

## 計量テキスト分析を用いた小説の視点研究 伊坂幸太郎を例として

山下 夏実                      鈴木 崇史  
東洋大学 社会学部          東洋大学 社会学部

作家によって物語の描き方はそれぞれであり、その個性は小説のもつ魅力のひとつである。中でも小説の語り手が誰であるかは、読み手に与える印象に影響する重要な要素である。そこで本研究では、小説内で語り手が変化する作品の視点ごとのテキストの特徴を分析する。分析対象は、伊坂幸太郎の「オーデュボンの祈り」、「グラスホッパー」の2作品とする。視点別の節数と述べ語数を計量したのち、すべての形態素の相対頻度とランダムフォレスト機械学習法を用いてテキスト分類を行い、また、分類に有効な特徴量を抽出する。その結果、「オーデュボンの祈り」では、視点を主人公の伊藤、他の登場人物をまとめたその他の2つとしたところ、視点ごとの明確な特徴はみられなかった一方、「グラスホッパー」は視点を鈴木、鯨、蟬の3つとしたところ、視点ごとの明確な特徴がみられ、特徴語にそれぞれの視点ごとのキャラクターを反映した語調や、特有なシーンが含まれることが示された。

## 1. はじめに

近年自然言語処理技術の進展により、計算機を用いたテキスト研究は進み、脅迫状の文体分析 [6] や企業の注力技術比較 [5]、またコールセンターの対話音声分析 [7] など、犯罪捜査やマーケティング戦略といった様々な分野に応用されている。

小説は作家によって書き方に個性が出るテキストであり、その個性が魅力のひとつである。その作家の個性に焦点を当てた研究には、作家の生きた時代や環境といった要因から作品が書かれた背景を探る方法がある。一方、計量的文体分析を用いて、文章の著者を推定し、著者の文章の特徴を明らかにする方法もある [3; 6]。この場合、著者の特徴情報として、文や単語の長さの分布や単語の使用頻度、品詞の使用率などがよく用いられている [3; 6]。

小説を書く際、物語をどの視点から描くかは、読み手に与える印象に影響する重要な要素である。物語を語る視点については、以前より研究があり、形式的に大きく一人称視点と三人称視点に分けられる。石丸 [2] は、「一人称視点」を「語り手が主役で登場し一人称で語る文章」とし、「三人称視点」を「三人称の主役に身をよせ主役の周辺に生起する事件を語る文章」と「作中に登場せず全知全能の『神の視点』から語られる文章」としている。つまり、語り手が

主観的に「私」の見たこと感じたことを語る文章が一人称視点であり、語り手が主人公または作中の一人物、あるいは、登場人物以外の、人物や事象を全て知り尽くした者として、観察的に語る文章が三人称視点である。このような視点の設定の仕方は、作家の個性や作品の描かれ方によって様々なタイプがあり、作品内で視点が変わる特徴をもつ作家も存在するため、重要な研究テーマとなり得る。

このような背景のもと、本研究では、小説内で語り手が変化する作品の視点ごとのテキストの特徴を分析する。本研究の分析対象とする伊坂幸太郎は、主人公以外の登場人物の視点からも物語を描く傾向があり、とりわけ、工藤ら [4] が、「並行形式小説」とよぶ、意図的に複数の登場人物の視点を交差させる形式をしばしば用いる。並行形式小説を描く作家として村上春樹も挙げられるが、村上の作品は奇数章と偶数章で主人公が入れ替わるという形式が主である。一方で伊坂の描く並行形式小説の場合、3人以上の視点を用いることもあり、入れ替わり方も作品によって違いがある。登場人物の種類や入れ替わり方に多様性がみられる伊坂を分析対象として選ぶことで、本研究では、より多様な、視点ごとのテキストの特徴を分析する。

## 2. データと分析手法

### 2.1 データ

分析の対象は文庫化されている伊坂の作品で主人公以外の視点からも書かれており、執筆時期も初期、中期と比較的離れている「オーデュボンの祈り」, 「グラスホッパー」の 2 作品とした。テキストデータは、文庫本をスキャンした PDF の画像データを OCR でテキストファイルに変換した後、手作業でルビの削除と OCR の際のエラーの修正を行ったものを作成し、使用した。テキストデータは節ごとに作成し、テキストの整形をおこなった後、MeCab<sup>1</sup>による形態素解析を適用した。

### 2.2 分析手法

まず、作品のそれぞれの節について、手作業で視点を割り当てた。どの登場人物の視点かは、心理描写（「…だと…は思った」など）から判断した。節の切り替わりは節の頭にあるマークとし、マークがない場合は行間があっても同じ節とした。

「オーデュボンの祈り」では、視点の分け方を主人公である伊藤と、伊藤以外の人物（その他）とした。これは、伊藤以外の節を人物ごとに抽出すると、それぞれの節数が極端に少なくなるためである。節数は伊藤が 56、その他が 22 であった。また、伊藤の節が終わっても次の節で伊藤の視点が続いていることがあることから、節の変わり目が登場人物の切り替えではない。

「グラスホッパー」は、主人公の鈴木、自殺屋の鯨、ナイフ使いの蟬の 3 つの視点から構成されており、この 3 つの視点を用いた。節数は、鈴木が 17、鯨が 15、蟬が 10 であった。また、視点の切り替わりは、鈴木、鯨、蟬の順で固定されており、物語中で人物が死亡すると抜けていくという規則性がみられた。

作成されたテキストについて、まず、視点別に節数と述べ語数を観察した。つづいて、全形

態素の相対頻度を特徴量としてランダムフォレスト機械学習法[1]を用い、テキスト分類を行った。実験の評価には、エラー率を用いた。この分類結果によって、分類性能が高く特別な特徴をもつ視点と、分類性能が低く特徴の少ない視点を明らかにした。また、分類に有効な特徴量上位 20 を取り出し、視点ごとの特徴を分析した。

## 3. 結果と考察

### 3.1 基礎的な観察結果

表 1. 「オーデュボンの祈り」基礎データ

	節数	延べ語数			
		総数	平均	標準偏差	変動係数
伊藤	56	118119.00	2109.27	1697.57	0.80
その他	22	23316.00	1013.74	1285.46	1.21

表 1 は、「オーデュボンの祈り」の視点ごとの節数と、延べ語数をまとめたものである。節数は伊藤が 7 割以上を占めており、標準偏差は伊藤のほうが高い一方、変動係数は伊藤の方が低い。伊藤が、節数も、一節の平均語数も高いため、標準偏差の値が高くなっているが、平均値で統制した変動係数をみると、多数の登場人物の視点を含むその他の方が高い値をとることが示された。

表 2. 「グラスホッパー」基礎データ

	節数	延べ語数			
		総数	平均	標準偏差	変動係数
鈴木	17	51256.00	3015.06	1873.70	0.62
鯨	15	33472.00	2231.47	1209.66	0.54
蟬	10	27168.00	2716.80	898.45	0.33

表 2 は、「グラスホッパー」の視点ごとの節数、延べ語数をまとめたものである。物語の進行上、若干鈴木の前節が多く、標準偏差、変動係数も、鈴木が最も高いことが示された。

<sup>1</sup> [mecab.sourceforge.net](http://mecab.sourceforge.net)

### 3.2 ランダムフォレストを用いた分類実験

表 3. 「オーデュボンの祈り」分類結果

	伊藤	その他	エラー率
伊藤	56	0	0.00
その他	18	4	0.82

表 3 は、「オーデュボンの祈り」の分類結果である。伊藤の節はすべて伊藤と分類されているのに対し、その他の節が正しく分類されているのは 2 割に満たない結果となった。その他の多くが伊藤として分類されているのは、伊藤の節数が圧倒的に多いことが影響していると推察される。

表 4. 「グラスホッパー」分類結果

	鈴木	鯨	蟬	エラー率
鈴木	16	1	0	0.06
鯨	0	15	0	0.00
蟬	0	4	6	0.40

表 4 は、「グラスホッパー」の分類結果である。主人公である鈴木よりも、鯨の節にエラーが少ない（一つもない）ことが特徴としてみられた。また、蟬の節が鯨の節と分類されることはあっても、鯨の節が蟬の節として分類されることはなかった。蟬より鯨の節数が多いことが影響しているが、最も節数の多い鈴木は影響を及ぼしていないことが注目される。

### 3.3 分類に有効な特徴量

表 5. 「オーデュボンの祈り」特徴語

語	MeanDecreaseAccuracy
1 轟音	0.03116
2 だ	0.00993
3 ん	0.00859
4 皮膚	0.00832
5 一部分	0.00654
6 いや	0.00509
7 でっちあげ	0.00500
8 止まっ	0.00394
9 物理	0.00340
10 よ	0.00332

11 カッ	0.00318
12 声	0.00290
13 千	0.00248
14 」	0.00247
15 ニ	0.00226
16 長い	0.00215
17 内容	0.00211
18 私	0.00199
19 かも	0.00192
20 苦しむ	0.00189

表 5 は、「オーデュボンの祈り」のテキスト分類の際に有効な特徴量上位 20 語である。これらの重要語からは伊藤とその他の判別に明白なものがなく、その他の節が伊藤の節として分類された結果と対応するものである。これは、伊藤の節数が多いことに加え、視点の分類の際に伊藤以外をその他にまとめているため、共通した特徴が少ないことによると推察される。

表 6. 「グラスホッパー」特徴語

語	MeanDecreaseAccuracy
1 つう	0.02137
2 ねえ	0.01130
3 喋ら	0.01000
4 煎じ	0.00920
5 え	0.00829
6 噴射	0.00692
7 身の毛	0.00677
8 脈やか	0.00653
9 とろ	0.00652
10 ワイパー	0.00647
11 提	0.00623
12 症状	0.00615
13 問い	0.00463
14 よ	0.00452
15 は	0.00435
16 瞳孔	0.00435
17 比	0.00415
18 結婚	0.00408
19 です	0.00402
20 やり	0.00348

表 6 は、「グラスホッパー」のテキスト分類の際に有効な特徴量上位 20 語である。蟬の喋り方にてでくる「つう」、「ねえ」や、鈴木

「です」という語調が反映されており、キャラクターの語尾や語調の特徴が、分類に影響していることが明らかになった。また、「症状」は鯨の節に共通して出てくる語であり、鯨が時折感じる眩暈のことを指している。この症状も分類に強く影響していることが明らかになった。

#### 4. おわりに

本研究では、伊坂幸太郎の作品である「オーデュボンの祈り」と「グラスホッパー」を対象として、小説の視点ごとの特徴を明らかにするため、テキストの計量的分析をおこなった。

「オーデュボンの祈り」では、視点を主人公の伊藤、他の登場人物をまとめたその他の2つとしたところ、視点ごとの明確な特徴は示されなかった。「グラスホッパー」では視点を鈴木、鯨、蟬の3つとしたところ、特徴語にそれぞれの視点ごとのキャラクターを反映した語調や、特有なシーンが含まれており、視点ごとの特徴が明確に示された。

本研究では伊坂幸太郎の「オーデュボンの祈り」、「グラスホッパー」を分析対象とし、異なる結果を得た。「グラスホッパー」では視点ごとの特徴が明らかになった一方、「オーデュボンの祈り」では明確な特徴がみられなかったことから、今後、節数を統制するなどして、特徴を明らかにすることを目指したい。

現在も伊坂は作品を生み出し続けているため、今後も対象作品を増やすことが可能である。より多くのデータを用い、検証することも今後の課題としたい。

#### 謝辞

本研究は、科研費若手(B)「計算文体論による多種メディアテキスト解析(研究代表者:鈴木崇史, 研究課題番号:23700288)」より、一部支援を受けています。ここに記して謝意を表します。

#### 文献

- [1] Breiman L. Random forests, *Machine Learning*, Vol.45, pp.5-23, 2001.
- [2] 石丸晶子. 文章における視点, 日本語学, 4(12), 明治書院, 22-31, 1985.
- [3] 金明哲. 日本語における単語の長さの分布と文章の著者, 社会情報, 5(2), 13-21, 1996.
- [4] 工藤彰・村井源・往住彰文. 共通語の布置と変化に基づく並行形式小説の物語構造, 情報知識学会誌, 22(3), 187-202, 2012.
- [5] 宮田道生・福川忠昭・高橋正子. 特許戦略のための特許情報の計量化: テキストマイニングを活用した企業の注力技術比較の試み, 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, 78-79, 2006.
- [6] 村上征勝. シェークスピアは誰ですか?: 計量文献学の世界, 文春新書, 東京, 2004.
- [7] 小野寺佐知子・落谷亮. テキスト情報と韻律情報を利用したコールセンター対話の分析, 言語・音声理解と対話処理研究会, 40, 45-50, 2004.