

『源氏物語』 第三部の成立に関する計量的な考察

土山 玄

村上 征勝

同志社大学大学院 文化情報学研究科 同志社大学文化 情報学部

日本学術振興会 特別研究員

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、第三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三部を採り上げ、計量的な手法を用い第三部の成立過程について分析を行った。分析においては、品詞構成比率・語の頻度・語の長さという項目を用いた。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3つのグループに分類された。すなわち、第三部は現行の巻序通りではあるが、連続的に執筆されたのではなく、段階的に成立したと考えられる。

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

Gen Tsuchiyama
Graduate School of Culture
and Information Science
Doshisha University

Masakatsu Murakami
Faculty of Culture and Information
Science
Doshisha University

Research Fellow of Japan Society
for the Promotion of Science

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 third part, which is the final 13 volumes (from Niomiya, volume 42, to Yume no Ukihashi, volume 54). Following a principal component analysis of the relative word frequency, word length distribution, and proportion of the parts of speech, these 13 volumes could be classified into three groups, i.e., volumes 42-44 (Niomiya to Takekawa), 45-49 (Hashihime to Yadorigi), and 50-54 (Azumaya to Yume no Ukihashi). Therefore, based on the comparison of these results for the three groups, we conclude that the third part was written over a period of time as a series of individual stories rather than as a coherent story.

1. はじめに

本研究は計量文献学の方法を用い『源氏物語』第三部の成立過程について検討を加える計量的な研究である。『源氏物語』全54巻は一般に三部構成であると考えられており[1]、第三部とは第42巻「匂宮」から第54巻「夢浮橋」までの13巻である。一般に、「匂宮」から第44巻「竹河」までは匂宮三帖、第45巻「橋姫」以降の10巻は宇治十帖と称される。なお、第一部は第1巻「桐壺」から第33巻「藤裏葉」までであり、第二部は第34巻「若菜上」から第41巻「幻」までの8巻である。

『源氏物語』の成立過程を考察する成立論において、第三部が採り上げられることは少ない。成立論の考察対象は第一部の33巻に集中しており、『源氏物語』の成立過程に関する研究の見解の1

つとして、登場人物の出現状況の調査に基づく客観的なデータから、第一部には「紫上系」と称される17巻と「玉鬘系」と称される16巻の2系統が内在しているという説が論じられている[2]。これは、初出が「紫上系」となる登場人物は「玉鬘系」においても登場するが、初出が「玉鬘系」である人物は「紫上系」に登場しないという指摘である。

また、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[3]や、内容の考証に基づき第38巻「鈴虫」の後記説[4]および第39巻「夕霧」の後記説[5]が報告されている。

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[6]。

そこで、本研究では計数可能な文体的特徴、すなわち表現形式を分析項目として採り上げ、これを集計したデータについて多変量解析を行い、計量的な観点から第三部の 13 巻の成立過程について検討を加えた。

このような文章の量的側面を対象とした計量分析では、語の長さ・語の頻度などが分析項目として採り上げられることが多い[7]。また、古典文学作品を対象とした研究では、品詞別の構成比率が採り上げられることもあり、ジャンル別に品詞構成比率が相違することが報告されている[8]。本研究では品詞構成比率・語の頻度・語の長さを分析項目として採り上げた。

2. データ

本研究では、『源氏物語』の写本系統の 1 つである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』[9]および『源氏物語語彙用例総索引 付属語編』[10]を電子化したデータベースを分析に利用した。

また、『源氏物語語彙用例総索引』は『源氏物語』の本文すべてについて、形態素解析を行ったものである。なお、形態素解析については、『源氏物語大成 索引篇』[11]の単語認定基準に準拠している。

3. 関連研究

日本における文章の量的側面に注目した初期の研究として、大野(1956)は著名である[8]。大野(1956)では、『万葉集』『枕草子』『徒然草』『方丈記』『紫式部日記』『土佐日記』『讃岐典侍日記』『竹取物語』『源氏物語』を対象とし、名詞の比率が減少するにつれて、動詞、形容詞、形容動詞の構成比率が増加することを明らかにした。なお、これは「大野の法則」と称され、「大野の法則」は水谷(1965)において定式化されている[12]。

次に、安本(1957)は『源氏物語』を分析対象とし、『源氏物語』を宇治十帖とその他の 44 巻に二分し、宇治十帖における他作者説を検討するために、統計的検定を行っている[13]。検定に用いた項目は、各巻のページ数・和歌の使用度・直喩の使用度・声喩の使用度・心理描写・文の長短・色彩語の使用度・名詞の使用度・用言の使用度・助詞の使用度・助動詞の使用度・品詞数の 12 項目である。検定の結果、宇治十帖の文体は作り物語、用言的、緊密かつ連続的な構想による詳細な描写を特徴とし、一方、他の 44 巻の文体は歌物語的、体言的、飛躍的、断続的な構想による直感的描写を特徴とすると指摘し、結論として、宇

治十帖の作者は他 44 巻の作者と同一人物であるとは言い難い、と論じている。

『源氏物語』を対象とし、多変量解析の手法を用いた本格的な研究としては村上・今西(1999)がある。日本語の機能語の 1 つである助動詞についての語の頻度を採り上げ、数量化Ⅲ類を用いて『源氏物語』の執筆巻序の推定を行っている。分析の結果、現行の巻序と各巻の成立順序が同様ではない可能性を指摘した[14]。

また、古典文における作者の識別に統計手法が有効であることを明らかにした研究として土山・村上(2011)がある[15]。土山・村上(2011)は、平安時代に成立した長編物語である『源氏物語』と『宇津保物語』を分析対象とし、語の頻度について主成分分析を行っている。分析において名詞・動詞・形容詞・形容動詞・副詞・助詞・助動詞といった品詞の語の頻度についての分析の結果、『源氏物語』と『宇津保物語』との間に語の出現傾向の顕著な相違が認められることが報告されている。したがって、古典文を分析対象とする場合も現代文と同様に、作者の相違を見いだせると考えられる。

4. 分析

本研究における分析対象は『源氏物語』第三部の 13 巻である。これら 13 巻は各巻の延べ語数が同様ではないため、頻度ではなく各巻の延べ語数に対する相対頻度、すなわち出現率を用いた。また、全語彙を分析するのではなく、名詞・代名詞・動詞・補助動詞・形容詞・形容動詞・副詞・連体詞・助詞・助動詞の 10 品詞について分析を行った。

分析においては、後述するように第三部の 13 巻について分析を加えた結果、宇治十帖は表現形式の量的傾向が相違する 2 つのグループに分類されることが推測されたため、対象を宇治十帖に限定しあらためて分析を行った。すなわち、本研究では第三部 13 巻についての分析、宇治十帖と称される 10 巻についての分析という 2 段階をもって構成される。また、どの分析項目に対しても主成分分析を行った。

4.1 第三部についての分析および結果

4.1.1 品詞構成比率

まず、品詞構成比率を採り上げ主成分分析を行った。品詞構成比率とは、分析対象となる文献における各品詞の占める割合である。図 1 は上述した 10 品詞の各巻における比率に対する主成分分析の結果であり、横軸を第 1 主成分の主成分得点、縦軸を第 2 主成分の主成分得点とした散布図である。また、図 1 における実線と破線で示した楕円は主成分得点から推定される匂宮三帖と宇治

十帖の 95%信頼楕円である。図 1 において、宇治十帖の信頼楕円の内部に句宮三帖は配置されておらず、両グループは異なった品詞構成比率の傾向を有していると考えられる。

4.1.2 語の頻度

語の頻度とは各語の対象における出現頻度である。本研究では出現率を用い、品詞別に分析を行った。分析結果を先に述べると、名詞・代名詞・動詞・形容詞・形容動詞・助動詞において、句宮三帖と宇治十帖の間に量的傾向の相違が認められ、他の品詞では顕著な相違は認められなかった。以下においては、名詞・代名詞・動詞・形容詞・形容動詞・助動詞の分析結果について報告する。

第三部における名詞の延べ語数は 20687、異なり語数は 2798 であり、出現頻度上位 102 語を分析に用いた。該当する語は頻度が 31 以上の語であり、上位 102 語までの累積頻度は 11397 となり、名詞の延べ語数の 55.1%を占める。主成分分析の結果は図 2 に示す通りであり、宇治十帖の信頼楕円の内部に句宮三帖は配置されず、両グループの間に語の出現傾向の相違が認められる。

代名詞の延べ語数は 842、異なり語数は 35 である。代名詞は異なり語数が名詞に比べ少ないことから、頻度が 10 以上の語について主成分分析を行った。これは出現頻度上位 18 語が該当し、累積頻度は 795、延べ語数の 94.4%を占める。主成分分析の結果、図 1 や図 2 と異なり両グループの信頼楕円は重複する。しかし、おおよそ第 1 主成分の正に句宮三帖、負に宇治十帖が付置される傾向が認められた。

動詞の延べ語数は 20932、異なり語数は 2835 である。出現頻度上位 101 語について分析を行った。該当するのは頻度が 28 以上の語であり、累積頻度は 11969 となり、延べ語数の 57.2%を占める。分析結果は図 3 に示す通りであり、宇治十帖の信頼楕円の内部に句宮三帖は配置されず、両グループの間に量的傾向の相違が認められる。

形容詞の延べ語数は 7111、異なり語数は 544 である。出現頻度上位 99 語について主成分分析を行った。これは頻度 16 以上の語が該当し、累積頻度は 5329 であり、形容詞の延べ語数の 74.9%となる。図 4 は主成分分析の結果であり、句宮三帖と宇治十帖は混在せず、両グループの語の出現傾向には相違が認められる。

形容動詞の延べ語数は 3037、異なり語数は 425 である。出現頻度上位 99 語について主成分分析を行った。これは頻度 7 以上の語が該当し、累積頻度は 2368 であり、形容動詞の延べ語数の 78.0%を占める。分析結果は図 5 に示す通りであり、句宮三帖と宇治十帖との間において、語の出現傾向に顕著な相違が認められる。

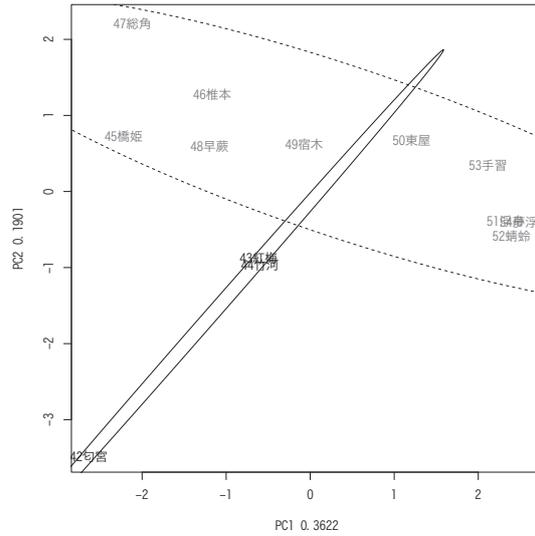


図 1 第三部における品詞構成比率の分析結果

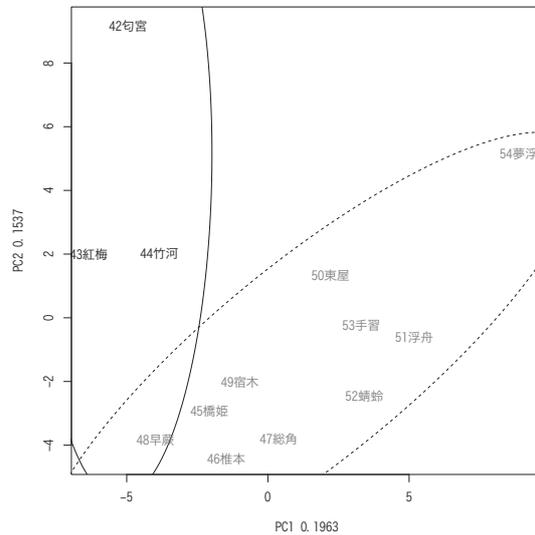


図 2 第三部における名詞 102 語の分析結果

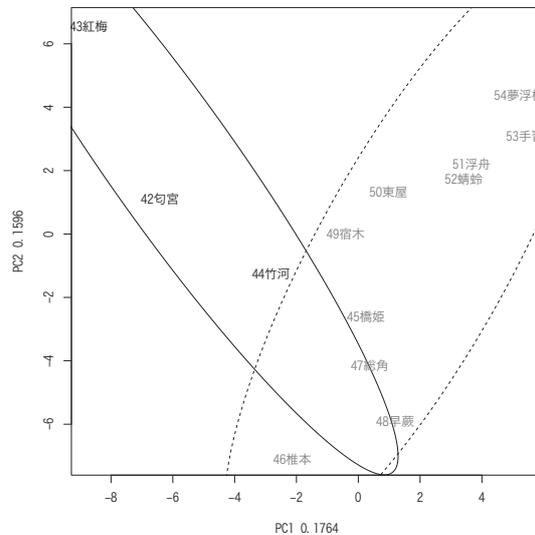


図 3 第三部における動詞 101 語の分析結果

最後に、助動詞の延べ語数は 15188、異なり語数は 25 である。助動詞は助詞と同様に機能語であり、異なり語数は少なく延べ語数は多い。出現頻度が 10 以上となる出現頻度上位 22 語について主成分分析を行った。累積頻度は 15183 であり、延べ語数に対する割合はおよそ 100%となる。図 6 は主成分分析の結果であり、宇治十帖の第 48 巻「早蕨」と句宮三帖の第 43 巻「紅梅」が接近して位置するが、第 1 主成分の正負に句宮三帖と宇治十帖が分離して配置される傾向があると考えられることから、語の出現傾向に相違があると推測される。

語の頻度の分析において、句宮三帖と宇治十帖との間に名詞・動詞・形容詞・形容動詞における語の出現傾向に相違が認められ、代名詞・助動詞においては両グループの巻が接近して付置されるものの、相違する傾向が認められた。

4.1.2 語の長さ

語の長さは対象において出現する語の文字数を計数し、文字数毎に出現する語の頻度を集計したものである。なお、本研究では各文字数を水準と称する。分析結果を先に述べると、名詞・補助動詞・助詞・助動詞の 5 品詞において量的傾向の相違が認められ、その他の品詞では顕著な相違は認められなかった。以下においては、上述の 5 品詞の分析結果について報告する。

名詞は長さ 2 から長さ 6 までの 5 水準において総度数の 92.3%となりこれら 5 水準について主成分分析を行った。分析結果は図 7 に示す通りであり、句宮三帖と宇治十帖との間において、語の長さの傾向に顕著な相違が認められる。

補助動詞は長さ 3 から長さ 5 までの 3 水準において総度数の 99.7%となり、これら 3 水準について主成分分析を行った。分析の結果、両グループの信頼楕円は重複し、句宮三帖の第 44 巻「竹河」が宇治十帖に接近して位置するが、第 2 主成分の正負において、両グループの間に語の長さの傾向に相違が認められると考えられる。

助詞は長さ 1 から長さ 3 までの 3 水準において総度数の 99.9%となるため、これら 3 水準について主成分分析を行った。分析の結果、宇治十帖の第 48 巻「早蕨」が宇治十帖の他の巻から外れて付置されるが、分析対象はおおよそグループ別にまとめられ配置されていると考えられる。

助動詞は長さ 1 から長さ 3 までの 3 水準において総度数の 99.4%となるため、これら 3 水準について主成分分析を行った。分析結果は図 8 に示す通りであり、信頼楕円は重複するが、句宮三帖と宇治十帖との間に、第 1 主成分において語の長さの傾向に相違が認められると考えられる。

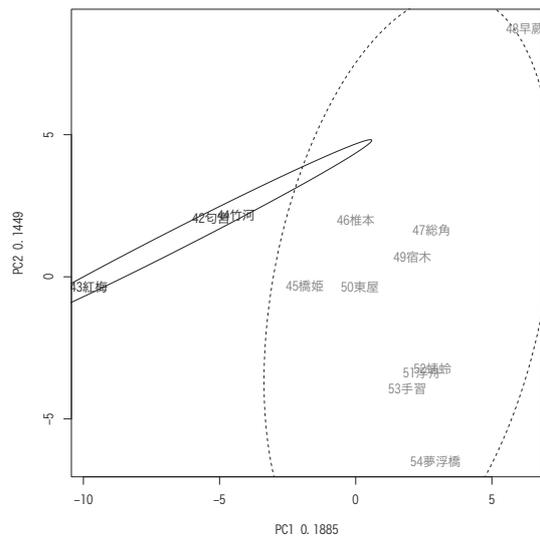


図 4 第三部における形容詞 99 語の分析結果

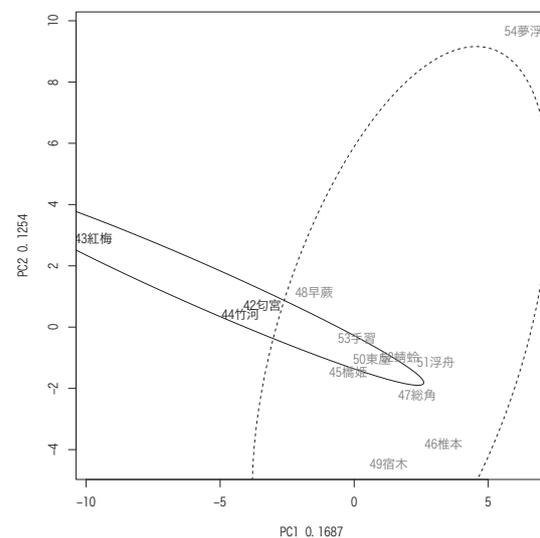


図 5 第三部における形容動詞 99 語の分析結果

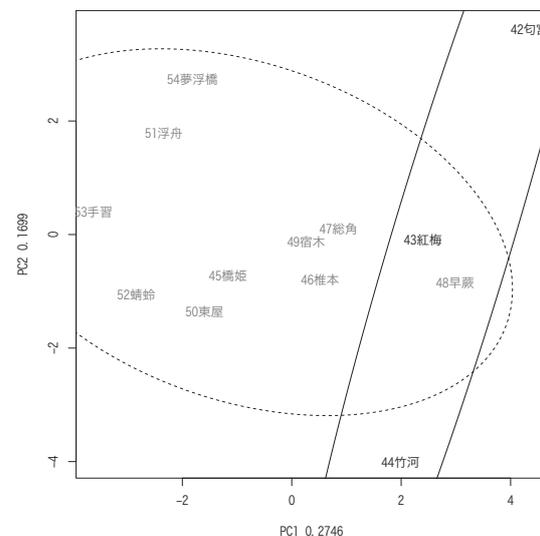


図 6 第三部における助動詞 22 語の分析結果

4.1.3 考察

土山・村上(2011)[15]によれば、作者が相違する場合、助詞・助動詞の両方に語の出現傾向の顕著な相違が認められることが報告されていることから、本研究においては句宮三帖と宇治十帖の作者が相違するとは結論づけられない。

しかし、本研究の分析結果において、品詞構成比率・語の頻度・語の長さにおいて、量的傾向の相違が認められる品詞が複数あることから、『源氏物語』の第三部は句宮三帖と宇治十帖の2つのグループに分類されると言える。

4.2 宇治十帖についての分析および結果

第三部を対象とした分析において、宇治十帖の前半5巻と後半5巻では表現形式の量的傾向が相違することを推測させる分析結果が得られた。このような推測は、図1、図2、図3、図6などにおける宇治十帖の諸巻の配置において、認められると考えられる。そこで、本研究では分析対象を宇治十帖に限定して計量的な検討を加えた。

4.2.1 品詞構成比率

まず、品詞構成比率について分析を行った。図9は10品詞の構成比率に対する主成分分析の結果であり、宇治十帖の前半5巻と後半5巻は第1主成分において分離して配置される。実線と破線の楕円はそれぞれ前半5巻と後半5巻の95%信頼楕円であり、重複しない。したがって、品詞構成比率においては宇治十帖の前半5巻と後半5巻との間において、顕著な量的傾向の相違が認められる。

4.2.2 語の頻度

分析結果を先に述べると、名詞・代名詞・動詞・形容詞・助詞・助動詞の6品詞において、宇治十帖の前半5巻と後半5巻の間に量的傾向の相違が認められ、他の品詞では顕著な相違は認められない。以下に上述の6品詞の分析結果を報告する。

宇治十帖における名詞の延べ語数は18186、異なり語数は2541である。出現頻度上位99語について主成分分析を行った。これに該当するのは頻度が28以上の語である。上位99語までの累積頻度は10211であり、延べ語数の56.1%を占める。図9は分析結果であり、宇治十帖の前半5巻と後半5巻は混在せず、両グループの間に語の出現傾向に顕著な相違が認められる。

代名詞の延べ語数は764、異なり語数は34である。頻度が10以上の語彙をについて主成分分析を行った。これは出現頻度上位17語が該当し、累積頻度は711であり、延べ語数の93.1%となる。分析結果は図11に示す通りであり、図9と同様に、宇治十帖の前半5巻と後半5巻との間において、語の出現傾向に相違が認められる。

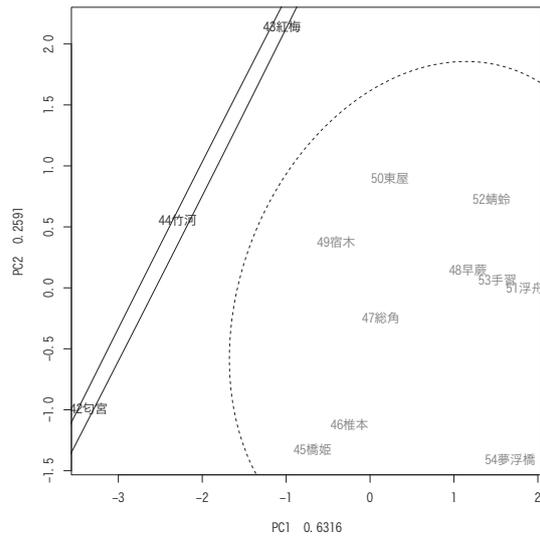


図7 第三部における名詞の長さの主成分分析

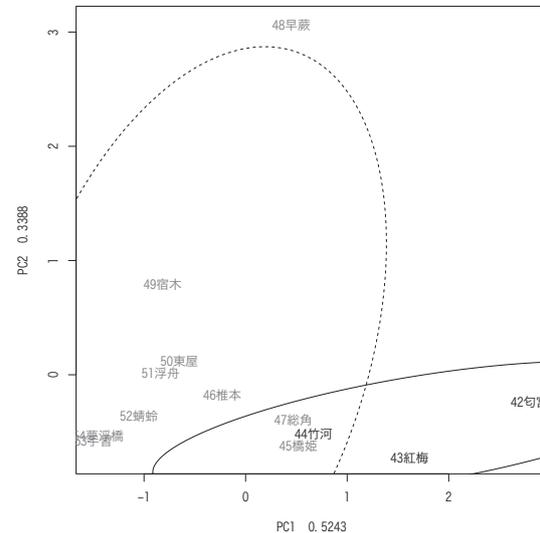


図8 第三部における助動詞の長さの主成分分析

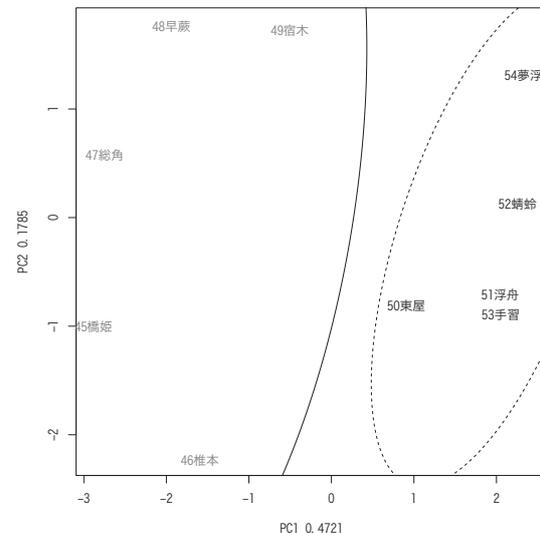


図9 宇治十帖における品詞構成比率の主成分分析

動詞の延べ語数は 18767、異なり語数は 2660 である。動詞の出現頻度上位 99 語に該当するのは頻度が 25 以上の語であり、累積頻度は 10735 となり、これは延べ語数の 57.2% を占める。分析結果は図 12 に示す通りであり、宇治十帖の前半 5 巻と後半 5 巻との間において、第 1 主成分に語の出現傾向に相違が認められる。

形容詞の延べ語数は 6386、異なり語数は 517 である。出現頻度上位 99 語は頻度 14 以上の語が該当し、累積頻度は 4807 であり、延べ語数の 75.3% となる。分析の結果、両グループの信頼楕円は重複するが、第 1 主成分の正負に前半 5 巻と後半 5 巻が分離して付置される傾向が認められる。

助詞の延べ語数は 35785、異なり語数は 54 である。頻度が 10 以上となる出現頻度上位 43 語の累積頻度 35745 であり、延べ語数に対して 99.9% となる。分析の結果、両グループの信頼楕円は重複するが、前半 5 巻と後半 5 巻はグループ別におよそまとまって付置された。よって、語の出現傾向に顕著な相違は認められないが、グループ内においては同様の語の出現傾向を有していると考えられる。

助動詞の延べ語数は 13733、異なり語数は 25 である。出現頻度が 10 以上となる出現頻度上位 23 語について、主成分分析を行った。累積頻度は 13731 であり、総度数に対する割合もおおよそ 100% となる。分析結果は図 13 に示す通りであり、第 1 主成分の正負において前半 5 巻と後半 5 巻が分離して配置しており、語の出現傾向に相違が認められる。

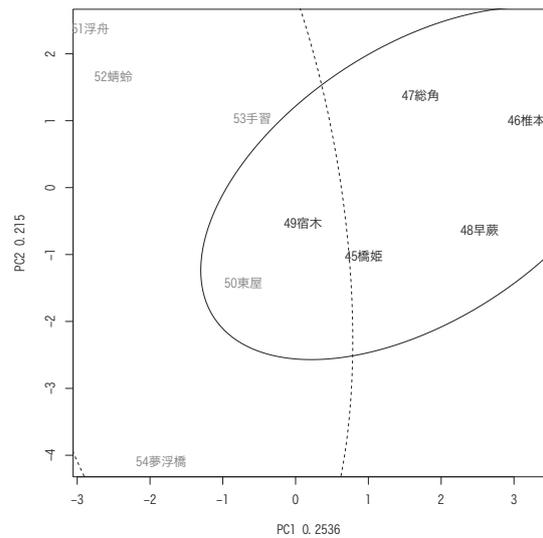


図 11 宇治十帖における代名詞 17 語の主成分分析

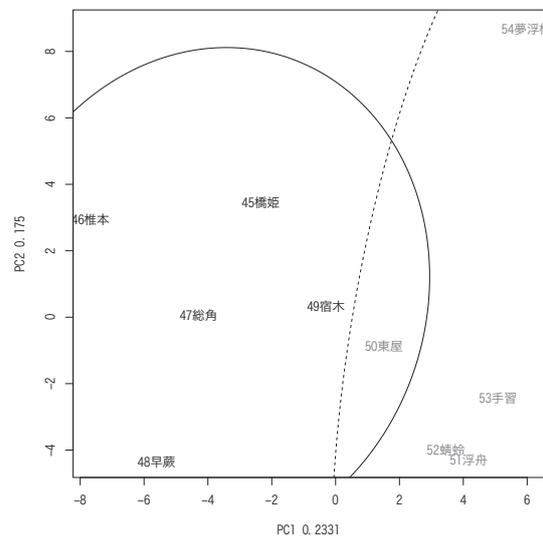


図 12 宇治十帖における動詞 99 語の主成分分析

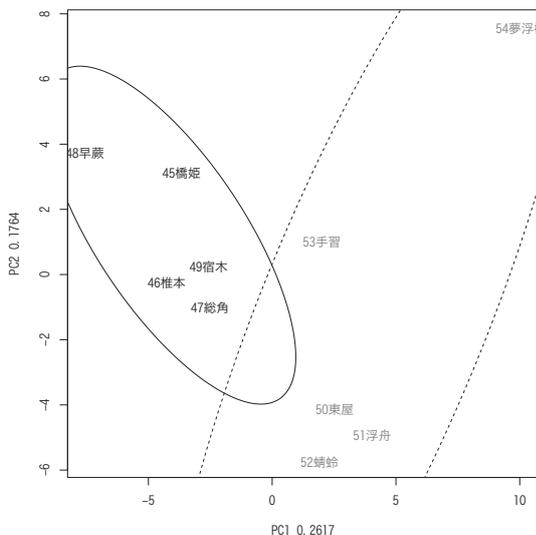


図 10 宇治十帖における名詞 99 語の主成分分析

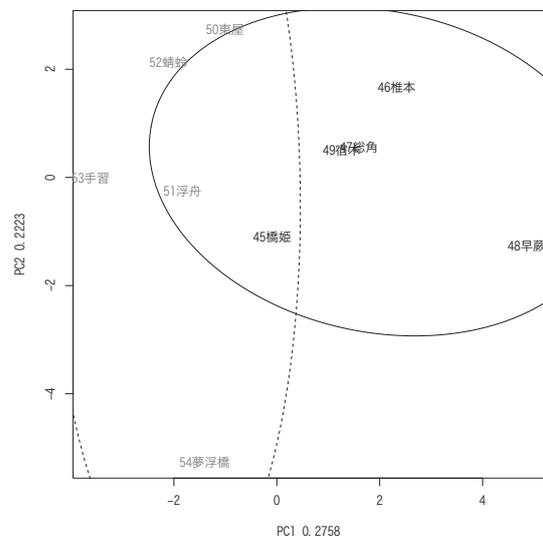


図 13 宇治十帖における助動詞 23 語の主成分分析

4.2.2 語の長さ

分析結果を先に述べると、名詞・代名詞・動詞・助動詞において量的傾向の相違が認められ、その他の品詞では顕著な相違は認められなかった。以下においては、名詞・代名詞・動詞・助動詞の分析結果について報告する。

ただし、連体詞の語の長さの分析において、連体詞は2水準になることから、宇治十帖を対象とした分析においては、連体詞を除く9品詞の語の長さの分布について検討を加えた。

名詞は長さ1から長さ6までの6水準において総度数の97.0%となり6水準について主成分分析を行った。図14は分析結果であり、両グループの95%信頼区間は重複するが、第1主成分において前半5巻と後半5巻との間に語の長さの量的傾向に相違が認められる。

代名詞は長さ1から長さ3までの3水準において総度数の95.5%となり、これら3水準について主成分分析を行った。分析結果は図15に示す通りであり、第1主成分において前半5巻と後半5巻との間における量的傾向に相違が認められる。

動詞は長さ1から長さ6までの6水準において総度数の96.4%となり、これら6水準について主成分分析を行った。図16に示す通り、前半5巻に属する第49巻「宿木」は後半5巻に接近するが、第1主成分において前半5巻と後半5巻との間において、語の長さの傾向に相違が認められると考えられる。

助動詞は長さ1から長さ3までの3水準において総度数の99.4%となり、これら3水準について主成分分析を行った。分析結果は図17に示す通り、前半5巻と後半5巻との間に、第1主成分において語の長さの傾向に相違が認められると考えられる。

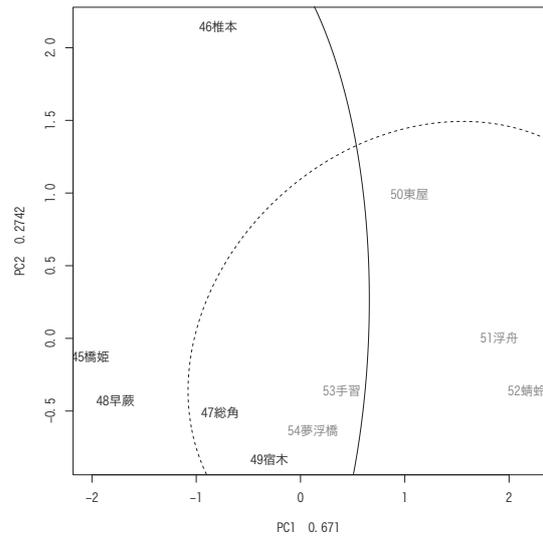


図 15 宇治十帖の代名詞の長さの主成分分析

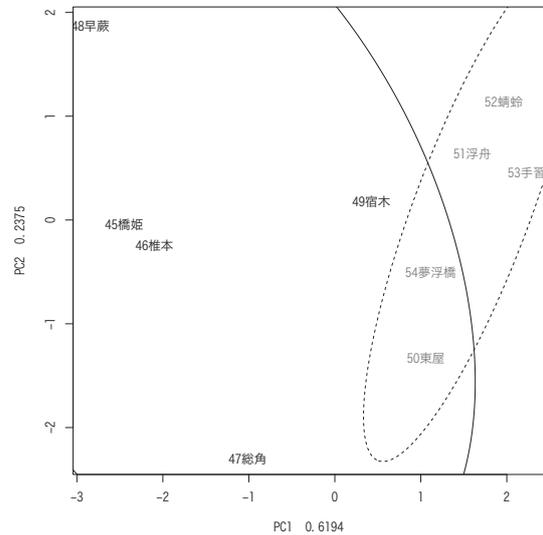


図 16 宇治十帖における動詞の長さの主成分分析

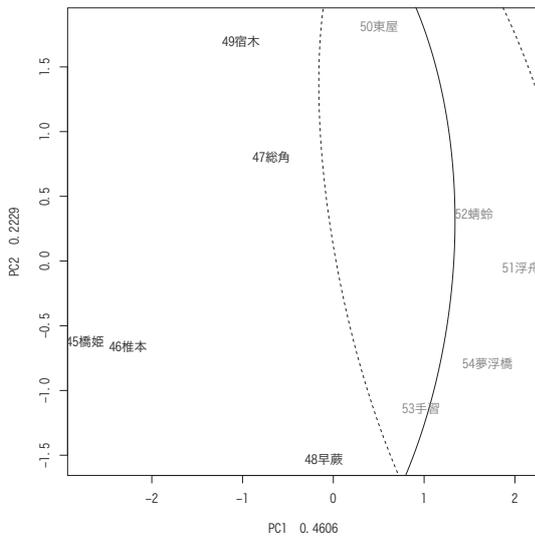


図 14 宇治十帖における名詞の長さの主成分分析

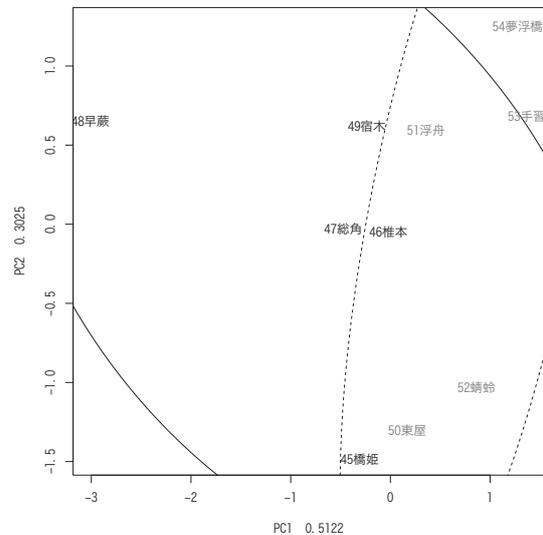


図 17 宇治十帖の助動詞の長さの主成分分析

4.2.3 考察

このように、宇治十帖の前半5巻と後半5巻との間において、品詞構成比率・語の頻度・語の長さといった3つの分析項目に量的傾向の相違が散見された。なお、語の頻度についての分析では名詞・代名詞・動詞・形容詞・助詞・助動詞に、語の長さの分布についての分析では、名詞・代名詞・動詞・助動詞において相違する量的傾向が認められた。したがって、品詞別に分析を行った語の頻度と語の長さにおいては、名詞・代名詞・動詞・助動詞の4品詞は、前半5巻と後半5巻との間に共通して量的傾向の相違が認められと言える。

このように、匂宮三帖と宇治十帖との間に認められた相違と同様の量的傾向の相違が、宇治十帖内部の前半5巻と後半5巻との間に認められると考えられる。

5. 結論

本研究においては、『源氏物語』の第三部と称される13巻について、品詞構成比率・語の頻度・語の長さという3つの計数可能な分析項目を採り上げ、これについて統計的に分析を行った。先にふれたように、第三部についての成立過程に関する考察は第一部に比べ十分に展開しているとは言えない。そこで、本研究における分析の結果、『源氏物語』の第三部には匂宮三帖・宇治十帖前半5巻・宇治十帖後半5巻という3つの異なる量的傾向を有するグループが存在することが明らかになった。

また、上述のように第三部の13巻は3つのグループに分類されることから、匂宮三帖・宇治十帖前半5巻・宇治十帖後半5巻は、それぞれがまとまって成立したと考えられる。すなわち、第三部の13巻は連続的に執筆されたのではなく、段階的に成立した可能性が本研究の分析結果から推測される。

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) 藤井貞和: 源氏物語論, 岩波書店 (2000).
- 4) 池田亀鑑: 新講源氏物語 (下), 至文堂 (1951).
- 5) 藤村潔: 源氏物語の構造, 桜楓社 (1966).
- 6) 加藤昌嘉, 中川照将: “『源氏物語』はどのように出来たのか?” を考えるために (『紫上系と玉鬘系—成立論のゆくえ』所収), pp.1-24 (2010).
- 7) Grieve, J.: Quantitative authorship attribution: An evaluation of techniques. *Literary and linguistic computing*, Vol.22, No.3,

pp.251-270 (2007).

8) 大野晋: 基本語彙に関する二三の研究-日本の古典文学作品に於ける, 国語学, Vol.24, pp.34-46 (1956).

9) 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).

10) 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).

11) 池田亀鑑: 源氏物語大成 索引篇, 中央公論社 (1985).

12) 大野晋: 基本語彙に関する二三の研究-日本の古典文学作品に於ける, 国語学, Vol.24, pp.34-46 (1956).

13) 水谷静夫: 大野の語彙法則について, 計量国語学, Vol. 35, pp. 1-13 (1965).

14) 安本美典: 宇治十帖の作者-文章心理学による作者推定-, 心理学評論, Vol. 2, No. 1, pp.147-156 (1957).

15) 土山玄, 村上征勝: 源氏物語と宇津保物語の語の使用傾向について, 人文科学とコンピュータシンポジウム論文集, Vol. 2011, No. 8, pp.125-132 (2011).