

# 『枕草子』長単位データを用いた相の類の分析

富士池 優美 (国立国語研究所 コーパス開発センター) †

## The Adjectives and the Adverbs in *Makura-no-Soushi*: An Analysis Based on Long-unit-word

Yumi Fujiike (Center for Corpus Development, NINJAL)

### 1. はじめに

国立国語研究所では現在、「日本語歴史コーパス」の準備が進められている。日本語歴史コーパスの形態論情報については、言語単位として「短単位」「長単位」の2種類を採用し、それぞれに代表形・品詞等の情報を与える<sup>1</sup>。

富士池 (2012b) では『枕草子』長単位データを用いて、随想的章段・類聚的章段・日記的章段の章段分類を別ジャンルの文章と見立て、品詞比率の比較を行った。その結果、名詞率・形容詞率は類聚的章段で高く、日記的章段で低く、随想的章段はその中間となった。また、副詞率は章段分類の別なくほぼ一定であった。一般に形容詞率や副詞率といった相の類の比率は名詞率と負の相関関係が認められるものだが、負の相関関係が認められないことが確認された。これは形容詞や副詞といった相の類が『枕草子』で多用されていることの現れと考えられる。

本発表では『枕草子』の相の類に注目し、どのように用いられているのか、実態を探る。

### 2. 問題の所在

#### 2. 1 富士池 (2012b)

まず、『枕草子』長単位データを用いて、随想的章段・類聚的章段・日記的章段の章段分類を別ジャンルの文章と見立て、品詞比率の比較を行った富士池 (2012b) について、概要を示す。

##### (1) 調査対象

調査にあたり、準備中の『枕草子』長単位データを用いた。ここで、言語単位の概要を説明したい。日本語歴史コーパスでは言語単位として、短単位・長単位の2種類を採用している<sup>2</sup>。短単位は言語の形態的側面に着目して規定された言語単位であり、意味を持つ最小の単位 (最小単位) を規定した上で、その最小単位を短単位認定規定に基づいて結合さ

---

† yfujiike@ninjal.ac.jp

<sup>1</sup> 短単位データについては「日本語歴史コーパス 平安時代編 (先行公開版)」として、『枕草子』を含む仮名文学作品 10 作品を公開中。 [http://www.ninjal.ac.jp/corpus\\_center/chj/](http://www.ninjal.ac.jp/corpus_center/chj/)

<sup>2</sup> 「日本語歴史コーパス」中古和文の単位認定基準は「現代日本語書き言葉均衡コーパス」で用いた認定基準を中古和文に対応できるように変更・拡張したものである。短単位の認定基準については小椋秀樹・須永哲矢 (2012)、長単位の認定基準ほか概要については富士池 (2012a) を参照。

せる（または結合させない）ことにより、短単位を認定する。これに対して長単位は構文的側面に着目して規定された言語単位であり、文節を規定した上で、文節を長単位認定規定に基づいて自立語と付属語に分割することにより、長単位を認定する。例えば「小野-小町」「たへ-がたし」「うつくし-げ」は短単位では「-」の位置で切り離されるが、長単位では一まとまりとなる。短単位と長単位は同じ品詞体系を持つ。

章段分類の認定については、池田亀鑑（1963）の分類に基づき、「随想」「随筆」となっているものを随想的章段、「分類」となっているものを類聚的章段、「日記」となっているものを日記的章段とした<sup>3</sup>。これらが混在する章段もある。新編全集で「一本」とある章段については分類が不明であったため、調査対象からは除外した。

## （2）調査結果

従来行われてきたテキストを特徴付ける指標として、主要な品詞比率を章段分類別に求めたほか、樺島忠夫・寿岳章子（1965）で提案された品詞比率に基づく指標である MVR（後述）を用いて、各章段分類の文体的特徴の把握を試みた。資料規模があまりに小さい章段と章段分類が混在する章段は除外し、随想的章段・類聚的章段・日記的章段のいずれかから延べ語数の多い 50 章段を取り出し、調査対象とした。50 章段で全体の約 6 割に当たる。

### 品詞比率

図 1 に示した名詞率・動詞率・形容詞率・副詞率は「各品詞の延べ語数／総語数」である。「Z」は随想的章段、「R」は類聚的章段、「N」は日記的章段を示し、ひげの下端が最小値、上端が最大値、箱の下端が第 1 四分位点、横線が中央値、上端が第 3 四分位点、ひげから外れた点は軽度の外れ値である。

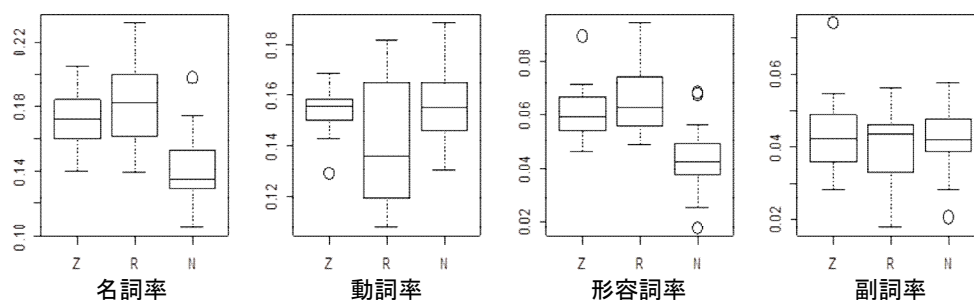


図 1 章段分類別の品詞比率

図 1 から、日記的章段の名詞率が他の章段分類と比較して低い様子が見てとれる。また、類聚的章段では章段による名詞率のばらつきが大きい。動詞率については、類聚的章段における章段によるばらつきの大きさが名詞率以上に目立つ。日記的章段の動詞率がやや高くなっている。形容詞率は名詞率と似た傾向を示しており、日記的章段の形容詞率は他の章段分類と比較して低い。副詞率はばらつきの差がややあるものの、章段分類による大きな差異は見られなかった。小磯ほか（2008）や富士池ほか（2010）は現代書き言葉を対象

<sup>3</sup> 章段分類の認定に関しては諸説あり、定番と言えるような基準がないようである。今回の調査では、内容を重視し、池田（1963）に基づくこととした。

とした調査であるが、名詞率と動詞率・形容詞率・副詞率は負の相関にあった。これに対して富士池（2012b）の調査結果では形容詞率が名詞率と正の相関を示しており、副詞率も負の相関は見られなかった。

## MVR

名詞の比率は文章の特質を表し、名詞の比率に応じて他の品詞もある傾向を持って変化する、つまり文章のジャンルによって品詞の割合が決定されると考えられる。自立語について、品詞をその機能によって、体（名詞類）・用（動詞）・相（形容詞・形状詞・副詞・連体詞）・他<sup>4</sup>の四つに分類したとき、体の類と、用・相それぞれの類の関係を見るにあたり、樺島・寿岳（1965）は MVR という指標を提案した。MVR は「 $MVR = 100 \times \frac{\text{相の類の比率}}{\text{用の類の比率}}$ 」の式で表される。体の類の比率（以下、名詞率とする）は、一般に要約的な文章で大きく、描写的な文章で小さいとする。また、MVR の値が大きいほどありさま描写的であり、MVR の値が小さいほど動き描写的と考えられるとし、名詞率と MVR の組み合わせから以下のような文体的特徴が見出せるとした<sup>5</sup>。

名詞率：大、MVR：小	要約的な文章
名詞率：小、MVR：大	ありさま描写的な文章
名詞率：小、MVR：小	動き描写的な文章

この指標を用いて、品詞比率から見る文体的特徴の把握を試みた。横軸に名詞率、縦軸に MVR を取った散布図を図 2 に示す。左下から右上にかけて、概ね日記的章段（黒）、随想的章段（白）・類聚的章段（灰）の順で並んでいる様子が見てとれる。文体的特徴としては、日記的章段は名詞率が小さく MVR も小さい「動き描写的」な文章、随想的章段は日記的章段と重なるが MVR がやや高く「ありさま描写的」な文章であり、類聚的章段は樺島・寿岳（1965）では示されていない、名詞率が大きく MVR も大きい文章という

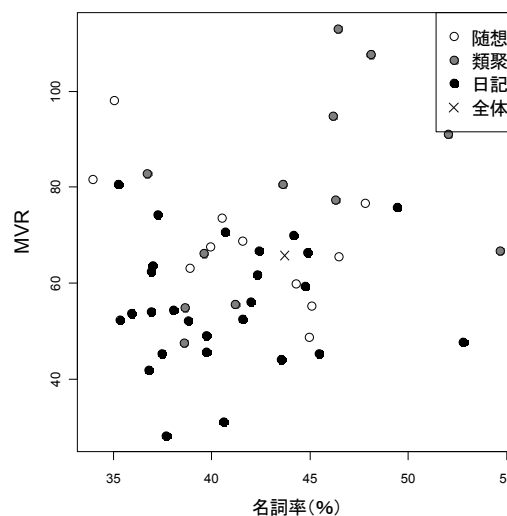


図 2 名詞率と MVR (『枕草子』章段分類別)

ことになる。類聚的章段は名詞率が高く、動詞率が低く、形容詞率が高いということで文体的特徴を「動き描写の少ない」文章としておく。「×」は『枕草子』全体の名詞率と MVR を示すものであり、ほぼ中心に位置している。これまで『枕草子』の品詞比率と考えられてきたものは、名詞率・MVR 共にばらつきがあるものが集約された結果であると言えるだ

<sup>4</sup> 樺島忠夫（1950）の分類による。樺島・寿岳（1965）にならい、「他」については数が少ないため省略した。

<sup>5</sup> 樺島・寿岳（1965）p.30-36。単位の長さについては言及がないが、品詞の説明に「大きな区分として自立語と附属語との二種に分かれる」（p.27）とあることから、文節に基づく長い系列の単位、つまり長単位相当と推測される。

ろう。

このような『枕草子』の各章段の名詞率・MVRの分布は中古和文の中でどのような位置付けになるのだろうか。他の作品と比較を試みるため、参考程度ではあるが、図2に『古典対照語い表』に基づく同様の散布図を重ね合わせたものを図3に示す。「\*」が『古典対照語い表』に基づくもの（作品の略称を付した）で、白・灰・黒・×の点が図2の点である。『古典対照語い表』所収の他作品と比較すると『枕草子』は名詞率が44.2%、MVRが69.5となっており、名詞率が小さくMVRが高い「ありさま描写的」な文章と見える。『蜻蛉日記』は名詞率が低くMVRもやや低めで、『枕草子』の日記的章段と同傾向に見え、似た品詞構成から成る可能性がある。その一方で類聚的章段のようにMVRが高いものは『古典対照語い表』所収作品には見られなかった。各章段分類と他作品との関係については、本来、単位を揃えて慎重に検討すべき問題であり<sup>6</sup>、他作品の長単位データ整備後の課題としたい。

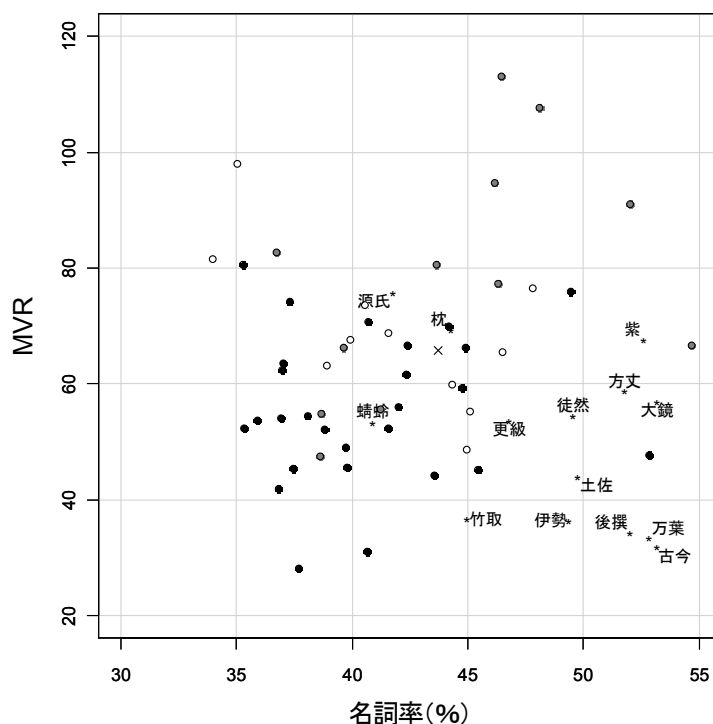


図3 名詞率とMVR (図2+『古典対照語い表』)

## 2. 2 問題の所在

『枕草子』はこれまで、名詞率に対してMVR、つまり動詞に対する形容詞類の割合が高いと考えられてきたが、富士池(2012b)では名詞率・MVRが共に高い類聚的章段、名詞率・MVRが共に低い日記的章段、名詞率・MVRが共に中間的(他作品と比較するとMVRが高い可能性がある)随想的章段という、異なる品詞比率を持つ文章の集合体であることが明らかになった。特に類聚的章段は、『古典対照語い表』所収の他作品との比較の限りでは、他に類を見ない品詞比率に見える。MVRは動詞に対する形容詞類の割合であるが、富士池(2012b)では動詞率は名詞率と負の相関があり、一般的な傾向を示していた。そこで、本稿では相の類(形容詞・形状詞・副詞)に注目し、どのように用いられているのか、章段分類別に見ることで実態を探る。

<sup>6</sup> 『古典対照語い表』の単位は文節に基づくものであり「長単位」に近いものであるが、一部接辞を認めていないため(「御」を冠する語はそれを除いた形式を1単位として認める等)、短単位と対応するものもある。

### 3. 『枕草子』相の類の分析

#### 3. 1 調査概要

調査にあたっては『枕草子』全体の対象とした長単位データを使用する<sup>7</sup>。随想・類聚・日記のいずれかの要素が混在する章段を「混在」章段とし、新編全集で「一本」とある章段については「不明」とした。ここで、『枕草子』の資料規模として、各章段分類の章段数・延べ語数（長単位）を表1で確認しておく。

表1 各章段分類の章段数・延べ語数

	随想	類聚	日記	混在	不明	計
章段数	97	140	50	12	28	327
延べ語数	17643	15094	34604	6392	1497	75230

『枕草子』における相の類の用いられ方を見るにあたって、章段分類別に相の類の頻度・比率、名詞・動詞・形容詞・形状詞<sup>8</sup>・副詞の相関関係から、相の類の各品詞が各章段分類でどのように用いられているのかを見る。また、高頻度語やコレスポネンス分析（対応）を通して、各章段分類で相の類のどのような語が用いられているのかを確認する。

#### 3. 2 調査結果

##### (1) 品詞比率

章段分類別相の類の頻度と比率を表2<sup>9</sup>に示す。比率は「各品詞の延べ語数／総語数」で、2. 1節の図1には表示されていない平均値を示したものである。図1と表2から日記的章段は相の類の比率が随想・類聚と比較して低く、形容詞においてそれが顕著であることがわかる。また、類聚で副詞の中央値は高いが平均比率は低く、他の章段分類と比べて副詞の比率が低い章段が多い様子がうかがえる。このように、相の類の中でも章段分類により品詞構成に差がある様子が確認できる。

表2 各章段分類の相の類の頻度・比率

	随想	類聚	日記	計
形容詞	907	820	1293	3435
	5.14%	5.43%	3.74%	4.57%
形状詞	236	184	246	774
	1.34%	1.22%	0.71%	1.03%
副詞	710	520	1431	2969
	4.02%	3.45%	4.14%	3.95%
計	1853	1524	2970	7178
	10.50%	10.10%	8.58%	

##### (2) 品詞比率の相関関係

2. 1節では名詞率とMVRの相関を見たが、MVRは動詞と相の類の比率を見る指標であり、動詞及び相の類に属する各品詞の比率が集約されてしまい、前項(1)で見たような章段分類による品詞構成の差が捉えきれない。そのためここでは各品詞の相関関係を見る。図4に普通名詞・動詞・形容詞・形状詞・副詞それぞれの比率（自立語中の割合）の相関を示す。図4の点はそれぞれ随想的章段（白）・類聚的章段（灰）、日記的章段（黒）、混在章段・不明（濃い灰）である。

<sup>7</sup> 富士池（2012b）は2012年8月時点の長単位データを使用したが、本稿では2012年12月時点のものを使用した。全体を対象としたため極端な結果を示す章段もある。例えば第19段「たちは」の場合、「たちは たまつくり。」のみであり、自立語について見たときの品詞比率は名詞率100%となる。

<sup>8</sup> 形状詞は形容動詞語幹に相当する。

<sup>9</sup> 計は随想・類聚・日記のほか、混在・不明を含む。

普通名詞との相関に注目すると、動詞は左上に類聚（灰）、右下に日記（黒）・随想（白）となっており、負の相関があるように見えるが、形容詞・形状詞・副詞はそれぞれ異なる様相を見せている。形容詞は類聚（灰）が負の相関を示しているが全体的に比率が高く、日記（黒）はある一定の低めの範囲に固まり、随想（白）は一部が正の相関を示しているように見える。形状詞は名詞の割合とは関係なくある一定の範囲にある。副詞は類聚（灰）が形容詞と同様に負の相関を示しており、日記（黒）・随想（白）はあ

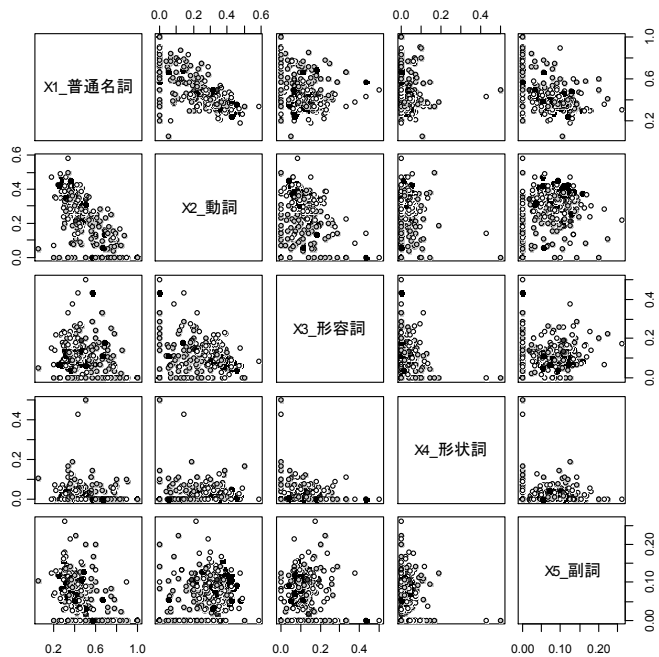


図4 普通名詞・動詞・形容詞・形状詞・副詞の相関

る一定の範囲にあるが、形状詞と比べるとばらつきが大きい。ここから、形容詞・副詞において名詞と負の相関関係が見出せなかったのは、日記（黒）・随想（白）の品詞構成によるものであることが確認できる。「猫は、上のかぎり黒くて、腹いと白き。」（第 50 段・猫は）など類聚的章段に多く、モノとその描写という性格を持ち動詞を必ずしも必要としないことから、名詞率に関わらず形容詞率が高い傾向にある。また、「冬は、いみじう寒き。夏は、世に知らず暑き。」（第 114 段・冬は<sup>10</sup>）、「坤元録の御屏風こそ、をかしうおぼゆれ。漢書の屏風はおぼしくぞ聞えたる。月次の御屏風もをかし。」（第 278 段・坤元録の御屏風こそ、をかしうおぼゆれ）のような比較的短い随想的章段で、名詞率と形容詞率が共に高くなっていた。その一方で日記的章段を中心に名詞率・形容詞率が共に低い章段がある。「大進生昌が家に、宮の出でさせたまふに、東の門は四足になして、それより御輿は入らせたまふ。」とはじまる第 6 段などが代表的なものであるが、日記的章段は他章段と比べて名詞率が低く、動詞率が高い傾向にある。『枕草子』中に、名詞率・形容詞率が共に高い章段と、共に低い章段が混在するため、結果として名詞率と形容詞率が負の相関を示さなかったことがわかる。

### （3）章段分類別頻度上位語

ここまで比率を見てきたが、ここでは各章段分類でどのような語が使われているのかを確認する。表 3 に形容詞・形状詞・副詞の章段分類別頻度上位 10 語を示す。全章段分類頻度上位 10 語に含まれない語に網掛けを付した。網掛けした語を見ると、「若し」は類聚に 7

<sup>10</sup> 池田（1963）では随想的章段と分類されていたが、「～は」型であり類聚的章段と類似の品詞構成を持つものと考えられる。章段分類には再考の余地があるものとする。

例、日記に 13 例あり、随想にやや多い程度である。一方で「疾し」のように全 63 例中 43 例が日記と出現章段分類に偏りがあるものもあり、比率だけではなく語に関しても、章段分類による差がある様子がうかがえる。

表 3 形容詞・形状詞・副詞の章段分類別頻度上位 10 語

形容詞				形状詞				副詞									
随想	907	類聚	820	日記	1293	随想	236	類聚	184	日記	246	随想	710	類聚	520	日記	1431
可笑し	166	可笑し	93	いみじ	159	衰れ	19	衰れ	29	衰れ	16	いと	215	いと	172	いと	250
いみじ	86	無し	61	可笑し	109	清気	17	流石	8	然様	8	然	45	然	27	然	95
良し	45	いみじ	55	無し	105	鮮やか	8	可笑し気	7	漫	8	猶	39	猶	21	唯	75
無し	44	憎し	32	めでたし	66	然様	8	更	7	無下	7	少し	31	唯	19	猶	75
白し	26	めでたし	29	疾し	43	忍びやか	7	憎気	7	大き	6	唯	27	必ず	14	皆	66
めでたし	25	良し	25	怪し	37	細やか	7	清気	6	可笑し気	6	又	17	少し	14	如何で	49
憎し	18	白し	20	憎し	32	可笑し気	6	大き	4	清気	6	数多	15	又	14	え	49
近し	16	近し	18	良し	29	更	6	汚気	4	顕証	6	え	15	え	13	又	45
若し	16	多し	16	多し	27	大き	5	心殊	4	まめやか	6	皆	14	良く	11	少し	39
多し	15	怪し	14	近し	24	艶やか	5	唯	4	密か	6	まいて	13	未だ	10	など	39
高し	15					無下	5	徒然	4					皆	10		

#### (4) コレスポンデンス分析

ここでは、コレスポンデンス分析（対応分析）で章段分類と相の類の高頻度語との対応を確認する。相の類の頻度上位 20 語のコレスポンデンス分析結果の散布図を図 5 に示す。分析には統計分析パッケージ R の ca パッケージの ca 関数を用いた。第 1 次元の寄与率は 80.33%、第 2 次元の寄与率は 19.67%であった。図 5 では第 1 次元の正の方向に日記、負の方向には随想・類聚が布置されており、第 1 次元は日記とその他を分ける軸、第 2 次元は随想と類聚を分ける軸と見ることができる。第 1 次元を見ると、正の方向に、日記と共に「如何に」「如何で」「猶」「然」「唯」「皆」「又」「え」といった副詞、「いみじ」「怪し」「めでたし」といった形容詞が布置された。副詞はほぼ正の方向に布置されており、動き描写にあたり用いられる語と言える。例外である「いと」は程度の甚だしい様を表すものであり、主に形容詞が布置された負の方向に共に布置され、ありさま描写に用いられたと見ることができる。第 2 次元を見ると、正の方向に随想及び「可笑し」「良し」が、負の方向に類聚及び「衰れ」「憎し<sup>11</sup>」が布置された。「めでたし」「可笑し」「良し」「衰れ」は共に肯定的な評価を表す

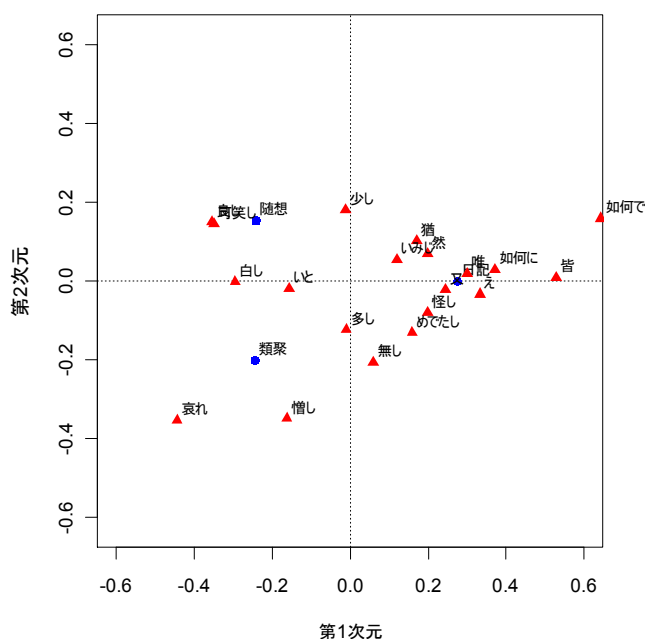


図 5 相の類頻度上位 20 語の散布

<sup>11</sup> 「憎し」は第 26 段「にくきもの」での多用が影響したものと考えられる。「形容詞+もの」型章段は多くあるが、その形容詞が章段中に多用されることは少なく、「にくきもの」は例外と言える。

語であるが、「めでたし」は日記を、「可笑し」「良し」は随想を、「哀れ」は類聚をそれぞれ特徴付ける語と考えられる。

#### 4. 終わりに

『枕草子』長単位データを用いて品詞構成とその相関関係について章段分類別に見たところ、日記的章段で動詞率が高いこと、類聚的章段や短い随想的章段で動詞率が極端に低く名詞率・形容詞率が共に高いことが確認された。『枕草子』中に、名詞率・形容詞率が共に高い章段と、共に低い章段が混在するため、結果として名詞率と形容詞率が負の相関を示さなかったと言えるだろう。また、相の類の頻度上位語とそのコレスポネンス分析から、日記的章段で副詞が動き描写に用いられ、随想・類聚的章段では形容詞がありさま描写に用いられたことや、肯定的な評価を表す形容詞類の中でも章段分類により差がある様子がうかがわれた。

#### 付 記

本稿は、国立国語研究所共同研究プロジェクト「通時コーパスの設計」（プロジェクトリーダーは近藤泰弘客員教授）の成果の一部である。また、用いた新編全集『枕草子』の電子テキストは、小学館から上記プロジェクトのために提供されたものである。

#### 文 献

池田亀鑑（1963）『全講枕草子』至文堂

小椋秀樹・須永哲矢（2012）「中古和文 UniDic 短単位規定集」、平成 21（2009）—平成 23（2011）年度科学研究費補助金基盤研究（C）「和文系資料を対象とした形態素解析辞書の開発」研究成果報告書 2（[http://dl.dropbox.com/u/73297026/report/unidic-EMJ\\_rulebook2012.pdf](http://dl.dropbox.com/u/73297026/report/unidic-EMJ_rulebook2012.pdf) よりダウンロード可能）

樺島忠夫（1950）「類別した品詞の比率に見られる規則性」『国語国文』24-6

樺島忠夫（1988）『日本語はどう変わるか —語彙と文字—』岩波書店

樺島忠夫・寿岳章子（1965）『文体の科学』綜芸舎

小磯花絵・小木曾智信・小椋秀樹（2008）「短単位情報に基づくジャンル間の文体に関する分析」『特定領域研究「日本語コーパス」平成 20 年度全体会議予稿集』、pp.99-106

富士池優美・小西光・小椋秀樹・小木曾智信・小磯花絵（2010）「長単位に基づく媒体・カテゴリ間の品詞比率に関する分析」、特定領域「日本語コーパス」平成 22 年度公開ワークショップ予稿集、pp. 273-280

富士池優美（2012a）「中古和文における長単位の概要」、第 2 回コーパス日本語学ワークショップ予稿集、pp.51-58

富士池優美（2012b）「枕草子の語彙——章段分類に注目して——」、第 101 回国語語彙史研究会（2012 年 9 月 29 日）口頭発表資料

宮島達夫編（1971）『古典対照語い表』笠間索引叢刊 4、笠間書院