

森鷗外の文体的特徴の変化に関する計量的な考察

土山 玄

1. はじめに

文学作品を対象とし、計量的な手法を用いて文献に検討を加える研究は計量文献学と称される。計量文献学は、著者の文体に関わる習慣的、形式的特徴を統計的に分析することで、著者の識別や推定、文献の成立年代、あるいは成立の順序を推定する学問分野⁽¹⁾である。古典文学作品などの歴史的な文献には、著者や成立時期が議論の対象となっている場合が少なからず存在する。このような問題を扱うとき、従来は記述内容の検討や成立に関する歴史的事実の考証という観点から研究を行うのが主たる方法であったが、計量文献学はこのような方法とは一線を画し、文章から得られる計量的なデータを収集し、これを分析することによって当該文献の文体的特徴を把握して、結論を導き出す。しかし、文体という概念は多様であり、学問分野によって文体という概念が指す内容は異なる。文章の計量分析における文体は文字や単語の頻度、単語や文の長さの平均値などの文章を構成する量的な要素である。

日本語で記述された文献を対象とした研究では、古典文学作品である『源氏物語』が研究対象となることが多い。このような研究では、単語の出現率を特徴量とし、多変量解析を行うことで『源氏物語』の最終10巻の文体が他の巻と異なること⁽²⁾や、『源氏物語』の作者が単独である可能性が高いことが報告されている⁽³⁾。また、現代文を対象とした研究では、井上靖、三島由紀夫、中島敦の著作を対象とし、読点の使用法に著者の特徴があらわれることを明らかにした研究が有名である⁽⁴⁾。つまり、計量文献学では形式的な文体的特徴に、著者の個性が反映されると考えられるのである。このように、著者の識別を目的とした研究は広く行われている。他方、文献の成立順序の推定を目的とした研究は、芥川龍之介の文章について分析を行った研究が有名である。これは芥川龍之介の文章309編について分析を行った結果、係助詞の「は」及び格助詞の「に」「を」「の」の出現率が継時的に増加し、反対に格助詞の「が」「と」や接続助詞の「て」の出現率が減少しているこ

表1 森鷗外の小説と発表年

タイトル	発表年	タイトル	発表年
うたかたの記	1890	かのように	1912
舞姫	1890	興津弥五右衛門の遺書	1912
文づかい	1891	鼠坂	1912
そめちがへ	1897	佐橋甚五郎	1913
キタ・セクスアリス	1909	護持院原の敵討	1913
半日	1909	阿部一族	1913
鶉	1909	堺事件	1914
あそび	1910	大塩平八郎	1914
普請中	1910	安井夫人	1914
木精	1910	栗山大膳	1914
杯	1910	じいさんばあさん	1915
沈黙の塔	1910	二人の友	1915
牛鍋	1910	余興	1915
独身	1910	山椒大夫	1915
花子	1910	最後の一句	1915
里芋の芽と不動の目	1910	津下四郎左衛門	1915
青年	1910	魚玄機	1915
食堂	1910	伊沢蘭軒	1916
カズイステカ	1911	壽阿彌の手紙	1916
妄想	1911	寒山拾得	1916
心中	1911	梶原品	1916
百物語	1911	渋谷抽斎	1916
雁	1911	高瀬舟	1916
		細木香以	1917

とを明らかにした⁽⁵⁾。しかし、このような文体的特徴の継時的な変化についての研究は広くなされておらず、まだ十分に展開しているとは言えない。

そこで、本研究では近現代を代表する文豪である森鷗外の小説を分析対象として採り上げ、その文体の量的な変化について検討を加える。

2. データ

本研究に用いた森鷗外の小説は表1に示した1890年から1917年までに発表された47作品である。また、これらの小説のテキストデータはwebサイトの

の青空文庫 (<http://www.aozora.gr.jp/>) から入手した。次に、テキストデータに対し、形態素解析によって単語に品詞のタグ付けを行った。形態素解析はフリーソフトのMeCab ver. 0.996を、形態素解析の際に用いる辞書はUniDic ver. 2.0.1を用いた。なお、形態素解析に辞書としてUniDicを使用すると文語であるかどうか識別が可能であり、本研究ではこの情報を分析に使用した。

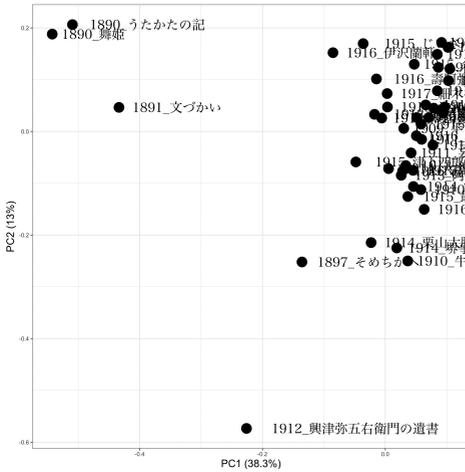


図2 47作品を対象とした助動詞の主成分分析の結果

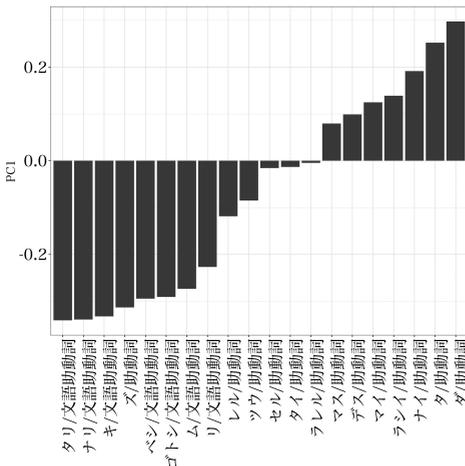


図3 47作品を対象とした助動詞の主成分負荷量

結果であり、図2では『うたかたの記』『舞姫』『文づかい』『興津弥五右衛門の遺書』の4作品が他の作品から大きく離れて位置している。特に、『うたかたの記』『舞姫』『文づかい』の3作品は第1主成分によって他の作品と分離されていると考えられる。つまり、上掲の3作品は第1主成分の主成分得点が他の作品に比べて小

析を助詞、助動詞と品詞別に行った。

3.2 分析結果

まず、助詞の出現頻度上位20語について主成分分析を行った。これら20語の累積頻度は助詞の総頻度の96.4%に該当する。分析の結果、図1に示すように『うたかたの記』『舞姫』『文づかい』『そめちがへ』『興津弥五右衛門の遺書』の5作品が他の作品から離れて付置されている。これら5作品のうち『興津弥五右衛門の遺書』を除く4作品は1900年以前の作品である。また、『興津弥五右衛門の遺書』は1912年に発表された小説であるが、森鷗外にとっての初の歴史小説とみなされている作品である。

次いで、助動詞の出現頻度上位20語について主成分分析を行った。分析に用いた20語の累積頻度は助動詞の総度数の97.9%となる。図2はその分析

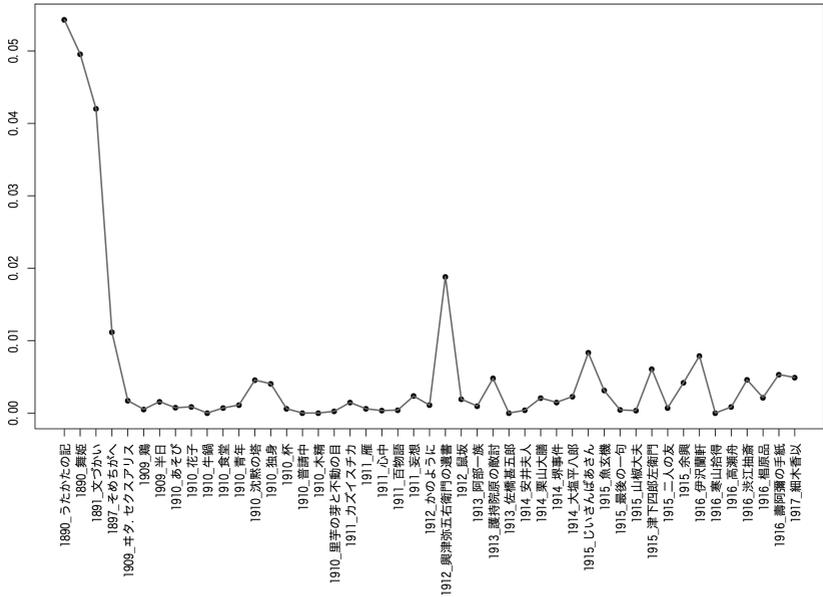


図4 文語助動詞の比率

さく、第1主成分に負の影響を与えている変数の出現傾向が他の作品に比べて高いことが推測される。図3は主成分分析によって求められた第1主成分の主成分負荷量をソートした棒グラフであり、主成分負荷量が小さい変数は「たり」「なり」「き」「べし」「ごとし」などである。これらの助動詞は形態素解析において文語と識別された助動詞であることから、森鷗外の初期の小説はそれ以降の小説に比べて文語表現が多く用いられてい

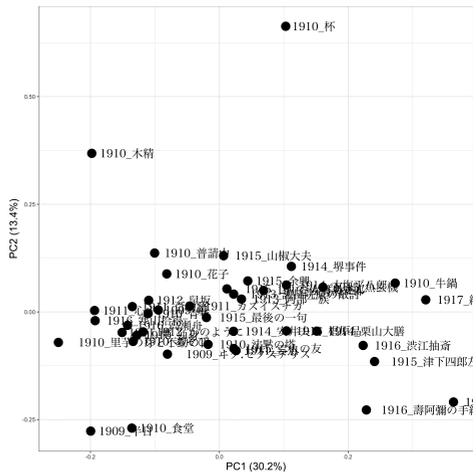


図5 42作品を対象とした助詞の主成分分析の結果

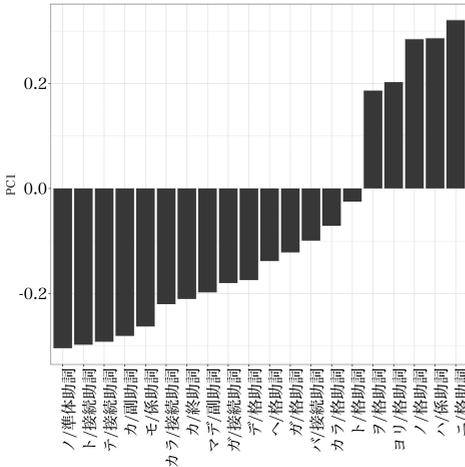


図6 42作品を対象とした助詞の主成分負荷量

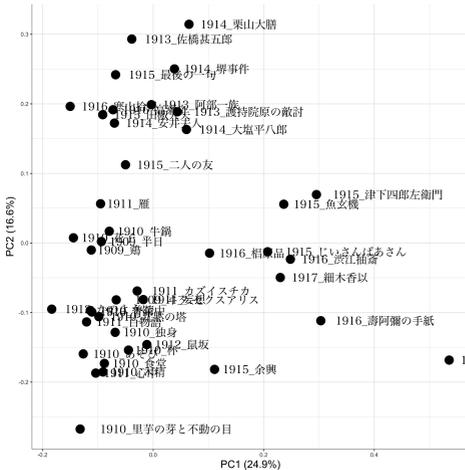


図7 42作品を対象とした助動詞の主成分分析の結果

で主成分分析を行った。図5に示すように、主成分分析の結果、早い時期の小説は第1主成分の主成分得点が負となり、1915年以降の小説は主成分得点が正となっていることが分かる。よって、助詞の出現傾向に継時的な変化が認められると考えられる。また、助詞の第1主成分の主成分負荷量は図6に示した通りであり、格助

たことが推測される。

そこで、文語と識別された助動詞のみを集計し、作品別の文語助動詞の延べ語数に対する比率を求めた。図4が文語助動詞の比率を視覚化したグラフである。図4より明らかであるように、『うたかたの記』『舞姫』『文づかい』の3作品の文語助動詞の比率が高く、それに『興津弥五右衛門の遺書』と『そめちがへ』が続く。また、先にふれたように、『興津弥五右衛門の遺書』は森鷗外にとっての初の歴史小説であり、『そめちがへ』は『うたかたの記』『舞姫』『文づかい』と同様に1900年以前に発表された初期の小説である。このような背景から、これらの5作品を分析対象から除外し、改めて主成分分析を行った。よって、分析に用いた小説は1909年から1917年までに発表された42作品である。

まず、先の分析と同様に、助詞の出現頻度上位20語について

詞の「に」「の」や係助詞の「は」といった単語の出現率が継時的に増加し、準体助詞の「の」や接続助詞の「と」「て」といった単語の出現率が継時的に減少していると考えられる。

次に、助動詞の出現頻度上位20語について主成分分析を行った。分析結果は図7のようになり、助詞の分析と異なり出現傾向が継時的に変化すると考えられる文体的特徴は認められない。図8は第1主成分の主成分負荷量、図9は第2主成分の主成分負荷量である。図8から明らかのように、第1主成分は文語助動詞であるか否かということの意味する合成変数である。文語助動詞の出現率が高いと考えられる作品を除けば図7において第2主成分の主成分得点が負になる作品は発表時期が早く、正になる作品は発表時期が遅いという傾向が読み取れる。そこで第2主成分の主成分負荷量を調査すると、図9より発表時期

の早い作品には「です」という助動詞の出現率が高く、発表時期の遅い作品には「ます」という助動詞の出現率が高くなるという傾向があることが推測される。

「です」や「ます」はどちらも文末表現に係わる助動詞であり、文末に現れる助動詞の出現傾向が継時的に変化するという可能性が推測される。そこで、助動詞の

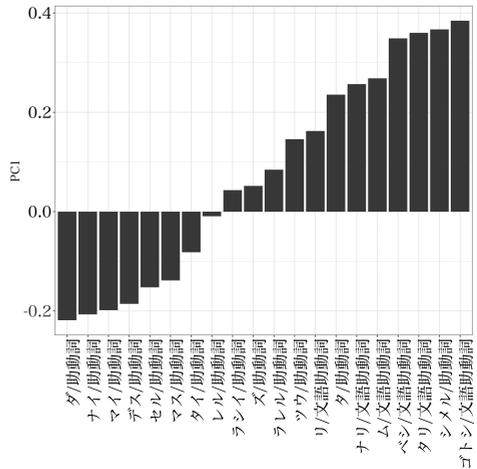


図8 助動詞の第1主成分の主成分負荷量

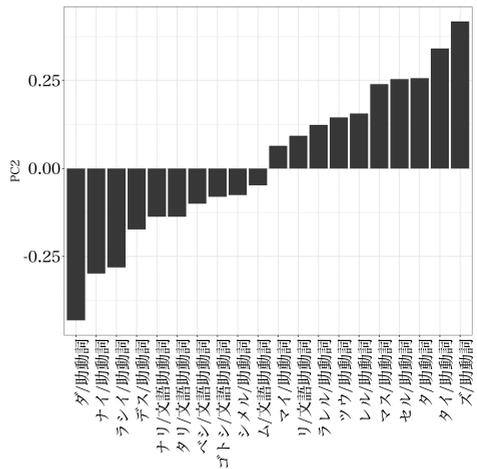


図9 助動詞の第2主成分の主成分負荷量

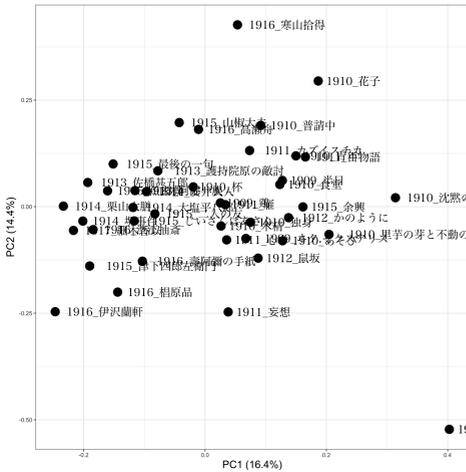


図10 助動詞と補助記号の主成分分析

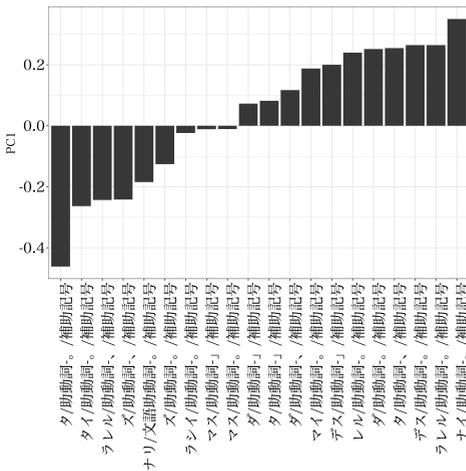


図11 助動詞と補助記号の主成分負荷量

出現率を単独で分析に用いるのではなく、助動詞とその助動詞に隣接する補助記号の組の出現率を求め、これを分析に用いた。本研究において補助記号と分類される要素は句点、読点、括弧の3種である。助動詞と補助記号の組を用いた主成分分析においても出現頻度上位20変数を分析に用いた。その結果、図10に示すように発表時期が早い小説は第1主成分の主成分得点が正に、発表時期が遅い小説は負になるという傾向が認められる。図11は第1主成分の主成分負荷量であり、この結果から「ない。」「られる。」「です。」という文末表現の出現率が継時的に増加し、「た。」「たい。」といった文末表現の出現率が継時的に減少していると考えられる。

4. 結論

本研究では森鷗外の小説47作品を対象に、計量的な観点から森鷗外の習慣的、形式的な文体的特徴の継時的変化について検討を加えた。分析の結果、計量的な判断に基づけば森鷗外の小説には継時的に出現傾向が変化する文体的特徴が認められた。まず、1900年以前に発表された『うたかたの記』（1890）、『舞姫』（1890）、『文づかい』（1891）、『そめちがへ』（1897）の4作品は他の小説に

比べ文語助動詞の出現率が高い。またこれら4作品に加え、森鷗外の最初の歴史小説である『興津弥五右衛門の遺書』も文語助動詞の出現率が高い。次に、これらの5作品を除き、森鷗外の小説における文体を考えると、助詞の「に」「の」「は」及び「た。」や「たい。」という文末表現の出現率が継時的に増加し、反対に助詞の「の」「と」「て」及び「ない。」「られる。」「です。」といった文末表現の出現率が継時的に減少するという傾向が認められる。この計量的な事実は、森鷗外の文体的特徴として指摘できるとともに、鷗外作品の成立過程を検討する際に一つの視点を提供するものと考えられる。

注

- (1) 村上征勝 (2002) 『文化を計る——文化計量学序説』朝倉書店
- (2) 安本美典 (1977) 「現代の文体研究」、『岩波講座 日本語』10、pp. 395-423、岩波書店
- (3) 土山玄・村上征勝 (2012) 「語の使用頻度の計量分析による宇治十帖他作者説の検討」『研究報告人文科学とコンピュータ (CH)』2012-5、pp. 1-8.
- (4) 金明哲 (1994) 「説点の打ち方と文章の分類」『計量国語学』19-7、pp. 317-330
- (5) 金明哲 (2009) 「文章の執筆時期の推定——芥川龍之介の作品を例として——」『行動計量学』36-2、pp. 89-103