

平安時代の文献における文の長さについての計量分析

土山 玄[†] 村上 征勝^{††}

本研究では、古典文学作品として著名な『源氏物語』と『宇津保物語』を分析対象とし、文の長さの分布を計量的に分析することで、両作品の各巻が作品別に分類されるのか検討を加えた。このような分析を行う背景に、欧米では古くは19世紀から文の長さを分析項目として、書き手の識別や同定を目的とした研究が行われていたが、その一方で、日本語で記述された文献については未だ十分に研究されていないことによる。本研究における分析の結果、文の長さの分布は日本語の文献においても、書き手の相違に起因すると考えられる文体的相違が認められた。

Quantitative Analysis of Sentence-Length in Classical Japanese Literatures

GEN TSUCHIYAMA[†] MASAKATSU MURAKAMI^{††}

We investigate some style differences between authors of Japanese classical literatures by using statistics method. We analyze sentence length, which is one of the classic textual measurements. For the analysis, we consider two works, i.e., "Tale of Genji" and "Tale of Utsubo," which were written in the Heian period (794-1185); both original manuscripts were lost and hence only passed on as transcriptions. Therefore, we use the revised scripts for the analysis. As the result of principal component analysis for sentence-length, we conclude that there are noticeable differences of sentence-length distribution between classical authors.

1. 研究背景と目的

計量文献学において、書き手の同定あるいは識別を目的とした研究は広く行われており、計量的な分析を行う上で、様々な文体的特徴が分析項目として採り上げられている。本研究で採り上げる文の長さは計量が容易な分析項目であることから、計量的な文体研究の草創期から分析に用いられている。文の長さをを用いた研究の歴史は古く、Sherman(1888)[1]まで遡り、書き手によって文の長さの平均値が相違することが報告されている。以降、欧米の言語で記述された文章においては、文献の書き手の同定を目的とした研究に文の長さを分析にたびたび用いられている。

一方、日本語で記述された文章においても、文の長さを分析項目として用いた計量的な研究は報告されている。しかし、書き手による文体的相違を明らかにするのはなくではなく、新聞や小説といった文章を対象とし、ジャンルによって相違する文体特徴としての文の長さの分布を採り上

げている研究や、文の長さの分布の正規性についての研究がその大部分を占めているのが現状である。このように、書き手の同定や識別のために文の長さをを用いた研究は管見に入らない[2]。

そこで、本研究は日本文における文の長さの分布を用いて書き手の識別を行うことを目的とする。分析対象は『源氏物語』および『宇津保物語』である。『源氏物語』は平安時代(794-1185)に成立した現存最古に類する和文体の長編物語であり、作者は紫式部(973-1014)であるとされる。『源氏物語』は54巻で構成され、各巻に1から54の番号が付されているが、これは物語において叙述される出来事の時系列に合うように、後人の手によってなされたものであり、各巻の成立順序に関しては様々な説が提起されている[3][4][5]。

また、一般に「宇治十帖」と称される、主人公である光源氏の死後を描いた『源氏物語』の最後の10巻や、「宇治十帖」の直前の「匂宮三帖」と称される3巻については、紫式部の娘である大式三位(生没年未詳)の作であるとする説も提起されている[6]。このように『源氏物語』には、その文体を対象とすることで、計量的なアプローチが有効な研究方法となり得る論題があると言える。

[†] 同志社大学大学院文化情報学研究所博士課程
日本学術振興会特別研究員
Doshisha University
Research Fellow (DC2), Japan Society for the Promotion of Science
^{††} 同志社大学文化情報学部
Doshisha University

本研究では、これらの『源氏物語』が内包する問題を議論する前に、上述の『宇津保物語』を比較対象とし、計量的な分析手法の有効性および妥当性を示す。『宇津保物語』を比較対象とする背景は、(1)平安時代に成立し、(2)和文体で記述された(3)長編物語である、という3つの共通性に基づく。なお、『源氏物語』に『宇津保物語』の引用があることから、成立は『源氏物語』より『宇津保物語』の方が早かったと考えられる。また、『宇津保物語』の作者については、源順とする説があるが不詳である[7]。

2. 資料

『源氏物語』の自筆原稿は散逸しており、現在においては写本によって受け継がれるのみである。一般に、『源氏物語』の写本系統は青表紙本系、河内本系、別本と3系統に大別される。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』[8]および『源氏物語語彙用例総索引付属語編』[9]を電子化したデータベースを分析に利用した。『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』[10]の単語認定基準に準拠している。

同様に、『宇津保物語』の自筆原稿も散逸しており、写本によって継承されている。本研究では尊経閣蔵前田家十三行本を底本とする『宇津保物語 全 改訂版』[11]を『源氏物語』と同一の単語認定基準を用いて構築されたデータベースを分析に使用した。

3. 関連研究

古典文には句読点が付与されておらず、文末が明確ではない。そこで文を規定する上で文末表現が非常に重要な要素となる。平安時代に著された文献における文末表現についての研究は、人文学においては数多くあるが、計量的なアプローチによる研究は十分に展開されているとは言えない。

古典文の文末表現についての計量分析は第30回日本行動計量学会大会の発表論文抄録集に所収されている「『源氏物語』の文頭文末の計量分析」[12]が本研究と関連する研究として挙げられる。「『源氏物語』の文頭文末の計量分析」[12]においては、『日本古典文学大系 源氏物語』[13]の本文に付与されている句点を参考にし、これによって規定される文頭の語彙と文末の語彙を用い数量化Ⅲ類を行っている。分析の結果、「宇治十帖」など内容につながりのある諸巻については文頭および文末の語彙の用法が類似していることを報告している。

また、『源氏物語』と『宇津保物語』を分析対象とし、計量的なアプローチによって、両作品の文体を言及した研究として、土山・村上[13]があげられる。土山・村上[13]では語彙の出現率について、本研究と同様に主成分分析を行っ

ている。分析に際しては、巻ごとに名詞・動詞・形容詞・形容動詞・副詞・助詞・助動詞といった7品詞の語彙の出現率を品詞別に集計し、主成分分析を行っている。分析の結果、『源氏物語』と『宇津保物語』の両作品の間において、語彙の使用傾向に相違が認められることを報告している。

4. 分析

4.1 分析項目と分析対象

先にふれたように、『源氏物語』をはじめとする古典文には句読点は付与されていない。しかし、『日本古典文学大系 源氏物語』(岩波書店)[14]、『新 日本古典文学大系 源氏物語』(岩波書店)[15]や『新編日本古典文学全集 源氏物語』(小学館)[16]、『新潮日本古典集成 源氏物語』(新潮社)[17]といった著名な出版社から刊行されている『源氏物語』の本文には読みやすさという観点から句読点が付与されているが、出版社によっては句読点の位置が異なることもしばしばある。これは『源氏物語』に限らず、他の古典文学作品についても同様である。

一例として、『日本古典文学大系 源氏物語』[14]および『新 日本古典文学大系 源氏物語』[15]における『源氏物語』の書出の巻である第1巻「桐壺」の冒頭を示す。

『日本古典文学大系 源氏物語』[14]

いつれの御時にか。女御・更衣あまたさぶらひ給ひけるなかに、いと、やむごとなき際にはあらぬが、すぐれて時めき給ふありけり。

『新 日本古典文学大系 源氏物語』[15]

いつれの御時にか、女御、更衣あまたさぶらひ給ひける中に、いとやんごとなき際にはあらぬがすぐれてときめき給ふ有りけり。

上掲の通り、前者は2文に分割されているが、後者は1文である。

そこで、本研究では動詞・形容詞・形容動詞・副詞・助動詞といった活用品詞の終止形、および終助詞・間投助詞を文末表現と定義する。したがって、これに基づいて、文の長さの分布を求めた。また、本研究では文の長さを1文に含まれる文字数とするのではなく、1文に含まれる単語数に基づき算出した。

『源氏物語』は54巻で構成され、他方『宇津保物語』は20巻で構成される。本研究では、主に1つの巻を1つの分析対象と捉え、計量的な分析を行った。分析対象となる各巻は、延べ語数に大きなばらつきがあることから、文の長さの分布を求める際に、傾度を用いるのではなく、延べ語数を分母とする比率を用いた。

4.2 分析方法

分析においては、主成分分析を用いた。主成分分析とは、

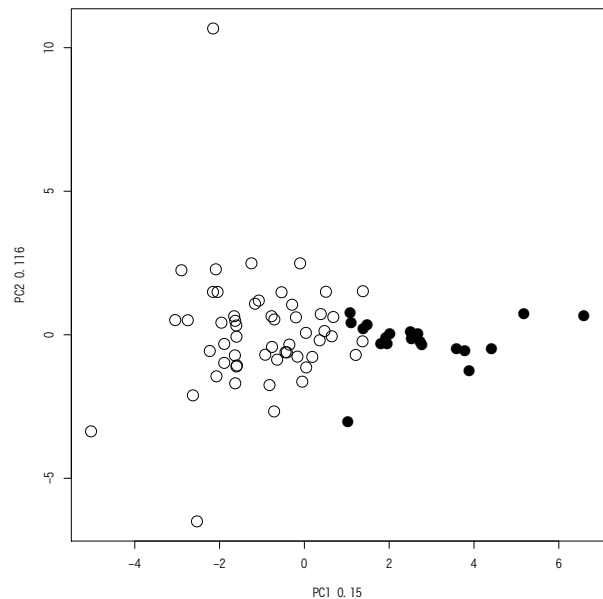
多次元データに対する次元縮約の手法であり、互いに無相関の合成変数である主成分を新たに求めることで、情報の縮約を行う手法である。各主成分にデータ全体の情報がどのくらい含まれているかは寄与率によって評価される。本研究では、情報量の多い第1主成分と第2主成分を用いた。また、主成分分析には、分析に用いる変数を標準化し分析を行う相関係数行列に基づく主成分分析と、標準化を行わない分散共分散行列に基づく主成分分析があるが、本研究では前者の主成分分析を行った。

ついで、文の長さの分布を分析するにあたって、文の長さを変数とする主成分分析を用いた。これはすなわち、単語数1の文の出現率、単語数2の文の出現率、といったようにこれらが分析に用いる変数となる。

4.3 結果と考察

『源氏物語』および『宇津保物語』の全巻を用いて主成分分析を行ったところ、図1および図2に示すように『源氏物語』の第11巻「花散里」や第27巻「篝火」が外れて付置された。図1および図2は出現率の高い30変数を用いて行った主成分分析であり、第1主成分の寄与率は0.150、第2主成分の寄与率は0.116である。また、変数を増減させて分析した場合においても、寄与率こそ変動するが、分析結果は大きく変わらない。なお、図1に描かれているベクトルは第1主成分と第2主成分の因子負荷量を可視化したものである。

語数がきわめて少なく、第11巻「花散里」が724語、第27巻「篝火」が653語である。このようなことから、2000語を閾値とし、延べ語数が2000語を下回る巻を分析対象から除外し、あらためて主成分分析を行う。該当する対象は第11巻「花散里」、第16巻「関屋」、第27巻「篝火」の3巻であり、第16巻「関屋」の延べ語数は934語である。なお、『源氏物語』および『宇津保物語』の両作品を通じて、これら3巻に次いで延べ語数が少ない対象は『源氏物語』の第3巻「空蟬」の2187語であるから、これら3巻が際立って延べ語数が少ないことが分かる。



○：源氏物語 ●：宇津保物語
 図2 全巻を用いた主成分分析の結果

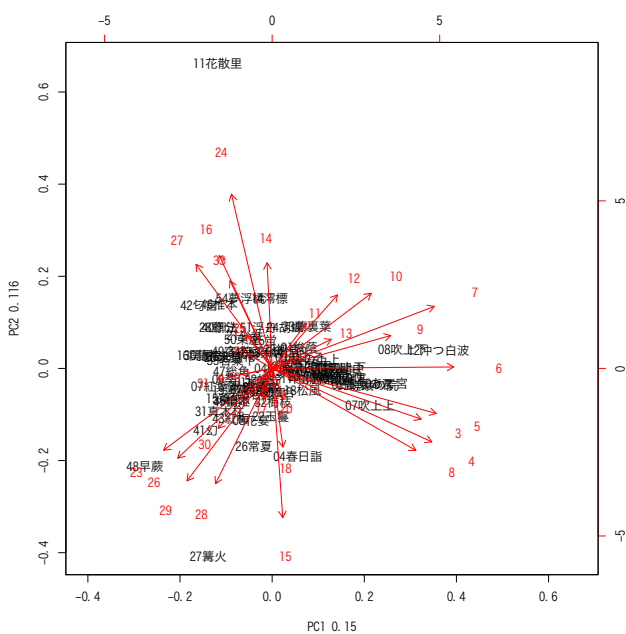


図1 全巻を用いた主成分分析の結果

次に、図2は図1の散布図の解釈を容易にするために分析対象のみを付置した散布図である。図2より、分析対象はおおよそ作品別に位置していることが認められ、作品によって文の長さが相違すると言える。

上記の『源氏物語』における2巻は他の巻に比べ、延べ

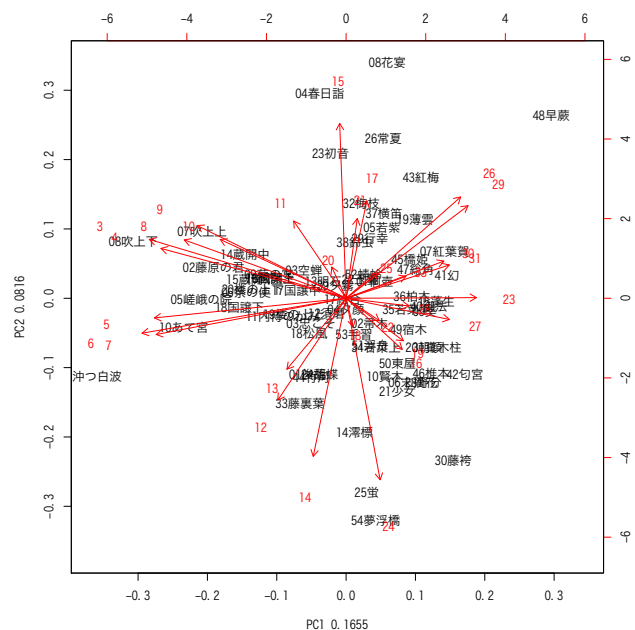


図3 出現率上位30変数に対する主成分分析の結果

図3は、図1および図2と同様に、出現率の高い30変数を用い、上記の3巻を削除し、すなわち71の対象に対して行った主成分分析である。第1主成分の寄与率は0.166、第2主成分の寄与率は0.082となっており、第2主成分までの累積寄与率は0.247となるから、図1および図2に示した主成分分析の結果よりやや累積寄与率が高い。

次に、図2と同様に分析対象のみ付置した散布図を図4に示す。

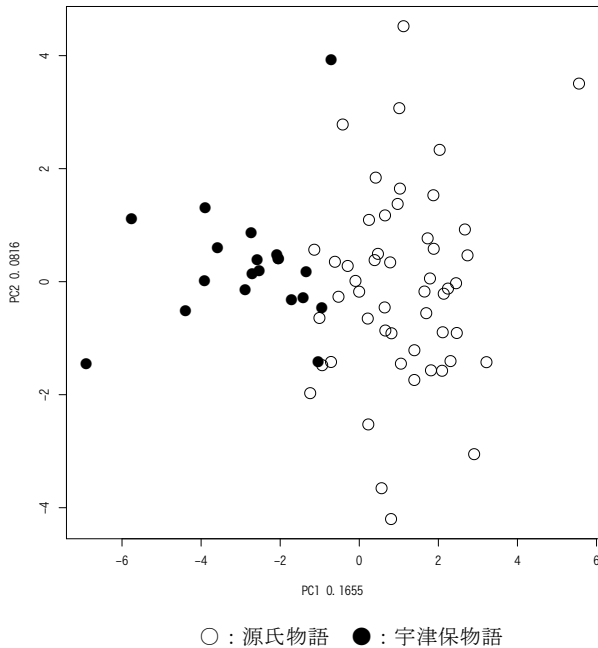


図4 出現率上位30変数に対する主成分分析の結果

図4から第1主成分において、両作品が別れていることが分かる。図3および図4で示した第1主成分と第2主成分の因子負荷量は表1に示す通りである。

表1 出現率上位30変数の因子負荷量

	PC1	PC2		PC1	PC2
3	-0.3291	0.1401	18	0.0119	-0.0679
4	-0.3096	0.1187	19	0.0958	-0.1013
5	-0.3205	-0.0468	20	-0.0240	0.0743
6	-0.3412	-0.0831	21	0.0186	0.1896
7	-0.3176	-0.0863	22	0.0557	-0.0534
8	-0.2705	0.1390	23	0.2181	0.0016
9	-0.2490	0.1732	24	0.0569	-0.4321
10	-0.2105	0.1401	25	0.0540	0.0600
11	-0.0876	0.1840	26	0.1916	0.2412
12	-0.1142	-0.2427	27	0.1727	-0.0505
13	-0.0990	-0.1689	28	0.1053	-0.0184
14	-0.0549	-0.3764	29	0.2038	0.2204
15	-0.0107	0.4164	30	0.1635	0.0886
16	0.0938	-0.1215	31	0.1726	0.0793
17	0.0347	0.2320	33	0.0999	0.0521

図3における因子負荷量のベクトルおよび表1の第1主成分の因子負荷量から、26, 27, 28, 29, 30, 31, 33といった長い文を表す変数が『源氏物語』を特徴づけ、3, 4, 5, 6, 7, 8, 9, 10という短い文を表す変数が『宇津保物語』を特徴づけていると考えられる。したがって、『源氏物語』は『宇津保物語』に比べ1文に含まれる単語量が多く、『宇津保物語』は1文に含まれる単語量が少ないと言える。

図3および図4に示した分析結果は出現率上位30変数を用いた主成分分析の結果であるが、変数を増減させた場合も分析結果は同様である。

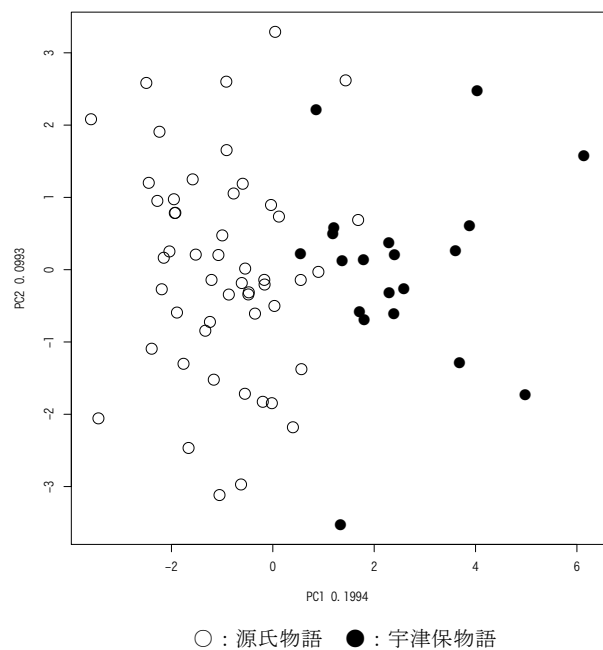


図5 出現率上位20変数に対する主成分分析の結果

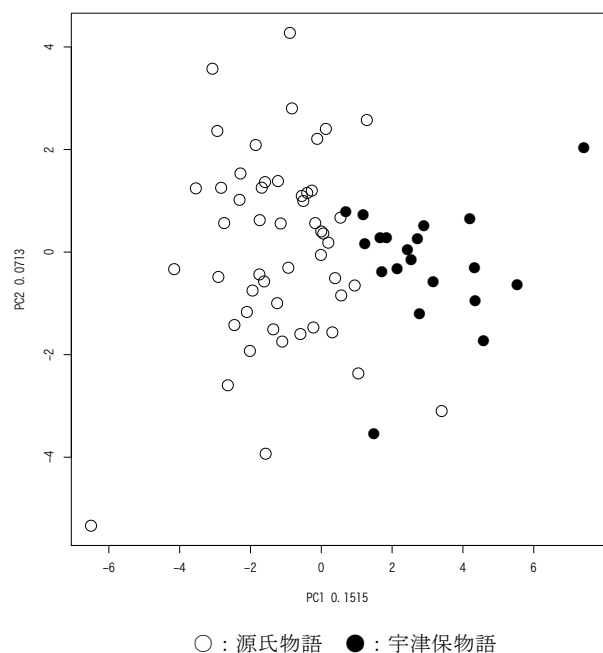


図6 出現率上位40変数に対する主成分分析の結果

出現率上位 20 変数を用いた主成分分析の結果を図 5, 出現率上位 40 変数を用いた主成分分析の結果を図 6 に示す。第 1 主成分と第 2 主成分の累積寄与率は順に 0.299, 0.223 となり, 出現率上位 30 変数を用いたときと大きく変わらない。図 5 および図 6 において、『源氏物語』の諸巻は図 4 と異なり第 1 主成分の負の領域に付置される。

次いで, 出現率上位 20 変数と出現率上位 40 変数を用いたときの主成分分析における第 1 主成分と第 2 主成分の因子負荷量を, 順に表 2, 表 3 に示す。

表 2 出現率上位 20 変数の因子負荷量

	PC1	PC2		PC1	PC2
3	0.3631	-0.1440	13	0.1048	0.3232
4	0.3468	-0.1145	14	0.0176	0.3051
5	0.3732	0.1488	15	0.0519	-0.5471
6	0.3861	0.1953	16	-0.1291	0.1089
7	0.3351	0.1733	17	-0.0368	-0.2610
8	0.3193	-0.1274	18	-0.0343	0.0653
9	0.3163	-0.1244	19	-0.1520	0.0085
10	0.2531	-0.1934	20	0.0339	-0.1531
11	0.0945	-0.1171	21	-0.0314	-0.2029
12	0.1014	0.3836	22	-0.0600	-0.0004

表 3 出現率上位 40 変数の因子負荷量

	PC1	PC2		PC1	PC2
2	0.2569	-0.0585	22	-0.0607	0.0404
3	0.2842	-0.0856	23	-0.1904	-0.0879
4	0.2798	-0.0861	24	-0.0488	0.3112
5	0.2641	0.0936	25	-0.0547	0.0467
6	0.3115	0.0623	26	-0.1569	-0.2458
7	0.3021	0.0332	27	-0.1536	0.0519
8	0.2364	-0.0904	28	-0.0939	0.0831
9	0.2208	-0.1154	29	-0.1794	-0.2826
10	0.1595	-0.0473	30	-0.1305	-0.0947
11	0.1163	-0.2223	31	-0.1472	-0.1701
12	0.1324	0.1163	32	-0.0949	0.0663
13	0.1022	0.0978	33	-0.0929	0.0079
14	0.0202	0.3941	34	-0.1371	0.0504
15	-0.0002	-0.2859	35	-0.1558	0.1359
16	-0.0652	-0.0396	36	-0.1406	0.0463
17	-0.0038	-0.2544	37	-0.1046	0.2943
18	0.0108	0.0418	38	-0.1978	-0.0758
19	-0.0641	0.1176	39	0.0182	0.0463
20	0.0288	-0.1030	40	0.0382	0.3250
21	0.0137	-0.1353	42	-0.1729	-0.0064

表 2, 表 3 において, 第 1 主成分が負の値になる変数は 1 文に含まれる単語量が多いことを意味する変数であること

から、『源氏物語』は長い文が多く, その一方で『宇津保物語』は短い文が多いと言える。

『源氏物語』の文が相対的に長いということについて, 『源氏物語』および『宇津保物語』の文末表現に着目すると, 『宇津保物語』は動詞の終止形で文が終わることが多い。『源氏物語』の文末表現における動詞の出現率は 19.8% に対し, 『宇津保物語』の文末表現における動詞の出現率は 25.7% に及ぶ。その一方で, 補助動詞の文末表現における出現率は『源氏物語』が 25.2%, 『宇津保物語』が 20.7% となり, 動詞の文末表現における出現率と大小関係は逆転する。補助動詞とは, 一般に他の動詞に後接する動詞であるから, 文末において補助動詞の出現率が高くなる場合, 必然的に文の長さが長くなることが推測される。したがって, ここに『源氏物語』と『宇津保物語』の文の長さについて相違が生じる根拠の 1 つを見出すことができる。

5. まとめ

『源氏物語』の文が『宇津保物語』に比べ長くなることについて, 文末表現が単純な形式ではないことが要因の 1 つであると考えられる。今後, 『源氏物語』の文末表現を計量的に分析することで, 十分に研究が展開されていない文に関わる『源氏物語』の, 計量的な観点による文体的特徴がより明らかになるものと考えられる。

参考文献

- [1] Sherman, L. A. Some Observations upon the Sentence-Length In English Prose. University [of Nebraska] Studies, Vol.1, pp.119-130 (1888).
- [2] 計量国語学会 (編): 計量国語学事典, 朝倉書店 (2009).
- [3] 青柳秋生: 源氏物語執筆の順序-若紫の巻前後の諸帖に就いて-, 国語と国文学, Vol.16, No.8-9 (1939).
- [4] 武田宗俊: 源氏物語の最初の形態, 文学, Vol.18, No.6-7 (1950).
- [5] 風巻景二郎: 源氏物語の成立に関する試論, 北海道大学文学部紀要, Vol.4 (1955).
- [6] 小林栄子: 源氏・伊勢物語新研究, 晋文館 (1935).
- [7] 稲賀敬二: 宇津保物語は合作か? (阿部秋生編, 『講座日本文学の争点(二) 中古編』所収), 明治書院 (1968).
- [8] 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).
- [9] 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).
- [10] 池田亀鑑: 源氏物語大成 索引篇, 中央公論社 (1985).
- [11] 室城秀之: うつほ物語 全 改訂版, おうふう

(1995).

[12] 村上征勝, 今西祐一郎, 谷岡典子 : 『源氏物語』 の文頭文末の計量分析 (言語の計量分析), 日本行動計量学会大会発表論文抄録集 Vol.30, pp.198-199 (2002).

[13] 土山玄, 村上征勝 : 源氏物語と宇津保物語における語の使用傾向について, 人文科学とコンピュータ論文集, Vol.2011, No.8, pp.125-132 (2011).

[14] 山岸徳平 : 日本古典文学大系 源氏物語, 岩波書店 (1958).

[15] 柳井滋, 室伏信助, 大朝雄二, 鈴木日出男, 藤井貞和, 今西祐一郎 : 新 日本古典文学大系 源氏物語, 岩波書店 (1993).

[16] 阿部秋生, 秋山虔, 今井源衛 : 日本古典文学全集 源氏物語, 小学館 (1970).

[17] 石田穰二, 清水好子 : 新潮日本古典集成 源氏物語, 新潮社 (1976).

謝辞

本研究の成果の一部は, 日本学術振興会科学研究費補助金(特別研究員奨励費)『源氏物語』の成立論および多作者説についての計量的研究(代表: 土山玄)(2013-2014 年度)によるものである。