統計数理（1996）
第44巻 第1号 127-131

公開講演会要旨

『源氏物語』の言葉を分析する

統計数理研究所　村　上　征　勝

（1995年11月1日，統計数理研究所　講堂）

1. はじめに

『源氏物語』の成立は西暦1004年頃といわれており，すでに1000年にならんとする長い研究の歴史がある．しかしながら，いまだ解決されていない問題も数多く存在する．たとえば，『源氏物語』54巻がすべて紫式部によって書かれたかどうかという基本的な問題を始めとして，54巻の成立順序や，書写者による部分的な補筆の可能性などに関する問題は，これまでの数多くの研究者の努力にもかかわらず残されたままである．現在『源氏物語』の研究は飽和状態にあるといわれているが，残されている多くの疑問を解決し，『源氏物語』研究を更に発展させるためには，従来とは異なる新しい視点からの研究法の導入が必要であると考えられる．そこで，『源氏物語』の言葉を数量的に分析するという方法によって，『源氏物語』研究の新たな道の開拓を試みている．現在，研究はデータベースがほぼ完成したという段階にあり，種々の疑問に対して解答を与えられるという状況ではないが，これまでに得られた結果を中間報告として述べることにする．

2. データベースの構築

『源氏物語』の言葉を計量的な視点から分析するには，そのような分析を可能にするデータベースが必要となる．しかし，計量分析可能な『源氏物語』のデータベースが存在しなかったため，データベースの作成から始めなければならなかった．データベース作成においては，底本として池田亀鑑編著『源氏物語大成』（中央公論社）を用いた．文章はこの『源氏物語大成』の索引に基づき，すべて“分かち書き”し，すべての言葉に品詞情報を付加した．また『源氏物語大成』の本文には句点がついていないが，文長等の計量分析の必要上，岩波書店の『古典大系』に従って句点を入れた．作成したデータベースは図1の様になっており，総語数で約38万語である．このうち，自立語（名詞，動詞，形容詞，形容動詞，副詞，接続詞，感動詞など，それ自身で一つの文節となる語）約22万語についてはデータ修正を終え，『源氏物語語彙用例総索引 —— 自立語篇 ——』全5巻として出版した．附属語（助詞，助動詞など単独では文節を構成することのない語）約16万語については，現在データを修正中で，今年度内に，『源氏物語語彙用例総索引 —— 附属語篇 ——』全5巻を出版する予定である．なお，現在データベースを修正中であるので，今回発表した種々の数値は今後多少変わる可能性がある．

3. 計量的観点からの『源氏物語』の全体像

まず『源氏物語』の全体像を計量的な観点から描き出してみる．『源氏物語』54巻の総語数は

0005-01
いつれ［代名］／の［助詞］／御時［名詞］／に［助詞］／か［助詞］／。／女御更衣［名詞］／あまた［副詞］／さふらひ［動詞］／給［補動］／ける［助動］／なか［名詞］／に［助詞］／いと［副詞］／やむことなき［形容］／きは［名詞］
0005-02
に［助詞］／は［助詞］／あら［動詞］／ぬ［助動］／か［助詞］／すくれ［動詞］／て［助詞］／時めき［動詞］／給［補動］／あり［動詞］／けり［助動］／。／はしめ［名詞］／より［助詞］／我［代名］／は［助詞］／と［助詞］／思あかり［動詞］／給へ［補動］／る［助動］／御方方［名副］
0005-03
めさましき［形容］／もの［名詞］／に［助詞］／おとしめ［動詞］／そねみ［動詞］／給［補動］／。／おなし［形容］／ほと［名詞］／それ［代名］／より［助詞］／下らう［名詞］／の［助詞］／更衣たち［名詞］
0005-04
は［助詞］／まして［副詞］／やすからす［連語］／。／あさゆふ［名副］／の［助詞］／宮つかへ［名詞］／に［助詞］／つけ［連語］／て［助詞］／も［助詞］／人［名詞］／の［助詞］／心［名詞］／を［助詞］／のみ［助詞］／うこかし［動詞］／うらみ［名詞］
0005-05
を［助詞］／おふ［動詞］／つもり［名詞］／に［助詞］／や［助詞］／あり［動詞］／けむ［助動］／いと［副詞］／あつしく［形容］／なりゆき［動詞］／もの心ほそけに［形動］／さとかちなる［形動］

図1．『源氏物語』のデータベース．

376232語で，このうち自立語は56.8%の213598語，附属語は43.2%の162634語である．また最も長い巻は（35巻）「若菜下」で20236語，最も短い巻は（27巻）「篝火」の651語である．最も短い巻と最も長い巻とでは，その長さにおいて30倍以上の差がある．巻毎に言葉の使用頻度などの統計量を比較する際にこのことが問題となる．

次に，最も出現頻度の多い自立語は副詞の「いと」で，出現回数は4225回であり，これは総語数の1.978%にあたる．つまり，「いと」という言葉は，平均すると100語に2回出現していることになる．また，最も出現頻度の多い名詞は「こと（事）」で，出現回数は3952回（1.8502%）である．

また，『源氏』54巻のすべての巻に出現する自立語は次に示す20語である．

［名詞］ かた（方） …694回， 心（こころ） …977回
御心（御こころ） …483回， こと（事） …3952回
人（ひと） …3441回， ほと（程） …1548回
もの（物） …1041回
［連体詞］ かの …727回， この …1442回
［副詞］ いと …4225回， え …660回， すこし …482回
［動詞］ あら（あり（有，在）の未然形） …1339回
おほし（おほす（思）の連用形） …622回
おもひ（おもふ（思）の連用形） …1170回
［動詞・補助動詞］ たまは（たまふの未然形） …1600回
たまひ（たまふの連用形） …3212回
たまふ（たまふの終止形） …3002回
たまふ（たまふの連体形） …2185回
たまへ（たまふの命令形） …2949回

この20語の中では，「心」，「御心」という言葉が全巻に出現することが興味深い．また，本居宣長の「もののあはれ」論以来，『源氏物語』は「もののあはれ」の文学ともいわれているので，「あはれ」という言葉に関して調べてみた．「あはれ」に関する言葉は，「あはれ」（名詞・感動詞），

「あはれがる」(動詞),「あはれさ」(名詞),「あはれなり」(形容動詞) など41種類『源氏物語』には現れる．どの巻にも，この41種の中のいずれかの語が出現しており，出現回数はこの41種類で計1036回である．

4.「匂宮」・「紅梅」・「竹河」の三帖について

『源氏物語』54巻は，通常（1巻)「桐壺」から（33巻)「藤裏葉」までの33巻を第一部，(34巻)「若菜上」から（44巻)「竹河」までの11巻を第二部，(45巻)「橋姫」から（54巻)「夢浮橋」までの10巻を第三部（「宇治十帖」）というように，3つのグループに分けられる．このなかで，第二部の（42巻)「匂宮」,（43巻)「紅梅」,（44巻)「竹河」の三巻については，古来より紫式部の作か否かに関し疑問が出されている．そこでこの三巻（以降「匂宮三帖」と呼ぶ）について言葉の出現頻度の観点から調べてみる．

この「匂宮三帖」のどの巻にも出現するが，他の51巻には出現しない語があるかどうか調べたところ，そのような言葉は存在しないことがわかった．少し条件をゆるめて，「匂宮三帖」の2巻以上に出現し，他の51巻には出現しない言葉を調べたところ，「かほる中将」(名詞),「御ことはり」(名詞) など8語ある事がわかった．ただこれらの言葉は，いずれも出現頻度が3巻を合わせても計2回と少なく，したがってこれらの言葉の出現頻度から「匂宮三帖」の他作家説を主張するのは難しいといえる．次に，「匂宮三帖」には出現しないが，他の51巻すべてには出現するような言葉があるかどうか調べたところ，そのような言葉も存在しないことがわかった．「匂宮三帖」には出現しないが，他の51巻の最も多くの巻に出現する自立語は「やうやう」(副詞) で，43巻に計135回出現する．しかし，それでも「匂宮」,「紅梅」,「竹河」以外にこの言葉が出現しない巻が8巻もあることから，この言葉からも他作家説を主張するのは難しい．

ただ，「匂宮三帖」に関しては付属語の「らむ」(助動詞) の用い方が少し気にかかる．助動詞の「らむ」は『源氏物語』54巻中で計669回出現するが，「匂宮」と「紅梅」の巻には「らむ」は一度も出現しない．「らむ」の出現率は約0.002であり，大体500語に1回の割合で出現する．もっとも短い「篝火」は651語であるが，この巻にも1回「らむ」は用いられている．「匂宮」,「紅梅」は，総語数がそれぞれ2698語，2516語であるので，5回程度は出現してもよさそうであるが，この2巻だけは「らむ」は一度も出現しない．書き手の文章のクセは無意識に書く部分に現れやすいというのが，これまでの文章の計量分析の研究で得られた一つの知見であるが，その点からすると，「匂宮三帖」他作家説を検討する際には，付属語の「らむ」の使われ方に関しては注目した方がよいと思われる．

ところで，「匂宮三帖」の中で「竹河」(8067語) には「らむ」は15回現れる．これはどのように考えるべきであろうか？「匂宮」,「紅梅」と「竹河」では書き手が異なるのであろうか．現時点ではこの点に関してはコメントすることはできない．ただ，「竹河」の巻は『源氏物語』のそれまであった話の蒸し返しが多く，似たような文章が多いという説がある．もし，他の人物が紫式部の文章を部分的に取り出し，つぎはぎしたならば，言葉の出現頻度だけの分析では作者が異なることを証明するのは困難となる．

いずれにしても，「匂宮三帖」に関しては附属語を中心とした更なる分析が必要である．

5. 第三部「宇治十帖」について

『源氏物語』の第三部「宇治十帖」と呼ばれる十巻が式部以外の手によるという説は古来よりあり，現在でもそのような疑いを持っている人は少なくない．

第一部，第二部という前半の44巻と，第三部「宇治十帖」の10巻とでは言葉の出現頻度等に

差がみられるのであろうか．平均文長や，語彙の豊富さを示すと考えられる延べ語数に対する異なり語数の比などには顕著な差はみられない．しかし自立語の割合を見ると，「宇治十帖」は前半の44巻に比べ自立語の割合が低い．このことは，名詞，助動詞の使用率に前半の44巻と「宇治十帖」で明らかに差が見られる（表1）という結果にも現れている．又，出現頻度の多い名詞の「ひと（人)」,「もの（物)」，動詞の「おもい（思い)」などの出現率にも差が見られる．

表1．『源氏物語』の前半（1～44巻）と後半（45～54巻）の品詞の使用率．

品詞	1～44巻	45～54巻	t 値
名詞	0.18189	0.16423	5.1624
動詞	0.16190	0.16505	−1.3581
形容動詞	0.02388	0.02426	−0.2963
助詞	0.31517	0.31872	−1.6130
形容詞	0.05907	0.05566	1.8597
助動詞	0.11309	0.12391	−5.0477
副詞	0.04000	0.04106	−0.8770

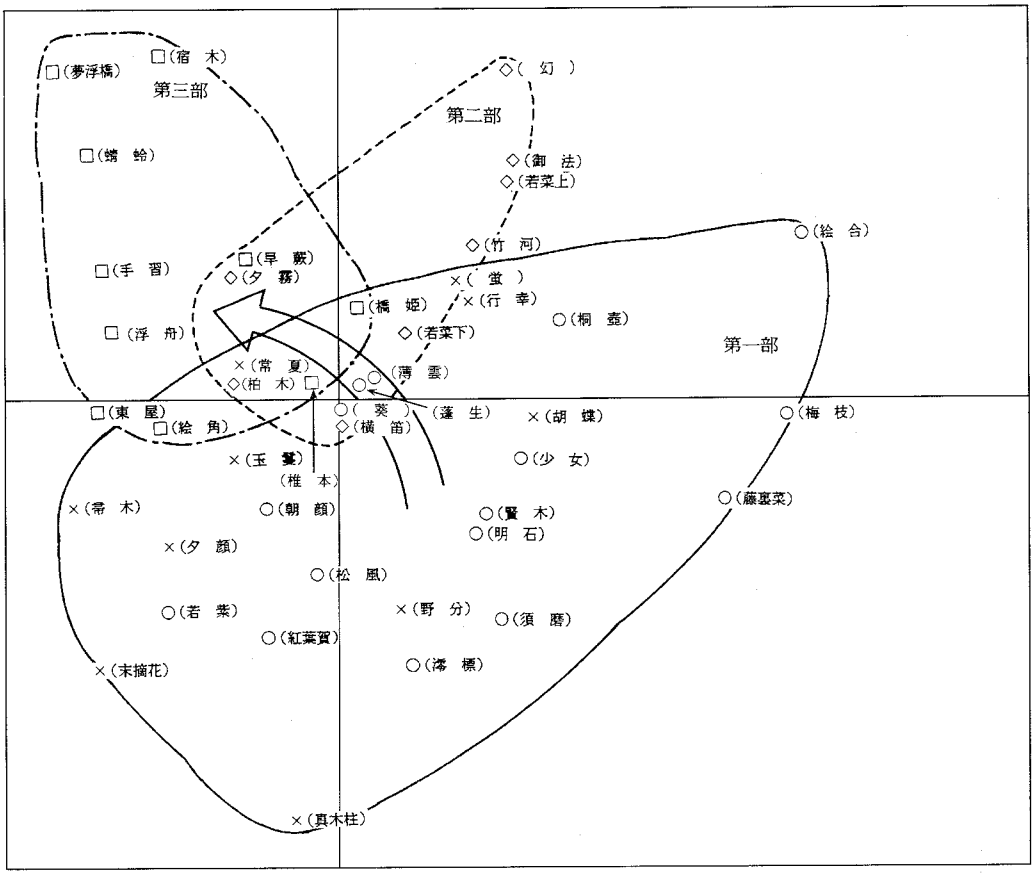

図2．主要8品詞による主成分分析結果（3000語以上の巻のみ）．

しかし平均出現率に差が見られることだけから，「宇治十帖」が紫式部以外の手によると結論づけるのは難しい．次の図2を見ていただきたい．この図2は名詞，動詞，形容動詞，形容詞，助動詞，副詞，敬語動詞，補助動詞の8品詞の使用率を主成分分析にかけ，これらの品詞の使用率の類似している巻を調べた結果である（ただし，54巻中3000語以下の短い次の10巻，「空蝉」，「花宴」，「花散里」，「関屋」，「初音」，「篝火」，「藤袴」，「鈴虫」，「匂宮」，「紅梅」は分析から外してある）．

これを見ると，第一部の巻は8品詞の使用率にかなりバラツキがあるのに対し，第二部，第三部（「宇治十帖」）と後の巻になる程，品詞の使用率が安定してきており，しかも，矢印で示したように，8品詞の使用率が第一部，第二部，第三部というように変化していくような傾向が見られる．同じようなことが，どの品詞の後にどの品詞が続いているかという，品詞の接続関係の主成分分析結果にも見られる．したがって，紫式部の文体が徐々に変化していったとも考えることができ，現段階では，「宇治十帖」の著者が紫式部であったか否かに関しても解答を与えることは出来ない．

6. おわりに

『源氏物語』の言葉の計量分析に関する研究を紹介しながら，『源氏物語』の著者複数説の中でも，「匂宮三帖」他作家説，および「宇治十帖」他作家説について，これまで得られている自立語に関する情報の範囲内でどのようなことが言えるかを述べてきた．いずれの説に関しても附属語に関する情報を加えた詳細な分析が不可欠で，現段階では何も明言できない．しかし，これまでの分析で得られた幾つかの知見は，今後の『源氏物語』の計量的分析に役立つものと思われる．