

# ライトノベルからショートショートへの変換法

山田真彦 天沼博 松澤和光

神奈川大学大学院 工学研究科

r202170165yp@jindai.jp amanuma@kanagawa-u.ac.jp

matsuk90@jindai.jp

## 概要

本稿では、ライトノベルをショートショートへ変換するための手法を提案する。この手法により、人間の物語生成を支援する事が出来ると期待する。まず、ライトノベルから特定の語句を含む文章などを除外し、残りの文章数から3分割する。次にそれぞれからショートショートを構成できそうな文章を選び、その文章に含まれる固有名詞を抜き出す。最後にそれぞれの文章群から抜き出した固有名詞を含む文章をさらに抜き出し、文章の登場順に並べ完成系とする。作品によって抜き出す文章量が一定ではないため、文章を抜き出すための新たな基準を設ける事が今後の課題である。

## 1 はじめに

最近、AIによる文章生成に大きな技術進歩があり、「AIのベリすと」や「ChatGPT」が大変な話題になっている。これらのシステムでは、利用者の簡単な入力情報を元にして自然な文章による「簡易な物語」が自動生成される[1]。ただし、これらの仕組みは基本的にDeepLearning系の技術であり、膨大なテキストデータを元にして文章(の流れ・接続)として生起確率の高い、「いかにもそれっぽい」「ありがちな」文章を生成しているに過ぎない。真の意味での「言語理解」の研究は未だ発展途上にあり、現在のAIは言語を理解して物語を生成しているわけではない。

確かに、生成された文章によっては物語的な「匂い」が感じられるものもあるようだが、それは読み手が勝手に自分なりの意図を投影して、無理に読み取っているだけだと推測される。物語を生成する上で、AIに何らかの創造的な意図が存在しているわけではないのだ。そんな「空虚な物語」を「読む意味」が果たしてあるのだろうか。

物語を生成する主体は、現状ではやはり人間でしかないと考える。その上で、その創造過程をAIに

様々な形で手伝い協力してもらうことは大いに有用だろう。この考えに基づいて、本研究では人間の物語生成を支援する新しい手法を提案する。具体的には、既存のライトノベルから「物語(的な要素)」を多様に抽出し、これをショートショートの形に変換する手法である。俗にライトノベルと呼ばれるジャンルでは、純文学に見られるような明確なテーマに基づいた作品よりも、作者の奔放な創作意欲のままに「荒削り」に量産された作品が多いと言われる。そうした作者本人にとってさえ未整理・多彩・多重に織り込まれた物語要素を、個々の完結した小さな物語(ショートショート)として抜き出すことにより、今後の新たな創作や他の作者との共同制作、複数作品のリミックス等に役立つことが期待できる。

ショートショート生成の既存研究としては、ショートショートを模範利用した新たなショートショートの自動作成や、ショートショートの会話を抜き出し他の作品に置換して自動生成を行うものがある[2, 3]。しかしこれらは、既存のショートショートの表現を一部変更し別の作品とする研究であり、ライトノベルのような長編作品から物語を抽出しショートショートに変換する研究やシステム開発は存在しない。

作品をショートショートに変換するにあたり、どのような観点から変換すべきかと考えた際に、落語の三題噺と呼ばれる形態に着目した。三題噺とは、3つの題目から1つの話を作るというものである。

本研究では作品の全文を3分の1ずつに分け、それぞれからショートショートを構成の元とできそうな文章に含まれる単語を抜き出し、それを含む他の文章をそれぞれの文章群から抜き出し繋ぎ合わせる事でショートショートとする手法を提案する。

## 2 提案手法

### 2.1 手法の概要

提案手法を図1に示す。まず、使用作品の本文(章タイトル、あとがき及び巻末の外伝作品等を除く)の内容全てを句点や感嘆符、疑問符などの文章の終わりごとに列に変換したテキストファイルを用意する。そして、この列の中から不要な文章とする条件を満たした文章を除外する。この不要な文章とする条件については2.2節にて示す。次に変換したテキストファイルの行数を数え、その行数の3分の1ずつに分割し3つの文章群(それぞれ前、中、後)とする。その後、それぞれの文章群からショートショート構成の元となる文章を抜き出し、抜き出した文章内の固有名詞を含む文章をすべて抜き出すことを前中後全てで行い、登場順に並べてショートショートとする。この時、構成元となる文章内に固有名詞が複数個存在する場合は人名、特に主人公や物語における重要人物の名前を優先的に選択する。本稿では構成元とする文章および文章内の固有名詞の選択においては、2.3節に示す条件及び方法を用いて、文章群の中から条件を満たす文章を事前に選択し、別のデータとして使用している。

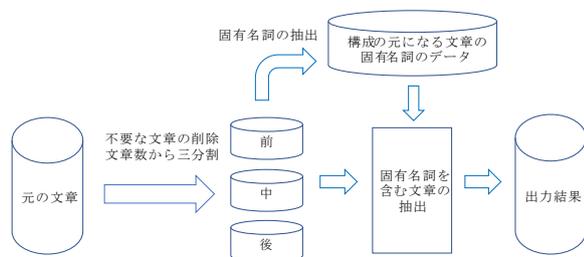


図1 提案手法の概要

### 2.2 不要な文章の条件

本研究において以上の処理の他に、文字数の制限及び一部語句、記号を含む文章の削除も行う。文章の削除の基準は以下のようになっている。

- 11[字]以下の文章
- 42[字]以上の文章
- 「、』『、(、《》の組み合わせの内片方のみの文章
- 「言」、「叫」、「呟」、「つぶや」を含む文章

文字数の制限については、試行を行った結果、極端に長い、もしくは短い文章はショートショート作成に不向きであると考えられるため、11[字](10[字]+句点、感嘆符、疑問符等)以下の文章及び42[字](41[字]+句点、感嘆符、疑問符等)以上の文章は削除を行った。

3つ目の条件については全ての作品において、会話文や心情描写において複数の文が含まれる事があり、以上の組み合わせのうち片方のみの文章が多く見られたため、このような条件を設けた。どちらかのみの文章を含めた場合、片方のみが抜き出された時に文章として成立せず、また片方のみの文章のほとんどが作品の内容に影響を及ぼさないため、抜き出す文章として不要であると考えたからである。

4つ目の条件については、3つ目の条件や会話文に固有名詞が含まれていない場合が多いため、直前の会話文が出力結果に存在しないにも関わらず、「言った」や「叫ぶ」、「つぶやく」といった文章が多く出力されていたため、直前の会話文に関連しているこれらの文章は不要であると判断し削除に至った。

### 2.3 構成元とする文章の基準

#### 2.3.1 構成元とする文章の条件

ショートショートを構成しそうな文章を抜き出す条件を以下に示す。

- 固有名詞(内容に深く関わるもの)を含む文章
- 文末が「た」、「だ」で終わる文章
- 疑問や問いかけの内容の文章

1つ目についてはこれまでに書いた通り、出力する内容は固有名詞を元に作成しているため、最も重要な条件である。しかし全ての固有名詞を対象にしているわけではなく、作品の内容に深く関わるものであり、内容に関係ない固有名詞は対象外である。

2つ目と3つ目に関しては、星新一のショートショートの作品パターンを参考にした。先行研究より、星新一のショートショートの展開パターンとして最も多いのは**真実の発覚**であり、謎の揭示、発覚及び解消を表すものである[4]。これに該当する文章は文末が「だった」、「なのだ」といった「た」や「だ」で終わるもの、もしくは「だろうか?」や「なのだろうか?」といった疑問や問いかけを示す表現が入っているためこの条件を設けた。

#### 2.3.2 文章を選択する方法

本研究では、2.3.1節の条件を元に前中後それぞれの文章群から構成元となる文章を事前に選択してい

る。選択方法は以下の通りである。まず条件を満たす文章を前の文章群は登場順に、中と後は文末から逆順に文章を探す。前の文章群のみ登場順に文章を探すのは、多くの作品において序盤に世界観の説明や、内容全体において重要となる疑問が書かれているため、登場順に確認した方が良いと考えたからである。また中と後に関しては、物語が進むにつれて情報が増えていき、特に終盤の文章は重要性が高い事が多いため、文末から文章を探すべきであると考えた。中の文章群においては、後のものよりも重要性は薄いショートショート作成には十分な文章を獲得できるのではないかと考え、後と同様の方法で文章の選択を行った。

また上記の方法に加えて、中と後の文章群の文章は前の文章群のものとは違う固有名詞を含む文章を選択するようにした(後の場合は中のものとも違う固有名詞を選択する)。つまり、上記の方法で文章を探した際に、他の文章群で選択された固有名詞のみを含む文章が先に登場した場合はその文章は無視し、条件に合った上で別の固有名詞を含む文章が登場するまで文章の探索を続ける。また見つけた文章が複数の固有名詞を含んでいた場合は、他の文章群で選んでいない固有名詞の中から選択する。この処理については、多くの作品においてこの処理を行わなかった場合、使用する固有名詞が同じ事が多く、作品によっては全て同じ固有名詞の場合もあった。これでは三題漸に着目した今回の手法とは合わないため、文章群ごとに固有名詞を変更するに至った。

### 3 評価実験

#### 3.1 使用する作品について

本研究ではライトノベル作品を 10 タイトル使用する[5, 6, 7, 8, 9, 10, 11, 12, 13, 14]。これらは有名な作品から無作為に選んだものである。

#### 3.2 出力結果の文字数の目標

本稿における出力結果の文字数の目標を 4000[字]とする。これは「坊っちゃん文学賞」というショートショートの文学賞の募集条件が 4000[字]以内であるからである[15]。

#### 3.3 実験結果

出力結果を表 1、図 2 に示す。表 1 より、文字数の目標に達した作品は[5]、[6]、[7]、[8]、[12]、[13]、

表 1 実験結果

番号	元々の文字数[字]	出力結果[字]
5	135356	2878
6	176767	3408
7	121260	2341
8	168404	1106
9	118549	4101
10	87719	4738
11	134012	8464
12	114657	2522
13	158253	3606
14	180510	2606

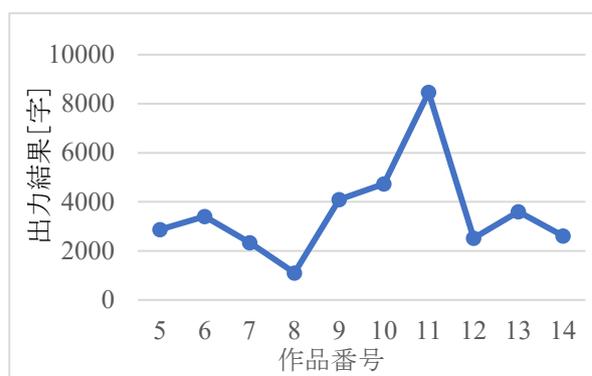


図 2 実験結果のグラフ

[14]の 7 作品であり、他の作品においては[9]、[10]が 4000[字]台であり、[11]に関しては 8464[字]と目標字数を大きく超えるという結果となった。

[11]の出力結果の文字数が多いのは、[11]の文章において代名詞がほとんど使用されておらず、登場人物の名前がそのまま使用されている場合が多いため、抽出される文章が他作品と比べて多くなったからであると考えられる。これに関しては、文章群ごとに抽出する文字数の上限を設け、それを上回った場合は重要ではない文章を探し削除する処理を行う必要があると考える。そのために固有名詞を含む文章の中から不要な文章を削除するための新たな基準の検討が今後の課題である。

[9]に関しては、数文削除すれば目標字数内に収まるため、[11]の場合と同様に不要な文章を削除するための基準を設ければよいと考える。[10]も同様であり、[9]よりも削除する文章は多いが、4000[字]台に留まっており、出力結果の内容においても、7 割近くの文章が前の文章群から抽出されたものである

ため、新たな基準によって前の文章群から抽出する文章の数を減らす事により解決可能であると考え。

また、目標字数以内に収まった作品においては[6]作品が 2000[字]から 4000[字]の範囲であるが、[8]は 2000[字]を下回り 1106[字]となっている。この原因は 2 点あると考える。1 つは、元とした固有名詞は重要であるが、それに関する出来事についての説明が主であるため固有名詞自体の登場する回数が少ない点である。これについては、本研究が固有名詞からショートショートを作成する手法を行っているために起きた状況であり、このような場合は固有名詞に関係のある文章と固有名詞を関連付け、こちらも抽出するといった処理が必要であると考え。もう 1 つは、他の文章群では多く登場するが、参照した文章群ではあまり登場しなかったという点である。今回の場合は後の文章群で選択した固有名詞に起きた状況であり、物語が進むにつれて他の要素が増えて話の流れが変化し、この固有名詞が触れられなくなったため登場しなくなったと考えられる。

### 3.4 出力結果の内容について

出力結果の内容を読んだ結果、ほとんどの内容に整合性がなく、物語というには不完全であった。また出力結果において、以下の 4 点が目立った。

- 指示詞の内容の変化、及び不在
- 内容の急な転換
- 会話文の発言者の変化
- 同じ発言者の会話文の連続

指示詞の内容の変化及び不在に関しては、指示詞が指している文章が除外され、それ以前の文章に続く形となるため、指示詞の内容が変化、もしくは存在しない内容に対して指示詞を使用しているといったものになっている。これについては、文章内の指示詞とその指す内容をセットとして扱うような事前処理が必要であると考え。

内容の急な転換は、主に次の文章群が切り替わった後、もしくは本文中において元となる固有名詞が次に使用されるのが十数行から数十行後であった場合に起こる。これは固有名詞を含まない文章を削除した事により、出力される文章の間での出来事や説明が取り除かれてしまい、次に出力される文章の時点では既に別の話題に変わっているという状況になるからであると考え。対処法として、同じ文章群内であれば次の文章までに間隔がある場合は、前後の内容に関係のある文章を出力される文章間から追

加で抽出し、文章群をまたぐ場合は、文章群の最後の文章と、関係のある次の文章群の序盤の文章をどちらも抽出するといった処理が必要であると考え。

会話文の内容や発言者の変化も指示詞の場合と同様に、発言者に関する情報を含む文章が除外され、抜き出された前後の文章の関係で別の発言者による会話のように捉えられてしまうといったものである。これに関しては、会話文の内容からどの人物による発言かを判別し、その人物に関する前後の文章とセットにするといった処理が必要であると考え。

同じ発言者の会話文の連続については固有名詞が人名の際に多く起こり、その人物を呼ぶ、考えるといった内容の会話文が連続して抜き出される事である。これは前後の文章内に人名が含まれていないため、文章が除外されてしまったために起きたものと考えられる。この場合、連続して出された会話文のほとんどが内容と関係がない文章となっているため、発言者の変化の時のように発言者を判別し、同じ発言者の会話文が連続した場合 1 文目以降を削除するという処理を行うのが最適であると考え。

## 4 おわりに

本研究では、三題噺の発想に着目しライトノベルからショートショートを作成する手法を提案した。結果としては 10 作品中 6 作品が 2000[字]から 4000[字]の範囲に収まったが、1 作品は 1000[字]台と少なく、3 作品は目標字数よりも多くなった。文字数が多い作品に関しては、文章群ごとに抽出する文字数の上限を設け、上限を超えた場合に不要な文章を削除するための新たな基準の検討が必要である。また、文字数が少ない場合の中で指定した固有名詞の登場回数が少なかった場合は、その固有名詞に関係する文章を関連付けて固有名詞を含む文章と共に抽出するという処理が必要であると考え。

また、作成されたショートショートの内容に整合性を持たせるために、指示詞とその内容の文章のセット化、離れた文章の内容をつなぐための文章の追加、会話文と発言者の関連付け、同じ発言者による会話文が連続した際に会話文を減らすための処理の方法の検討も今後の課題となる。

特に出力結果の内容の部分に多くの課題が残る形となってしまったが、今後の研究により実現されることに期待する。

## 参考文献

1. ー. 超話題の人工知能 ChatGPT に“小説”や“詩”を書いてもらい、“プログラム”は実行してみた. (オンライン) (引用日:2022年1月5日.) <https://weekly.ascii.jp/elem/000/004/116/4116274/>
2. 模倣と置換に基づく超短編小説の自動生成. 緒方健人他. 2014年度人工知能学会全国大会(第28回). 2014.
3. 会話を中心とした超短編小説の自動生成. 高木大生他. 2014年度人工知能学会全国大会(第28回). 2014.
4. 星新一ショートショート文学の物語パターン抽出. 佐藤知恵他. 情報知識学会誌, 2010, 第20巻. No.2. 123-128.
5. 大森藤ノ. ダンジョンに出会いを求めるのは間違っているだろうか. SBクリエイティブ株式会社, 2014.
6. 香月美夜. 本好きの下克上～司書になるためには手段を選んでられません～第一部 兵士の娘 I. TOブックス, 2020.
7. 上遠野浩平. 恥知らずのパープルヘイズ ―ジョジョの奇妙な冒険より―. 株式会社集英社, 2011.
8. 伏瀬. 転生したらスライムだった件①. 株式会社マイクロマガジン社, 2014.
9. 谷川流. 涼宮ハルヒの憂鬱. 株式会社角川書店, 2003.
10. 時雨沢恵一. キノの旅 the Beautiful World. 株式会社KADOKAWA, 2020.
11. 鎌池和馬. とある魔術の禁書目録. 株式会社KADOKAWA, 2018.
12. 佐島勤. 魔法科高校の劣等生① 入学編〈上〉. 株式会社KADOKAWA, 2019.
13. ソードアート・オンライン 1 アインクラッド. 株式会社KADOKAWA, 2020.
14. 丸山くがね. オーバーロード 1 不死者の王. 株式会社KADOKAWA, 2020.
15. ー. 坊っちゃん文学賞ーショートショート募集. (オンライン)(引用日:2022年12月18日.) <https://bocchan-shortshort-matsuyama.jp>.