

西鶴浮世草子における物語の内容別にみた 文章の特徴について

— 数量的観点からみた好色物、武家物、町人物、雑話物の文章の違い —

ウエサカ アヤカ
上阪 彩香

一. はじめに

江戸時代前期の浮世草子作者である井原西鶴（1642～1694）の浮世草子进行研究対象として、文章の特徴を数量的な側面から捉え、西鶴著とされる浮世草子24作品内での物語の内容（好色物、武家物、町人物、雑話物）における特徴把握を試みる。本研究では、一般に物語の内容に依存すると考えられる名詞、動詞、形容詞の出現率を中心に、主成分分析およびクラスター分析を用いて検討する。西鶴の浮世草子は一般的に好色物、武家物、町人物、雑話物に分類される。分類に関しては諸説あるが、本研究では宗政（1969）が挙げた以下の分類を用いる。

好色物：『好色一代男』『諸艶大鑑』『好色五人女』『好色一代女』『男色大鑑』『好色盛衰記』『色里三所世帯』『浮世栄華一代男』『西鶴置土産』

武家物：『武道伝来記』『武家義理物語』『新可笑記』

町人物：『日本永代蔵』『世間胸算用』『西鶴織留』

雑話物：『西鶴諸国はなし』『懷硯』

その他：『本朝二十不孝』『本朝桜陰比事』『西鶴俗つれづれ』『西鶴名残の友』

二. 分析対象作品

西鶴の文章データは、西鶴著とされる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であり、本文と自立語は『新編西鶴全集』の本文篇、自立語索引篇として出版されている。本研究では、西鶴の文章データの中から、浮世草子作品を分析に用いた。表1に、作品名、出版年、底本、延べ語数を示す。延べ語数は見出し及び本文を合わせたものである。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。新編西鶴全集編集委員会における単語の認定基準は、校訂を担当した研究者によって作品ごとに検討され、異なる部分は新編西鶴全集編集委員会の会議により、決定された。

表1 西鶴の浮世草子

作品名	出版年	底本	延べ語数
好色一代男	1682	東洋大学本（荒砥屋孫兵衛可心版）	36,781語
諸艶大鑑	1684	京都大学附属図書館本	45,753語
椀久一世の物語	1685	『賞奇楼叢書』第二期第五集・第六集	7,702語
好色五人女	1686	国会図書館本	20,184語
好色一代女	1686	国会図書館本（第六は真山青果記念館本）	26,581語
西鶴諸国はなし	1685	霞亭文庫本	16,444語
本朝二十不孝	1686	国会図書館本	18,419語
男色大鑑	1687	早稲田大学本	50,452語
武道伝来記	1687	早稲田大学本	49,019語
好色盛衰記	1688	大阪府立中之島図書館本	20,866語
懷硯	1687	早稲田大学本	22,839語
日本永代蔵	1688	東洋大学本	29,547語
色里三所世帯	1688	フランス国会図書館本	11,895語
武家義理物語	1688	東洋大学本	21,456語
嵐は無常物語	1688	上巻は天理図書館本、下巻は神原文庫本	8,727語

新可笑記	1688	東洋大学本	25,157語
本朝桜陰比事	1689	花月文庫本	26,466語
世間胸算用	1692	国会図書館本	21,260語
浮世栄花一代男	1693	京都大学附属図書館本	22,576語
西鶴置土産	1693	大阪府立中之島図書館本	17,204語
西鶴織留	1694	天理図書館本	29,617語
西鶴俗つれづれ	1695	東洋大学本	13,966語
万の文反古	1696	東洋大学本	16,940語
西鶴名残の友	1699	ケンブリッジ大学本	12,380語

表2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。データ化されている項目は、本文、品詞、活用等の17項目である。また作品に出現する品詞は、名詞、助詞、動詞、助動詞、形容詞、副詞、連体詞、接続詞、形容動詞、感動詞等の14品詞である。

表2 西鶴のデータの一例（『好色一代男』の冒頭部分）

作品名	巻	本文	品詞	活用形	ルビ	現代がな終止	現代がな活用
一男	巻一	桜	名詞		サクラ	さくら	さくら
一男	巻一	も	助詞		△	も	も
一男	巻一	ちる	動詞	連体	△	ちる	ちる
一男	巻一	に	助詞		△	に	に
一男	巻一	歎き	動詞	連用	ナゲー	なげく	なげき
一男	巻一	月	名詞		△	つき	つき
一男	巻一	は	助詞		△	は	は
一男	巻一	かぎり	名詞		△	かぎり	かぎり
一男	巻一	あり	動詞	連用	△	あり	あり
一男	巻一	て	助詞		△	て	て

三. 分析について

品詞の構成比を用いて検討した結果について考察する。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総単語数に対して占める割合である。分

析に用いたデータでは、すべての単語が14品詞に分類されているが、本研究の分析では出現頻度上位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。

表 3 10品詞における構成比

作品名	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
武道伝来記	0.336	0.318	0.177	0.08	0.026	0.027	0.019	0.005	0.002	0.002
武家義理物語	0.336	0.327	0.168	0.079	0.026	0.025	0.024	0.003	0.003	0.001
新可笑記	0.348	0.315	0.163	0.083	0.027	0.023	0.026	0.004	0.003	0.001
日本永代蔵	0.367	0.338	0.155	0.07	0.028	0.018	0.016	0.005	0.003	0.001
世間胸算用	0.351	0.344	0.148	0.072	0.031	0.024	0.016	0.006	0.003	0.002
西鶴織留	0.349	0.346	0.155	0.07	0.027	0.023	0.017	0.006	0.003	0.001
好色一代男	0.33	0.346	0.173	0.068	0.028	0.025	0.017	0.002	0.002	0.002
諸艶大鑑	0.341	0.35	0.164	0.069	0.027	0.021	0.016	0.003	0.002	0.001
好色五人女	0.33	0.335	0.166	0.078	0.034	0.027	0.018	0.003	0.002	0.002
好色一代女	0.334	0.334	0.162	0.08	0.034	0.025	0.017	0.002	0.003	0.001
男色大鑑	0.336	0.34	0.164	0.074	0.029	0.022	0.017	0.005	0.003	0.001
好色盛衰記	0.351	0.338	0.151	0.076	0.03	0.022	0.019	0.005	0.002	0.002
色里三所世帯	0.37	0.337	0.158	0.061	0.026	0.02	0.017	0.003	0.002	0.002
浮世栄花一代男	0.332	0.339	0.16	0.08	0.028	0.025	0.017	0.006	0.003	0.002
西鶴置土産	0.342	0.346	0.154	0.074	0.029	0.024	0.018	0.005	0.003	0.002
西鶴諸国はなし	0.325	0.343	0.173	0.072	0.028	0.023	0.02	0.005	0.002	0.001
懷硯	0.335	0.329	0.173	0.074	0.028	0.026	0.021	0.004	0.004	0.002

表 3 に各々の作品のなかで用いられている品詞の構成比を示した。例えば、『武道伝来記』では、名詞が約33.6%、助詞が約31.8%の割合で用いられているということが読み取れる。しかしながら、これらの数値のみでは、データの全体像を把握することは難しいため、主成分分析とよばれる分析手法を用いて検討する。主成分分析とは、変数が持つ情報を可能な限り失わずに、主成分と呼ばれる少数個の合成変数に縮約し、個体の分類を試みる多変量解析の手法である。相関係数行列と分散共分散行列を用いる方法があるが、本研究では相関係数行列に基づく主成分分析を用いる。元の変数の情報を多く含む主成分から順

に第1主成分、第2主成分とよび、各々の主成分が元の全変数の情報の何%を含んでいるかは寄与率で表わされる。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。

表3の品詞の構成比を用いて、主成分分析を行った散布図を図1に示す。主要10品詞の情報をを用いて分析を行った結果、各々の作品が内容別にまとまりをもって位置している。

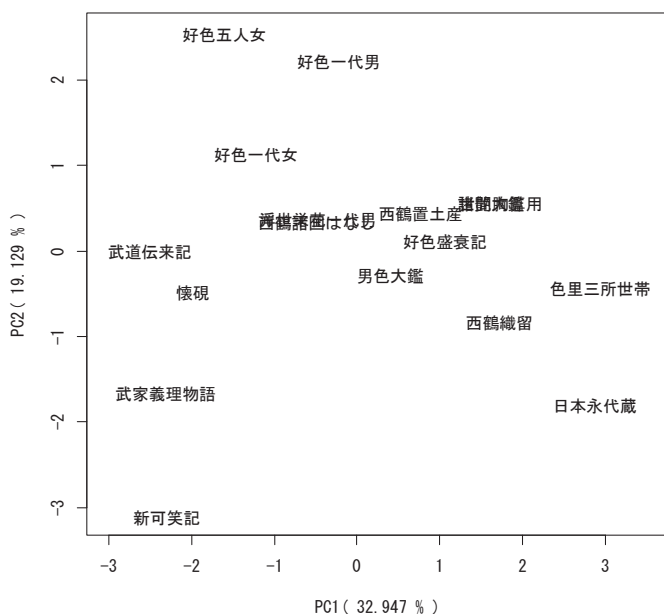


図1 主成分分析の結果

つぎに、物語の内容と文章の数量的特徴の間に関連があるか否かを、品詞のなかでも一般に物語の内容に依存すると考えられる名詞、動詞、形容詞の出現率を、主成分分析およびクラスター分析を用いて検討する。

クラスター分析は、個体の分類に関して、外的基準（分類基準）がない場合に用いる分類法である。本研究では名詞、動詞、形容詞の各々の情報が似ている作品のグループ化（クラスター化）を試みる。クラスター化する際に、個体

(あるいはクラスター)の非類似度として、ユークリッド平方距離、マハラノビス距離、ミンコフスキー距離等が提案されている。またクラスターの連結法として、単連結法、完全連結法、群平均法、重心法、メディアン法、ウォード法等のアルゴリズムが提案されている。本研究では、非類似度距離として、Kullback-Leibler 距離、クラスターの連結法としてウォード法を用いた分析を行った。ウォード法はグループ内の分散が小さく、かつグループ間の分散が大きい組み合わせでグループ分けする方法である。個体がどのようにグループ化されるかはデンドログラム(樹形図)で示される。

表4は、出現頻度上位25語までの名詞をトピック別にまとめたものである。この表をみると、出現頻度上位25位のなかで武家物にのみ現れた名詞は「ふたり、かたき、しさい」、町人物にのみ現れた名詞は「いえ、かね、とし、こ、むかし、みな」、好色物にのみ現れた名詞は「じょう、め」、雑話物にのみ現れた名詞は「あと、ひ、なか、ひとり」であった。

表4 トピック別の出現頻度上位25位の名詞

	武家物	町人物	好色物	雑話物		武家物	町人物	好色物	雑話物
1	こと	こと	こと	こと	13	ほど	ほか	ころ	み
2	これ	ひと	これ	もの	14	うち	いえ	うち	それ
3	ひと	これ	ひと	これ	15	のち	ころ	きょう	ころ
4	もの	もの	もの	ひと	16	おとこ	かね	それ	われ
5	み	よ	み	よ	17	いのち	とし	じょう う	こ
6	とき	ほど	よ	とき	18	ほか	きょう	われ	のち
7	ところ	とき	ほど	ところ	19	こ	なに	ほか	あと
8	よ	み	おとこ	いま	20	なに	のち	かた	いのち
9	ころ	ところ	おんな	うち	21	かた	き	こ	なに
10	いま	うち	とき	おんな	22	ふたり	こ	なに	かた
11	おんな	いま	いま	ほど	23	かたき	むかし	き	ひ
12	それ	おとこ	ところ	おとこ	24	しさい	それ	のち	なか
					25	われ	みな	め	ひとり

つぎに、分析に用いた17作品のなかで出現頻度が上位50位までの名詞を用いて主成分分析およびクラスター分析を行った。表5の主成分負荷量をみると、第1主成分においては、「かね、よう、ひとつ、きょう、ほど、とし、き、みな、もの、ほか、おとこ、じょうろ、やど、こと、たゆう」が正の方向に、「こころ、とき、いのち、みち、ゆめ、こえ、かた、のち、うえ」が負の方向に影響を与え、第2主成分では「こい、かた、め、はな、おんな、あと、おとこ、われ、こえ、うえ、て、ゆめ、きょう、たゆう、それ、ひ、かみ、ここ」が正の方向に、「いえ、こ、ひと、こと、よ、とき」が負の方向に影響を与えている。図2を見ると、各々の作品が内容別にまとまりをもって位置していることが見てとれる。さらに、好色物は中央から上にかけて位置していることから「こい、かた、め、はな」等の名詞が多く用いられ、武家物は左下に位置していることから「とき、のち」等、町人物は右下に位置していることから「いえ、こ、こと」等が多く用いられているということがわかる。

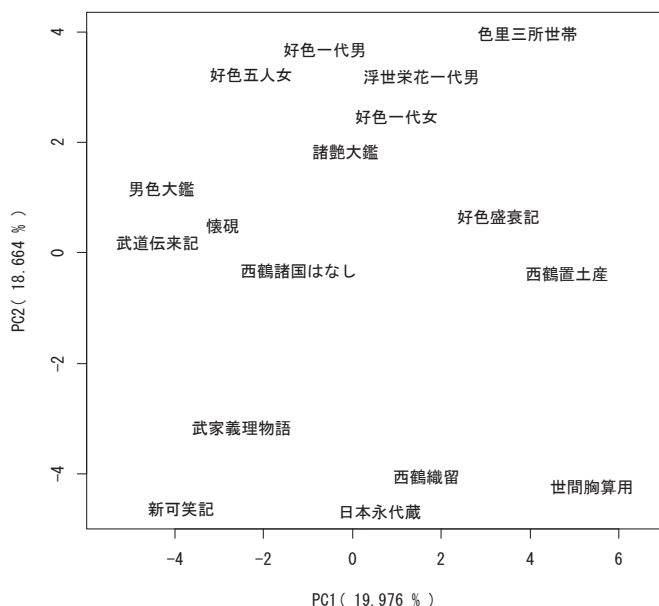


図2 主成分分析の結果

表5 主成分負荷量

	PC 1	PC 2		PC 1	PC 2
こと	0.45	-0.47	め	0.36	0.74
ひと	-0.10	-0.54	むかし	0.14	-0.30
これ	0.13	-0.01	なか	-0.03	-0.11
もの	0.57	-0.29	とし	0.62	-0.34
み	-0.29	0.06	じょうろ	0.53	0.39
よ	-0.21	-0.46	ひとり	0.38	0.17
とき	-0.70	-0.45	いのち	-0.64	0.13
ほど	0.66	0.38	ひ	0.13	0.44
ところ	-0.07	-0.13	ひとつ	0.76	-0.12
いま	-0.19	0.24	かね	0.86	-0.19
おとこ	0.53	0.59	よう	0.83	-0.11
おんな	0.01	0.68	いえ	0.04	-0.82
こころ	-0.72	0.03	いずれ	0.18	0.30
うち	0.31	0.26	こい	0.14	0.85
それ	-0.35	0.46	はな	-0.02	0.73
きょう	0.68	0.48	みな	0.59	-0.02
ほか	0.57	-0.11	みち	-0.63	-0.06
われ	-0.08	0.52	うえ	-0.40	0.51
のち	-0.41	-0.27	こえ	-0.47	0.52
なに	0.38	-0.11	こ	0.15	-0.69
ここ	-0.12	0.40	かみ	0.24	0.44
かた	-0.45	0.75	やど	0.47	0.17
き	0.62	0.32	たゆう	0.43	0.47
て	-0.14	0.50	みやこ	0.38	0.12
あと	-0.33	0.64	ゆめ	-0.53	0.49

図3は、クラスター分析の結果（ウォード法、Kullback-Leibler 距離）のデンドログラムで、大まかに3つのクラスターに分かれており、左側のクラスターに好色物、中央のクラスターに町人物、右側のクラスターに武家物、雑話物、好色物が分類されている。

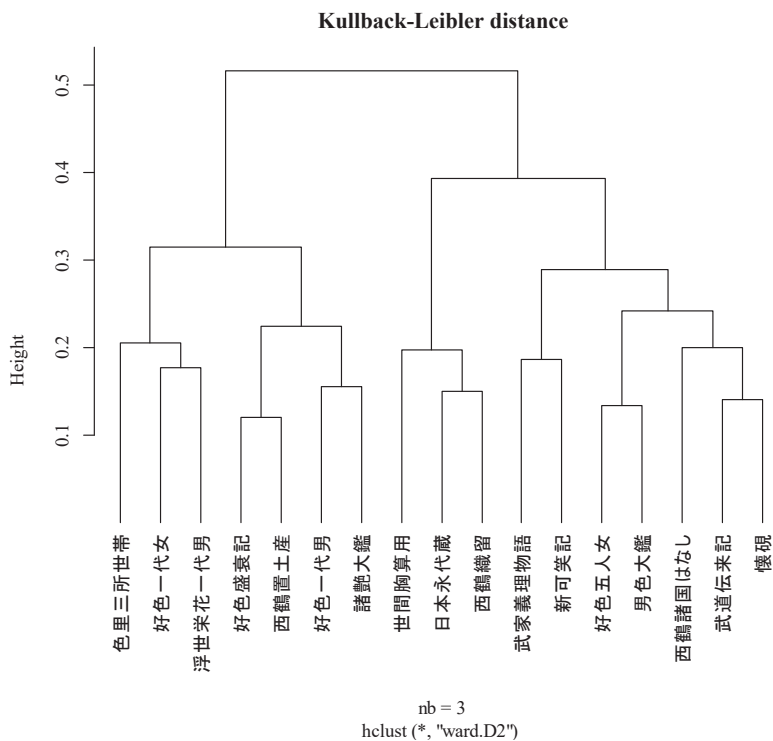


図3 クラスター分析の結果

表6は、出現頻度上位25語までの動詞をトピック別にまとめたものである。この表をみると、出現頻度上位25位のなかで武家物にのみ現れた動詞は「たずぬ、さす」、町人物にのみ現れた動詞は「もつ、うる、いたす、すむ」、好色物にのみ現れた動詞は「あう、すぐ」、雑話物にのみ現れた動詞は「のこる」であった。

表6 トピック別の出現頻度上位25位の動詞

	武家物	町人物	好色物	雑話物		武家物	町人物	好色物	雑話物
1	す	す	す	す	13	いる	たつ	つく	きく
2	あり	あり	あり	あり	14	つく	もうす	きく	とる
3	いう	いう	いう	いう	15	かたる	たまう	たつ	いる
4	なる	なる	なる	なる	16	ゆく	ゆく	とる	かえる
5	たまう	みる	みる	みる	17	かえる	かえる	あく	かく
6	もうす	とる	もうす	もうす	18	きく	もつ	みゆ	つく
7	うつ	つく	おもう	たまう	19	なす	あく	あう	なす
8	みる	いる	いる	おもう	20	かわる	きく	かえる	いず
9	しる	しる	たまう	ゆく	21	たずぬ	かわる	かたる	みゆ
10	かく	おく	しる	たつ	22	さす	うる	なす	のこる
11	たつ	おもう	かく	かたる	23	とる	きる	かわる	うつ
12	おもう	かく	ゆく	しる	24	いず	いたす	きる	おく
					25	みゆ	すむ	すぐ	あく

分析に用いた17作品のなかで出現頻度が上位25位までの動詞を用いて主成分分析およびクラスター分析を行った。表7の主成分負荷量をみると、第1主成分においては「かたる、もうす、たまう、うつ、きく」が正の方向に、「す、とる、おく、なる、つく、いう、あく」が負の方向に影響を与え、第2主成分では「おもう、みる、いる、あく、ゆく、あう、きく、かく、みゆ」が正の方向に、「うつ、たつ」が負の方向に影響を与えている。図4を見ると、各々の作品が内容別にまとまりをもって位置していることが見て取れる。さらに、好色物は中央から上にかけて位置していることから「おもう、みる、いる、あく、ゆく、あう、きく、かく、みゆ」等の名詞が多く用いられ、武家物は右下に位置していることから「かたる、たまう、うつ」等、町人物は右下に位置していることから「す、たつ」等、雑話物は中央から右側にかけて位置していることから「かたる、もうす、たまう、うつ、きく」等が多く用いられているということがわかる。

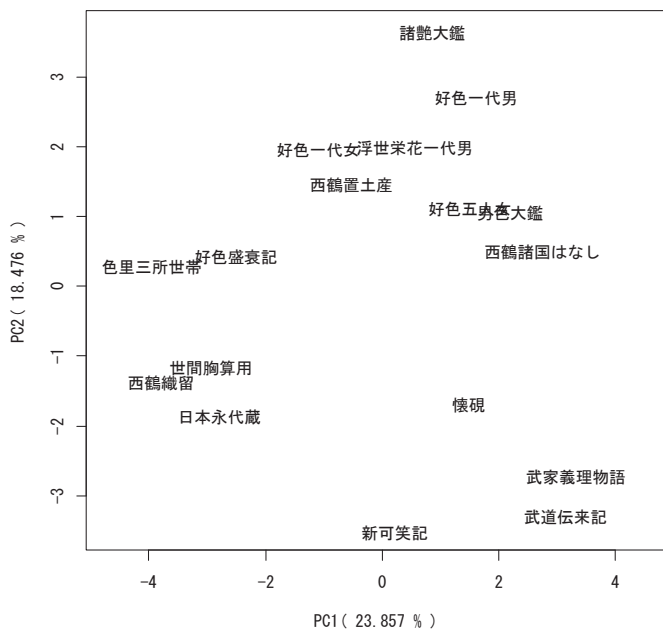


図 4 主成分分析の結果

表 7 主成分負荷量

	PC 1	PC 2		PC 1	PC 2
す	-0.86	-0.19	つく	-0.64	0.06
あり	-0.11	0.12	たつ	-0.12	-0.49
いう	-0.52	0.25	とる	-0.76	-0.06
なる	-0.68	-0.02	きく	0.48	0.56
みる	-0.24	0.75	かえる	0.29	-0.04
もうす	0.63	0.30	かたる	0.89	-0.09
たまう	0.60	-0.18	みゆ	0.30	0.44
おもう	-0.15	0.82	なす	-0.29	-0.04
いる	-0.32	0.70	あく	-0.43	0.65
しる	0.21	0.30	おく	-0.69	-0.17
かく	0.29	0.47	かわる	-0.23	0.13
ゆく	0.37	0.65	あう	0.08	0.64
			うつ	0.52	-0.53

図5は、クラスター分析の結果（ウォード法、Kullback-Leibler 距離）のデンドログラムで、大まかに3つのクラスターに分かれており、左側のクラスターに町人物、好色物、雑話物、武家物、中央のクラスターに武家物、右側のクラスターに雑話物、好色物が分類されている。

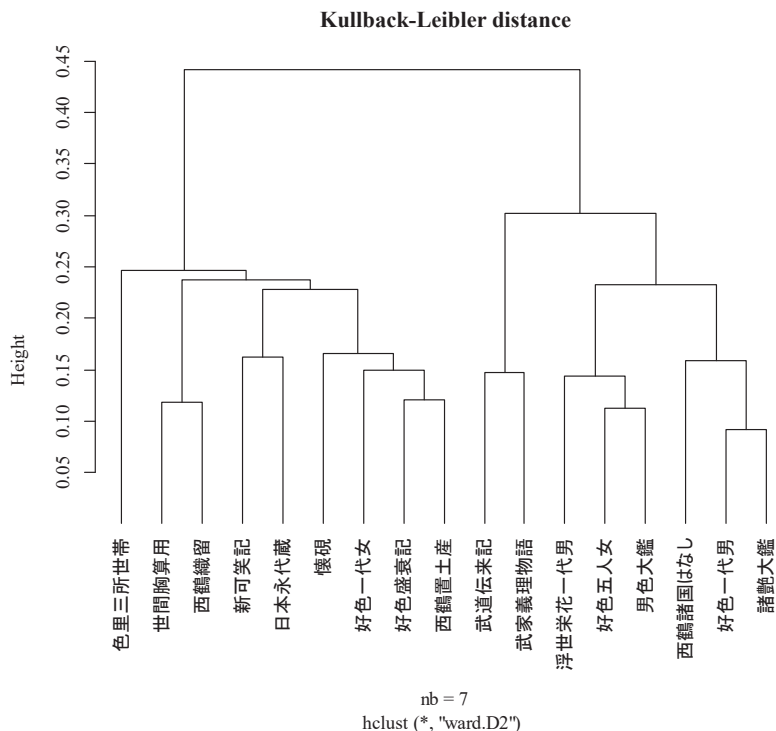


図5 クラスター分析の結果

表8は、出現頻度上位25語までの形容詞をトピック別にまとめたものである。この表をみると、出現頻度上位25位のなかで武家物にのみ現れた形容詞は「いさぎよし、たのもし、かいなし、ただし、つよし、かくれなし」、町人物にのみ現れた形容詞は「むづかし、わかし、ちいさし、わるし」、好色物にのみ現れた形容詞は「にくし、いやし」、雑話物にのみ現れた形容詞は「あさまし、やすし、おそし」であった。

表8 トピック別の出現頻度上位25位の形容詞

	武家物	町人物	好色物	雑話物		武家物	町人物	好色物	雑話物
1	なし	なし	なし	なし	13	いさぎよし	むつかし	やさし	おもしろし
2	かたし	よし	よし	よし	14	ちかし	くちおし	おし	あさまし
3	よし	かたし	おかし	ふかし	15	うし	ひさし	にくし	おそろし
4	ふかし	おなじ	ふかし	おなじ	16	たのもし	おおし	うつくし	ありがたし
5	おなじ	かしこし	かたし	かたし	17	おもしろし	おし	かしこし	ひろし
6	くちおし	おかし	おなじ	ひさし	18	かいなし	わかし	さびし	くちおし
7	かなし	かなし	ちかし	おかし	19	ただし	ちいさし	いやし	さびし
8	ありがたし	おそろし	おもしろし	ちかし	20	つよし	はやし	はやし	うし
9	うれし	ひろし	かなし	かなし	21	おかし	ふかし	ひさし	たかし
10	ぜひなし	あし	うれし	はやし	22	かしこし	せわし	くちおし	やさし
11	よろし	よろし	せわし	うつくし	23	あし	ぜひなし	あし	やすし
12	ひさし	たかし	おそろし	おおし	24	おし	おもしろし	ひろし	うれし
					25	かくれなし	わるし	たかし	おそし

分析に用いた17作品のなかで出現頻度が上位25位までの形容詞を用いて主成分分析およびクラスター分析を行った。表9の主成分負荷量をみると、第1主成分においては「やさし、ふかし、ちかし、うれし、にくし」が正の方向に、「ひろし、かしこし、なし、よし」が負の方向に影響を与え、第2主成分では「おもしろし、おかし、たかし、よし、おなじ、にくし、さびし」が正の方向に、「かたし、くちおし、よろし、ありがたし」が負の方向に影響を与えている。図6を見ると、各々の作品が内容別にまとまりをもって位置していることが見て取れる。さらに、好色物は中央から上にかけて位置していることから「おもしろし、おかし、たかし、よし、おなじ、にくし、さびし」等の名詞が多く用いられ、武家物は中央から下側にかけて位置していることから「かたし、くちおし、よろし、ありがたし」等、町人物は左側に位置していることから「ひろし、かしこし、なし、よし」等が多く用いられているということがわかる。

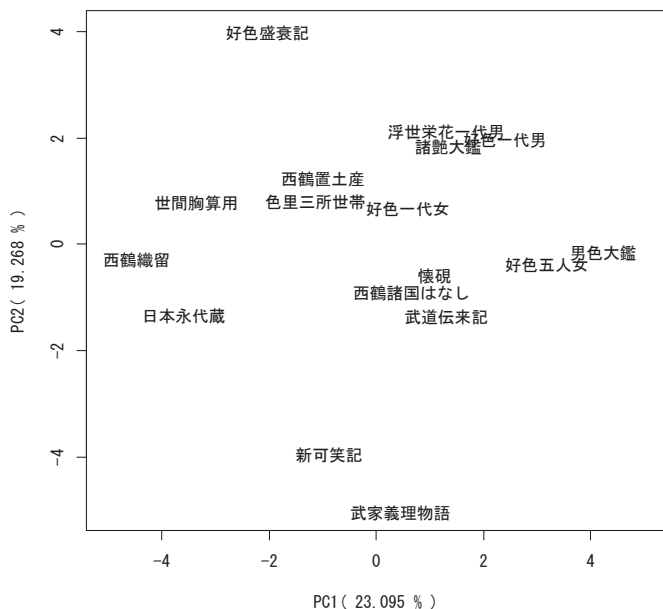


図6 主成分分析の結果

表9 主成分負荷量

	PC 1	PC 2		PC 1	PC 2
なし	-0.793	-0.201	ひさし	-0.161	-0.110
よし	-0.652	0.618	くちおし	-0.110	-0.692
かたし	-0.245	-0.752	おし	0.225	0.330
ふかし	0.757	-0.388	はやし	0.218	-0.101
おなじ	-0.324	0.580	ひろし	-0.834	0.169
おかし	0.316	0.653	あし	-0.108	0.087
ちかし	0.692	0.239	せわし	0.101	0.316
かなし	0.152	-0.251	やさし	0.803	0.008
おもしろし	-0.050	0.684	さびし	0.075	0.548
うれし	0.673	-0.021	にくし	0.642	0.580
かしこし	-0.797	0.008	ありがたし	0.347	-0.503
おそろし	-0.105	-0.042	たかし	-0.325	0.643
			よろし	-0.366	-0.525

図7は、クラスター分析の結果（ウォード法、Kullback-Leibler 距離）のデンドログラムで、大まかに3つのクラスターに分かれており、左側のクラスターに好色物、中央のクラスターに町人物、右側のクラスターに武家物、雑話物が分類されている。

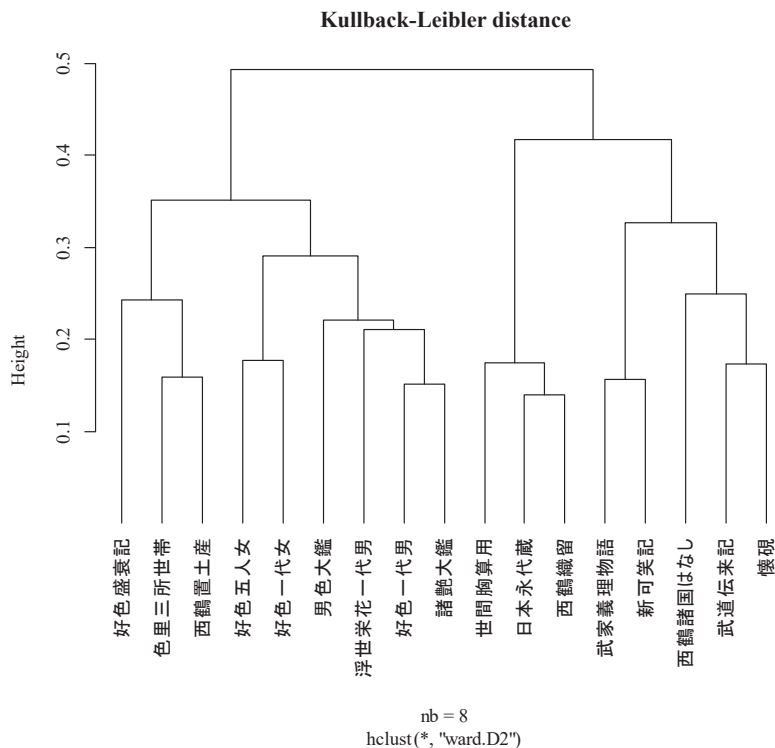


図7 クラスター分析の結果

四. おわりに

好色物、武家物、町人物、雑話物ごとに名詞、動詞、形容詞の出現頻度上位25語を各々まとめると、出現頻度の高い単語（1～12位）は共通して用いられていたが、出現頻度が低い単語（13～25位）では好色物、武家物、町人物、雑話物ごとに特徴的に用いられている単語が散見された。さらに、同様の情報を

用い、主成分分析およびクラスター分析で分析した結果、好色物、武家物、町人物、雑話物ごとにまとまって位置する傾向を示した。これらのことから、同一作者の作品内でも物語の内容によって、特徴の違いがみられるということが明らかになった。

参考文献

- 宗政五十緒、(1969).『国文学解釈と鑑賞』34(11). pp.26～35.「仮名草子から浮世草子へ」、至文堂。
浅野晃、他編、(2000).『新編西鶴全集』第1巻、勉誠出版。
浅野晃、他編、(2000).『新編西鶴全集』第2巻、勉誠出版。
浅野晃、他編、(2003).『新編西鶴全集』第3巻、勉誠出版。
浅野晃、他編、(2004).『新編西鶴全集』第4巻、勉誠出版。
上阪彩香、「西鶴浮世草子の文章に関する数量的研究——遺稿集を中心とした著者の検討——」、同志社大学博士論文、2016
村上征勝・金明哲・土山玄・上阪彩香、『計量文献学の射程』、勉誠出版、2016（著書）

* 討論要旨

野本忠司氏は、動詞を分析した場合に物語のジャンルに沿った結果が出にくいということは、動詞のほうが名詞よりも作者の個性が表れやすいと考えることもできるのではないかと指摘した。発表者は、今回のデータには分析手法上の課題が残されているため、現時点では即断できない、と回答した。野本氏はまた、分析手法がやや古いのではないかと発言した。発表者は、今回用いたKullback-Leibler情報量とワード法にはユークリッド平方距離では拾えないような情報を拾えるという利点があり、テキストマイニングにおいて有効な方法だとされている、と説明した。

板坂則子氏は、物語のジャンルに応じて文体が変化するという今回の結論は予想の範疇であるため、そうした傾向におさまらないデータにこそ西鶴の文体的特徴を見出すことができるのではないかと指摘した。板坂氏はまた、団水と西鶴のデータを比較することによって、どのような違いが見られたかと質問した。発表者は、団水と西鶴のデータを比較すると、助動詞に関してはあまり差が見られなかったが、助詞の組み合わせや名詞など多くの項目でそれぞれの作者に固有の特徴が見られた、と回答した。