鏡花の語彙と文体
ー計量言語学的アプローチー

間淵　洋子

1. はじめに

本研究では、言語計量の手法を用いて、泉鏡花作品の層化を試みる。

明治から昭和初期にかけて、近代の新たな文学・文体の、模索・成立期に活動した鏡花の作品には、テーマ性、文体、作風において変容が見られ、その多様性が鏡花文学の魅力の一つにもなっている。そこで本研究では、計量言語学的手法に基づき鏡花の小説作品を分析し、その多様性・重層性の構造について検討してみたい。

2. 研究方法

2.1 調査方針と対象資料

まず第1段階として鏡花の語彙・文体を大掴みに捉えることを目指し、鏡花作品の本文テキストをできるだけ多く収集し、計量言語学的手法による網羅的な語彙調査・分析を実施する。資料性への問題意識や本文研究的視点とは異なる視座からアプローチするため、大規模かつコンピュータで文字列を扱える形で電子化されていることを調査資料の要件として、整合的な形式で鏡花の多くの作品を電子化している『青空文庫』(https://www.aozora.gr.jp)のテキストデータを用いる。なお、鏡花作品には戯曲や随筆も多く含まれるが、本研究では小説のみを対象として分析を行う。「小説」の認定においては、『青空文庫』が各作品のテキストデータと共に提供する書誌情報のアーカイブページ(「図書カード」)から、「作品データ」の「分類」項目に記載されているNDC情報(「日本十進分類法」〈Nippon Decimal Classification: NDC〉に基づく分類番号)を用いることとし、「NDC913」(日本文学-小説. 物語)の情報を持つ作品を、本研究における「小説」として選定する。

2.2 調査対象作品の選定

前節に示した調査方針に基づき、2023年2月1日時点で『青空文庫』に公開されている全ての泉鏡花作品から、NDC913の全147作品を選定し、調査対象とした。下表に、初出年(和暦アルファベット略称による)ごとに作品名を示す。なお、対象とした147作の中には、『鏡花全集』(岩波書店、1940～1976年)において、小説とは別に小品、紀行、雑記等を収集した第27巻、第28巻に収録されている作品が35作含まれている。これらの作品は、小説として扱うことへの妥当性に欠ける可能性も考えられるが、ひと

表1　調査対象作品一覧

年	作品名(『青空文庫』の表記による)	年	作品名(『青空文庫』の表記による)
M26	活人形,金時計	T1	霰ふる,浅茅生,印度更紗
M27	義血侠血,黒壁	T2	艶書,二た面,陽炎座,崑崙本
M28	外科室,貧民倶楽部,旅僧,取舵,夜行巡査,妖怪年代記▲	T3	第二崑崙本,革鞄の怪,魔法罐,日本橋,みつ柏▲
M29	紫陽花,化銀杏,琵琶伝,海城発電,竜潭譚,照葉狂言	T5	人魚の祠,白金之絵図,浮舟
M30	誓之巻,凱旋祭,化鳥,清心庵,十万石▲,迷子▲	T7	茸の舞姫,大阪まで▲
M31	蛇くひ,星あかり	T9	売色鴨南蛮,伯爵の釵,瓜の涙,唄立山心中一曲,月令十二態▲
M32	黒百合,三尺角,湯島詣,弥次行▲	T10	銀鼎,雪霊記事,雪霊続記,雪の翼,続銀鼎
M33	葛飾砂子,高野聖,三枚続,湯女の魂,怪談女の輪▲	T11	妖魔の辻占,十六夜▲
M34	註文帳,蠅を憎む記,木精,政談十二社,処方秘箋,森の紫陽花▲	T12	鶴狩,みさごの鮨,婦人十一題▲,間引菜▲,祭のこと▲
M35	妖僧記,山の手小景▲	T13	夫人利生記,小春の狐,光籃,眉かくしの霊,露萩,春着▲,二、三羽▲
M36	伊勢之巻,二世の契,薬草取,草あやめ▲	T14	甲乙,怨霊借用
M38	悪獣篇,海異記,女客,わか紫	S2	河伯令嬢,卵塔場の天女,十和田湖▲
M39	式部小路,春昼,春昼後刻	S3	ピストルの使い方,鳥影▲
M40	縁結び,婦系図,聞きたるまゝ▲	S5	木の子説法
M41	星女郎,草迷宮,沼夫人,花間文字▲,妙齢▲	S6	古狢,貝の穴に河童の居る事,木菟俗見▲
M42	神鑿,海の使者,一寸怪▲,怪力▲,銭湯▲	S7	菊あわせ,白花の朝顔
M43	国貞えがく,歌行灯,画の裡▲,松の葉▲,麦搗▲	S8	開扉一妖帖,神鷺之巻,灯明之巻
M44	露肆,貴婦人,朱日記,月夜,妖術,吉原新話,夜釣,一景話題▲,甲冑堂▲,鑑定▲,人参▲	S12	薄紅梅,雪柳
M45	片しぐれ,南地心中,三人の盲の話,一席話▲,唐模様▲	S14	遺稿,縺紅新草

まず全て調査対象に含めることとし、他と区別するために、以下のリストで「▲」の印を付す。

2.3　本文の選定

選定した作品について分析を行う際には、どのような本文を用いるかが重要である。『青空文庫』では、同一作品に底本の異なる複数のテキストデータが存在する場合があるが、この場合は旧字体より新字体を、旧仮名遣いより新仮名遣いを優先して使用テキストを選択する。これは、機械による形態素解析を施す際、新字体・新仮名遣いのものの方が解析精度が高くなるためである。なお、鏡花の文学表現においては、各語がどう表記されるかという点[1)]が、重要な研究課題になるが、本研究では語彙を中心に分析を行い、表記の問題は扱わない。

2.4　形態素解析

本文選定後、語彙調査のために形態素解析を行う。形態素解析とは、文を単語に相当する一定の言語単位に区切り、読みや品詞といった単語情報を付加する技術を指す。形態素解析器と解析辞書を組み合わせて実行されるもので、使用する解析器・辞書により解析結果が異なるが、本研究では非現代日本語や表記・語形のバリエーションに強く、斉一な単語分割が可能な国立国語研究所開発の「UniDic」(小木曽他2013、https://clrd.ninjal.ac.jp/unidic/) を用いることとし、Web上で簡便にUniDicと形態素解析器MeCabを用いた形態素解析を実施できる「Web茶まめ」(堤・小木曽2023、https://chamame.ninjal.ac.jp) を利用した。「Web茶まめ」は、Webブラウザ上で、解析対象テキストの入出力、テキストの定型的加工処理、解析辞書の選択などが可能なツールで、本研究では、『青空文庫』の定型のルビタグ (例1下線部) を削除し、踊り字を展開する処理を選択し、文体に応じて「近代文語」「旧仮名口語」「近現代口語小説」の3つから解析辞書を選んで解析を実施した。図1に解析結果の例を示す。

図1　「Web茶まめ」による解析結果の例

辞書	文境界	書字形（＝表層形）	語彙素	語彙素読み	品詞	活用型	活用形	発音形出現形	仮名形出現形	語種	書字形(基本形)	語形(基本形)
近代文語	B				空白					記号		
近代文語	I	制帽	制帽	セイボウ	名詞-普通名詞-一般			セーボー	セイボウ	漢	制帽	セイボウ
近代文語	I	の	の	ノ	助詞-格助詞			ノ	ノ	和	の	ノ
近代文語	I	庇	庇	ヒサシ	名詞-普通名詞-一般			ヒサシ	ヒサシ	和	庇	ヒサシ
近代文語	I	の	の	ノ	助詞-格助詞			ノ	ノ	和	の	ノ
近代文語	I	下	下	シタ	名詞-普通名詞-一般			シタ	シタ	和	下	シタ
近代文語	I	に	に	ニ	助詞-格助詞			ニ	ニ	和	に	ニ
近代文語	I	ものすごく	物凄い	モノスゴイ	形容詞-一般	文語形容詞-ク	連用形-一般	モノスゴク	モノスゴク	和	ものすごし	モノスゴシ
近代文語	I	潜め	潜む	ヒソム	動詞-一般	文語四段-マ行	命令形	ヒソメ	ヒソメ	和	潜む	ヒソム
近代文語	I	る	り	リ	助動詞	文語助動詞-リ	連体形-一般	ル	ル	和	り	リ
近代文語	I	眼光	眼光	ガンコウ	名詞-普通名詞-一般			ガンコー	ガンコウ	漢	眼光	ガンコウ
近代文語	I	は	は	ハ	助詞-係助詞			ワ	ハ	和	は	ハ

例1）制帽の庇《ひさし》の下にものすごく潜める眼光は、機敏と、鋭利と厳酷とを混じたる、異様の光に輝けり。[2)]　　—『夜行巡査』M28（1895）年

2.5　調査項目

2.5.1　基礎統計

単語頻度の集計に基づき、データの総語数を表す延べ語数（Token）と、何種類の単語がデータ内に現れたかを示す異なり語数（Type）を求める。また、作品ごとの語数計測結果から、147作品の基礎統計（平均値、最小値、最大値、中央値。句読点や鉤括弧、空白などの記号類を除く。以下同様）を求める。その上で、出現度数（以下「頻度」）の高い語を観察する。

2.5.2　特徴語

他の近代小説作品と比較した際に見いだせる、鏡花作品に特徴的な語を求める。本研究では、コーパスから特徴語を抽出する際に用いられる手法の一つである「対数尤度比」（Log-likelihood ratio；LLR）を使用する。LLRでは、対象コーパスA（ここでは鏡花作品のデータ）と参照コーパスBにおける単語頻度比較により特徴語を導くが、本研究では、明治中期から大正末期にかけて刊行された主要な小説21作品をコーパス化した国立国語研究所（2021）『日本語歴史コーパス　明治・大正編Ⅳ近代小説』（以下、「CHJ近代小説」）を参照コーパスとして用い、以下の式により求めた。

$$LLR = 2(a\,log(a)+b\,log(b)+c\,log(c)+d\,log(d)-(a+b)\,log(a+b)-(a+c)\,log(a+c)$$
$$-(b+d)\,log(b+d)-(c+d)\,log(c+d)+(a+b+c+d)\,log(a+b+c+d)) \quad \cdots 数式（1）$$

a：　鏡花データでの当該語の度数　　b：　CHJ近代小説での当該語の度数

c：　鏡花データの延べ語数-a　　d：　CHJ近代小説の延べ語数-b

※ad-bc<0の場合は-1を乗じる

2.5.3　文体指標

鏡花の小説作品を語彙的特徴に基づき分析するために、文章を特徴付ける指標（以下「文体指標」）について検討する。文体指標としては樺島・寿岳（1965）における、名詞の比率、MVR（Modifying words and Verb Ratio）、字音語の比率、文の長さ（1文あ

たりの自立語数)、引用文の比率、接続詞を持つ文の比率、現在止めの文の比率、指示詞の比率、色彩語の比率、表情語の比率などがよく知られている。本研究ではこれを参照し、形態素解析結果を利用して求めることが可能な、①各品詞(名詞・代名詞・動詞・形容詞・形状詞・副詞・連体詞・接続詞・感動詞)の比率、②MVR、③各語種(和語・漢語・外来語・混種語)の比率、④平均文長(1文あたりの自立語数)、⑤1文あたりの読点数を、作品ごとに調査する。「MVR」は相の類(形容詞・形容動詞[3)]・副詞・連体詞)の語数合計に対する用の類(動詞)の語数の比率により表される値で、以下の式により求められる。MVR値が高ければ形容表現優位で「ありさま」を描写する傾向、低ければ動詞表現優位で「動き」を描写する傾向を表すものとされる。

MVR=(形容詞+形容動詞+副詞+連体詞)×100/動詞　…数式(2)

2.5.4　語彙・文体特徴に基づく作品のクラスタリング

前節に示した文体指標値を用いて、鏡花小説147作品の層化・分類を試みる。分類には、データ間の距離に基づき類似度を計り分類を行うクラスター分析を用いることとし、ユークリッド距離、ウォード法を使用した階層クラスタリングを実施する。なお、分析には統計分析用フリーソフト「R」(https://www.r-project.org、ver.4.3.0)を用いた。

3. 分析と考察

3.1　基礎統計

調査対象とした鏡花小説147作品の延べ語数、異なり語数、および、1作品あたりの語数における基礎統計を表2に示す。延べ語数・異なり語数については、比較対象とする『CHJ近代小説』の語彙量をあわせて示す。また、頻度の高い上位10語(自立語のみ)を頻度と共に表3に掲出する。なお、頻度は、計測値そのものである「粗頻度」と、これを100万語あたりに調整した「PMW」(Per Million Words)の2値で示す。

表2　調査データの基礎統計

	延べ語数	異なり語数	平均値	最小値	最大値	中央値
鏡花小説	1,582,793	32,148	10841.4	472	97,781	6,928
CHJ近代小説	697,215	24,733	—	—	—	—

表3　鏡花小説作品における頻度上位10語(自立語のみ)

順位	語彙素	語彙素読み	品詞	粗頻度	PMW	参考:CHJ	CHJ:粗頻度	CHJ:PMW
1	為る	スル	動詞	21,845	13,802	為る	12,409	17,798
2	有る	アル	動詞	12,913	8,159	居る	7,811	11,203
3	言う	イウ	動詞	12,126	7,662	有る	6,138	8,804
4	其の	ソノ	連体詞	9,907	6,260	言う	6,099	8,748
5	無い	ナイ	形容詞	8,858	5,597	無い	4,740	6,798
6	居る	イル	動詞	7,425	4,691	事/コト/名詞	4,373	6,272
7	様	ヨウ	形状詞	7,157	4,522	其の	3,946	5,660
8	成る	ナル	動詞	6,823	4,311	成る	3,888	5,576
9	物	モノ	名詞	5,833	3,685	様	3,814	5,470
10	何	ナニ	代名詞	5,720	3,614	其れ/ソレ/代名詞	3,336	4,785

表2より、分析対象として約160万語程度のデータが確保できており、計量分析に耐えるデータ量が得られていること、1作品あたりの語数には500語弱から10万語弱まで幅があり、ばらつきが大きいことが分かる。また、表3からは、頻度上位の語彙においては、鏡花の使用語彙と近代小説の使用語彙の間に大きな差が見られないことも見て取れる。

3.2 特徴語

LLR値10.83以上を基準として抽出された特徴語を表4に示す。「婦人」「女房」「肩」「袖」のように、女性や人物・衣服の一部を指す語、「薄い」「露わ」など様子・属性を形容する表現、「紅」「青い」「真っ白」「柔らか」など色彩、触覚などの五感表現が多く抽出されている点が注目されるほか、「殺す」「化ける」といった怪異小説を多く手がける鏡花ならではの語や、発話内に多く現れる「御座る」「為さる」「遊ばす」といった敬語表現が抽出されている点は興味深い。

表4　鏡花の特徴語（LLR値上位30位までを抜粋）

名詞			動詞			形容詞			形状詞		
語彙素	語彙素読み	LLR	語彙素	語彙素読み	LLR	語彙素	語彙素読み	LLR	語彙素	語彙素読み	LLR
雪	ユキ	434.3	御座る	ゴザル	797.8	白い	シロイ	281.3	同一	ドウイツ	94.0
婦人	フジン	301.6	飛ぶ	トブ	250.7	青い	アオイ	133.2	婀娜	アダ	43.0
袖	ソデ	276.5	込む	コム	243.5	薄い	ウスイ	115.9	唐突	トウトツ	42.1
影	カゲ	276.1	申す	モウス	220.1	等しい	ヒトシイ	62.4	真っ白	マッシロ	38.6
背後	ハイゴ	262.7	掛かる	カカル	168.3	細い	ホソイ	59.5	判然	ハンゼン	36.2
紅	クレナイ	249.0	為さる	ナサル	155.6	暗い	クライ	58.9	ばらばら	バラバラ	34.5
手	テ	227.0	冴える	サエル	146.7	凄い	スゴイ	51.6	恍惚	コウコツ	33.8
明かり	アカリ	194.9	存ずる	ゾンズル	140.8	怖い	コワイ	51.1	朦朧	モウロウ	32.2
婆	ババ	183.1	掛ける	カケル	135.7	黒い	クロイ	45.5	真っ暗	マックラ	31.9
褄	ツマ	175.4	消える	キエル	123.1	冷たい	ツメタイ	43.6	寂然	ジャクネン	25.6
水	ミズ	173.7	開く	ヒラク	122.5	危うい	アヤウイ	41.8	柔らか	ヤワラカ	21.2
蝶	チョウ	169.1	覗く	ノゾク	108.9	濃い	コイ	39.6	無事	ブジ	18.3
色	イロ	168.5	抱く	イダク	104.2	優しい	ヤサシイ	29.3	きざ	キザ	16.8
女房	ニョウボウ	166.8	付ける	ツケル	99.1	気高い	ケダカイ	27.6	雑	ザツ	16.8
緋	ヒ	166.6	潜る	モグル	98.4	慌ただしい	アワタダシイ	26.7	陰々	インイン	16.1
雲	クモ	163.5	さす	サス	94.4	怪しい	アヤシイ	22.2	茫	ボウ	15.3
膝	ヒザ	159.2	着く	ツク	94.3	淡い	アワイ	20.8	露わ	アラワ	15.1
肩	カタ	149.9	遊ばす	アソバス	88.1	強い	コワイ	19.5	然様	サヨウ	15.0
姉	ネエ	147.9	撥ねる	ハネル	87.3	若い	ワカイ	18.5	真っ黒	マックロ	15.0
紫	ムラサキ	139.6	透く	スク	82.9	堆い	ウズタカイ	18.3	強か	シタタカ	14.7
柳	ヤナギ	136.3	吐く	ハク	79.2	美しい	ウツクシイ	17.8	如何	イカガ	14.1
波	ナミ	132.2	打つ	ウツ	74.7	煩い	ウルサイ	16.3	自若	ジジャク	14.1
襟	エリ	127.1	退く	シリゾク	73.3	高い	タカイ	15.4	茫然	ボウゼン	14.0
呼吸	コキュウ	121.6	上る	ノボル	73.0	清い	キヨイ	13.7	艶やか	アデヤカ	14.0
縁	エン	121.3	縋る	スガル	71.3	赤い	アカイ	12.8	邪険	ジャケン	13.5
草	クサ	118.6	据える	スエル	71.2	心無い	ココロナイ	12.2	陰気	インキ	12.4
腰	コシ	116.2	殺す	コロス	70.6				高らか	タカラカ	11.5
九	キュウ	114.5	見える	ミエル	70.5						
此処	ココ	113.8	立つ	タツ	69.1						
片手	カタテ	112.3	化ける	バケル	66.4						

3.3 文体指標

3.3.1 各品詞の比率とMVR

作品ごとの品詞比率とMVRについて、基礎統計を表5に示す。樺島・寿岳（1965）では、近代小説100作品のサンプリングに基づき、各文体指標に対して5段階尺度で評価するための基準値を示しているが、これを適用して鏡花作品に見られる文章特徴を

表5　鏡花小説の品詞比率とMVR

	名詞率	代名詞率	動詞率	形容詞率	形状詞率	副詞率	連体詞率	接続詞率	感動詞率	MVR
平均値	50.5%	3.5%	28.5%	4.4%	2.4%	5.9%	2.4%	0.8%	1.4%	53.51
最小値	37.2%	1.2%	21.8%	1.6%	0.2%	3.1%	1.1%	0.1%	0.0%	28.87
最大値	63.9%	8.6%	37.7%	7.3%	4.4%	10.2%	4.2%	2.0%	3.6%	65.95
中央値	50.7%	3.5%	28.4%	4.4%	2.4%	5.9%	2.4%	0.8%	1.4%	54.02

総合的に評価すると、名詞率平均50.5%、MVR平均53.51は、いずれも「普通」評定であり、特異性があるとは言い難い。一方で、個々の作品の品詞比率に目を向けると、その特徴が見えてくる。

例2）霑いたる緑の黒髪は颯と乱れて、背と胸とに振分けたり。想うに、谷間を流るる一条の小川は、此処に詣ずる行者輩の身を浄むる処なれば、婦人も彼処にこそ垢離を取れりしならめ。　—『黒壁』M27（1894）年、名詞率：57.84%

例3）（前略）手術台なる伯爵夫人は、純潔なる白衣を絡いて、死骸のごとく横たわれる、顔の色あくまで 白く、鼻高く、頤 細りて手足は綾羅にだも堪えざるべし。脣 の色少しく褪せたるに、玉のごとき前歯かすかに見え、眼は固く閉ざしたるが、眉は思いなしか顰みて見られつ。わずかに束ねたる頭髪は、ふさふさと枕に乱れて、台の上にこぼれたり。　—『外科室』M28（1895）年、MVR：63.37

例4）汀の蘆に潜むか、と透しながら、今度は心してもう一歩。続いて、がたがたと些と荒く出ると、拍子に掛かって、きりきりきり、きりりりり、と鳴き頻る。／熟と聞きながら、うかうかと早や渡り果てた。
—『海の使者』M42（1909）年、副詞率：10.22%

樺島・寿岳（1965）は、品詞構成と文章の特徴との関連性について「名詞比率Nが大きい文章は要約的，小さい文章は描写的の傾向がある．またMVRが大きい文章はありさま描写的，小さい文章は動き描写的である．」と述べる。鏡花作品においては、名詞率が高い例2『黒壁』、MVRの高い例3『外科室』は、いずれも初期作品で、装飾性・修辞性の高い、いわゆる美文調の文体であり、副詞率の高い例4『海の使者』は、口語体だがオノマトペを多用した装飾的な文体となっている。注目したいのは、名詞率の高さが要約性や事柄の密度に寄与するのではなく、例2に見えるように、描写を詳細にするファクターとして機能している点であり、品詞構成の特徴がいずれも「装飾性」「描写性」に働く可能性を見いだせる。換言すれば「装飾」「描写」を様々な品詞の表現で行っているということで、鏡花の文章表現の多様性の表れとして捉えることができる。

3.3.2　各語種の比率

作品ごとの各語種の比率について、基礎統計を表6に示す。樺島・寿岳（1965）の指

標に基づけば、鏡花作品は字音語率（本調査における「漢語率」）が極めて低い。例5『紫陽花』に見える和語を主体とした擬古文体は鏡花の特徴と言えようが、加えて漢語率が最も高い前出例2『黒壁』（漢語率：20.58%）のように、仏語や漢語形状詞を多く含む作品ですら、近代小説においては平均的な比率に過ぎない。注1に指摘した、漢語文字列に対して振り仮名で和語の読みを付した表現を多用する点も含めて、鏡花の語用意識の投影と捉えられよう。

例5）色青く光ある蛇、おびたゞしく棲めればとて、里人は近よらず。其野社（そののやしろ）は、片眼の盲ひたる翁ありて、昔より斉眉（かしず）けり。／　其（その）片眼を失ひし時一たび見たりと言ふ、几帳の蔭に黒髪のたけなりし、それぞ神なるべき。

—『紫陽花』M27（1896）年、漢語率：5.71%

表6　鏡花作品の語種構成

	和語率	漢語率	外来語率	混種語率
平均値	85.5%	11.4%	0.2%	1.1%
最小値	75.2%	5.3%	0.0%	0.3%
最大値	93.9%	20.6%	1.0%	2.3%
中央値	85.5%	11.0%	0.2%	1.1%

表7　鏡花作品の文長と読点数

	文長	読点数／文
平均値	10.38	2.32
最小値	5.98	1.04
最大値	26.68	6.43
中央値	9.50	2.15

3.3.3　平均文長、1文あたりの平均読点数

作品ごとの平均文長、平均読点数について、基礎統計を表7に示す。ばらつきが非常に大きいのが特徴的であり、文長が長く読点の多いものは、例6のような擬古文体、文長が短く読点数の少ないものは、例7のように短い会話文が多用される作品である。文体の変遷、あるいは小説形式の多様性の反映と捉えられる。

例6）此処に予め遊蝶花（あらかじ）、長命菊、金盞花（きんせんくわ）、縁日名代の豪のもの、白、紅、絞、濃紫（こむらさき）、今を盛に咲競ふ、中にも白き花紫雲英（はなげんげ）、一株方五尺に蔓（はびこ）り、葉の大なること掌（たなそこ）の如く、茎の長きこと五寸、台（うてな）を頂く日に二十を下らず、蓋（けだ）し、春寒き朝、めづらしき早起の折から、女形とともに道芝（みちしば）の霜を分けてお濠（ほり）の土手より得たるもの、根を掘らんとして、袂に火箸を忍ばせしを、羽織の袖の破目（やぶれめ）より、思がけず路に落して、大（おほい）に台所道具に事欠（ことかき）し、経営惨憺仇（あだ）ならず、心なき草も、あはれとや繁りけん。　—『草あやめ』M36（1903）年、文長：26.68語

例7）「生（い）きてるかい、」／「へゝゝ。」／「確乎（しつかり）しろ。」／「へゝゝ、おめでたう、へゝゝへゝ。」／
「可加減（いいかげん）にしねえな。おい、串戯（じようだん）ぢやねえ。お前（まへ）の前（まへ）だがね、悪女（あくぢよ）の深情（ふかなさけ）つてのを通越（とほりこ）して居（ね）るから、鬼（おに）に喰（く）はれやしねえかツて、皆友達（みなともだち）が案（あん）じて居（ね）るんだ。お前（まへ）の前（まへ）だがね、おい、よく辛抱（しんぼう）して居（ね）るぢやねえか。」

—『鑑定』M44（1911）年、文長：5.98語

3.4 作品のクラスタリング

前節までで示した文体指標を用いて階層クラスタリングを実施した結果を図2に示す。高さ30の位置でAからGの7つのクラスタに分割し、各クラスタの特徴を捉えて

図2　鏡花作品のクラスター分析

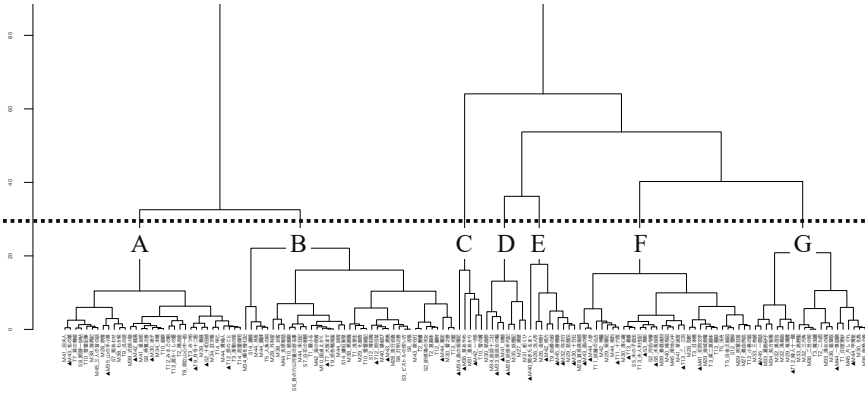


表8　クラスタ別作品一覧

クラスタ	作品
A	M28_旅僧, M30_迷子▲, M34_註文帳, M34_木精, M35_山の手小景▲, M38_海異記, M38_わか紫, M39_式部小路, M39_春昼, M41_星女郎, M41_沼夫人, M42_銭湯▲, M43_松の葉▲, M44_貴婦人, M45_三人の盲の話, T2_陽炎座, T3_革鞄の怪, T3_みつ柏▲, T7_茸の舞姫, T9_瓜の涙, T9_唄立山心中一曲, T9_月令十二態▲, T10_銀鼎, T10_雪霊記事, T12_みさごの鮨, T12_祭のこと▲, T13_眉かくしの霊, T14_甲乙, T14_怨霊借用, S2_十和田湖▲, S7_菊あわせ, S8_開扉一妖帖, S8_神鷺之巻
B	M28_外科室, M28_夜行巡査, M29_化銀杏, M30_誓之巻, M34_蠅を憎む記, M38_悪獣篇, M38_女客, M40_縁結び, M42_海の使者, M43_国貞えがく, M43_歌行灯, M43_麦搗▲, M44_露肆, M44_朱日記, M44_月夜, M44_妖術, M44_吉原新話, M44_鑑定▲, T1_霰ふる, T1_浅茅生, T2_艶書, T2_莨蒻本, T5_人魚の祠, T7_大阪まで▲, T9_売色鴨南蛮, T10_雪霊続記, T10_続銀鼎, T12_鷭狩, T12_間引菜▲, T13_光籃, S2_卵塔場の天女, S3_ピストルの使い方, S6_古狢, S6_貝の穴に河童の居る事, S7_白花の朝顔, S8_灯明之巻, S12_薄紅梅, S14_遺稿, S14_縷紅新草
C	M31_星あかり, M34_森の紫陽花▲, M36_草あやめ▲, M42_一寸怪▲, T10_雪の翼
D	M27_黒壁, M28_妖怪年代記▲, M30_凱旋祭, M31_蛇くひ, M33_怪談女の輪▲, M34_政談十二社, M35_妖僧記, M41_妙齢▲
E	M26_活人形, M26_金時計, M29_琵琶伝, M30_十万石▲, M32_弥次行▲, M40_聞きたるまゝ▲, M42_怪力▲, M45_唐模様▲, T9_伯爵の釵
F	M27_義血侠血, M28_貧民倶楽部, M28_取舵, M29_紫陽花, M29_海城発電, M29_照葉狂言, M30_清心庵, M39_春昼後刻, M40_婦系図, M41_草迷宮, M41_花間文字▲, M42_神鑿, M43_画の裡▲, M44_夜釣, M44_人参▲, M45_南地心中, T3_第二莨蒻本, T3_日本橋, T5_白金之絵図, T5_浮舟, T11_妖魔の辻占, T11_十六夜▲, T13_夫人利生記, T13_小春の狐, T13_露萩, T13_春着▲, T13_二、三羽▲, S2_河伯令嬢, S3_鳥影▲, S5_木の子説法, S6_木菟咎見▲, S12_雪柳
G	M29_竜潭譚, M30_化鳥, M32_黒百合, M32_三尺角, M32_湯島詣, M33_葛飾砂子, M33_高野聖, M33_三枚続, M33_湯女の魂, M34_処方秘箋, M36_伊勢之巻, M36_二世の契, M36_薬草取, M44_一景話題▲, M44_甲冑堂▲, M45_片しぐれ, M45_一席話▲, T1_印度更紗, T2_二た面, T3_魔法罎, T12_婦人十一題▲

表9　クラスタ別文章指標平均値

	MVR	名詞率	代名詞率	動詞率	形容詞率	形状詞率	副詞率	連体詞率	接続詞率	感動詞率	和語率	漢語率	外来語率	混種語率	文長	読点/文
A	55.7	50.9%	3.4%	27.9%	4.6%	2.6%	5.9%	2.5%	0.9%	1.3%	86.1%	10.8%	0.3%	1.1%	9.7	2.3
B	60.8	49.3%	3.7%	27.7%	4.8%	2.8%	6.7%	2.5%	0.9%	1.6%	86.1%	10.8%	0.3%	1.0%	8.9	2.1
C	51.2	53.2%	2.5%	28.7%	4.8%	2.0%	5.3%	2.5%	0.6%	0.4%	86.9%	10.5%	0.1%	1.0%	22.2	5.1
D	41.6	51.7%	3.4%	30.8%	3.7%	1.7%	5.0%	2.4%	0.7%	0.6%	83.6%	14.0%	0.2%	1.1%	16.0	2.9
E	40.6	51.9%	3.6%	30.2%	3.4%	1.7%	5.1%	2.0%	0.8%	1.3%	80.9%	14.8%	0.2%	1.3%	9.6	1.7
F	50.6	50.5%	3.8%	28.7%	4.1%	2.3%	5.7%	2.4%	0.8%	1.7%	84.9%	11.7%	0.2%	1.1%	8.4	1.9
G	51.6	50.4%	3.4%	29.2%	4.5%	2.3%	5.8%	2.4%	0.7%	1.2%	86.4%	10.6%	0.2%	1.1%	12.6	2.8

みたい。各クラスタに所属する作品を表8に、文体指標の平均値を表9に示す。

クラスタA・BはMVRが高い点が特徴的で、Bでより顕著である。AはMVRに関する指標を除き平均的であるが、Bは名詞比率の低さ、文の短さ、読点の少なさなどに特徴がある。A・Bと対照的なのがクラスタD・Eで、MVRが低い。両者は漢語率が高い点でも類似しているが、文長・読点数においてDが高くEが低いという対比が見られる。一方クラスタCは、品詞構成において顕著に名詞比率が高く、語種では和語率が極めて高く、文長・読点数が顕著に高い層で、他と異なっている。クラスタF・Gは、品詞比率や語種構成が平均的である点に類似性が見られ、文長・読点数において、Fで低くGでやや高いという対比が見られる。

3.5 鏡花作品の層化とその意味

前節で試みた鏡花作品の層化の結果は、「鏡花の文体」についてどのような事柄を表しているだろうか。

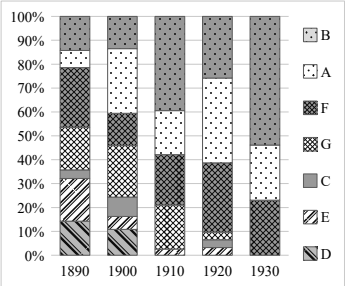
まず、各クラスタは、それぞれ異なる文章特徴を持つことが分かったが、そこに所属する作品の発表年代、文体（文語文・口語文等）、ジャンル（観念小説・怪異小説等）に明確な対応関係は見られなかった。これは、鏡花の文体が多様性に富むことの表れと捉えることができる[4)]。

一方で、各クラスタを特徴付ける文体指標において、特に数値が高い・低いといったクラスタの典型例とでも言えるような作品は、早い時期の発表作品に偏っている傾向が見られた[5)]。そこで、10年を単位とする年代別にクラスタの分布を確認したところ、初期ではクラスタ分布が多様であり、後年ほど分布の偏りが見られること、特にクラスタC・D・Eはほぼ1900年代以前の発表作品に限られることが明らかになった（表10、図3）。これは、初期作品において、さまざまな特徴的を持つ多様な文体が試みられ、次第に収束し平均的な様相を見せつつ、後年に向けてMVRが高く、文が短く簡潔で、和語を中心に構成された文体として確立されていったことを示唆するものと考えられる。

表10　発表年代別クラスタ分布

発表年	D	E	C	G	F	A	B	総計
1890	4	5	1	5	7	2	4	28
1900	4	2	3	8	5	10	5	37
1910		1		7	8	7	15	38
1920		1	1	1	9	11	8	31
1930					3	3	7	13
総計	**8**	**9**	**5**	**21**	**32**	**33**	**39**	**147**

図3　発表年代別クラスタ分布



4. おわりに

本研究では、泉鏡花の小説作品を対象とした大規模な語彙調査を行い、その結果に基づき鏡花作品の層化を試みた。異なる文章特徴を持つ7つのカテゴリに作品を分類した結果、初期の作品では語彙的な特徴がより強く見られ、文体的バリエーションが豊富だが、次第に平均的な様相を見せ、和語を中心とした短く簡潔な文で、形容表現の豊富な描写性の高い文体に収束していく様が見て取れた。ここから、鏡花の文体が極めて多様であること、そして、様々な語り方を意欲的に試みながら、文体を確立していった可能性があることを述べた。

今回の文体調査は一部に留まっており、分析も不十分である。貴重な言語資源を活用し、近代の文体形成の一端を担った「鏡花の文体」の実態・変遷をより明瞭に把握する必要がある。今後も検討を続けたい。

注

1)「壮佼(わかもの)」(『義血侠血』) のように、本行の漢字表記に一般的ではない読みを振り仮名で示すような表現が、鏡花作品には多用される。これは、鏡花の言語表現の特徴として重要なファクターであるが、別途検討することする。

2) 引用は『青空文庫』による。下線は引用者による。改行を「／」で置き換える場合がある。

3) UniDicによる形態素解析では、「形容動詞」を「形状詞＋助動詞「だ」」と解析するため、「形容動詞」を「形状詞」と読み換えて用いる。

4) シンポジウム時の吉田遼人氏・富永真樹氏との議論による。

5) シンポジウム時の田島優氏の御指摘による。記してお礼申し上げる。

参考文献・使用データ

小木曽智信・小町守・松本裕治 (2013)「歴史的日本語資料を対象とした形態素解析」『自然言語処理』20(5), 727-748.

樺島忠夫・寿岳章子 (1965)『文体の科学』綜芸舎

国立国語研究所(2021)『日本語歴史コーパス　明治・大正編Ⅳ近代小説』(短単位データ1.0) https://clrd.ninjal.ac.jp/chj/meiji_taisho.html#shosetsu

堤智昭・小木曽智信 (2023)「複数のUniDic辞書による形態素解析支援ツール『Web茶まめ』の実装と運用」『情報処理学会論文誌』64 (3), 749-757.

宮島達夫・近藤明日子 (2011)「古典作品の特徴語」『計量国語学』28(3), 94–105.

付記　本研究は、和洋女子大学研究奨励費、および、国立国語研究所共同研究プロジェクト「多様な語彙資源を統合した研究活用基盤の共創」による研究成果の一部である。

(和洋女子大学)