

単語情報に基づく源氏物語の計量分析

村上征勝 (統計数理研究所)

上田英代 (古典総合研究所)

樺島忠夫 (神戸学院大学)

今西祐一郎 (九州大学)

上田裕一 (もとぶ野毛病院)

筆者達は、池田亀鑑編著『源氏物語大成』(中央公論社)をテキストとして「源氏物語」の品詞情報付きデータベースを構築中であり、このデータベースを用いて紫式部の文体を調べ、「源氏物語」の複数著者説、書写者による補筆の可能性等の問題の解明を試みている。

この小論では、単語の出現頻度の情報に基づくいくつかの分析結果について報告する。

On word frequency in Genji Monogatari for analyzing the style of Lady Murasaki

Masakatsu Murakami (The Institute of Statistical Mathematics)

Hideyo Ueda (Japanese Classic Literature Research Institute)

Tadao Kabashima (Kobe Gakuin University)

Yuichiro Imanishi (Kyushu University)

Yasuichi Ueda (Motobu Noge Hospital)

We are building a full-text database of Genji Monogatari with codes for parts of speech. Using this database, we are now analyzing the style of Lady Murasaki to solve the authorship problem for Genji Monogatari.

In this paper, we report some results of the analysis from the standpoint of word frequency.

1. はじめに

日本古典文学の最高峰といわれる『源氏物語』は、その成立からすでに1000年近くを経ようとしている。しかし、紫式部の直筆が現存せず、写本で伝承されてきたこともあり、記述内容や文体の詳細な検討から生じた、複数作家説、成立過程における幾つかの巻の後期挿入説、書写者の部分的な補筆の可能性など、研究すべき課題は数多く残されたままである。

たとえば、複数作家説にしても、

- ・鎌倉時代頃から主張されている、紫式部と関白藤原道長との合作説
- ・年記・官位などの矛盾から、「匂宮」、「紅梅」、「竹河」の3巻を別人の作とする説
- ・41巻の「幻」までが紫式部で、42巻「匂宮」以後は娘の大貳の三位が書いたという説
- ・「藤裏葉」までが紫式部で、「若菜」以降は大貳の三位の作とする説
- ・後半の10巻（「宇治十帖」）は大貳の三位の作とする説

など諸々の説があり、未だ明確な結論は出されていない。

著者達は、数年前より、文の長さ、単語の出現頻度、品詞の出現頻度、語彙量などの、文章の数量的な性質に注目し、それらを分析することによって、『源氏』に関する諸々の課題を解決すべく研究を続けている。

この小論では、単語の出現頻度等の情報を用いた、『源氏物語』の計量分析について紹介する。

2. テキスト及びデータベース

「源氏物語」については数多くの校訂本文が存在するが、本研究では、

- 1) 写本の系統が明記してある
- 2) 他系統の本文との校異が精密である
- 3) 語彙索引が完備している

の3つの理由から、池田亀鑑編著『源氏物語大成』（中央公論社）をテキストとして採用した。

いつれの御時にか女御更衣あまたさふらひ給けるなかにいとやむことなきは
にはあらぬかすくれて時めき給ありけりはしめより我はと思あかり給へる御方
くめさましきものにおとしめそねみ給おなしほとそれより下らうの更衣たち
はましてやすからすあさゆふの宮つかへにつけても人の心をのみうこかしうら
みをおふつもりにやありけむいとあつくなりゆきもの心はそけにさとかな
るをいよくあかすあはれなる物におもほして人のそしりをえはゝからせ給
はず世のためしにもなりぬへき御もてなし也かたちめうへ人なともあいなく
めをそはめつゝいとまはゆき人の御おほえなりもろこしにもかゝることのおこ
りにこそ世もみたれあしかりけれとやうくあめのしたにもあちきなう人のも
てなやみくさになりて楊貴妃のためしもひきいてつへくなりゆくにいとほした
なきことおほかれとかたしけなき御心はへのたくひなきをたのみにてましらひ

11 10 9 8 7 6 5 4 3 2 1

図1

ただし、『源氏物語大成』は図1のように、句読点が付いていないので、写本の系統は異なるが、日本古典文学大系『源氏物語』（岩波書店）を参考にして、句読点を付けてデータベース化を試みた。

作成中のデータベースでは図2に示したように、すべての文章は分ち書きされて単語に分割され、品詞情報が付加されている。また単語の出現位置が把握できるように、『源氏物語大成』のページと行番号も付加されている。データベースの大きさは、単語数にして約38万語である。なお、このデータベースを用いて、自立語約22万語に関する『源氏物語語彙用例総索引－自立語編』（勉強社、5500頁）を昨年出版したが、付属語約16万語に関する語彙用例総索引も年内に出版する予定である。

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[動詞]/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[動詞]/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[動詞]/る[助動]/御方方[名詞]/

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[動詞]/。/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[名詞]/の[助詞]/宮つかへ[名詞]
/に[助詞]/つけ[動詞]/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]/

図 2

3. 出現単語に基づく文体のクセの把握

文体という観点から文献の著者に関する疑問を
解明する場合には、書き手の文章のクセをうまく
把握できるか否かが鍵となる。

ここでは、文献に現れる単語に注目して、どの
ようにして書き手のクセを計量的にとらえようと
しているかを紹介する。

3.1 著者の好みの単語

著者がどのような単語を好んで用いているかと
いうことは、複数著者説や、書写者による部分的
な補筆の可能性を調べる場合に、有効な情報を与
えてくれると考えられる。

たとえば、『源氏物語』54巻の中には約2万
5千種の単語が用いられているが、この中で54
巻すべてに出現する自立語は、次に示す20語だ
けである。(表記のうち一番多い形で示す。)(
内は『索引編』の見出し語)

[名詞]

かた(かた[方])、心(こころ)、御心
(御こころ)、こと(こと[事])、人(ひと)、
ほと(ほと[程])、もの(もの[物])

[動詞]

あら(「あり(有・在)」の未然形)

おほし(「おほす(思)」の連用形)

おもひ(「おもふ(思)」の連用形)

[動・補動]

たまは(たまふー動・補動四段未然形)

たまひ(たまふー動・補動四段連用形)

たまふ(たまふー動・補動四段終止形)

たまふ(たまふー動・補動四段連体形)

たまへ(たまふー動・補動四段命令形)

[連体] かの、この

[副詞] いと、え、すこし

「人」「もの」のような名詞は誰でも用いるこ
とが考えられるが、もし、源氏物語が紫式部一人
によって書かれたのであれば、例えば、「心」と
というような単語は式部が好んで用いた言葉といえ
るのではないだろうか。

次に、『源氏物語』と『源氏物語』を模倣して
書かれたとされる本居宣長の『手枕』とを比較し、
『手枕』に用いられているが『源氏物語』には用
いられていない単語をリストアップしたのが表1
である。宣長が『源氏』の文体を研究し、紫式部
の文体に似せる努力をしたにもかかわらず、『源
氏』に用いられていない単語をいくつか用いてい

単 語	初稿	再稿	単 語	初稿	再稿
いさきよく	○	○	みかとかね	○	○
*うちかすみ	○	○	ゆきかくれ	○	○
うちまけかたく	○	○	見え行	○	○
うちましりゆく	○	○	*御かしつきくさ	○	○
おもほしめさ	○	○	*御みかうし	○	○
*かゝやかしう	○	○	*御ゆるされ	○	○
くちはて	○	○	手まくら	○	○
そゝ	○	×	忍ひより	○	×
*なけの御筆つかひ	○	○	*恋し	○	○

*の単語は、活用形の異なった形や接頭語の無い形では『源氏物語大成』に存在している。

表 1

ることがわかる。特に、「いさきよく」という単語は、いわゆる男言葉で、中古の和文にはみられない単語とされている。

3.2 表記の違い

単語によっては、どの部分を漢字で書くか等でいくつかの異なる表記の仕方があるものがある。

たとえば、『源氏物語大成』には「ものおもひ」という名詞は59回出現するが、この言葉の場合、次の7通りの方法で表記されている（カッコ内は出現度数）。

ものおもひ （25）
 ものをもひ （1）
 もの思 （9）
 もの思ひ （8）
 物おもひ （10）
 物思 （2）
 物思ひ （4）

また、「らむ」という助動詞（終止形・連体形）は、「らん」、「らむ」、「覽」の3通りの方法で表記されているが、漢字による表記が見られるのは『源氏』54巻の中で、「(35)若菜下」、「(39)夕霧」、「(49)宿木」、「(51)浮舟」のわずか4巻だけである。（なお、54巻中、「(42)匂宮」と「(43)紅梅」の2巻には、「らむ」という助動詞（終止形・連体形）は、いずれ

の表記法でも出現しない。）

このように同じ言葉でも表記が異なる原因として

- ・同じ著者が異なる表記を用いた
- ・著者が異なる
- ・書写者が表記を変えた、あるいは補筆した

などが考えられる。これまで『源氏物語』に関して表記の違いを詳細に調べた研究はないように思えるが、種々の言葉について2.2のように計量的な観点から分析を行うことで、複数筆者説や書写者による補筆部分、あるいは写本の系統などが明らかになる可能性がある。

3.3 高頻度単語の各巻での出現率

単語の出現頻度に注目して計量分析を行う場合、出現頻度の大きい単語は出現率の信頼性が高いという意味で分析に適している。

表2は、『源氏物語』54巻における使用頻度の多い自立語、上位20語とその出現率を示したものである。ここでは、最も使用頻度の高い自立語「いと」（副詞）について調べてみる。この「いと」の使用率が、紫式部の場合に0.01978であったと仮定して、各巻での「いと」の出現率の99%信頼区間を推定し、実際に、各巻の「いと」の使用率がこの区間内に入っているか否かを調べた結果が図3である。

順位	語(索引項目の表記)	品詞	度数	度数／自立語総数 (千分率)
1	こと(事)	詞	4225	19.780
2	ひと	詞	3952	18.502
3	ひふへふえは	副詞	3441	16.110
4	ひふへふえは	補名	3212	15.038
5	ひふへふえは	補名	3002	14.055
6	ひふへふえは	補名	2949	13.806
7	ひふへふえは	補名	2185	10.230
8	ひふへふえは	補名	1877	8.788
9	ひふへふえは	補名	1600	7.491
10	ひふへふえは	補名	1548	7.247
11	ひふへふえは	補名	1442	6.751
12	ひふへふえは	補名	1339	6.269
13	ひふへふえは	補名	1203	5.632
14	ひふへふえは	補名	1190	5.571
15	ひふへふえは	補名	1170	5.478
16	ひふへふえは	補名	1053	4.930
17	ひふへふえは	補名	1041	4.874
18	ひふへふえは	補名	977	4.574
19	ひふへふえは	補名	961	4.499
20	ひふへふえは	補名	937	4.387

表 2

「いと(副詞)」の巻ごとの使用率と99%信頼区間

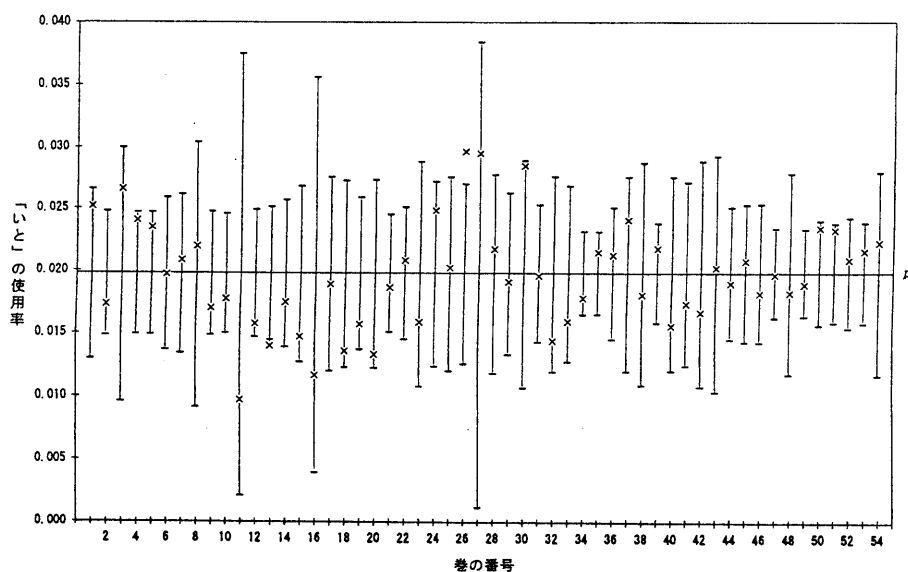


図 3

この図を見ると、「(13) 明石」と「(26) 常夏」の2巻の「いと」の使用率が、99%信頼区間に入っていないことがわかる。

もし、ある特定の巻においてどの単語に関してもその使用率が99%信頼区間に入らないというような事が生じたなら、その巻の著者は紫式部で

はない可能性が出てくる。

そこで同様に、二番目に出現頻度の大きい単語「こと(事)」(名詞)、三番目に出現頻度の大きい単語「ひと」(名詞)の使用率に関して、99%信頼区間を求め、この区間内にそれぞれの使用率が入らない巻を求めてみると、

「こと（事）」に関しては

「(5) 若紫」「(6) 末摘花」
「(7) 紅葉賀」「(14) 澤標」
「(19) 薄雲」「(23) 初音」
「(28) 野分」「(29) 行幸」
「(52) 蜻蛉」

「ひと」に関しては

「(10) 賢木」「(18) 松風」
「(34) 若菜上」「(44) 竹河」
「(50) 東屋」「(51) 浮舟」
「(52) 蜻蛉」「(53) 手習」

となっている。このように、出現頻度の大きい3つの自立語に関しては、単語ごとに99%信頼区間外に落ちる巻が異なっている。

従って、この3語の分析では、『源氏物語』の中のある特定の巻が紫式部以外の手によるという可能性は見いだせない。

ただ99%信頼区間外に使用率が落ちる巻が理論的には1巻以下であるはずなのに、「いと」に関しては2巻、「こと（事）」に関しては9巻、「ひと」に関しては8巻と異常に多い点については、信頼区間を求める際の二項分布の正規近似に問題があるのか、あるいは、かなり多くの巻が他の人物によって書かれている等が原因で、紫式部の「いと」、「こと（事）」、「ひと」の使用率が用いた数値と大きく異なっているというようなことなのか検討を要する。

3.4 「宇治十帖」と前半の44巻との比較

前半の44巻の中で、15巻以上の巻には用いられているが「宇治十帖」には現れない単語（自立語）は表3の通りである。また、後半の10巻、いわゆる「宇治十帖」には用いられているが、前半の44巻には現れない単語（自立語）は表4の通りである。

「宇治十帖」他作家説を検討する際には、このような単語に関する情報の分析が必要となる。

単語	品詞	44巻までの出現回数	出現する巻数
おととの君	名詞	34	15
御方方	名詞	54	23
けう	名詞	25	15
そこら	副詞	26	18
御たいめん	名詞	19	15
になく	形容詞	21	15

表3

4. おわりに

『源氏物語』の計量分析を進める上で、どのような単語情報に注目しているかを紹介してきたが、付属語に関するデータが完全になれば、さらに詳細な検討が可能になる。各付属語の出現度数は大きく、文体研究にとって付属語の使用法の検討は不可欠と考えられるので、付属語の情報を加えてさらに分析を進めたいと考えている。

	あな(穴)[名詞]	あらましき(荒)[形容詞]	うち(宇治)[名詞]
前半の44巻	0	0	0
45	0	3	3
46	1	0	4
47	0	2	0
48	1	0	0
49	2	2	3
50	0	1	3
51	1	1	9
52	0	0	0
53	1	0	2
54	0	0	1
計	6	9	25

表4