

『源氏物語』第二部の文体的特徴についての計量的な検討

土山 玄（お茶の水女子大学）

平安時代に著された『源氏物語』の全54巻は現行の巻序とは異なる順序で成立したという見解があり、第一部と称される33巻については従来から議論されている。しかし、第二部に関してはおよそ検討が加えられていない。そこで、本研究においては第二部を採り上げ、第二部を特徴づける単語の抽出を行った。分析ではカイ二乗分析、および残差分析を行った。その結果、「こと」や「ひと」といった出現頻度の高い語の一部が第二部に偏って頻出していることが明らかになった。

A Quantitative Analysis on the Formation of the Second Part of “The Tale of Genji”

Gen Tsuchiyama (Ochanomizu University)

The formation of “The Tale of Genji,” which is a Japanese classic comprising 54 volumes, has not been well understood. We have investigated the formation of the second part. Therefore, in this study, we picked up the second part and extracted the words that characterize the second part. Chi-square analysis and residual analysis were performed in the analysis. As a result, it became clear that some of the frequently-appearing words such as “koto” and “hito” appear frequently in the second part.

1. はじめに

『源氏物語』は平安時代に成立した長編物語である。この物語は54巻から構成され、一般に『源氏物語』は三部に分割されると考えられている[1]。第一部には第1巻「桐壺」から第33巻「藤裏葉」までが属し、第二部は第34巻「若菜上」から第41巻「幻」の8巻が、第三部は第42巻「匂宮」から第54巻「夢浮橋」の13巻が属するとされる。この第一部の33巻には成立順序が現行の巻序と異なるとする見解があり、これら33巻は「紫上系」と「玉鬘系」という2つの群に分類されると論じられている[2]。

また、第三部においては、第45巻「橋姫」、第49巻「宿木」、第53巻「手習」の3巻の冒頭は共通して「そのころ」という発語によって開始されることから、第三部は構造上4つのブロックに分類される可能性が指摘されている[3]。

詳細は後述するが、これらの第一部及び第三部における各巻の関係や他の部との関係は計量的な観点から検討が加えられている。しかし、第二部においても、「玉鬘系」の人物が登場するのは第34巻「若菜上」、第35巻「若菜下」、第36巻「柏木」の3巻のみであるという指摘[4]や、内容の考証に基づき第38巻「鈴虫」の後記説[5]および第39巻「夕霧」の後記説[6]が報告されている。

そこで、本研究では第二部を主な研究対象とし、第一部及び第三部と第二部を比較したときに、第二部を特徴づける単語を計量的な観点から明らかにすることが目的である。また、分析では名詞、動詞、形容詞、副詞の単語を採り上げる。

2. 関連研究

第一部の成立過程については、すでに計量的な検討が加えられている。村上・今西 (1999) は『源氏物語』の全文を電子化したテキストデータを用い、多変量解析を行った初の本格的な研究である[7]。村上・今西 (1999) では、各巻の助動詞の出現率を求め数量化 III 類を行っている。分析の結果、紫上系は玉鬘系よりむしろ第二部の8巻と助動詞の出現傾向は類似していることを明らかにした。従って、第1巻「桐壺」を起筆の巻と仮定するのであれば、紫上系が成立した後に第二部が成立し、その後に玉鬘系が成立した可能性を論じている。

小野 (2015) は村上・今西 (1999) と同一のデータを用いて、分散安定化変換を行い、階層的クラスタ分析を用いて『源氏物語』の成立過程について検討を加えている。分析の結果、『源氏物語』の諸巻は紫上系及び第二部、玉鬘系及び宇治十帖という2群に分類されることを報告した[8]。

また、土山 (2016) では助動詞以外の品詞についても巻別の出現率を求め主成分分析を行っている。その結果、紫上系と玉鬘系との間において出現傾向が異なっている品詞は助動詞のみであることを指摘し、紫上系の諸巻は玉鬘系の諸巻に比べて使役・尊敬の助動詞が頻出していることを明らかにした。またこれに加えて、動詞の未然形に接続する受身・尊敬・可能・自発の助動詞である「る」は紫上系に、「らる」は「玉鬘系」に多く出現することを報告した[9]。

次いで、土山・村上 (2014) は『源氏物語』の第三部を採り上げ、計量的な手法を用いて分析を行った。分析では、品詞の構成比率・語の頻度・語の長さの3つの項目を用いている。分析の結果、『源氏物語』の第三部の13巻は、匂宮三帖(第42巻「匂宮」から第44巻「竹河」)・宇治十帖前半5巻(第45巻「橋姫」から第49巻「宿木」)・宇治十帖後半5巻(第50巻「東屋」から第54巻「夢浮橋」)という計量的な特徴の異なる3つのグループに分類されることを報告している[10]。

3. データ

本研究では青表紙本系の大島本を主な底本とする『源氏物語語彙用例総索引 自立語編』[11]及び『源氏物語語彙用例総索引 付属語編』[12]を電子化したテキストデータを分析に利用した。『源氏物語語彙用例総索引』の『自立語編』及び『付属語編』はどちらも『源氏物語』の本文を『源氏物語語彙用例総索引』の著者らの手によって『源氏物語大成 索引篇』[13]の単語認定基準に準拠し、形態素解析を行ったものである。従って、形態素解析器を用いていない。本研究で用いたテキストデータではすべての語に、表記形、表記形仮名読み、終止形、終止形仮名読み、品詞コード、活用形コード、意味コードなどの情報が付与されている。このテキストデータは村上・今西 (1999)、土山 (2014)、土山 (2016) で使用したデータと同一のデータである。

4. 分析

本研究では第二部に偏って出現する単語、すなわち特徴語を抽出するためにカイ二乗検定を行った。カイ二乗検定とは2つの項目の間における独立性の検定に使用される手法であり、2つの項目間に関連があるのであればp値が小さくなる。特徴語の抽出に際しては『源氏物語』に出現するすべての単語を品詞別及び部ごとに集計し、ある語がその語を除くすべての語の出現傾向と同様であるのか検定した。従って、カイ二乗検定を行うときには表1のような2行3列のクロス表を作成した。表1のようなクロス表は品詞ごとに作成し、品詞別に分析を行った。なお、表中のa, b, c, d, e, fはそれぞれ頻度をあらわす。

表1 カイ二乗検定を行うときのクロス表

	第一部	第二部	第三部
単語Aの頻度	a	b	c
それ以外の単語の合計頻度	d	e	f

次いで、本研究では第二部に特徴的に出現している単語を抽出するために、カイ二乗検定の下位

分析である残差分析を行った。この残差分析は2行2列より大きいクロス表に対して行う分析であり、表中のa, b, c, d, e, fの観測度数のうちいずれが期待度数と有意に異なるかを検討する分析手法である。本研究では残差分析を行い、調整済み残差を求め、有意水準が5%となるように特徴語を抽出した。すなわち、第二部における単語Aの調整済み残差の値の絶対値が1.96より大きい単語を特徴語であるとして抽出した。また、カイ二乗検定を行う際のクロス表に、期待度数に5以下となるものが1つでも含まれる場合はカイ二乗検定が不正確になるため、そのような単語は本研究における分析から除外した。

まず、名詞を対象としてカイ二乗検定を行い、第二部において偏って出現する単語を抽出した。表2はその結果である。表2は残差分析を行った結果であり、第二部の調整済み残差の絶対値が1.96以上となった単語を『源氏物語』全54巻における総度数の降順に並び替えた上位15単語である。また、表中のp値はカイ二乗検定で求められたp値である。

表2 名詞の特徴語

	第一部	第二部	第三部	総度数	p値	第二部の調整済み残差
コト	0.064	0.074	0.071	4496	0	3.095
ヒト	0.048	0.043	0.067	3463	0	-5.703
ミ	0.008	0.013	0.010	630	0	4.472
キミ	0.008	0.006	0.009	508	0	-3.339
トコロ	0.008	0.004	0.008	452	0	-4.513
ワザ	0.004	0.007	0.006	318	0	3.192
ナニゴト	0.003	0.005	0.005	269	0.001	1.988
ココチ	0.003	0.005	0.005	266	0	2.244
キン	0.003	0.010	0.002	251	0	13.089
ヨノナカ	0.003	0.005	0.004	247	0.014	2.477
カギリ	0.003	0.005	0.003	231	0	4.076
カタチ	0.004	0.002	0.003	218	0	-3.719
トノ	0.003	0.001	0.003	189	0	-4.349
ココロザシ	0.002	0.003	0.002	168	0.068	2.314
オヤ	0.003	0.001	0.002	155	0.006	-2.609

「こと」や「ひと」は総度数が3000を超えることから『源氏物語』の全巻を通じて数多く出現する単語であると言えるが、有意水準5%で「こと」は第二部に偏って出現しており、その一方で「ひと」は第二部に多く用いられていない。なお、これらの単語の頻度を対象にして行ったカイ二乗検定のp値も0.05を下回っている。

また、上掲の2語に比べると総度数は少ないが「みん」の調整済み残差は13.089となっており、これは名詞を対象とした分析では調整済み残差の絶対値が最も大きい単語である。

図1は「こと」、図2は「ひと」、図3は「みん」の各部における出現率を棒グラフにして可視化

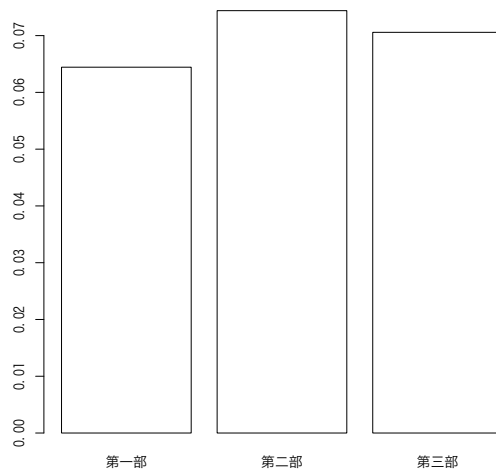


図1 「こと」の出現率

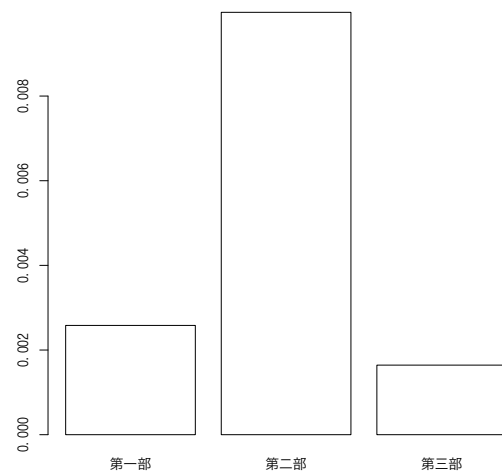


図3 「みんな」の出現率

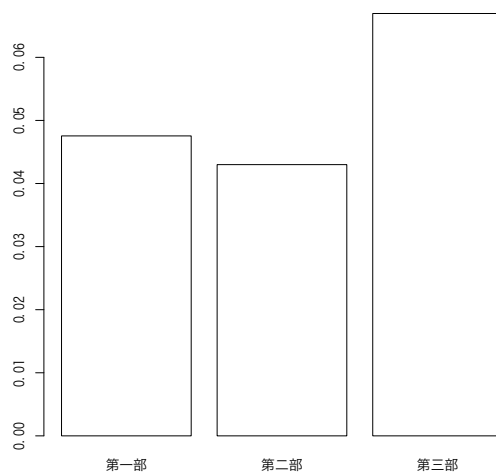


図2 「ひと」の出現率

表3 動詞の特徴語

	第一部	第二部	第三部	総度数	p値	第二部の調整済み残差
アリ	0.055	0.065	0.059	3686	0	3.451
オモフ	0.035	0.029	0.046	2357	0	-5.475
キコユ	0.022	0.024	0.019	1339	0.011	2.143
イフ	0.013	0.011	0.022	995	0	-4.637
シル	0.008	0.007	0.010	532	0.007	-2.467
ワタル	0.006	0.010	0.004	374	0	6.329
ソフ	0.005	0.006	0.004	311	0.036	2.233
マウス	0.003	0.002	0.004	201	0.001	-2.512
カハル	0.003	0.004	0.002	159	0.002	3.028
カヘル	0.002	0.001	0.003	152	0.042	-2.37
ウラム	0.002	0.002	0.003	150	0.038	-2.099
キコシメス	0.002	0.004	0.002	137	0	4.402
ニル	0.002	0.001	0.003	133	0.011	-2.82
ユルス	0.002	0.003	0.002	129	0.064	1.987
カタラフ	0.002	0.001	0.003	128	0.003	-2.433

したグラフである。出現率は部ごとに集計し、出現率の分母は各部の感動詞の延べ語数となる。カイ二乗検定及び残差分析の結果の通り、「こと」は第二部の出現率が最も高く、「ひと」は第三部における出現率が最も高いが第二部の出現率は第一部よりも低いことが分かる。図3では、「みんな」の第二部における出現率が第一部や第三部の出現率よりも顕著に高いことが示されている。

次に、動詞について分析を行った。表4が分析の結果であり、表2と同様に総度数で降順に並び替えた上位15語である。動詞の分析では名詞のように調整済み残差の絶対値が10を超える単語は存在せず、「わたる」の調整済み残差の絶対値が最大となり、「おもふ」の調整済み残差の絶対値がこれに次いで大きくなる。「あり」及び「おもふ」、そして「わたる」の部ごとの出現率は

順に図4、図5、図6に示す通りである。これらの図より、「わたる」という動詞が顕著に第二部に頻出していることが分かる。

次いで、形容詞について検討を加えた。表4が分析の結果であり、降順に並び替えた上位15語である。形容詞の分析では「をかし」の調整済み残差の絶対値が最大となる。表4における上位3語はいずれもカイ二乗検定のp値が0.05を下回っており、調整済み残差の絶対値も他の形容詞に比べ大きいと言える。表4における上位2単語である「をかし」及び「あやし」の出現率を可視化すると図7、図8に示す通りとなる。図7より「をかし」という形容詞は顕著に出現率が低いことが分かる。その一方で「あやし」は第三部の出現率が高く第一部と第二部の出現率は大きく乖離していないことが分かる。

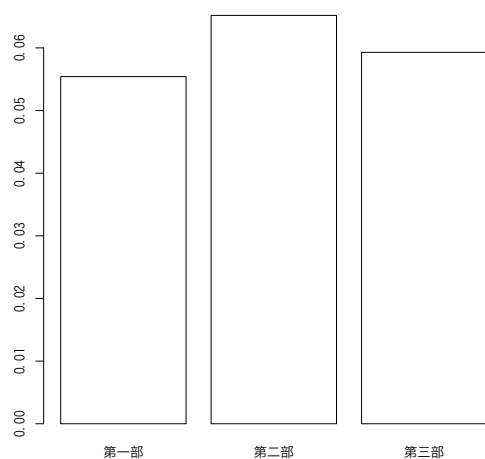


図4 「あり」の出現率

表4 形容詞の特徴語

	第一部	第二部	第三部	総度数	p値	第二部の調整済み残差
ヲカシ	0.026	0.014	0.027	536	0	-5.152
アヤシ	0.020	0.016	0.032	508	0	-3.65
フカシ	0.014	0.026	0.019	402	0	4.473
ココログルシ	0.012	0.016	0.013	288	0.126	1.981
ウレシ	0.009	0.006	0.013	213	0.001	-2.89
メデタシ	0.011	0.007	0.008	211	0.012	-1.978
アサマシ	0.009	0.007	0.010	201	0.092	-2.015
ワリナシ	0.008	0.005	0.008	161	0.13	-2.003
アリガタシ	0.005	0.008	0.005	121	0.013	2.93
ウツクシ	0.005	0.007	0.004	118	0.061	1.968
ハシタナシ	0.006	0.002	0.005	104	0.002	-3.273
イカメシ	0.003	0.006	0.002	76	0.001	3.208
オモシ	0.003	0.005	0.003	75	0.048	2.432
メザマシ	0.004	0.005	0.001	68	0.001	2.687
ネタシ	0.004	0.002	0.003	66	0.096	-2.019

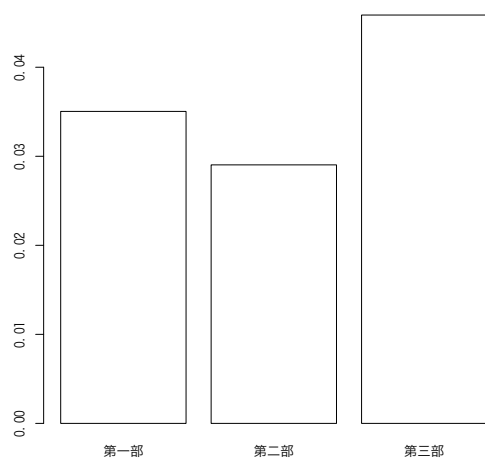


図5 「おもふ」の出現率

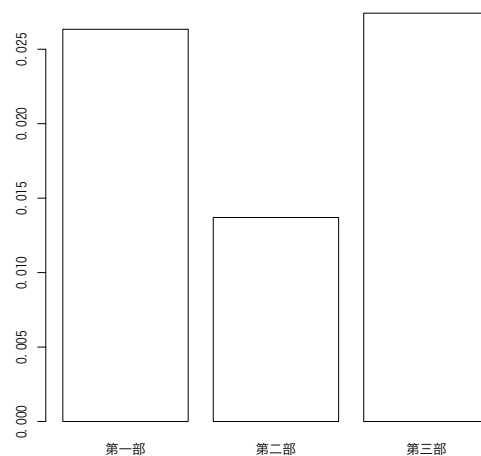


図7 「をかし」の出現率

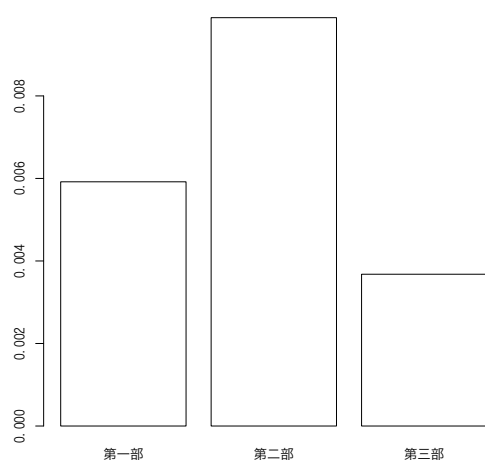


図6 「わたる」の出現率

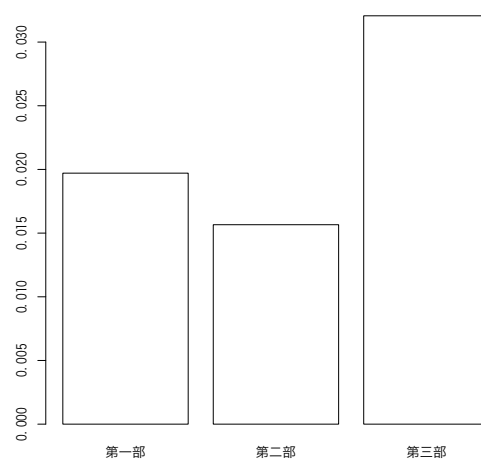


図8 「あやし」の出現率

表 5 副詞の特徴語

	第一部	第二部	第三部	総度数	p値	第二部の調整済み残差
カク	0.058	0.082	0.071	1054	0	3.778
ナホ	0.047	0.061	0.054	818	0.007	2.626
イトド	0.021	0.016	0.026	338	0.006	-2.331
イカガ	0.017	0.011	0.018	253	0.031	-2.613
イカデ	0.012	0.007	0.016	187	0.002	-2.971
ヲサヲサ	0.005	0.008	0.004	84	0.059	2.318
トミニ	0.003	0.006	0.003	56	0.035	2.54
ヤヲラ	0.004	0.001	0.003	53	0.071	-2.176
マイテ	0.002	0.001	0.004	34	0.001	-2.244
トリワキテ	0.000	0.005	0.002	27	0	4.73

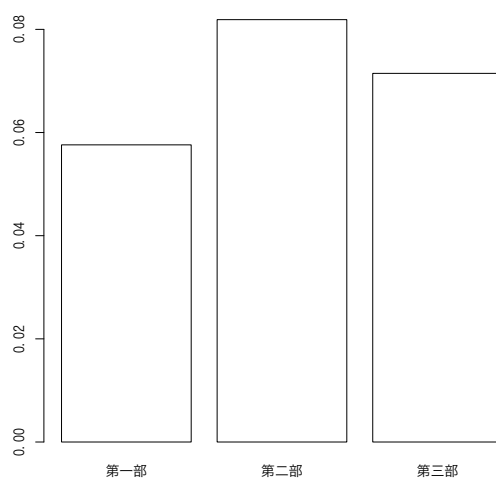


図 9 「かく」の出現率

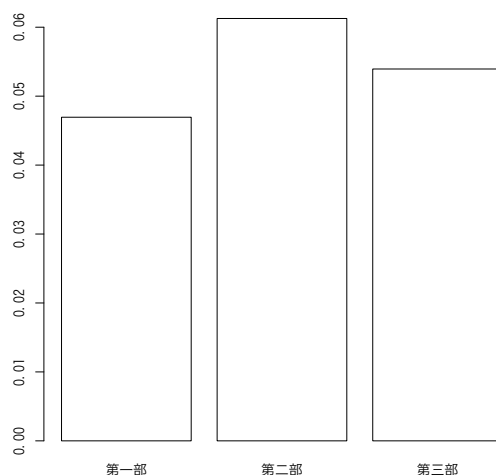


図 10 「なほ」の出現率

最後に、副詞について検討を加えた。表 5 が分析の結果であり、これら 10 語が本研究の分析で抽出された特徴語のすべてである。「とりわきて」

の調整済み残差は 4.730 となっており、副詞の中では調整済み残差の絶対値が最大となるが、総度数は 27 と小さいことから、1 度も出現していない巻もある。これは「とりわきて」という単語が第一部における出現率が 0.000 であることに起因すると考えられる。それに対し、「かく」や「なほ」と言った上位 2 語は総度数も多く、なおかつ調整済み残差の絶対値も大きい。これら 2 語の出現率は図 9 及び図 10 に示す通りである。両語ともに同様の傾向を示しており、第二部、第三部、第一部の順で出現率が高い。

5. 関連研究

本研究の目的は統計的な手法を用いて、『源氏物語』の第二部において偏って出現する単語の抽出を試みた。分析では名詞、動詞、形容詞、副詞の単語を採り上げ、『源氏物語』全 54 巻を通じて頻出される単語を中心に検討を加えた。その結果、部ごとの計量的な特徴の相違が明らかになったと言える。今後の展望として、単語の n-gram を特徴量として第二部の計量的な文体的特徴を明らかにすること、また機械学習の手法を用いてより詳細な検討を加えることで、より包括的に第二部の計量的な文体的特徴の相違を明らかにすることである。

謝辞

本研究は JSPS 科研費（課題番号：19K20627）の助成を受けたものです。

参考文献

- [1] 池田亀鑑. 源氏物語の構成 (新講源氏物語 (上) 所収). 至文堂, 1951.
- [2] 武田宗俊. 源氏物語の研究. 岩波書店, 1954.
- [3] 加藤昌嘉, 中川照将. “『源氏物語』はどのように来たのか?” を考えるために, 紫上系と玉鬘系—成立論のゆくえ, 1-24. 勉誠出版, 2010.
- [4] 藤井貞和. 源氏物語論. 岩波書店, 2000.
- [5] 池田亀鑑. 新講源氏物語 (下). 至文堂, 1951.
- [6] 藤村潔. 源氏物語の構造. 桜楓社, 1966.
- [7] 村上征勝, 今西祐一郎. 源氏物語の助動詞の計量分析. 情報処理学会論文誌, 40(3), 774-782, 1999.
- [8] 小野洋平. 源氏物語成立論の統計科学的再考察: 村上・今西(1999)を中心に. 計量国語学, 29(8), 296-312, 2015.
- [9] 土山玄. 『源氏物語』における「紫上系」と「玉鬘系」における語の使用傾向について, 第 21 回公開シンポジウム「人文科学とデータベース」発表論文集, 11-16. 2016.
- [10] 土山玄, 村上征勝. 『源氏物語』 第三部の

成立に関する計量的な考察. じんもんこん 2014
論文集 2014.3, 213-220. 2014.

[11] 上田英代, 村上征勝, 今西祐一郎, 樺島忠夫,
上田裕一: 源氏物語語彙用例総索引 自立語編,
勉誠社, 1994.

[12] 上田英代, 村上征勝, 今西祐一郎, 樺島忠夫,
上田裕一, 藤田真理: 源氏物語語彙用例総索引
付属語編, 勉誠社, 1996.

[13] 池田亀鑑: 源氏物語大成 索引篇, 中央公
論社, 1985.