

テキストマイニングによる読書感想文の新たな可能性

—読書感想文の省力化を目指して—

変態学研究所

夏休みにおける定番の宿題といえば、読書感想文である。しかし、書籍を選び、読み、理解し、文章として感想文を作成することには大変な労力を要する。それを踏まえて、テキストマイニングを用いて読書感想文を作成することが可能であるかを検討した結果、テキストマイニングの結果から読書感想文を書くことは可能であることが分かった。作成においては、「ワードクラウド」、「単語出現頻度」から物語中のキーワードを、それらの関係性を「共起キーワード」や「階層的クラスタリング」を用いて同定して物語の骨子を組み上げ、他の書籍ではみないような特徴的な単語や、形容詞から物語内の情緒を汲み取って肉付けすることで物語を独自に再構成し、それに対する読書感想文を書くという手順を踏めばよいと考えられる。

1. 背景

夏休み等の長期休暇における宿題の定番であり、多くの児童・生徒にとって負担が大きいものの代表に、読書感想文がある。青少年読書感想文全国コンクールのホームページによると、読書感想文を書く目的は、「書くことによって考えを深め」ることであり、「読書感想文を書くことを通して思考の世界へ導かれ」、「わからなかったことを解決できる」という。

さらに、同ホームページは、読書感想文の書き方について、「本を読んで自分がどこに感動したのか、なぜ感動したのかを考え」、「自分の生き方や経験と本の世界とを照らし合わせ」、「感じたこと、思ったこと、連想したことなどを」、「順番を入れ替えたり内容を補ったりして、同自分の心の動きにぴったりするか、それがうまく人に伝わるかを考えましょう」としている。すなわち、読書感想文とは、本の内容を踏まえ、自身が連想したことを、他者に伝えるように書いた文章であることが分かる。長期休暇という長いようで短い期間において、自身が読むべき書籍を探し、それを読み、その感想を文章にするというステップは大変時間が掛かり、多大な労力を要する。とりわけ、書籍を読むというステップが最も時間が掛かるものと考えられる。本リサーチでは、読書感想文において、テキストマイニングという新た

な視点を提供し、書籍を読むというステップの省力化を図ることで、読書感想文の負担を軽くし、児童・生徒諸君が、各々の変態性を高めることに注力できるようにすることを最終目標とする。

2. 目的及び方法

本リサーチでは、実際に文章を読むというステップを、テキストマイニングを用いて文章を分析するステップに置き換えた上で、読書感想文を作成することが可能か否かを検討することを目的とする。

テキストマイニングとは、小木(2015)によると、「常々明確な定義がないことは様々な場所で言われているが、基本的にはテキスト(文章)をマイニング(情報発掘)することであり、定性的な特徴をもつテキストを定量的に分析することである」とされる。すなわち、文章に対して様々な処理を行うことで、その文章が持つ情報を数量的に要約するツールであるといえる。また、テキストマイニングのツールはインターネット上に複数あるが、本リサーチでは、ユーザーローカル(2007-2020)によって無償で提供されているテキストマイニングツールによって分析を行う。これはユーザーローカルのツールが、ダウンロード等の必要性がなく、ブラウザ上で使えるため、多くの児童・生徒諸君が使えるためである。

そして、実際に読書感想文の作成にあたっては、2名の研究協力者に対して、「書籍の題名」、「著者」、「テキストマイニングによって得られた結果」を提示し、それらのみで、読書感想文の作成を求める。読書感想文の作成様式等については、青少年読書感想文全国コンクールの高等学校の部に準拠し、本文は2000字以内とし、句読点はそれぞれ1文字に数え、また、改行のための空白箇所も字数として数え、題名・氏名は字数に数えない。応募区分は自由読書（自由に選んだ図書）とする。具体的な書籍については、使用するツールの制約から、青空文庫にある10万字以下の文章の中から、研究協力者が読んだことのないものを選択する。また、青空文庫にある有名な文章を選択した場合、提出先である教諭がその本を読んだことがあり、作成した文章と実際の物語との乖離が露見してしまう危険性が高い。他方、教諭が全く知らない著者を選択した場合、知らないが故にインターネット等であらすじ等を検索する可能性があり、その結果として感想文と物語の乖離が露見するおそれもある。そのため、青空文庫にある文章のうち、一定程度の知名度のある著者であり、さほど内容が一般に知られていないものの中から選ぶこととする。文章量と知名度等の関係から、本リサーチでは、有島武郎による『一房の葡萄』を題材とする。

3. 結果

3.1 テキストマイニング

実験協力者に提示したテキストマイニングの結果をFigure 1からFigure 5に示す。また、実験協力者から提出された読書感想文は付録に示す。Figure 1は、ワードクラウドである。これは、ユーザーローカルによると、「スコアが高い単語を複数選出し、その値に応じた大きさを図示しています。単語の色は品詞の種類で異なっており、青色が名詞、赤色が動詞、緑色が形容詞、灰色が感動詞を表しているものであるという。

なお、ここでいう「スコア」について、ワードクラウドは、『一般的な文書でよく出る単語は、重要ではないため、重み付けを軽くする』、いっぽう『一般的な文書ではあまり出現しないけれど、調査対象の文書だけによく出現する単語は重視する』仕組みを取り入れ（中略）出現回数だけでなく、重要度を

加味した値が『スコア』で」とあるとしている。本リサーチの、他のテキストマイニングの結果に記載されるスコアもすべて同じ定義とする。

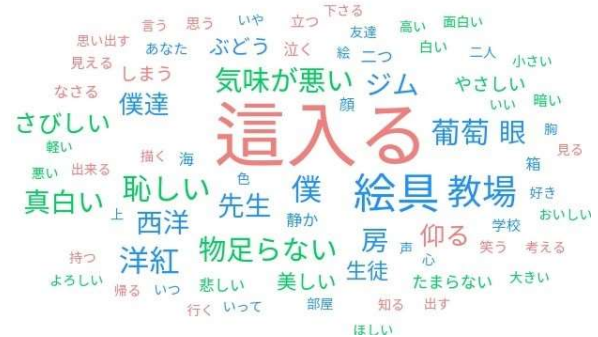


Figure 1 ワードクラウド

Figure 1 から、「這入る」という言葉が最も大きいことが高いことが分かる。そのほかに、「絵具」や、「気味が悪い」、「僕」、「教場」といった単語も這入るに次いで大きいことが見て取れる。

Figure 2 は、単語出現頻度である。これは、ユーザーローカルによると、「文章中出现する単語の頻度を表にし」たものであり、「通常はその単語の出現回数が多いほどスコアが高くなるが、「どの文書にもよく現れる単語についてはスコアが低めになる」ものであるという。

Figure 2 の名詞の表では、出現頻度は、「僕」、「先生」、「ジム」といった単語が多かった一方、スコアは、出現頻度がそれらよりも低い「絵具」が最も高く、それに「教場」や「僕」、「葡萄」といった単語が続いた。

動詞の表では、「しまう」、「思う」、「這入る」が出現頻度の上位であった一方、スコアでは「這入る」のみが約 135 と顕著に高く、それ以外は 1 未満であった。

形容詞の表では、「いい」、「楽しい」、「悲しい」が出現頻度の上位であった一方、出現頻度はさほど多くない「恥かしい」、「気味が悪い」、「さびしい」のスコアが高かった。

■ 名詞	スコア	出現頻度	■ 動詞	スコア	出現頻度
僕	38.37	97	しまう	0.83	23
先生	7.99	34	思う	0.23	20
ジム	12.70	22	這入る	134.75	12
絵具	143.37	20	行く	0.11	12
顔	1.00	16	見る	0.06	9
葡萄	36.06	10	泣く	0.34	8
教場	76.46	9	出来る	0.17	8
学校	0.62	9	知る	0.16	8
眼	7.09	8	見える	0.19	6
生徒	4.31	7	笑う	0.18	6
箱	1.24	7	言う	0.03	6
部屋	0.42	7	なさる	0.44	5
上	0.26	7	描く	0.13	5
房	17.84	6	出す	0.10	5
静か	1.54	6	持つ	0.07	5

■ 形容詞	スコア	出現頻度	■ 感動詞	スコア	出現頻度
いい	0.13	13	—	—	—
美しい	1.20	7	—	—	—
悲しい	0.40	6	—	—	—
ほしい	0.11	6	—	—	—
やさしい	0.98	3	—	—	—
たまらない	0.58	3	—	—	—
高い	0.05	3	—	—	—
恥しい	10.45	2	—	—	—
気味が悪い	5.92	2	—	—	—
さびしい	2.12	2	—	—	—
暗い	0.21	2	—	—	—
白い	0.20	2	—	—	—
よろしい	0.08	2	—	—	—
軽い	0.08	2	—	—	—
小さい	0.08	2	—	—	—

さらに単語の出現頻度を表示する

Figure 2 単語出現頻度

Figure 3 は、共起キーワードである。これは、ユーザーローカルによれば、「文章中出现する単語の出現パターンが似たものを線で結んだ図」であり、「出現数が多い語ほど大きく、また共起の程度が強いほど太い線で描画されるものであるという。

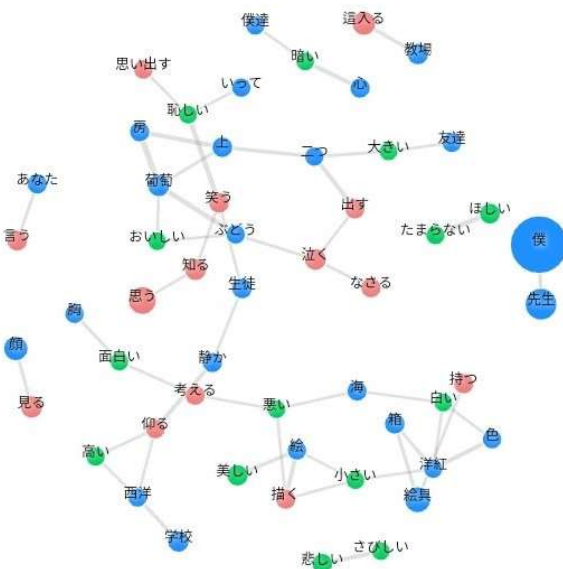


Figure 3 共起キーワード

Figure 3 では、大きく描かれているのは、「僕」であり、それと線で結ばれているのは「先生」であった。また、線が目立って太いのは「房」と「葡萄」の間であった。僕と先生がつながりをもって文章中で描かれていること、房と葡萄が似た出現パターンであることが分かる。

Figure 4 は、2次元マップである。これは、ユーザーローカルによれば、「文章中での出現傾向が似た単語ほど近く、似ていない単語ほど遠く配置され」、「距離が近い単語はグループにまとめ、色分け」された図である。



Figure 4 2次元マップ

Figure 4 から、単語は「泣く」や「いい」といった緑の群、「這入る」や「眼」といった青の群、「先生」や「思う」といった紫の群、「ジム」や「知る」といったオレンジの群、「見える」や「笑う」といった赤の群の4グループに分けられることが分かった。

Figure 5 は、階層的クラスタリングである。これは、ユーザーローカルによると、「文章中での出現傾向が似た単語をまとまりとしてとらえられるよう樹形図で表したものである」、「グループは色分けして表示」されている。

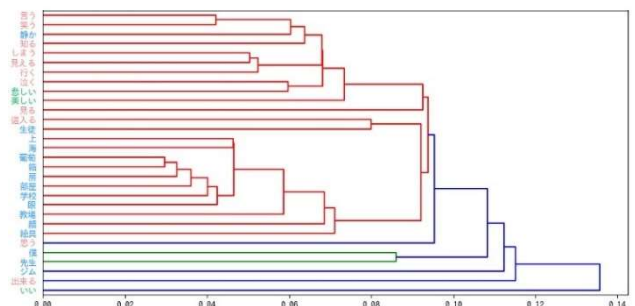


Figure 5 階層的クラスタリング

Figure 5 から、出現する単語は、「言う」や「笑う」といった赤の群、「僕」と「先生」のみからなる緑の群、「いい」や「出来る」といった青の群の3グループに分けられることが分かった。

3.2 読書感想文

著書名や作者、テキストマイニングの結果を用いて、読書感想文を作成することは可能であった。文章内容も、支離滅裂になることはなく、一定程度まとまりのある文章となっていた。そして、その内容について共通点及び相違点が認められた。

まず、共通点は、「僕」を主人公と解釈したこと、その主人公が何かしら後ろめたいことをしたことの二つであった。

次に、相違点は、作者について言及したか、「先生」の存在を文中に導入したかの二つであった。

4. 考察

4.1 テキストマイニング

Figure 1 と Figure 2 の結果を踏まえると、「這入る」や「絵具」、「教場」、「葡萄」、「気味が悪い」といった単語のスコアや重要度が高く、注目すべき要素であると考えられる。また、Figure 3 から、「這入る」は「教場」、絵具は「持つ」や「箱」、「葡萄」は「房」とつながりが深いことが分かる。それに加えて、「僕」と「先生」のつながりが深いことも分かる。

ここまでのことから、主人公である「僕」が「教場」に「這入る」ことが想定される。また、そうした行動の背景に「葡萄」や「絵具」と関わりがあると考えられる。「先生」や「気味が悪い」という言葉について、教場に入る僕は学生であることが想定されるため、学生と深い存在である先生が頻出であることは想像に難くない一方、文章中で気味が悪いとしているのは何であるのかは不明確である。教場に這入ることや主人公が学生であろうことは、Figure 4 において、学校や生徒という単語が取り上げられている点、教場と這入るの距離が近く、同じグループに分類されている点からも見て取れる。

さらに、Figure 5 を見ると、「這入る」の単語の下位には、感情に関する言葉が多く含まれており、「這入る」ことに関して、その主体は様々な感情の中でその行為に及んだと考えられる。また、「教場」の下

位には、「葡萄」や「箱」といった言葉が含まれており、教場に葡萄があることが想定できる。そして、それらが上位で結び付いていることから、様々な感情を抱えた中で、葡萄のある教場に這入るものと考えられる。

4.2 共通点と相違点の検討

感想文を書くことができたという結果から、テキストマイニングを用いることで読書感想文を作詞することができることが示された。そのため、提出された感想文間の共通点・相違点を取り上げることで、テキストマイニングの結果から、読書感想文を作成するためのエッセンスを抽出する。

協力者から提供された読書感想文間の共通点と相違点は、テキストマイニングの結果及びその解釈、書籍情報の使用方法によるものと考えられる。

テキストマイニングの結果及びその解釈について、出現頻度及びスコアの高い単語が、感想文の主軸になることが推察されるため、共通点が生まれたと考えられる。すなわち、「僕」や「這入る」、「葡萄」、「絵具」、「教場」という単語を軸にして、物語の内容を想像し、それに対する感想文を記載したのである。まず、一般的に、小説において「僕」等の一人称で表記されることが多い登場人物が主人公であると推察されるため、協力者らは「僕」が主人公であると設定したと考えられる。また、出現頻度が高かったりスコアが高かったりした単語のうち、「這入る」は特徴的である。通常であれば「入る」を使用するところを、あえて「這って」・「入る」と書く「這入る」という単語を使っている。その上、スコアが高い形容詞は「恥かしい」や「気味が悪い」である。これらのことから、協力者らは、「後ろめたいことをしている」事態を想定して、感想文を書いたと考えられる。さらに、様々な単語間の関係性は、共起キーワードや階層的クラスタリングから推察した結果と思われる。

他方、相違点が生じた理由は、テキストマイニングの結果の取り扱い方であると思われる。わかりやすい違いは、出現頻度の高い「先生」という単語を含めるか否かである。

「先生」という単語は、出現頻度が高い一方でスコアが低い。すなわち、よく出てくるが、文章中での重要度は低いのである。出現頻度が多いことを重視

するならば、感想文中に取り上げる方が自然であるし、そうでなければあえて取り上げる必要性はない。この点は、協力者の個性といえる部分であり、どちらが良いというものではない。ただ、文章量を増やすという点では、「先生」という要素を含めた方が有効であろう。

4.3 まとめ

以上から、テキストマイニングの結果を用いて読書感想文を作成する場合、まずは、ワードクラウドや単語出現頻度から、物語の軸となるキーワードを決め、それらを共起キーワードや階層的クラスタリングを参考にしつつ、つなげることで物語の骨子を想像しているようである。そして、「這入る」のような特徴的な動詞や、スコアの高い形容詞等から、事実や出来事の連なりである物語を主人公や著者がどのように解釈しているのかを想像し、物語に肉付けしている。そうしてテキストマイニングの結果から物語を独自に再構成し、それに対する読書感想文を書けば、実際に書籍を読まずとも、統計結果のみを用いて、読書感想文を作成することが可能となる。

また、同じ統計結果であっても、その統計結果をどのように解釈するか、どのように重要度を考慮するかによって感想文の内容が変わることも分かり、仮に学生・生徒諸君が同じ書籍を題材にテキストマイニングを用いて読書感想文を書いても内容が重複することはないことも確かめられた。

そのほか、著者の情報を記載するという方略も有効であることが示唆された。テキストマイニングの結果のみを用いたが文章量が足りない際には、付随する情報として著者を感想文中に入れ込むことで、文章量を増やすことが可能であるといえる。

なお、感想文間の違いについて、著者の個性に由来する相違点も確かに存在するが、それは書籍を読んだ後の感想が個々人で異なるように、提示された統計結果等からそれぞれが連想した内容が異なったことにより生じたと考えられ、その違いを考察することは本リサーチの目的ではないため、取り上げない。

5. 最後に

本レポートは、テキストマイニングを用いて読書感想文を作成することができるかを探索的に検討したものであり、抽出された作成手法は唯一の方法ではない。本レポートは、あくまでもスタート地点であることに留意する必要がある。より多数の者が作成を行えば、より効率的な手法が見いだされるであろう。

また、本レポートは、書籍を読む意味を否定するものではない。読書感想文は本来、本を読み、その感想を記載するものである。変態性を高めるために少しでも時間が欲しいというわけではなければ、本レポートで示した手法は用いず、しっかりと本を読み、その世界に触れるべきであることは明言しておく。

引用文献

- 小木 しのぶ (2015). テキストマイニングの技術と動向 計算機統計学, 28, 31-40.
全国学校図書館協議会・毎日新聞社 (2001-2020). 感想文 Q&A (<https://www.dokusyokansoubun.jp/qa.html>)
ユーザーローカル (2007-2020). AI テキストマイニング by ユーザーローカル (<https://textmining.userlocal.jp>)

謝辞

感想文を作成いただいた協力者に心より感謝を申し上げます。