

## 源氏物語成立過程に関するコンピュータ解析

琉球大学 工学部 上 田 裕 一  
 統計数理研究所 村 上 征 勝  
 上 田 英 代

近年、国語学・国文学の分野でのコンピュータ利用はめざましく、各研究機関でも様々な文献の本文データ・ベース作りや、語彙索引の自動作成など試みられている。又、古来より「源氏物語」に関する諸研究は夥しい数にのぼっている。こうしたこれまでの研究成果をふまえた上で、「源氏物語」の成立過程、複数作家説をコンピュータにより解析、検証を行ってみたい。今回の発表は、その途中経過報告である。

## 〈コンピュータ解析の目的〉

「源氏物語」は、作品構成、人物描写、心理描写等々の点に於て、他に類をみない長編であるため、古来より紫式部一人による作品ではなく複数の作家によって書かれたとする説があり、特に作品中の和歌が巧みになる宇治十帖に関しては娘の大式三位著作説などが有力である。これに関しては最初の統計分析が昭和32年、安本美典氏（現産能大）によって行われた。

更に和辻氏は、作品自体の構図の弱さを指摘して、同一作家によることへの疑問や、執筆順序への疑問より、「源氏物語」が序々に拡大された物語だとした。又作品への人物の登場の仕方や心理描写の違いなどを分析することによって各巻の執筆順序が、従来の巻順通りではないことを指摘した青柳説、武田説などについて論議がかわされたが、未だ結論はでていない。又、この論を発展させ、単に巻毎の執筆順序を検討するのみでなく『夕顔』の巻の構成を調べることによって巻の内部ですら執筆順が違うことを述べた甲斐説等もある。我々も巻内部の各節の執筆順が異なることを主張している。

これらの諸説について、コンピュータによる詳細な計量分析を行うことによって結論を求めてみたい。

## 〈解明の方法と途中経過〉

## イ) 機械可読文献の作成

データ入力に際してはOCR（富士電機 XP-50S）を使って読み取り、テキストは池田亀鑑編著「源氏物語大成」を使った。「源氏物語大成」を使った理由は、諸本による校異が整っていること、語彙索引が完備していること、平がなによる表記が多く原文に近いこと等による。

機械読み取りのミスはワープロで修正し、多くの空行の削除、余分な部分の読み取りの削除等も行った。

次に「日本古典文学大系」を参照して、テキストに句点をつけていった。文の長さが重要な情報の一つとなるからである。

## ロ) 品詞分解とデータ解析

イ) でできたデータベースを単語分割してゆき、更に別の修正エディタにかけて、一語ずつ品詞コードをつけていった（図1, 2, 表1）。

これを解析プログラムにかけた結果が、図3, 4, 5, 6である。これを全部の巻にかけてゆく。

## 〈今後の展開〉

前記までで、各巻に必要な文法情報は整備されるので、これをもとに、各巻の特徴を計量分

ねられたまはぬまゝには我はかく人にくまれてもならばぬをこよひなむはし  
 めてうしとよをおもひしりぬれははつかしくてなからふましようこそおもひなり  
 ぬれなどのたまへはなみたをさへこほしてふしたりいとらうたしとおほすてさ  
 くりのほそくちいさきほとかみのいとなかゝらさりしけはひのさまかよひたる  
 もおもひなしにやあはれなりあなちにかゝつらひたとりよらむも人わろかる  
 へくまめやかにめさましとおほしあかしつゝれいのやうにものたまひまつはさ  
 す夜ふかういてたまへはこのこはいといとをしくさうしと思ふ女もなみ  
 ならずかたはらいたしと思ふに御せうそもたえてなしおほしこりにける  
 と思にもやかてつれなくてやみ給なましかはうからまし。しみていとをしき御ふ  
 るまひのたえさらむもうたてあるへし。よきほとにかくてとちめてんとおもふも  
 のからたゝならすなめかちなりきみは心つきなしとおほしなからかくてはえ  
 やむましよう御こゝろにかゝり人わろくおもほしわひてこきみにいとつらうも  
 れたうもおほゆるにしみておもひかへせと心にしもしたかはすくるしきをさり  
 ぬへきおりみてたいめむすへくたはかれとのたまひわたればわつらはしけれと

14 13 12 11 10 9 8 7 6 5 4 3 2 1

図 1.

1 ね られ たまは ぬ まゝ に 我 は かく 人 に にくま れ て も なら は ぬ を こよひ なむ はしめて  
 2 う し と よ を おもひ しり ぬれ は はつかしく て なから ふ ましよう こそ おもひ なり  
 3 ぬれ など のたまへ は なみた を さへ こほし て ふし たり 。 いと らうたし と おほす 。 てさくり  
 4 の ほそく ちいさき ほと かみ の いと なかゝら さり し けはひ の さま かよひ たる  
 5 も おもひなし に や あはれなり 。 あなちにかゝつらひ たとり よら む も 人わろかる  
 6 へく まめやかに めさまし と おほし あかし つゝ れい の やう に も のたまひ まつはさ  
 7 す 。 夜 ふかう いて たまへ は この こ は いと いとをしく さうし し と 思ふ 。 女 も なみならず  
 8 ず かたはらいたし と 思ふ に 御 せうそこ も たえ て なし 。 おほし こり に ける  
 9 と 思 に も やかて つれなく て やみ 給 なまし かは うから まし 。 しみて いとをしき 御 ふるまひ  
 10 の たえ さら む も うたて ある へし 。 よき ほと に かくて とちめ てん と おもふ ものから  
 11 たゝなら す なめかち なり 。 きみ は 心つきなし と おほし なから かくて は え  
 12 やむ ましよう 御 こゝろ に かゝり 人わろく おもほし わひ て こきみ に いと つらう も うれたう  
 13 も おほゆる に しみて おもひ かへせ と 心 に し も したかは す くるしき を さり  
 14 ぬ へき おり み て たいめむ す へく たはかれ と のたまひ わたれ は わつらはし けれ と

図 2.

表1.

番号	行	品詞	語長	かな漢字	語長	カタカナ
1	1	動詞	1	ね	1	ネ
2	1	助動詞	2	られ	2	ラレ
3	1	動詞	3	たまは	3	タマハ
4	1	助動詞	1	ぬ	1	ヌ
5	1	普通名詞	2	まゝ	2	マ?
6	1	助動詞	1	に	1	ニ
7	1	代名詞	1	我	2	ワレ
8	1	助動詞	1	は	1	ハ
9	1	副詞	2	かく	2	カク
10	1	普通名詞	1	人	2	ヒト
11	1	助動詞	1	に	1	ニ
12	1	動詞	3	にくま	3	ニクマ
13	1	助動詞	1	れ	1	レ
14	1	助動詞	1	て	1	テ
15	1	助動詞	1	も	1	モ
16	1	動詞	3	ならは	3	ナラハ
17	1	助動詞	1	ぬ	1	ヌ
18	1	助動詞	1	を	1	ヲ
19	1	普通名詞	3	こよひ	3	コヨヒ
20	1	助動詞	2	なむ	2	ナム
21	1	副詞	4	はしめて	4	ハシメテ
22	2	形容詞	2	うし	2	ウシ
23	2	助動詞	1	と	1	ト
24	2	普通名詞	1	よ	1	ヨ
25	2	助動詞	1	を	1	ヲ
26	2	動詞	3	おもひ	3	オモヒ
27	2	動詞	2	しり	2	シリ
28	2	助動詞	2	ぬれ	2	ヌレ
29	2	助動詞	1	は	1	ハ
30	2	形容詞	5	はつかしく	5	ハツカシク

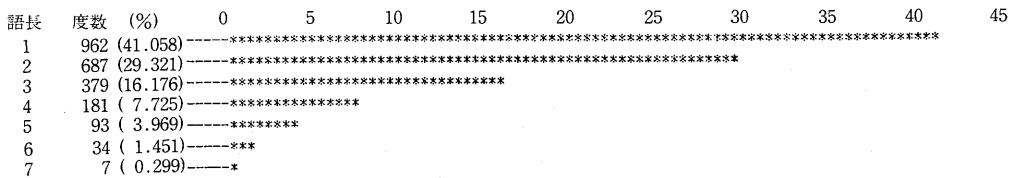


図3. 単語の長さ(文字の数, かな漢字)のヒストグラム. 全単語数=2343.

語数	度数 (%)	0	5	10	15	20	25	30	35
1 --- 5	10 ( 7.752)	*****							
6 --- 10	21 (16.279)	*****							
11 --- 15	33 (25.581)	*****							
16 --- 20	24 (18.605)	*****							
21 --- 25	18 (13.953)	*****							
26 --- 30	8 ( 6.202)	*****							
31 --- 35	7 ( 5.426)	*****							
36 --- 40	0 ( 0.0 )	-----							
41 --- 45	2 ( 1.550)	***							
46 --- 50	2 ( 1.550)	***							
51 --- 55	0 ( 0.0 )	-----							
56 --- 60	3 ( 2.326)	*****							
61 --- 65	0 ( 0.0 )	-----							
66 --- 70	1 ( 0.775)	**							

図4. 文の長さ(単語の数)のヒストグラム. 文の数=129.

番号	度数	かな漢字	カタカナ	(%)	0	10	20	30	40
1	1	いよ	イヨ	( 7.143)	*****				
2	1	伊与の介	イヨノスケ	( 7.143)	*****				
3	1	うつせみ	ウツセミ	( 7.143)	*****				
4	2	きのかみ	キノカミ	(14.286)	*****				
5	3	こきみ	コキミ	(21.429)	*****				
6	4	こ君	コキミ	(28.571)	*****				
7	1	にしの君	ニシノキミ	( 7.143)	*****				
8	1	二条院	ニジャウイン	( 7.143)	*****				

図5. 固有名詞. 最大度数=4. 総度数=14.

品詞	度数	0	5	10	15	20	25	30	
名詞	436 (0.186)	*****							
動詞	525 (0.224)	*****							
形容動詞	60 (0.026)	*****							
助詞	699 (0.298)	*****							
接頭語	31 (0.013)	***							
接尾語	5 (0.002)	-----							
形容詞	157 (0.067)	*****							
感動詞	2 (0.001)	-----							
助動詞	270 (0.115)	*****							
連体詞	30 (0.013)	***							
副詞	124 (0.053)	*****							
接続詞	4 (0.002)	-----							

図6. 品詞の度数(接?語を除かない, 名詞を区別しない). 総度数=2343.

析する。前半44帖と宇治十帖の違いを明らかにし、宇治十帖他作家説、安本説を検証する。更に、青柳、武田説の検証、甲斐、上田説の検証を行う。その際に必要とされる文法情報は更に整備してゆくつもりである。