

## ジャンルの観点から西鶴作品の文章の数量分析

上阪彩香 (同志社大学大学院文化情報学研究科・日本学術振興会特別研究員 DC)

本研究では、江戸時代前期の俳諧師・浮世草子作者である井原西鶴 (1642?~1693) の作品の文章の特徴について、数量的な観点から検討を行った。西鶴作品は我が国の文学史における重要性から、多くの国文学者によって思想や記述内容の検討、成立に関する歴史的考証が続けられてきたが、西鶴作品の中でも浮世草子には、著者や成立年代等について今なお疑問が出され、解明すべき問題が残されている。西鶴は浮世草子のほかにも作品を残しており、役者評判記『難波の貌は伊勢の白粉』、浄瑠璃『暦』『凱陣八島』、地誌『一目玉鉾』の4作品は西鶴著である。これらの作品は浮世草子と同じ散文で書かれていることから、浮世草子に提出されている著者問題を検討する際に有効な情報となる可能性がある。本研究では、浮世草子、役者評判記、浄瑠璃、地誌とジャンルの異なる作品においても、著者が西鶴とされている作品であれば、同一の文章の特徴を示すのかを主成分分析を用いて検討した。

## Quantitative Comparative Analysis for Saikaku's Works

Ayaka Uesaka (Graduate school of Culture and Information Science, Doshisha University ·  
Research Fellow of the Japan Society for the Promotion of Science)

In this article, we focus on Saikaku Ihara's writing style. Saikaku Ihara (c. 1642~93) is one of the most famous writers of the Genroku period (1688~1704) in Japan. Saikaku's novels are known for their significance of developing Japanese novels today. However, it remained unclear which novels were really written by Saikaku. Meanwhile, he also wrote other genres of work: *Yakusha Hyoubanki* (actor's critique), *Jyoururi* (Japanese ballad drama) and *Chishi* (topography). We examined his novels and other genres of work using Principal Component Analysis. These results revealed that genres had more influence compared with a writer's style.

### 1. まえがき

江戸時代前期の俳諧師・浮世草子作者である井原西鶴 (1642?~1693) の作品は、日本文化に多大な影響を与えた古典籍と位置づけられている。西鶴はそれまでにない新しい内容・文体・方法を備えた浮世草子という文学分野 (谷脇, 1991) を確立し、当時の文学界、さらには明治以降、幸田露伴や芥川龍之介など写実主義・自然主義を論じた多くの作家に影響を与えた。このことから元禄の文豪との評価を受け、近松門左衛門 (1653~1724)、松尾芭蕉 (1644~1694) と並んで元禄文学を代表する作家とされる (谷脇, 1991)。

このように西鶴作品は我が国の文学史における重要性から、多くの国文学者によって思想や記述内容の検討、成立に関する歴史的考証が続けられてきた。しかしながら、西鶴作品の中でも浮世草子には、著者や成立年代等について今なお疑問が出され、解明すべき問題が残されている。

西鶴は浮世草子のほかにも作品を残しており、役者評判記『難波の貌は伊勢の白粉』、浄瑠璃『暦』『凱陣八島』、地誌『一目玉鉾』の4作品は西鶴著であり、これらの作品は浮世草子と同じ散文で書かれていることから、浮世草子に提出されている著者問題を検討する際に有効な情報となる可能性がある。

文体の定義は様々あり、一意に定義することは難しいが、「個人文体」と「ジャンル文体」に大別される。「個人文体」とは、個人の個性や資質等によってもたらせる文章の特徴であり、「ジャンル文体」とは一般的に非個人的な、ジャンルの違いによってもたらされる文章の違いである (安本, 1982)。

小林・小木曾 (2013) は、『源氏物語』『紫式部日記』『更級日記』の助詞と助動詞の使用傾向を、クラスター分析等を用いて検討し、書き手による文体差 (個人文体) よりもジャンルによる文体差 (ジャンル文体) の方が大きいと指摘している。そのため、本研究の対象である西鶴の作品

においても、ジャンルが異なった場合、個人的な文章の特徴よりもジャンルの影響の方が大きく表れることが考えられる。そこで本研究では、浮世草子、役者評判記、浄瑠璃、地誌といったジャンルの異なる作品においても、著者が西鶴とされている作品であれば、同一の文章の特徴を示すのかを検討する。

## 2. データベースについて

本研究には、西鶴のデータベース（以後、新編西鶴全集データベースと記す）を用いた。新編西鶴全集データベースは、西鶴作と考えられる120作品（浮世草子24作品、俳書80作品、発句1作品、浄瑠璃2作品、役者評判記1作品、地誌1作品、歌謡1作品、書簡7通、その他3作品）の文章を電子化及び形態素解析したデータベースで、本文と自立語は『新編西鶴全集』巻1～巻5の本文篇・自立語索引篇として出版されている。

文章の数量分析には、形態素解析された文章データが必要となるが、本研究の研究対象である近世文学に対応する精度の高い形態素解析辞書は存在しない。そのため形態素解析については、『新編西鶴全集』を編集した新編西鶴全集編集委員会における単語の認定基準に従うこととした。

表1 新編西鶴全集データベースの一例  
（『好色一代男』の冒頭部分）

作品名	巻	本文	品詞	活用形	現代がな終止
一男	巻一	桜	名詞		さくら
一男	巻一	も	助詞		も
一男	巻一	ちる	動詞	連体	ちる
一男	巻一	に	助詞		に
一男	巻一	歎き	動詞	連用	なげく
一男	巻一	月	名詞		つき
一男	巻一	は	助詞		は
一男	巻一	かぎり	名詞		かぎり
一男	巻一	あり	動詞	連用	あり
一男	巻一	て	助詞		て
一男	巻一	入佐山	名詞		いるさやま
一男	巻一	爰	名詞		ここ
一男	巻一	に	助詞		に
一男	巻一	但馬の国	名詞		たじまのくに
一男	巻一	かね	名詞		かね
一男	巻一	ほる	動詞	連体	ほる
一男	巻一	里	名詞		さと
一男	巻一	の	助詞		の
一男	巻一	辺	名詞		ほとり
一男	巻一	に	助詞		に
一男	巻一	浮世	名詞		うきよ
一男	巻一	の	助詞		の
一男	巻一	事	名詞		こと
一男	巻一	を	助詞		を
一男	巻一	外	名詞		ほか
一男	巻一	に	助詞		に

表1には、形態素解析済みのデータの一部を示した。すべての文章が単語に分割され、それぞれの単語には数量分析を行う際に必要となる品詞や活用形等の情報が付けられている。

## 3. 分析について

### 3.1. 分析対象

本研究では、新編西鶴全集データベースに収録されている120作品の中から表1に示した28作品（浮世草子24作品、役者評判記1作品、地誌1作品、浄瑠璃2作品）を用いた。表に示した延べ語数は、見出しの語数と本文の語数を合わせたものである。

表2.a 西鶴の浮世草子(作品名と延べ語数)

好色一代男 (36,781 語)	諸艶大鑑 (45,753 語)	椀久一世の物語 (7,702 語)
好色五人女 (20,184 語)	好色一代女 (26,581 語)	西鶴諸国はなし (16,444 語)
本朝二十不孝 (18,419 語)	男色大鑑 (50,452 語)	武道伝来記 (49,019 語)
好色盛衰記 (20,866 語)	懐硯 (22,839 語)	日本永代蔵 (29,547 語)
色里三所世帯 (11,895 語)	武家義理物語 (21,456 語)	嵐は無常物語 (8,727 語)
新可笑記 (25,157 語)	本朝桜陰比事 (26,466 語)	世間胸算用 (21,260 語)
浮世栄花一代男 (22,576 語)	西鶴置土産 (17,204 語)	西鶴織留 (29,617 語)
西鶴俗つれづれ (13,966 語)	万の文反古 (16,940 語)	西鶴名残の友 (12,380 語)

表2.b 西鶴の他のジャンルの作品(作品名と延べ語数)

難波の貌は伊勢 の白粉 (4,401 語)	曆 (5,941 語)	凱陣八島 (8,843 語)
一目玉鉾 (17,110 語)		

以下に示すのは、『新編西鶴全集』に収録されている分析対象の作品の冒頭部分の文章である。

#### 浮世草子：『好色一代男』

桜もちるに歎き、月はかぎりありて。入佐山、爰に但馬の国、かねほる里の辺に、浮世の事を外になして、色道ふたつに、寝ても覚ても、夢介と、かえ名よばれて、名古や三左、加賀の八などと、七ツ紋のひしにくみして、身は酒にひたし、一条通り、夜更て戻り橋、或時は若衆出立、姿をかえて墨染の長袖、又はたて髪かつら、化物が通るとは、誠に是ぞかし。

#### 役者評判記：『難波の貌は伊勢の白粉』

都に猿らかんと云印籠、有はあれども買ものなし。

世中広しといへ共、又狭しと云沙汰、難波江の誰やらが音に聞て、後藤極め三十両に求たり。是も都に闇の夜の瓢箪、十五両て堀江の浪に取寄た。せばいやうで広ひは大坂、西南東北のうかれ舟、銭の湊の道頓堀があるぞ。

#### 浄瑠璃：『暦』

けんこんひらけばんぶつしやうず、けいしやうゆたかなるときつづくに。抑仁皇四十一代は、持統天皇としゆくし世の御まつりごとただしく、くはんくはこどくをあはれみ、ひりうざんじつをすくはせ給へば、しよてんのめぐみひさかたの太上天皇とはじめてあがめ奉る。

#### 浄瑠璃：『凱陣八島』

扱も其後、序げげんも思ひより出、けどうもはかりことよりおこる。これみなじんしんのなせるところ。ここに人王七十七代、ごしらかはのみんとあがめ奉りゆゆしきせいしゆおはします。

#### 地誌：『一目玉鉾』

東の沖浪しつかに、毎朝くれなゐの影ゆたかに、久かた日の本爰也。

○日の出の浜

我国は天照神の末なれば日の本としも云にそ有ける天空空替らす照す日の本の国静なる御代そかしこき

このようなジャンルの異なる作品においても、著者が同一であれば、類似した文章の特徴を示すのかを文章の数量的特徴に着目し、検討する。

### 3. 2. 分析項目

本研究において取り上げた分析項目を大別すると、①品詞の構成比、②単語の出現率、③品詞別単語の出現率、④bigramの出現率の4つである。分析には、作品の長さの影響を取り除くため、構成比と出現率を用いた。

#### 3. 2. 1. 品詞の構成比

品詞の構成比とは、それぞれの品詞に属する単語が作品の延べ語数に対して占める割合である。分析に用いたデータベースでは、すべての単語が名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞・接続詞・形容動詞・感動詞・連語・接頭語・接尾語・補助動詞の14品詞に分類されているが、本研究の分析には、出現頻度上位10品詞(名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞・接続詞・形容動詞・感動詞)の構成比を用いること

とした。

#### 3. 2. 2. 単語の出現率

単語の出現率とは、それぞれの単語が作品の延べ語数に対して占める割合である。単語の出現率においても、データベースに収録されている14品詞のなかから出現頻度上位10品詞(名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞・接続詞・形容動詞・感動詞)の出現率を用いた。

#### 3. 2. 3. 品詞別単語の出現率

品詞別単語の出現率とは、それぞれの品詞に属する各単語がその品詞の延べ語数に対して占める割合を作品ごとに集計したものである。本研究では、上位7品詞(名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞)の品詞別単語の出現率を分析に用いた。これは出現頻度が8番目以下の品詞(接続詞・形容動詞・感動詞・連語・接頭語・接尾語・補助動詞)の出現頻度が低いためである。

#### 3. 2. 4. bigramの出現率

n-gramとは、n個の記号の度数を集計する方法で、nとは集計を行うために切り取った隣接している記号列の長さを表しており、nが1のときunigramと呼ぶ。①品詞の構成比、②単語の出現率、③品詞別単語の出現率はunigramに関する情報である。nが2のときbigramと呼ぶ。n-gramを用いた著者推定に関する初期の研究にFucks(1952)がある。また金(2002)は、一般人が書いた短い文章(日本語の日記文)の著者推定に、助詞のn-gramを用いる方法を提案し、その有効性を示した。

本研究では、n-gramのなかでもbigramを用いて検討することとした。bigramの度数を計算する場合、名詞をはじめとした異なり語数の多い品詞では、多くの組み合わせが発生し、出現頻度0が多く含まれたデータとなるため、異なり語数の少ない品詞・助詞・助動詞のbigramの出現率を用いた検討を行うこととした。

### 3. 3. 分析手法

本研究では、相関係数行列に基づく主成分分析を用いて検討を行った。分析結果を散布図で示す際には、横軸を第1主成分、縦軸を第2主成分とした。また図の横軸には、第1主成分の寄与率、縦軸に第2主成分の寄与率、上部に第2主成分ま

での累積寄与率を示した。

一般に主成分分析の結果は、分析に用いる変数によって異なるので、分析に用いる単語数等によって結果が大きく異なることがないかを確認するため、①品詞の構成比を用いた場合には、出現頻度の低い品詞を順に分析から外し、10品詞の構成比を用いた分析から4品詞の構成比を用いた分析までの計7種類の分析を行った。また②単語の出現率、③品詞別単語の出現率、④bigramの出現率を用いた検討の場合には、出現頻度が9回以下の変数を others という項目にまとめて用いた分析、others を除いた出現頻度が10回以上の変数を用いた分析、出現頻度上位200語までの変数を用いた分析、150語までの変数を用いた分析、100語までの変数を用いた分析、50語までの変数を用いた分析、25語までの変数を用いた分析の計7種類のうち、出現頻度が上記の条件を満たした項目で分析を行った。

#### 4. 分析結果

##### 4. 1. 品詞の構成比を用いた分析

表3には、西鶴の浮世草子・浄瑠璃・地誌・役者評判記の主要10品詞（名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞・接続詞・形容動詞・感動詞）の構成比をジャンル別にまとめた。浮世草子24作品では、名詞の構成比が0.342と最も高く、続いて助詞(0.334)、動詞(0.165)、助動詞(0.074)が0.05以上の構成比をもつ変数であり、上位4位までの品詞の構成比の合計は0.915である。役者評判記1作品では、名詞の構成比が0.361と最も高く、続いて助詞(0.334)、動詞(0.157)、助動詞(0.079)が0.05以上の構成比をもつ変数であり、上位4位までの品詞の構成比の合計は0.931である。浄瑠璃2作品は、助詞の構成比が0.320と最も高く、続いて名詞(0.316)、動詞(0.176)、助動詞(0.073)が0.05以上の構成比をもつ変数であり、上位4位までの品詞の構成比の合計は0.885である。地誌1作品は、名詞の構成比が0.471と最も高く、続いて助詞(0.282)、動詞(0.131)、助動詞(0.071)が0.05以上の構成比をもつ変数で、上位4位までの品詞の構成比の合計は0.955である。

浮世草子24作品、役者評判記1作品、地誌1作品では名詞の構成比の値が最も大きい、浄瑠璃2作品では助詞の構成比の値が最も大きくな

っている。また地誌1作品の名詞の構成比は0.471と、浮世草子24作品(0.342)、役者評判記1作品(0.361)、浄瑠璃2作品(0.316)と比較して大きな値になっており、それに対して助詞(0.282)、動詞(0.131)、助動詞(0.071)の構成比は小さな値になっている。

表3 西鶴の浮世草子・浄瑠璃・地誌・役者評判記の品詞の構成比

	浮世草子 (24 作品)			役者評判記 (1 作品)		
	異なり 語数	延べ語 数	構成 比	異なり 語数	延べ語 数	構成 比
名詞	25718	195714	0.342	1016	1587	0.361
助詞	164	191297	0.334	44	1468	0.334
動詞	4228	94686	0.165	340	693	0.157
助動詞	133	42276	0.074	27	346	0.079
形容詞	567	16003	0.028	38	83	0.019
副詞	911	13437	0.023	67	98	0.022
連体詞	72	10607	0.019	11	56	0.013
接続詞	76	2321	0.004	11	15	0.003
形容動詞	156	1466	0.003	9	13	0.003
感動詞	101	778	0.001	15	23	0.005
その他	63	3646	0.006	3	19	0.004
合計	32,189	572,231	1	1,581	4,401	1

	浄瑠璃 (2 作品)			地誌 (1 作品)		
	異なり 語数	延べ語 数	構成 比	異なり 語数	延べ語 数	構成 比
名詞	2122	4674	0.316	3968	8057	0.471
助詞	56	4729	0.320	45	4819	0.282
動詞	734	2600	0.176	457	2234	0.131
助動詞	32	1077	0.073	24	1218	0.071
形容詞	133	385	0.026	77	262	0.015
副詞	241	592	0.040	59	120	0.007
連体詞	14	179	0.012	8	267	0.016
接続詞	26	126	0.009	4	26	0.002
形容動詞	16	28	0.002	17	49	0.003
感動詞	52	197	0.013	1	1	0
その他	5	197	0.013	4	57	0.003
合計	3,431	14,784	1	4,664	17,110	1

次に、主成分分析（相関係数行列）を用いて分析し、浮世草子24作品と他のジャンル4作品の文章の特徴を検討する。図2は、浮世草子24作品、役者評判記1作品、浄瑠璃2作品、地誌1作品の計28作品の主要10品詞（名詞・助詞・動詞・助動詞・形容詞・副詞・連体詞・接続詞・形容動詞・感動詞）の構成比を主成分分析で分析した結果(第2主成分までの累積寄与率54.668%)の散布図である。図の中に描かれている黒色の楕円は、浮世草子24作品の第1主成分と第2主成分の主成分得点から推定した西鶴浮世草子の95%が入ると考えられる95%信頼楕円である。『嵐は無常物語』と『万の文反古』は95%信頼楕円から外れて位置しているが、他の浮世草子22作品はある程度まとまって位置しており、浮世草子24作品の文章の特徴は類似した傾向にあ

るといことがわかる。また役者評判記『難波の貌は伊勢の白粉』は、西鶴浮世草子 24 作品の 95% 信頼楕円に含まれた。一方で、他の 3 作品『暦』『凱陣八島』『一目玉鉾』は、西鶴浮世草子 24 作品の 95% 信頼楕円には含まれなかった。

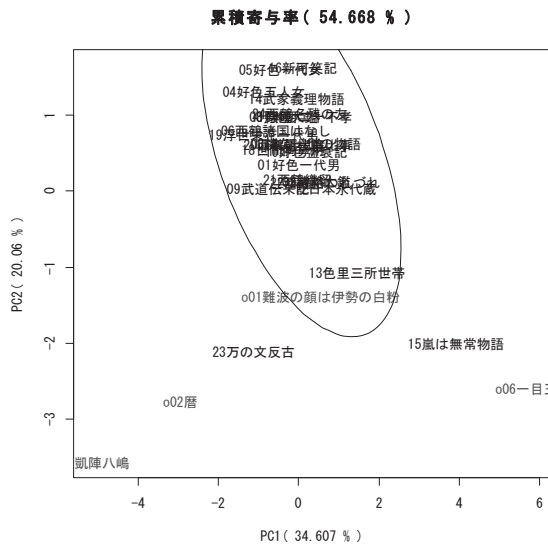


図 1 主要 10 品詞の構成比の主成分分析の結果

表 4 主要 10 品詞の構成比の主成分負荷量

	PC1	PC2
名詞	0.882	-0.364
助詞	-0.178	0.489
動詞	-0.569	-0.138
助動詞	-0.327	0.518
形容詞	-0.488	0.591
副詞	-0.951	-0.153
連体詞	0.097	0.642
接続詞	-0.773	-0.258
形容動詞	-0.100	0.351
感動詞	-0.680	-0.610
固有値	3.460	2.005
寄与率	0.346	0.201
累積寄与率	0.346	0.547

表 4 の主成分負荷量をみると、第 1 主成分においては名詞が正の方向に、副詞・接続詞・感動詞・動詞・形容詞が負の方向に影響を与え、第 2 主成分では連体詞・形容詞・助動詞・助詞が正の方向に、感動詞が負の方向に影響を与えている。

そのほかに、分析項目を変えて同様の主成分分析を品詞の構成比に行った。第 1 主成分と第 2 主成分を用いて描いた西鶴浮世草子の 95% 信頼楕円のなかに、他のジャンル 4 作品が含まれたのかを各々調べた結果を表 5 に示した。表には、浮世草子の 95% 信頼楕円に含まれた他のジャンルの作品名を記した。浮世草子の 95% 信頼楕円に含まれる作品がなかった場合は「なし」とした。

この結果をみると、品詞の構成比を用いた 7 回の分析では、地誌『一目玉鉾』ほどの分析でも西鶴浮世草子と異なった特徴を見せた。一方、役者評判記『難波の貌は伊勢の白粉』は 7 回、浄瑠璃『暦』は 6 回、『凱陣八島』は 4 回が西鶴浮世草子 24 作品の 95% 信頼楕円に含まれた。

表 5 品詞の構成比を用いた分析結果

分析に用いた品詞数	95% 信頼楕円に含まれた浮世草子以外の作品
10 品詞	『難波の貌は伊勢の白粉』(注: 分析結果は図 1 に表示)
9 品詞	『難波の貌は伊勢の白粉』, 『暦』
8 品詞	『難波の貌は伊勢の白粉』, 『暦』
7 品詞	『難波の貌は伊勢の白粉』, 『暦』
6 品詞	『難波の貌は伊勢の白粉』, 『暦』
5 品詞	『難波の貌は伊勢の白粉』, 『暦』, 『凱陣八島』
4 品詞	『難波の貌は伊勢の白粉』, 『暦』, 『凱陣八島』

#### 4. 2. 単語の出現率を用いた分析

単語の出現率を用いた分析結果を示す。図 2 は、浮世草子 24 作品、役者評判記 1 作品、浄瑠璃 2 作品、地誌 1 作品の計 28 作品の全出現単語の中から、出現頻度上位 25 語の出現率を主成分分析で分析した結果(第 2 主成分までの累積寄与率 49.071%) の散布図である。

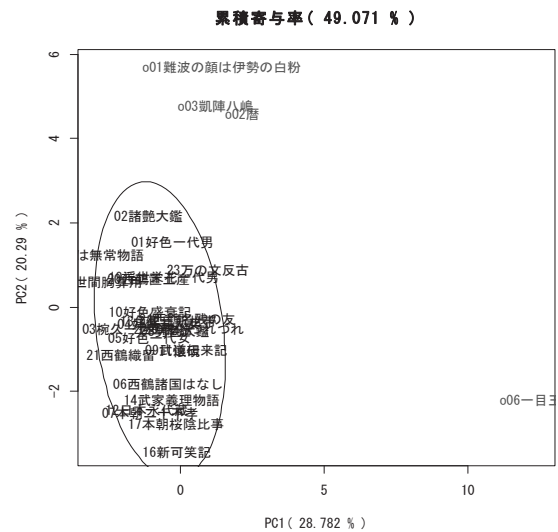


図 2 出現頻度上位 25 語の出現率の主成分分析の結果

図に描いた楕円は浮世草子 24 作品の 95% 信頼楕円である。この図をみると、浮世草子 24 作品はある程度まとまって位置しており、浮世草子 24 作品の文章の特徴は類似した傾向にあるということがわかる。一方で、他のジャンル 4 作品は、西鶴浮世草子 24 作品の 95% 信頼楕円に含まれず、出現頻度上位 25 語の単語の出現率からは、浮世草子 24 作品の文章と役者評判記 1 作品、浄

瑠璃 2 作品, 地誌 1 作品の文章は, 異なった特徴を持つことがわかる.

表 6 の主成分負荷量をみると, 第 1 主成分においては「あり・より・なり・の・これ」が正の方向に, 「ず・こと・なし・もの・て・は・す・も・ひと・なる・を・その」が負の方向に影響を与え, 第 2 主成分では「が・は・と」が正の方向に, 「に・この・その・き・を・これ・なる・ひと・より・けり・なし」が負の方向に影響を与えている.

表 6 出現頻度上位 25 語の出現率の主成分負荷量

	PC1	PC2		PC1	PC2
の	0.520	-0.050	この	0.272	-0.750
に	-0.210	-0.869	なし	-0.692	-0.403
を	-0.494	-0.589	けり	-0.199	-0.405
て	-0.629	-0.158	が	-0.305	0.689
は	-0.628	0.500	いう	-0.120	0.123
と	-0.342	0.434	これ	0.404	-0.583
も	-0.578	0.225	ひと	-0.551	-0.447
ず	-0.827	-0.017	その	-0.439	-0.648
なり	0.578	-0.379	より	0.656	-0.419
ば	-0.398	0.253	なる	-0.541	-0.475
す	-0.615	-0.242	もの	-0.673	0.027
き	0.151	-0.590	固有値	7.193	5.072
こと	-0.719	-0.290	寄与率	0.288	0.203
あり	0.876	-0.122	累積寄与率	0.288	0.491

表 7 は, 分析項目を変えて同様の主成分分析を単語の出現率を用いて行った結果である. この結果をみると, 単語の出現率を用いた 7 回の分析で, 地誌『一目玉鉾』, 役者評判記『難波の貌は伊勢の白粉』, 浄瑠璃『暦』『凱陣八島』の文章は, 西鶴浮世草子とは異なる特徴を示した.

表 7 単語の出現率を用いた分析結果

分析に用いた単語数	95%信頼楕円に含まれた浮世草子以外の作品
全て*	なし
出現頻度 10 語以上	なし
上位 200 語	なし
上位 150 語	なし
上位 100 語	なし
上位 50 語	なし
上位 25 語	なし (注: 分析結果は図 2 に表示)

#### 4. 3. 品詞別単語の出現率を用いた分析

品詞別単語の出現率を用いた結果を示す. 図 3 は, 出現頻度上位 50 語の助詞の出現率を主成分分析で分析した結果である. この図の浮世草子 24 作品が位置する範囲を 95%信頼楕円で囲うと, 浮世草子 24 作品はまとまりをもって付置されるが, 他のジャンルの 4 作品は浮世草子 24 作品とは離れて位置している. 助詞は付属語で, 物語の内容の影響を受けづらいため, 著者の検討を行う際に用いられることが多いが, 出現頻度上位 50 語の助詞の出現率を用いて行った分析でも, 浮世

草子 24 作品の文章と役者評判記 1 作品, 浄瑠璃 2 作品, 地誌 1 作品の文章は, 異なった特徴を持つ可能性が高い.

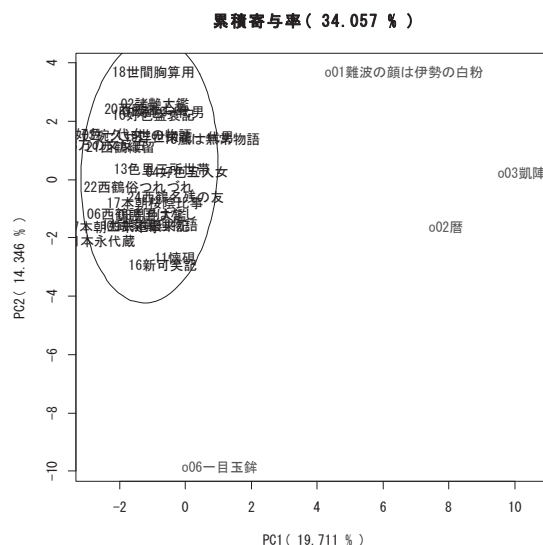


図 3 出現頻度上位 50 語の助詞の出現率の主成分分析の結果

表 8 の主成分負荷量をみると, 第 1 主成分においては, 「や・よ・な・こそ・だに・かな・か・そ・ぞ・やら・がな・が・と・とて・もって」が正の方向に, 「に・まで・て・を・して・ばかり」が負の方向に影響を与え, 第 2 主成分では「から・が・も・は・か・で・など・かし・やら・がな・て・い」が正の方向に, 「より・の・とて・に・なむ・そ・のみ・だに」が負の方向に影響を与えている.

そのほかに, 分析項目を変えて同様の主成分分析を品詞別単語の出現率に行った結果を表 9 にまとめた. 表の斜線は, 出現頻度が満たさなかった項目である. 例えば, 助動詞では, 分析に用いた 28 作品 (浮世草子 24 作品, 役者評判記 1 作品, 浄瑠璃 2 作品, 地誌 1 作品) の出現頻度が 9 回以下の助動詞を others にまとめたときの異なり語数が 36 であるため, 上位 200 語から 50 語までの単語を用いた分析を行うことができず, 表の上位 200 語から 50 語までの欄を斜線とした. この結果をみると, 品詞別単語の出現率を用いた分析では, 35 回の分析のうち地誌『一目玉鉾』は 3 回, 役者評判記『難波の貌は伊勢の白粉』は 2 回, 浄瑠璃『暦』は 5 回, 『凱陣八島』は 7 回が西鶴浮世草子 24 作品の 95%信頼楕円に含まれた.

表 8 出現頻度上位 50 語の助詞の出現率の主成分負荷量

	PC1	PC2		PC1	PC2
の	0.067	-0.582	さえ	-0.227	0.152
に	-0.732	-0.537	ど	0.156	0.219
を	-0.468	-0.191	よ	0.823	0.068
て	-0.501	0.434	ずつ	-0.398	0.344
は	0.225	0.568	のみ	-0.029	-0.445
と	0.489	0.214	な	0.795	0.229
も	-0.215	0.627	かな	0.734	-0.174
ば	0.378	0.189	やら	0.572	0.445
が	0.493	0.648	でも	-0.318	0.357
より	-0.205	-0.691	おいて	0.074	-0.027
にて	0.167	0.103	い	0.245	0.406
へ	-0.100	0.335	もがな	-0.323	-0.276
ぞ	0.611	0.124	なむ	0.165	-0.492
まで	-0.537	0.113	もって	0.423	-0.068
か	0.719	0.529	なり	-0.147	-0.077
や	0.869	-0.187	いで	-0.205	0.302
とて	0.474	-0.557	がな	0.519	0.441
して	-0.458	-0.173	だに	0.750	-0.441
ながら	-0.056	0.348	ほど	-0.219	-0.184
から	-0.229	0.722	しも	0.312	-0.349
ども	0.330	0.263	き	-0.103	-0.290
ばかり	-0.405	0.299	そ	0.703	-0.478
で	-0.291	0.522	まま	-0.130	-0.244
こそ	0.759	0.017	固有値	9.853	7.172
かし	-0.372	0.467	寄与率	0.197	0.144
など	-0.079	0.467	累積寄与率	0.197	0.341
とも	0.298	-0.094			

表 9 品詞別単語の出現率を用いた分析結果

分析に用いた単語数	95%信頼楕円に含まれた浮世草子以外の作品			
	名詞	助詞	動詞	
全て*	なし	なし	なし	
出現頻度 10 語以上	なし	なし	なし	
上位 200 語	なし		なし	
上位 150 語	なし		なし	
上位 100 語	『凱陣八島』		なし	
上位 50 語	『凱陣八島』		『難波の貌は伊勢の白粉』, 『凱陣八島』	
上位 25 語	なし	なし (注: 分析結果は図 3 に表示)	『難波の貌は伊勢の白粉』, 『凱陣八島』	

分析に用いた単語数	助動詞	形容詞	副詞	連体詞
	全て*	『曆』	なし	『一目玉鉾』
出現頻度 10 語以上	『曆』	なし	『一目玉鉾』	なし
上位 200 語			『一目玉鉾』	
上位 150 語		なし	『曆』, 『凱陣八島』	
上位 100 語		なし	なし	
上位 50 語		『曆』	なし	
上位 25 語	『曆』, 『凱陣八島』	『凱陣八島』	なし	

#### 4. 4. bigram の出現率を用いた分析

図 4 は、出現頻度上位 25 位の助詞の bigram の出現率を主成分分析で分析した結果 (第 2 主成分までの累積寄与率は 45.848%) の散布図である。助詞の bigram を用いた分析の結果からも、他のジャンルの 4 作品は西鶴浮世草子の 95%信頼楕円のなかに含まれなかった。このことから、

助詞の bigram を用いた分析でも、浮世草子 24 作品の文章と役者評判記 1 作品、浄瑠璃 2 作品、地誌 1 作品の文章は、異なった特徴を持つということが明らかになった。

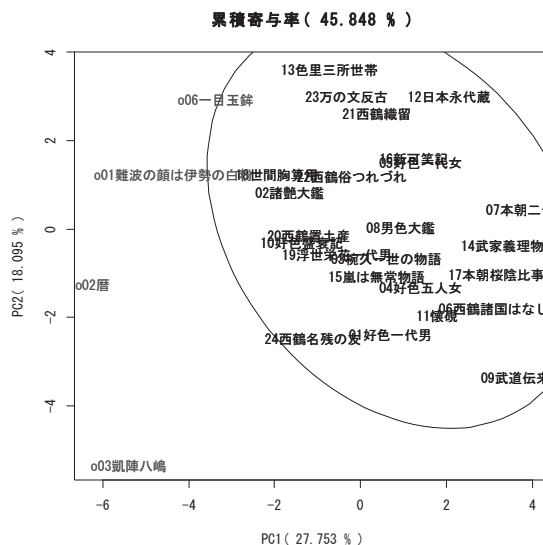


図 4 出現頻度上位 25 位の助詞の bigram の出現率の主成分分析の結果

表 10 の主成分負荷量をみると、第 1 主成分においては、「て-を・に-を・を-に・て-に・に-て・を-て・を-の・の-を・に-に・に-も」が正の方向に、「の-は・は-の・の-の-」が負の方向に影響を与え、第 2 主成分では、「は-に・に-は・も-の・の-も・に-の・は-の・の-に・に-も・の-を・の-の・に-に」が正の方向に、「と-に・に-と・と-の」が負の方向に影響を与えている。

表 10 出現頻度上位 25 位の助詞の bigram の出現率の主成分負荷量

	PC1	PC2		PC1	PC2
の_に	0.072	0.457	は_に	0.248	0.667
の_の	-0.465	0.417	て_を	0.912	-0.189
に_の	0.092	0.507	に_も	0.406	0.445
に_て	0.665	0.009	の_と	-0.125	-0.216
の_を	0.629	0.435	も_の	-0.326	0.560
て_の	0.326	0.134	と_に	0.320	-0.775
を_て	0.632	-0.188	も_に	0.299	0.346
を_に	0.869	0.242	と_の	-0.033	-0.474
は_の	-0.613	0.504	の_も	-0.228	0.532
に_を	0.899	0.099	に_と	0.359	-0.558
て_に	0.805	-0.048	固有値	6.938	4.524
に_は	0.286	0.595	寄与率	0.278	0.181
に_に	0.493	0.407	累積寄与率	0.278	0.459
を_の	0.631	0.362			
の_は	-0.707	0.295			

表 11 は、分析項目を変えて同様の主成分分析を bigram の出現率を用いて行った結果である。

この結果をみると、bigram の出現率を用いて行った 18 回の分析のうち、地誌『一目玉鉾』ではすべての項目において浮世草子の 95%信頼楕円に含まれなかった。一方で、役者評判記『難波の貌は伊勢の白粉』では 6 回、浄瑠璃『暦』では 5 回、『凱陣八島』では 4 回の分析で浮世草子 24 作品の 95%信頼楕円に含まれた。

表 11 bigram の出現率を用いた分析結果

分析に用いた bigram の数	95%信頼楕円に含まれた浮世草子以外の作品		
	品詞の bigram	助詞の bigram	助動詞 bigram
全て*	なし	なし	なし
出現頻度 10 語 以上	なし	なし	なし
上位 200 語		なし	なし
上位 150 語		なし	『難波の貌は伊勢の白粉』、『暦』、『凱陣八島』
上位 100 語		なし	『難波の貌は伊勢の白粉』、『暦』、『凱陣八島』
上位 50 語	『難波の貌は伊勢の白粉』	なし	『難波の貌は伊勢の白粉』、『暦』、『凱陣八島』
上位 25 語	『難波の貌は伊勢の白粉』、『暦』	なし (注: 分析結果は図 4 に表示)	『難波の貌は伊勢の白粉』、『暦』、『凱陣八島』

## 5. 考察

本研究では、分析項目を変えて計 67 回の主成分分析を行った。このうち、41 回の分析で浮世草子の 95%信頼楕円のなかに他のジャンルの 4 作品は含まれなかった。また 4 作品全てが浮世草子の 95%信頼楕円に含まれる分析項目もなかった。一方、他のジャンルの作品が含まれた 26 回の分析のうち、『難波の貌は伊勢の白粉』は 15 回、『暦』は 16 回、『凱陣八島』は 13 回、『一目玉鉾』は 3 回の分析で浮世草子 24 作品の 95%信頼楕円に含まれた。西鶴浮世草子の 95%信頼楕円には、西鶴浮世草子と似た文章の特徴を持つ作品が、100 回の分析のうち約 95 回入ることが予想される。したがって、67 回の分析では統計的には 63 回から 64 回は含まれることになるが、最も多く含まれた『暦』でも 16 回であった。以上のことから、本研究での分析の結果からは、西鶴の浮世草子、役者評判記、浄瑠璃、地誌の文章は異なった特徴を持つということが考えられる。

## 6. あとがき

本研究では、西鶴の浮世草子、役者評判記、浄瑠璃、地誌における文章の特徴を、主成分分析を

用いて検討した。その結果、著者の特徴よりもジャンルの特徴の方が強く表れるという結果を得た。このことから、西鶴浮世草子の著者について検討する際には、西鶴の作品であったとしても浮世草子以外の作品を分析に用いる必要性をこの分析の結果からは、見出すことができなかった。

今後の課題としては、各々のジャンルにおける特徴的な表現の検討を行うこと、西鶴の浮世草子における著者問題を検討する際には、西鶴門弟の北条団水(1663~1711)をはじめとした他の作家の浮世草子との比較分析を行うことが挙げられる。

## 参考文献

- 1) 谷脇理史・吉行淳之介、『新潮古典文学アルバム 17 井原西鶴』, 新潮社(1991).
- 2) 安本美典、『講座日本語学 8 文体史 II』, 明治書院(1982).
- 3) 小林雄一郎・小木曾智信、『国立国語研究所論集』6, pp.29-43, 『中古和文における個人文体とジャンル文体—多変量解析による歴史的資料の文体研究—』, 国立国語研究所(2013).
- 4) 浅野晃, 他編:『新編西鶴全集』第 1 巻, 勉誠出版 (2000).
- 5) 浅野晃, 他編:『新編西鶴全集』第 2 巻, 勉誠出版 (2002).
- 6) 浅野晃, 他編:『新編西鶴全集』第 3 巻, 勉誠出版 (2003).
- 7) 浅野晃, 他編:『新編西鶴全集』第 4 巻, 勉誠出版 (2004).
- 8) 浅野晃, 他編:『新編西鶴全集』第 5 巻, 勉誠出版 (2007).
- 9) 波多野完治, 『文章心理学』, 三省堂(1935).
- 10) 安本美典, 『文学 語学』, 宇治十帖の作者 - 文章心理学による作者推定 -, 朝倉書店(1957).
- 11) Fucks, W, Biometrika39(1/2), pp.122-129, On mathematical analysis of style, Oxford University Press(1952).
- 12) 金明哲, 『計量国語学』23(5), pp.225-240, 「助詞の n-gram モデルに基づいた書き手の識別」, 計量国語学会(2000).