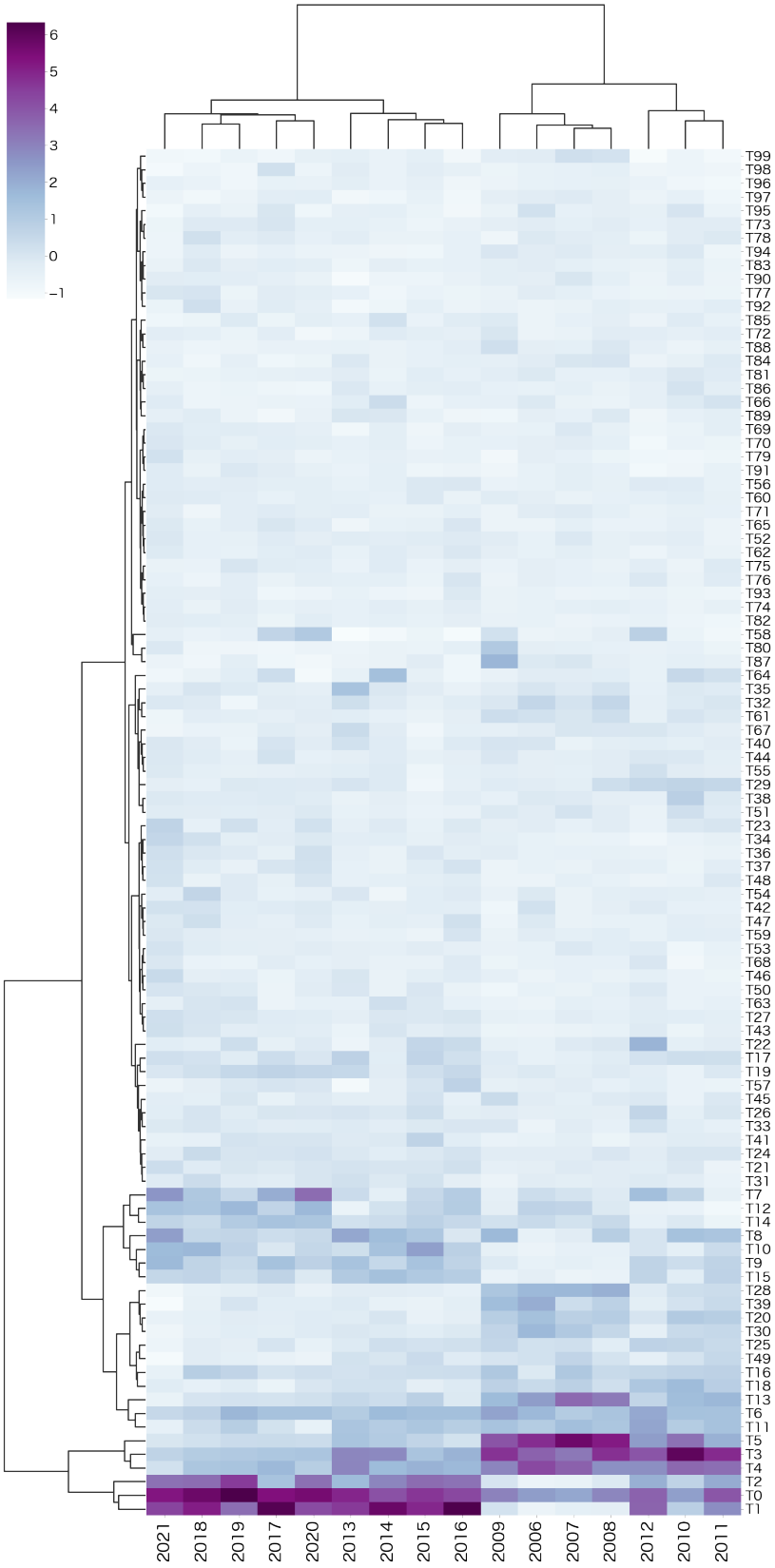


図 7.2　コメントの投稿年とトピックの関係図

コメントの習性に関しては、コメントの投稿年により影響されるのではないかと思われる。図 7.2 ではコメントの投稿年とそれぞれのトピックとの関係を表示している。2012 年以前のコメントは Topic3（幸せな気持ち）、Topic 5（感謝）[54]といったような賛美が含まれた発言のトピックと、Topic4 のようにキャラクター、文章構成、ストーリーのテンポや展開など内容全体に関する統括的なコメントが主導的であり、丁寧さが非常に感じられる。

図 7.3　Topic4 のワードクラウド

一方で、2013 年以降の主要トピックでは、Topic 0（面白い、楽しみ）、Topic1（スキルとアイテム等）と Topic2（貴族に関わる諸々の事情）がほとんどである。

ここでさらに、「良い点」、「一言」と「気になる点」の 3 項目に分けて細かく考察する。図 7.4 から分かるように、良い点として評価するトピックに関しては、Topic0 と Topic4 の傾向が非常に目立つ。2006 年から 2021 年にかけて、Topic4（全体評価）から Topic0（面白い）への重みの変化が見られ、特に 2017 年から 2021 年までの Topic0 の割合が圧倒的に高くて、近年ではより気軽な「面白さ」を評価軸とするコメントが頻出していることが考えられる。

[54] 付録 3 では出力された読者コメントの全てのトピックを表示している。それぞれのトピックの構成単語とサイズを参照することができる。

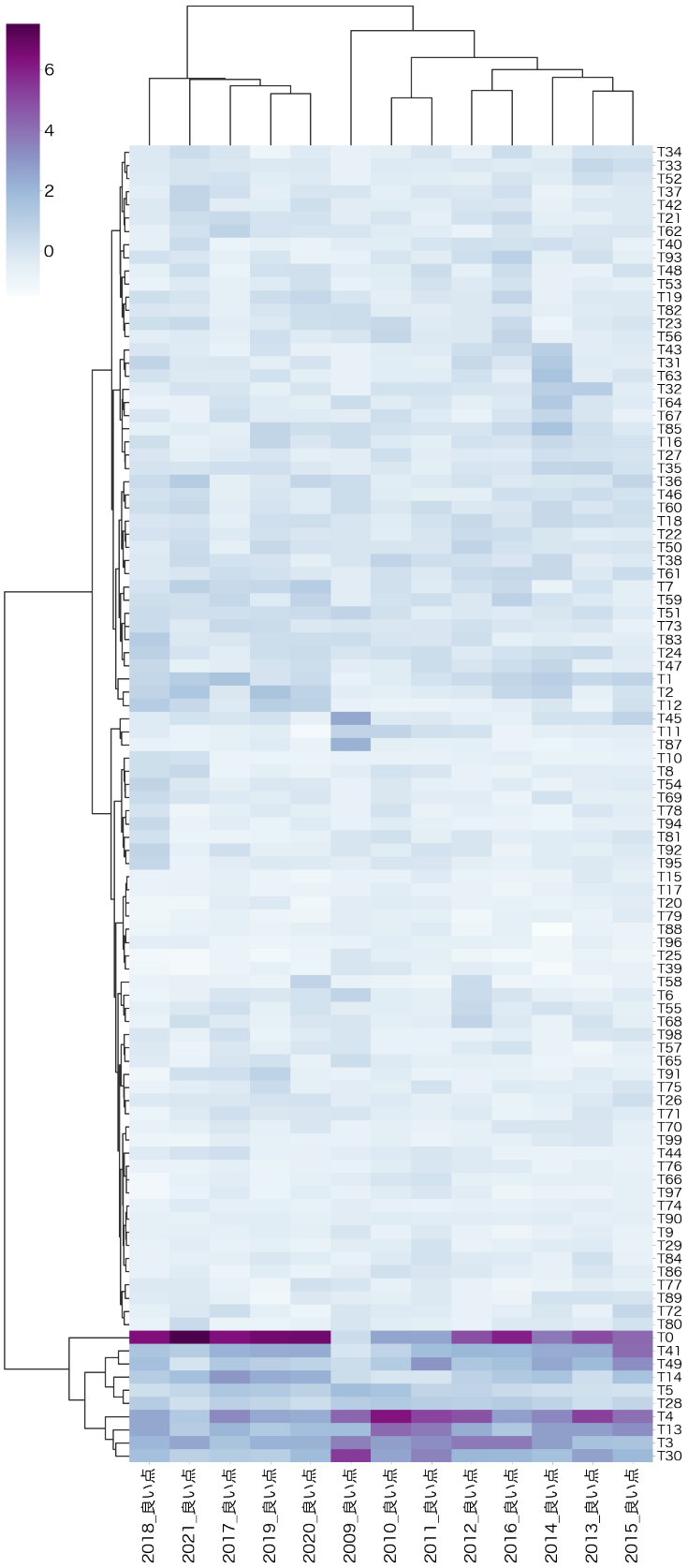


図 7.4　「良い点」コメントの投稿年とトピックの関係図

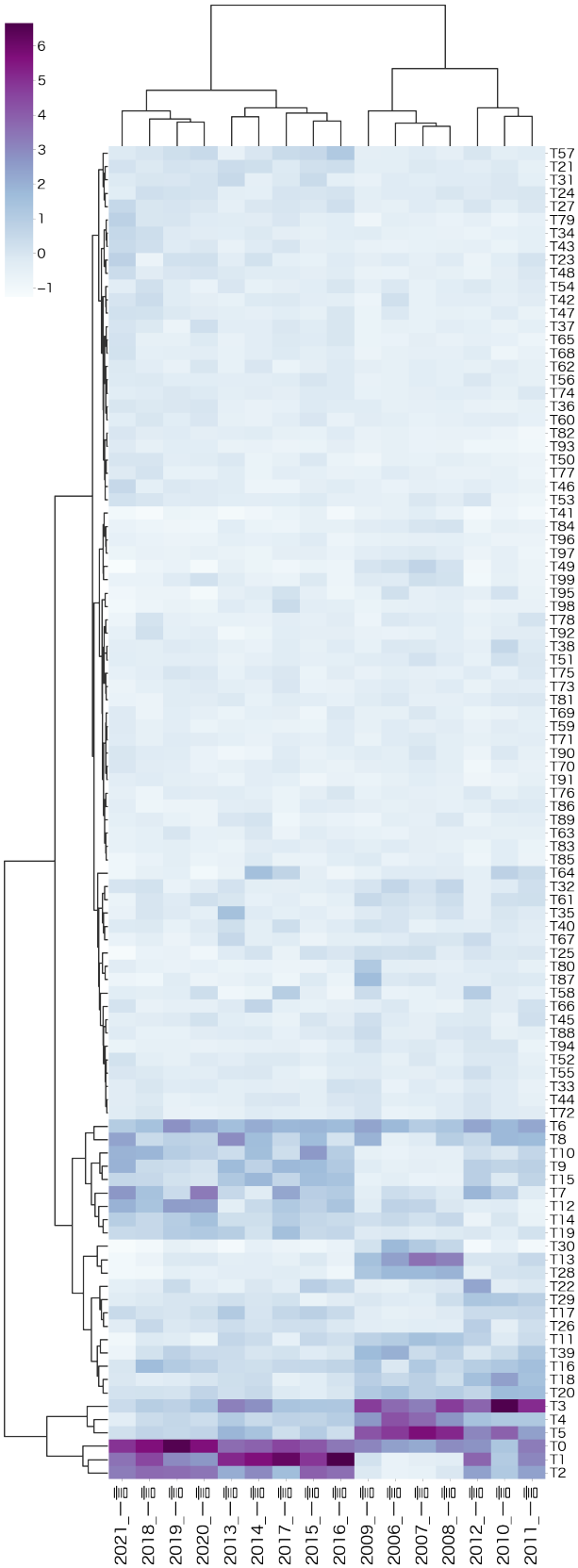


図 7.5　「一言」トピックの関係図

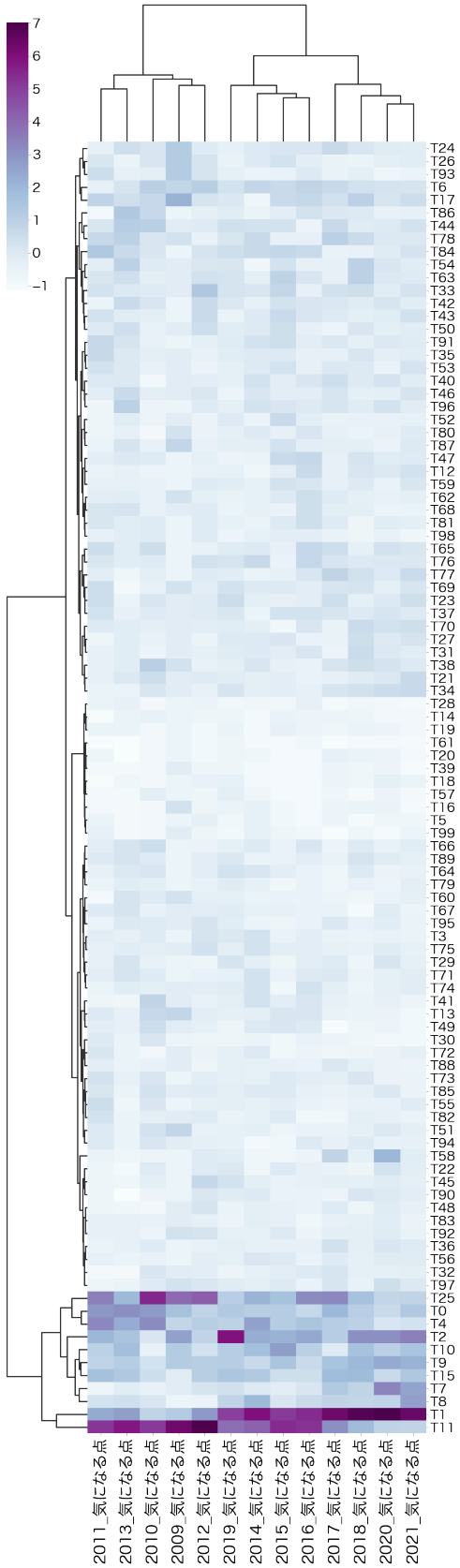


図 7.6　「気になる点」のトピック関係図

「一言」のヒートマップ図 7.5 は全体の傾向を表す図 7.2 と類似していることが分かる。図 7.6 の「気になる点」のヒートマップと合わせて見れば、貴族をめぐるコメントの Topic2 とスキルやアイテムなどファンタジーと SF 系の要素が含まれる Topic1 はどちらも近年において注目度が高いが、Topic2 と比べ、Topic1 はより論争が行われそうである。（気になる点として投稿されたコメントでは Topic1 の圧倒的なトレンドが見られるのに対して、Topic2 は割合が低い）

また、図 7.6 において Topic1 以外には、Topic11 も非常に特徴的なトピックだと読み取れる。Topic11 が文章の表現や文法などに関する指摘のトピックであり、意見コメントの類としても非常に特徴的である。(類似した傾向を持つ Topic25 は誤字脱字の指摘に関するトピック)　図 7.7 から明らかなように、改行問題と句読点の付け方が最も指摘されている点であろう。読者にとって視覚性が最も気になることなのではないかと考えられる。

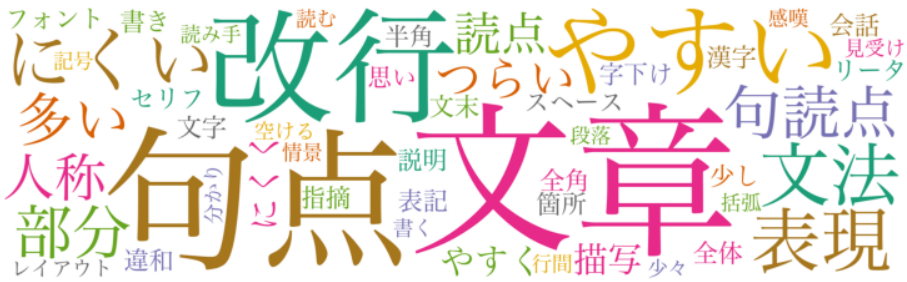


図 7.7　Topic11 のワードクラウド

先行研究において團（2020）は内容の表現など修正すべき点をはっきりと示すのではなく、「違和感がある」「しっくりくると思います」「気がします」など格下げすることで書き手の権威性を保ちながら意見を受け入れてもらうことについて言及したが、実際 Topic11 で指摘された問題のうち、婉曲表現を加えていないのも少なくない。ここではいくつか作者の返事がついている実例を挙げる。

（1）【『異世界を渡りし者』 |2010_ハイファンタジー】 (2010 年投稿)

読者のコメント：戦闘の場面をもう少し増やしたほうがいい。

作者の返事：御感想ありがとうございます 戦闘描写は苦手なんですが、なんとか期待に応えられるように頑張りたいと思います。最強・不死身・不老不死と反則の塊な小説ですが今後ともよろしくお願いいたします。

（2）【『最新のゲームは凄すぎだろ』 |2012_ VR ゲーム〔SF〕】 (2012 年投稿)

読者のコメント：視点変更の箇所がわかりづらく感じる

作者の返事：感想ありがとうございます。実は自分でも視点変更には悩んでまして。※で切り分けてみたり、空白行を増やしてみたりしたんですがわかりにくいですかね。とりあえず空白行をもう少し増やしてみますので、またなにかあればご指摘いただければと思います。

（3）【『ヘルモード　～やり込み好きのゲーマーは廃設定の異世界で無双する～』 |2019_ハイファンタジー】 (2019 年投稿)

読者のコメント：語尾にやたらと「のだ」を使用している点。また、主人公視点の語りの語尾には「ようだ」をあててみたり、語尾が特徴的という以上におかしいところ。

作者の返事：感想及びご指摘ありがとうございます。作者としましても前作より語尾についてはずっと悩んでいたところでございます（涙）今回のご指摘を参考に修正させていただきます。今後ともご愛読のほどよろしくお願いします。

（4）【『転生先で捨てられたので、もふもふ達とお料理します　～お飾り王妃はマイペースに最強です～』 |2019_異世界〔恋愛〕】 (2019 年投稿)

読者のコメント：クロナの「そんな～～」が　…　のほうが無表情だけど残念な感じが出ると思いました！

作者の返事：こんにちは、ミサンガさま。感想ありがとうございます。確かにそちらの方が、よりクロナの無念が伝わりそうですね。貴重なご意見あいがとうございます。修正させてもらいますね。

上記の 4 つの書き込みに関して問題点は様々であるが、改善すべき点を直接指摘し、格下げが見受けられないことが共通している。それに対して、作者はよく受け入れている姿勢を見せている。（1）、（2）については今後注意していくこと、（3）、（4）は修正を入れていくことで作者と読者のコミュケーションの相互行為がなされている。

もちろん、團（2020）で指摘された書き手の特権性を尊重しながら言葉遣いを気にする例もたくさん存在するが、やはり個人差が存在しており、ポリシーにはなっていないと思われる。書き手にとっては言葉の丁寧さよりも、アドバイスの合理性が優先されているのではないか。

8. おわりに

8.1 まとめ

日本のウェブ小説サイトにおいて、投稿数が年々上昇し、ウェブ小説への関心が高まっていることが数値データを通じて確証することができた。ウェブ小説のうち、異世界を舞台とするハイファンタジー物語は投稿の作品数が圧倒的に多く、高い人気を博している。ただし、2019 年からは異世界恋愛ジャンルがブームとなっていることが観察された。また VR ゲームは投稿数がそれほど多くないが、このジャンルに強い興味を抱くファンダムが存在し、人気度の高いジャンルであることは否定できない。

第 4 章ではウェブ小説のタイトルに注目し、近年だんだんと長いタイトルが用いられる傾向があることを明らかにした。とりわけ人気の高いファンタジージャンルにおいてその趨勢が一層目立つ。他の作品との差別化を図り、長いタイトルで読者を惹きつけるようとする著者たちの試みだと考えられる。語彙の多様性は年々減少していることが明らかとなり、特に人気度の高い作品ほどその傾向が強い。タイトルにおける言葉遣いの共通性が高くなっていることが原因だと考えられる。タイトルにおける語彙の使用はジャンルによる明らかな差異が観察され、投稿の年代による大きな違いも見受けられた。近年では悪役令嬢と追放ファンタジーが流行となっていることが明らかになった。

第 5 章では Top2Vec という新たなトピックモデルを用いてあらすじをトピックにまとめ、トピックを通してあらすじの時系列的な変化を考察することができた。内容的には切ない片思いに関するトピックが次第に減少し、2012 年からは vrmmorpg のトピックが流行となり、近年では皇族・貴族に関わる婚約破棄と、パーティーから追放されるファンタジーストーリー、女主人公が乙女ゲームに悪役令嬢として転生するようなトピックが非常に高い人気を博している。これらの現象に関しては、タイトルの語彙分析でも類似した特徴語が観察されており、この解釈を裏書きする要素となっている。小説の内容が含まれるトピック以外にも、2012 年からは著者と読者のコミュニケーションが緊密になり、著者からの誤字脱字の指摘願いに関わるあらすじが多くなり、また人気作品において、書籍化の宣伝や小説を応援したことへの感謝を表すトピックも多く出現していることが明らかになった。さらに、それぞれのジャンルにおける流行トピックの興廃変化も解明し、人気作品とそれ以外の作品のトピックの相違を詳細に検討した。

第 6 章では人気作品の第一話を分析対象として、同じく Top2Vec によるトピックモデリングを実行した。出力されたトピックをもとに、婚約破棄、追放ファンタジーおよび異世界転生・異世界転移をテーマとする作品の冒頭部における典型的な書き出しパターンを考察した。

最後に第 7 章では「小説家になろう」サイトから取り出した大量の感想コメントを Top2Vec で解析した。その結果、2013 年以降のコメントと比べ、2013 年以前のコメントは丁寧さが感じられるトピックが主導的である。近年では気軽な「面白さ」評価が目立ち、かつ SF とファンタジーの肝心な要素であるスキルやアイテムなどをめぐる議論と異世界恋愛の重要な登場人物である貴族の諸々の事情に関わるコメントが多く付与されている。文章の表現に関する指摘は改行と句読点のような視覚と関わる問題が大部分を占めている。また、読み手の修正意見の文体に関しては格下げという丁寧な態度を取るより、意見の合理性がより重視されている。

第 1 章で設置されたリサーチクエスチョンに対しては、全て答えを見つけることができた。さらに、出力された実行結果から、当初は予想しなかった考察も多数行うことができた。これがビッグデータを用いた量的研究の魅力でもあると思われる。

8.2 今後の課題と展望

第 6 章では人気作品の本文第一話を分析対象とし、Top2Vec を用いてトピックモデリングを実行したが、その的確性が疑問として残されている。長い文章が多いため、1 つのテキストファイルを 1 つのトピックでまとめることは情報をなくす可能性が高いため、今後は他の手法を試みて比較することを課題とする。

また、本論文は 2021 年（未満）までのウェブ小説に関わる様々な情報を取り上げて行った研究であるが、ウェブ小説の流行の興廃変化は早いと思われ、今後引き続き情報を収集し、発展的な研究を行う必要がある。

さらに、本研究の結果が中日のウェブ小説の異同を考察するための一助となることを期待する。

参考文献

（日本語）

飯田一史（2016）『ウェブ小説の衝撃—ネット発ヒットコンテンツのしくみ』筑摩書房.

飯田委哉, 伊東栄典, 佐嘉田悠樹（2018）「クラスタリングによるオンライン小説の多様性動向分析」『火の国情報シンポジウム論文集』1–7.

今田水穂（2018）「語彙多様性指標の可視化と単回帰分析による TTR の補正」『言語資源活用ワークショップ発表論文集』3: 519-530.

浦川隆寛, 伊東栄典（2013）「オンライン小説におけるキーワードの時系列傾向分析」『情報処理学会研究報告』B-3-2 https://catalog.lib.kyushu-u.ac.jp/opac_download_md/26537/QS2013-novel-trend.pdf

奥野陽, グラム・ニュービッグ, 萩原正人（2016）『自然言語処理の基本と技術』翔泳社.

川上量生（2014）『ネットが生んだ文化 誰もが表現者の時代』角川学芸出版.

クジラ飛行机（2016）『Python によるスクレイピング＆機械学習[開発テクニック]』ソシム株式会社.

黒田絢香（2017）「小説テクストの計量的分析 アーサー・コナン・ドイルの作品から」『言語文化共同研究プロジェクト 2016』23–41.

黄晨雯（2020）「中国のミステリー小説におけるトピック解析の試み」『言語文化学』（大阪大学言語文化学会）1–17.

黄晨雯（2021）「Top2Vec による小説の探索的研究 —程小青の作品解読を中心に—」『言語文化共同研究プロジェクト 2020』43–53.

小木曽智信, 小椋秀樹, 小磯花絵, 宮内佐夜香, 渡部涼子, 伝康晴（2010）「形態素解析辞書のベンチマークテスト—IPAdic・NAIST-jdic・UniDic のジャンル別精度比較—」『言語処理学会第 16 回年次大会発表論文集』326–329.

小林雄一郎（2017）『R による優しいテキストマイニング：機械学習編』オーム社.

堺雄之介, 伊東栄典（2020）「オンライン小説の流行語抽出」『情報処理学会全国大会講演論文集』82: 6T–01 https://catalog.lib.kyushu-u.ac.jp/opac_download_md/2557143/2557143.pdf

佐藤一誠（2015）『トピックモデルによる統計的潜在意味解析』コロナ社.

実崎直人, 伊東栄典（2019）「回帰分析によるオンライン小説の人気度推定」『情報処理学会全国大会講演論文集』81: 7S–08 https://catalog.lib.kyushu-u.ac.jp/opac_download_md/2244135/ipsj81-narou.pdf

田島ますみ, 深田淳, 佐藤尚子（2008）「語彙多様性を表す指標の妥当性に関する研究—日本人大学生の書き言葉コーパスの場合—」『中央学院大学社会システム研究所紀要』9(1): 51–62.

田畑智司（2004a）「コーパス言語学のための多変量解析入門」英語コーパス学会第 24 回大会ワークショップ http://www.lang.osaka-u.ac.jp/~tabata/JAECS2004/JAECS2004hand.pdf

田畑智司（2004b）「-ly 副詞の生起頻度解析による文体識別—コレスポンデンス分析と主成分分析による比較研究—」『電子化言語資料分析研究』97–114.

田畑智司 (2012) 「テクストマイニングからテクスト分析へ： Collins との共著作品における Dickens の文体」 『電子化言語資料分析研究 2011-2012』3–17.

田畑智司（2017）「FLOB コーパスの意味構造：確率論的トピックモデルによる言語使用域の特徴づけ」『言語文化共同研究プロジェクト 2016』5–21.

田畑智司（2020）「英国 Classic Fiction コーパスの潜在的トピック: LDA によるテクストクラスタリング」『言語文化共同研究プロジェクト 2019』47–58.

玉井建也（2020）「ウェブ小説に見る物語構造と虚構性ー『転生したらスライムだった件』を事例としてー」『東北芸術工学大学紀要』27: 1–12.

團康晃（2020）「協働としてのオンライン小説ーオンライン小説投稿サイトの感想欄に注目してー」『年報社会学論集』33: 24–31.

張強（2015）「近・現代の日本人の名前に関する言語学的研究」 博士論文　広島市立大学.

土村成美（2017）「トピックモデルによるアガサ・クリスティ作品の計量的分析」『じんもんこん 2017 論文集』177–182.

本田優也, 伊東栄典（2017）「利用者投稿型小説サイトにおけるキーワードの多様性分析」『火の国情報シンポジウム論文集』A2–4 https://www.ipsj-kyushu.jp/page/ronbun/hinokuni/1006/A2/A2-4.pdf

船山貴光（2021）「テキスト情報を含む大規模データの分析と可視化に関する研究」博士論文 東海大学大学院.

増田秀人（2020）『現場で使える！pandas データ前処理入門 機械学習・データサイエンスで役立つ高速処理手法』翔泳社.

水本篤 (2009)「コーパス言語学研究における多変量解析手法の比較: 主成分分析 vs. コレスポンデンス分析」統計数理研究所共同研究リポート 232『コーパス言語研究における量的データ処理のための統計手法の概観』53–64.

矢田竣太郎, 岩井美樹, 影浦峡（2017）「日本語書籍タイトルの形式的構造の分析」言語処理学会第 23 回年次大会 https://www.anlp.jp/proceedings/annual_meeting/2017/pdf_dir/P14-5.pdf

吉田拓真, 尾原颯（2020）『現場で使える！Numpy データ前処理入門 機械学習・データサイエンスで役立つ高速処理手法』翔泳社.

涌井良幸, 涌井貞美（2010）『統計解析がわかる』技術評論社.

涌井良幸, 涌井貞美（2011）『多変量解析がわかる』技術評論社.

（中国語）

陈义（2018）「文本挖掘在网购用户评论中的应用研究」 修士論文　浙江工商大学.

高寒凝（2020）「网络文学研究中的数字人文视野 ——以晋江文学城积分榜单及“清穿文”为例」『中国现代文学研究丛刊』8: 201–212.

黎嘤, 谢宗彦, 张公鹏, 郝志成, 向征（2017）「基于 LDA 的游客网络评论主题分类：以故宫为例」『情报工程』3, 55–63.

战玉冰（2019）「网络小说的数据法与类型论——以 2018 年的 749 部中国网络小说为考察对象 」『扬子江文学评论』5: 53–61.

周志雄（2014）「网络小说的类型化问题研究」『南京社会科学』3: 129–135.

（英語）

Angelov, D. (2020) TOP2VEC: Distributed Representations of Topics, *Computation and Language*, arXiv:2008.09470

Blei, D.M., Ng,A. and Jordan, M. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3: 993–1022.

Blei, D.M. (2012a) Probabilistic Topic Models, *Communications of the ACM*, 55(4): 77–84.

Blei, D.M. (2012b) Topic Modeling and Digital Humanities, *Journal of Digital Humanities*, 2(1): 8–11.

Culpeper, J. (2002) 'Computers, language and characterisation: An Analysis of six characters in Romeo and Juliet'. In: U. Melander-Marttala, C. Ostman and Merja Kytö (eds.), *Conversation in Life and in Literature: Papers from the ASLA Symposium*, Association Suedoise de Linguistique Appliquee (ASLA), 15. Universitetstryckeriet: Uppsala, 11–30.

Deerwester, S., Dumais, S., Landauer, T., Furnas, G., and R. Harshman (1990) Indexing by Latent Semantic Analysis, *Journal of 1he American Society of Information Science*, 41 (6): 391-407.

Hofmann, T. (1999) Probabilistic Latent Semantic Analysis, *UAI'99 Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 289–296.

Jiang, W. (2013) Measurements of development in L2 written production: The case of L2 Chinese, *Applied Linguistics*, 34(1): 1–24.

Jockers, M. and Mimno, D. (2013) Significant themes in 19th-century literature, *Poetics*, 41: 750–769.

Karsdorp, F.（2021）*Humanities Data Analysis: Case Studies with Python*, Princeton University Press.

Le, Q.V. and Mikolov, T.（2014）Distributed Representations of Sentences and Documents Distributed Representations of Sentences and Documents, *Proceedings of the 31st International Conference on Machine Learning*, PMLR 32(2): 1188–1196.

Leech, G. and Short, M.（2007）*Style in Fiction: A Linguistic Introduction to English Fictional Prose*, Second Edition, Routledge.

Mahlberg, M. and McIntyre, D. (2011) A case for corpus stylistics: Ian Fleming's Casino Royale, *English Text Construction*, 4(2): 204–227.

McIntyre, D. and Walker, B. (2019) *Corpus Stylistics: Theory and practice*, Edinburgh University.

Mikolov, T. et al. (2013a) Distributed Representations of Words and Phrases and their Compositionality, arXiv:1310.4546

Mikolov, T. et al. (2013b) Efficient Estimation of Word Representations in Vector Space, arXiv:1301.3781

Moody, C.E. (2016) Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec, arXiv:1605.02019

Moretti, F. (2013) *Distant Reading*, Verso.

O'Halloran, K. (2007) Corpus-assisted literary evaluation, *Corpora*, 2(1): 33–63.

Papadimitriou, C., Raghavan, P., Tamaki, H., and S. Yempala (1998) Latent Semantic lndexing: A probabilistic analysis, *PODS'98 Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART Symposium Principles of Database Systems*, 159-168. doi: 10.1145/275487.275505

Peinelt, N. et al. (2020) tBERT: Topic Models and BERT Joining Forces for Semantic Similarity Detection, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7047–7055.

Stubbs, M. (2005) Conrad in the computer: examples of quantitative stylistic methods, *Language and Literature*, 14(1): 5–24.

Tsuchimura, N. (2021) Exploring Stylistic Features in Agatha Christie's Works through Stylometric Approaches, Doctoral Dissertation, Osaka University.

付録 1「小説家になろう」の投稿ジャンル[55]

大ジャンル	番号	ジャンル	説明
恋愛	101	異世界	恋愛を主題とし、異世界を主な舞台とした小説。
	102	現実世界	恋愛を主題とし、現実に近しい世界を舞台とした小説。
ファンタジー	201	ハイファンタジー	現実世界とは異なる世界を主な舞台とした小説。
	202	ローファンタジー	現実世界に近しい世界にファンタジー要素を取り入れた小説。
文芸	301	純文学	芸術性に重きを置いた小説。
	302	ヒューマンドラマ	人と人との交流、人の一生、人間らしさ等を主題とした小説。
	303	歴史	過去を舞台にした小説。時代小説、タイムトラベルものを含む。
	304	推理	事件などを推理し、謎解きを行なう過程を主体とした小説。
	305	ホラー	読者に恐怖感を与えることを主題とした小説。
	306	アクション	戦闘描写、アクションシーンを主体とした小説。
	307	コメディー	読者を笑わせることを主題とした小説。
SF	401	VR ゲーム	VR 技術を利用したゲームが主体となる小説。
	402	宇宙	宇宙を舞台とした小説。
	403	空想科学	実在、非実在を問わず、何らかの技術・理論の要素を含む小説。
	404	パニック	天災・汚染・大事故・侵略などの危機的状況下を舞台とした小説。
その他	9901	童話	幼年、児童に向けた内容の読み物。
	9902	詩	言葉に美と響きを乗せて想いを表現している作品。
	9903	エッセイ	個人的観点により思想や物事を論じている作品。
	9904	リプレイ	ＴＲＰＧのプレイ結果を文章にて記録したもの。
	9999	その他	上記ジャンルに該当しない作品。

[55] 付録 1 は https://syosetu.com/man/toukoukoumoku/#genre より引用し一部改変

付録 2 あらすじのトピックを構成する主要単語表

トピック	サイズ	キーワード
Topic0	10,987	王太子 侯爵 公爵 婚約 殿下 破棄 令息 子爵 夜会 伯爵
Topic1	9,859	わかん ねー あー かぁ だろ はぁ なになに えー おぉ ねぇ
Topic2	9,854	脱字 文法 文才 誤字 コメント 初めまして 暖かい 処女 アドバイス 拙い
Topic3	9,096	つらつら 思い付い 思いつい 書きなぐっ 詩集 暇つぶし てきとー 文才 思いつき つまん
Topic4	9,041	vrmmorpg mmorpg vrmmo vrmmorpg vr プレイヤー オンライン ライン オン vrmmo
Topic5	8,657	役立た パーティ 初級 低級 パーティー ランク 組め 無能 足手 最弱
Topic6	8,217	統べ 幾千 分かた 終焉 産み落とさ 滅びゆく 災厄 齎し 戦乱 穢れ
Topic7	8,115	みや みず ざわ かんざき はる かわ たちばな りょう ゆい あき
Topic8	7,609	軍勢 統べ 滅ぼさ 魔神 魔族 魔王 打ち倒し 人魔 討ち 相打ち
Topic9	7,505	空虚 分かた 立ち止まる 色褪せ 狂おしい 淡く 蝕み 不幸せ 流転 セカイ
Topic10	7,383	夕暮れ 夕焼け 見上げる 薄暗く 照らさ 明かり 透き通っ アスファルト 降る 街灯
Topic11	7,132	ブサメン アラフォー サラリーマン ニート 童貞 社畜 チート イージー 過労 おっ
Topic12	6,835	悪役 乙女 令嬢 隠しキャラ 推し フラグ ゲー モブ 当て馬 断罪
Topic13	6,647	置き場 小話 こぼれ話 本編 ネタバレ 未読 掌編 スピン 系列 単品
Topic14	6,474	片想い 甘酸っぱい 彼氏 甘酸っぱく 告白 付き合っ 片思い 切なく 失恋 切ない
Topic15	6,105	はははははははは 爆笑 ボケ マラカス アニオタ はちゃめちゃ アニメオタク 県立 漫才
Topic16	6,065	苦しく 恋っ 悲しく ずっと きっと 怖かっ 出逢え 苦し 淡く 想っ
Topic17	6,009	変わり映え 思春 多感 馴染め 億劫 息苦しい 立ち直れ 紗江 空虚 見失い
Topic18	5,879	美咲 片瀬 告白 亜希 しょうた 彼氏 拓也 健太 終業 付き合っ
Topic19	5,761	竜使い 捨て子 カイン エルゼ アイク セリス ルリア ウルス 王都 アレン
Topic20	5,516	新卒 社員 入社 会社 残業 就職 倒産 出勤 面接 外資
Topic21	5,423	詫び 累計 pv くださり ありがとう ブックマーク 下さっ モチベ 休載 くださっ
Topic22	5,388	ノンケ セックス スパダリ 自慰 淫乱 公総 相姦 オメガバース フェロモン セフレ
Topic23	5,229	店内 店員 喫茶 店主 来店 常連 店長 開店 bar 裏通り
Topic24	5,153	www. .net series novel https .jp id show detail member
Topic25	5,080	内気 片想い 引っ込み 大人しい 見知り 席替え 奥手 おとなしい 甘えん坊 彼氏
Topic26	5,061	平等 枠組み 淘汰 普遍 強食 弱肉 万別 千差 弱者 貧富
Topic27	5,061	刑事 捜査 犯人 警部 警視 犯行 アリバイ 殺人 他殺 密室
Topic28	5,012	詫び 手違い 特典 女神 轢か チート 転生 貰い 手際 貰っ
Topic29	5,008	マンホール 足元 吸い込ま 草原 真っ白 見知ら 方陣 覚める 見覚え さめる
Topic30	4,924	憎み 憎い 赦さ 憎悪 憎しみ 復讐 憎ん 断頭 裏切ら 恨み

Topic31	4,919	再婚 引き取ら 亡くし 親戚 連れ子 離婚 母子 両親 祖父母 女手
Topic32	4,918	資料 ホライズン 百科 添付 準拠 史料 文献 テキスト wikipedia 散文
Topic33	4,852	父王 後宮 即位 側室 皇后 正妃 皇帝 崩御 世継ぎ 皇太子
Topic34	4,823	一人暮らし 引っ越し アパート 引っ越す 転勤 マンション 帰省 入居 独り暮らし 都会
Topic35	4,795	あいつ 恋っ たって アイツ ふふっ ずーっと なんて わかん あぁ 怖かっ
Topic36	4,788	小国 大国 帝国 諸侯 共和 北方 強国 列強 兵力 皇国
Topic37	4,750	アニオタ てきとー ワイワイ ヲタク コスプレ 手芸 ol ふわ マラカス 小森
Topic38	4,734	船長 旅人 遺跡 船乗り 船員 海辺 リティル 末姫 港町 海賊
Topic39	4,704	海軍 陸軍 艦隊 空母 ソ連 中隊 国防 日米 連隊 海兵
Topic40	4,696	王都 王国 ティア ラス セル ヴィア カイル エルゼ フォード 公国
Topic41	4,674	可愛く 恋っ 彼氏 付き合え かわいい 照れ 可愛い ちょー モテる イケメン
Topic42	4,617	彼氏 ol 社内 セフレ 不倫 やけ酒 浮気 既婚 振ら 付き合っ
Topic43	4,592	始業 学期 担任 教室 終業 進級 授業 公立 放課 学年
Topic44	4,461	人類 絶滅 死に絶え 滅び 荒廃 侵略 繁栄 侵攻 大戦 数百
Topic45	4,418	喰らう 狩る 異形 喰う 化物 化け物 喰らい 血肉 生き血 怪物
Topic46	4,374	力者 異能 発現 扱え アビリティ 行使 扱える 生まれ持っ 先天 言霊
Topic47	4,321	警視 部隊 極秘 特務 特捜 公安 警察 直轄 任務 隊員
Topic48	4,199	むかしむかし 昔々 ずきん 童話 ばあ 桃太郎 浦島 トナカイ おとぎ 白ウサギ
Topic49	4,187	淡く 咲い 花弁 花びら 咲き誇る 花々 咲く 儚く 舞い散る 儚い
Topic50	4,160	淡く 温もり 愛おしく 狂おしい 碧い 愛しく 寂し 出逢え 逢え 愛おしい
Topic51	4,094	東方 project 害さ 助長 オリキャラ 不快 ブラウザ 飲酒 上海 喫煙
Topic52	4,093	詩集 エッセイ 評論 つれづれ 散文 自伝 つらつら 書き連ね 徒然 日記
Topic53	4,070	借金 返済 日銭 払え 高額 日暮らし 多額 商人 小銭 稼ぐ
Topic54	4,056	ai 知能 人工 コンピューター 搭載 自律 チップ ネットワーク アンドロイド コンピュータ
Topic55	4,017	カリキュラム 国立 学院 王立 実技 養成 高等 首席 名門 特待
Topic56	4,007	猫耳 尻尾 モフモフ もふもふ トカゲ しっぽ 生え ケモミミ 獣人 コボルト
Topic57	3,997	玄関 トイレ ドア ソファー 夕飯 布団 風呂 シャワー チャイム ベランダ
Topic58	3,991	巫女 祀る 霊山 言い伝え 代々 豊穣 祀ら 神子 昔々 神事
Topic59	3,982	患者 病院 病気 病名 入院 奇病 難病 病棟 内科 患っ
Topic60	3,970	発車 見切り 行き当たり ばったり いきあたり 文才 書きなぐっ 思いつき 息抜き 生暖か
Topic61	3,954	怪人 戦隊 結社 ヒーロー 裁け ギャング ジャスティス マフィア ヴィラン 特撮
Topic62	3,915	きに 終結 停戦 休戦 人魔 国々 両国 協定 戦乱 争っ
Topic63	3,846	陰陽 妖怪 安倍 晴明 平安 式神 退魔 頼光 物の怪 魍魎
Topic64	3,843	キャ カースト イジメ いじめっ子 ぼっち 根暗 席替え リア充 ボッチ いじめ
Topic65	3,785	怪奇 怪異 神隠し 心霊 怪談 奇怪 不可思議 現象 オカルト 噂話

Topic66	3,739	俳優 芸能 声優 女優 プロダクション オーディション 共演 売れっ子 アイドル 番組
Topic67	3,731	感染 ウイルス ウィルス 人口 汚染 蔓延 政府 ワクチン パンデミック 温暖
Topic68	3,704	天寿 来世 看取ら 生まれ変われる 輪廻 生まれ変わる 老衰 全う 寿命 生まれ変わっ
Topic69	3,690	惑星 宇宙 恒星 航行 木星 衛星 異星 光年 探査 火星
Topic70	3,614	飛び降り 飛び降りよう やり直せる 投身 やり直す 自殺 やり直せ 死のう 飛び降りる
Topic71	3,585	思い出せ 思い出そう 記憶 思い出せる 失くし 断片 抜け落ち 暗闇 碧い 病室
Topic72	3,574	野球 甲子 強豪 選手 サッカー 投手 決勝 バスケットボール ピッチャー コーチ
Topic73	3,541	部員 入部 廃部 部活 部室 部長 同好 演劇 吹奏 顧問
Topic74	3,534	境界 迷い込む 裂け目 パラレル 幻想 歪み 霊夢 夢現 凡々 異界
Topic75	3,515	身長 短髪 髪型 メガネ 長身 黒縁 体型 茶髪 cm 三つ編み
Topic76	3,472	実話 ノンフィクション 詩集 自伝 エッセイ つらつら 赤裸々 散文 うつ病 脚色
Topic77	3,307	食べる 美味しく 食べ物 調味 食材 食べ 料理 美味しい 菓子 美味い
Topic78	3,294	優勝 競技 大会 決勝 対戦 出場 トレーディング アーケード デュエル 熱狂
Topic79	3,286	インストール アプリ 画面 パソコン ダウンロード スマホ タップ スマート インターネット
Topic80	3,284	迷宮 潜り 財宝 遺跡 千金 踏破 探索 一攫 トレジャー ダンジョン
Topic81	3,263	生徒 私立 中高 風紀 共学 全寮 会長 一貫 役員 委員
Topic82	3,261	戦国 織田 信長 武将 秀吉 家康 豊臣 天正 天下 大名
Topic83	3,217	進歩 科学 技術 飛躍 高度 粒子 物質 エネルギー 革新 発達
Topic84	3,214	あの世 閻魔 天国 生き返る 死後 冥界 三途 輪廻 地獄 現世
Topic85	3,207	重体 通り魔 重傷 刺さ 瀕死 取り留める 一命 轢か 庇い 庇っ
Topic86	3,196	薬草 調合 錬金 ポーション 採取 薬師 術師 採集 薬屋 鍛冶
Topic87	3,179	霊感 心霊 幽霊 除霊 怪談 霊能 肝試し オカルト 怪異 悪霊
Topic88	3,163	端麗 文武 両道 明晰 抜群 眉目 容姿 兼備 品行 秀麗
Topic89	3,155	トラック 轢か 撥ね 信号 跳ね はね 歩道 横断 車道 ひか
Topic90	3,144	探偵 助手 ホームズ 事務 シャーロック 刑事 ロンドン ワトソン 所長 密室
Topic91	3,099	his but this that and the of is in are
Topic92	3,097	死刑 監獄 脱獄 囚人 判決 牢屋 看守 牢獄 刑務 収監
Topic93	3,087	しれ また 本当 この 思う ちゃんと 思い これ 変え いく
Topic94	3,084	人型 機動 パイロット 操縦 機体 兵器 ロボット 愛機 装甲 搭乗
Topic95	3,009	音楽 演奏 ギター ピアノ 楽器 弾い コンクール 作曲 ボーカル 弾き
Topic96	3,008	幕末 江戸 浪士 新選 歳三 新撰 土方 戊辰 明治 会津
Topic97	2,865	乗客 電車 列車 停車 車内 乗車 トンネル 幹線 運転 線路
Topic98	2,772	差出 送り主 手紙 封筒 メール 届い レター 宛先 小包 宛て
Topic99	2,746	コミカライズ コミックス コミック ブックス 書籍 発売 受賞 オーバーラップ 刊行 mf

付録 3 コメントのトピックを構成する主要単語表

トピック	サイズ	キーワード
Topic0	3,269	おもろかっ 大当たり 読み入っ 出合え ウズウズ うおー おもろ 冬休み イール なつかし
Topic1	2,840	スキル ステータス レベル アイテム プレイヤー 装備 効果 システム 攻撃 取得
Topic2	2,138	王子 貴族 相手 婚約 立場 彼女 馬鹿 本人 令嬢 権力
Topic3	2,075	素敵 幸せ ありがとう 気持ち とても 本当 ござい でし よかっ 感動
Topic4	1,781	文章 作品 キャラクター ストーリー 物語 個性 小説 描写 非常 読者
Topic5	1,518	ありがとう ござい 読ん 読ま 完結 いただき 作品 一気 とても まし
Topic6	1,498	投稿 書い 作品 執筆 書き 下さい ください 感想 完結 申し訳
Topic7	1,439	織田 信長 大名 徳川 武田 家康 毛利 北条 家臣 戦国
Topic8	1,406	海軍 日本 戦争 アメリカ 技術 陸軍 資源 開発 大戦 イギリス
Topic9	1,247	移動 武器 可能 攻撃 必要 する できる 使用 火薬 程度
Topic10	1,201	貴族 立場 相手 権力 未遂 責任 王命 理由 無能 犯罪
Topic11	1,152	文章 句点 改行 やすい にくい 文法 表現 句読点 部分 多い
Topic12	1,123	書籍 発売 買い web 完結 作品 小説 コミカライズ 購入 コミック
Topic13	1,109	文章 テンポ ストーリー やすく やすかっ キャラクター 個性 とても 読み 丁寧
Topic14	1,070	ござい めでとう 完結 ありがとう 続編 嬉しい 番外 更新 読ま いただき
Topic15	1,048	値段 安い 価格 必要 銀貨 金貨 素材 原価 作る 銅貨
Topic16	964	更新 嬉しい 待ち 気長 ござい 読ま 願い いただき 応援 下さい
Topic17	876	kg http:// 食べ .jp もの would それ htm らしい guild
Topic18	873	佐倉 可愛い ちゃん バレー ぎゅー 爆笑 チロ 女の子 さん ハーヴェスト
Topic19	865	気長 待ち 更新 再開 めでとう 待っ 嬉しい 経とう おもろかっ 便り
Topic20	793	自愛 ください 執筆 下さい 体調 頑張っ いただき おり 更新 応援
Topic21	789	おおお ない それ まあ そんな 柚木 しか もん 悪い って
Topic22	778	幸せ ギデオン エリカ 気持ち 素敵 想い リサ 本当 ファリティナ フェリシテ
Topic23	777	なる だけ いる どう ちゃう この それ 状態 そう まあ
Topic24	766	作品 定期 小説 感想 書く なろう 書い 感じ 評価 物語
Topic25	765	脱字 漢字 誤字 箇所 部分 文章 変換 表現 にくい 指摘
Topic26	747	シメオン くる マリエル ワイ まで ない しかし でしょう まあ ミレーヌ
Topic27	739	カレー 風味 だけ いけ ない 意味 って なら 悪い しまう
Topic28	735	面白かっ 読ま おもしろい 読み 応援 一気 おもしろかっ とても 読ん 面白い
Topic29	735	旦那 きっと ヴィー ヴィオラ ちゃん サーシス 二人 っぷり あの 幸せ
Topic30	726	テンポ ストーリー 引き込ま 個性 文章 キャラクター やすく やすかっ とても 構成

Topic31	712	アイディア もん 後続 全裸 だけ ピッ 更新 ゲーム 競馬 ネタ
Topic32	712	円城 まし 書い でし 待ち 麗華 一番 喜び 最近 ます
Topic33	711	精神 ない 主人 ヒロイン いう こう だけ 意味 問題 でしょう
Topic34	705	だろう 人間 けど だけ する ない もの って 主人 相手
Topic35	702	水原 きっと あの でし ミーンミンミンミン もう ねー やっぱり そんな ちゃん
Topic36	688	レイン から そう 登板 とも 頑張れ 今回 言え やはり そして
Topic37	683	そう ない って なんか だろう なん 悪い それ あっ ちゃっ
Topic38	683	因果 すぎ 応報 そう だっ たい なん 気持ち られ よう
Topic39	683	執筆 ください 下さい 更新 頑張っ がんばっ 応援 ペース 自愛 気長
Topic40	681	ない ところ ちゃっ あっ なく いい なぁ こう 書い って
Topic41	681	大当たり ひと味 断トツ ブロド おもろかっ 読み入っ 煉瓦 ネタバラシ ずば抜け ハンデ
Topic42	680	そう それ ない たら 悪い から らしい だろう れる いう
Topic43	679	ない だろう たら いい やら 言っ それ そう いや まあ
Topic44	674	ユミエラ 少し 個人 非常 最強 文章 納得 ちょっと られる 感じ
Topic45	673	ヒロイン 悪い たい 女の子 言え 彼女 可愛い 桑島 アナスタシア 気持ち
Topic46	665	いい なん らしい どう でしょう たい そう たら わけ れる
Topic47	653	いい あっ って なぁ たら ところ ない 何故 いう 主人
Topic48	651	北伐 そう なんて から その 悪い それ ちゃい どう くる
Topic49	650	テンポ ストーリー サクサク 文章 読める 読め やすく 描写 とても 面白い
Topic50	646	ジェットン よう 悪い マガリ しまう もの そう だけ その 感じ
Topic51	645	ながら 気持ち 互い もっと 素敵 性格 たい まし すごく 雰囲気
Topic52	643	コルク 課金 いい 感じ こう なぁ 周り ハインド なる バトル
Topic53	643	アン デッド って なんて ねぇ なんか 出来る 流石 そう 以外
Topic54	641	まで しまう 本当 感じ つつ 気持ち 思い すぐ ポイズン ない
Topic55	640	サフィア アレク られ シェイラ つつ それ どんな いう わけ 流石
Topic56	638	こう いる ねー インフレ こそ 相手 ない いう また つまり
Topic57	632	なさい 更新 完結 嬉しい かえり 読み直し 読み返し ござい めでとう 久し
Topic58	631	試合 選手 野球 サッカー イニング チーム スポーツ 投手 高校 甲子
Topic59	624	感じ 展開 もっと ヒロイン こう いい まで なく いう いく
Topic60	621	展開 感じ いく しまう だけ こそ やはり もの たい まで
Topic61	617	いただき いただい 読ま 頂い 頂き 更新 楽しく 拝見 ござい 完結
Topic62	614	バルバトス 通行 よかっ ネリ ラーム しまう くれる めっぽう プギャー ネリエル
Topic63	613	半端 中途 感じ ヴェルナー 悪い マゼル ヒロイン 主人 納得 自覚
Topic64	611	セシル カイル エステル アシュレイ 気持ち ヒロイン 不憫 王子 自覚 一途
Topic65	598	ない だろ 言っ だろう 彼氏 できる その 言葉 じゃ それ

Topic66	594	って しかし そこ それ だけ だろう ない フィリ 意味 いう
Topic67	587	コルク ガレル クラスト 詩織 泣け まへ たい くれる ロッシュ ロッシェ
Topic68	585	どう なけれ 言っ 言え 頭身 それ ない だけ わけ する
Topic69	582	こう いっ ここ って だけ いう ちょっと なら いえ けど
Topic70	582	いけ 言っ それ こう なん そう 自分 そこ 意味 ねー
Topic71	577	でしょう いく そこ イメージ たい それ だけ 少し こう しまう
Topic72	574	貴女 プリシラ 言い ぐっ テポドン いけ 畳む ドゥエ じょ そんな
Topic73	574	素敵 読ん 気持ち 出会え 本当 まし ありがとう 読み 感謝 とても
Topic74	574	オリヴァー 正気 まあ 巨乳 って やっ 意味 言う ない 悪い
Topic75	571	エミーリア しよう some たい 吸っ 無茶 逃亡 デス って テンペスト
Topic76	570	わけ ない って だけ あれ それ たら だろう てる たり
Topic77	565	ない それ その もの いる なら はず しか 以外 あれ
Topic78	562	アルファ 読者 描写 感じ 少し 理由 ストーリー 行動 主人 思い
Topic79	561	降参 タル ムミネウシンム 発声 爆散 ボス 白旗 サプライズ 海月 カーズ
Topic80	559	ヘリオス アリア テリオス カーン そして ジュリアス ヴァン ウラ クロエ あの
Topic81	555	まで 書い あっ それ 引退 天気 なる です クリエーター ここ
Topic82	554	それ ない だけ どう ねぇ なん なんか 通り 賞味 だろう
Topic83	551	大地 もふもふ 恵み 次元 与えれ 封印 尊い 成長 さいこ たい
Topic84	551	物語 感じ 印象 文章 しっかり キャラ delta 読者 プロファイル 惚れ薬
Topic85	549	気持ち イザドル シャーリー 感じ とても 物語 素敵 魅力 両論 こそ
Topic86	549	vrmmo エーミール 赤井 過ぎる 語り 違和 感じ そこ ゲーム リアル
Topic87	543	いい しかも 貴様 精神 そんな なん ヒロイン まあ 女の子 可愛い
Topic88	542	新年 今年 明け 早々 終わっ いただき ざい めでとう 期待 失礼
Topic89	542	人間 魔神 つつ ヒロイン 和解 なっ 主人 なく 気持ち まで
Topic90	539	ねー こう なん ねぇ いい 吹き じゃ 食べ たー まあ
Topic91	539	出し 両親 なっ 生きる 思え まで 旦那 こそ まるで オウロ
Topic92	531	順調 ウサギ ぬるく 楽しみ ミーア 今後 いただき どの 次回 ます
Topic93	510	優花 そこ いぴき いう なん すら 課程 いっ 言っ それ
Topic94	509	マン リザード ナナ 日常 キャラ 登場 基礎 こう 尻尾 上限
Topic95	490	ラーニャ 魅力 作品 いただき 好み 恋愛 読ん ヴァイト シーン ヘレナ
Topic96	477	解る 大した 違和 いう 理解 ない 問題 しか 前編 デー
Topic97	474	蹂躙 くさい そこ 多く 過ぎる いる 歪み 意味 思考 クレア
Topic98	456	ウル イザベラ シュ チラリ ヒロイン 更新 展開 たー 応援 いい
Topic99	447	続き おもろかっ プリーズ 出逢える ぶっ通し 経とう 般若 来月 忙しかっ こうしん

謝辞

本論文を作成するにあたり、主指導教員の田畑智司先生、副指導教員のホドシチェクボル先生および岩居弘樹先生に懇切丁寧なご指導と貴重なご意見をいただきました。また、日本語表現のチェックをしてくださり、適切な助言もしてくださいました。先生方のおかげで、この博士論文を無事に完成することができました。心より厚く御礼を申し上げます。

本研究科に所属している先輩と後輩の皆様には常に暖かい励ましと応援をいただきました。特に藤田郁さんと徐勤さんには論文に関する貴重なアドバイスを多くいただき、心より感謝申し上げます。

最後になりましたが、いつも温かい目で見守ってくださった両親と親友には、心より感謝申し上げます。