

# 語の使用頻度の計量分析による宇治十帖他作者説の検討

土山玄<sup>†</sup> 村上征勝<sup>††</sup>

古くから他作者説が論じられている『源氏物語』の終わりの10巻である「宇治十帖」について、主成分分析およびランダムフォレストといった多変量解析を行い、「宇治十帖」の作者とそれ以外の諸巻の作者が同一であるかについて、計量的な側面から検討を加える。本研究では、語の使用頻度に対し分析を行い、その結果、「宇治十帖」の他作者説を支持する積極的な根拠は得られなかった。

## Research for the Author of “*Uji Jujo*” by Quantitative Analysis of Word Frequency

GEN TSUCHIYAMA<sup>†</sup> MASAKATSU MURAKAMI<sup>††</sup>

This paper focuses on the authorship of “*Uji Jujo*” which is the closing ten chapters of “*The Tale of GENJI*”. Besides, some researchers consider that their author was not same as the others. The result derived from Principal components analysis and Random forest by word frequency, and we will show that it is obvious different between the chapters and the others. Thus, we will conclude that it is highly possible that the chapters’ author was same as the others.

### 1. 問題の所在

平安時代に成立した『源氏物語』は光源氏を主人公とする全54帖から構成される長編物語であるが、第42巻「匂宮」以降の13巻は光源氏没後のストーリーが描かれている。これらの13巻は、一般に「宇治十帖」と称される諸巻と重なる。この「宇治十帖」について、小林(1935)によると、慶安3(1650)年に一華堂切臨によって著された『源義辨引證』において、「匂宮」より前の諸巻の作者とは異なるとされている。また、小林(1935)によれば、「宇治十帖」は文の調子が他の巻とは異なり、「宇治十帖」の趣向が第41巻「幻」以前と重複することの不自然さが論じられている。このようなことから、「宇治十帖」については他作者説が唱えられることがある。

この他作者説については、計量的な側面からも研究がなされている。「宇治十帖の作者-文章心理学による作者推定-」(安本, 1957)や、これを補説した『文章心理学の新領域』(安本, 1960)、「源氏物語・宇治十帖の作者問題：一つの計量言語学的アプローチ」(新井, 1997)などが先行研究としてあげられる。安本(1957・1960)によれば、「宇治十帖」の文体は他44帖とはやや異なるとされ、他方、新井(1997)によれば、「宇治十帖」の作者は他の諸巻の作者と別人であ

るとは考えられないと論じられている。これらの研究においては、統計的検定が用いられ、検討が加えられている。

そこで、本研究では語の使用頻度に注目し、多変量解析の手法を用いることで宇治十帖他作者説にアプローチする。使用する手法は主成分分析(Principal Components Analysis, PCA)およびランダムフォレスト(Random Forest, RF)である。なお、分析においては、分析手法および分析結果の考察についての妥当性を検証するために、『源氏物語』および同時代に成立した『宇津保物語』、『源氏物語』の続編あるいは補作であると考えられる『山路の露』と『雲隠六帖』を分析対象に加え、これらの分析結果の検討を行う。予め結論を述べると、本研究では「宇治十帖」における他作者説を支持する積極的な分析結果は得られなかった。

### 2. 資料

本研究では古典文について計量的な分析を試みるため、電子化されたデータベースを用いる。データベースのもととなったテキストを以下に示す。

- 源氏物語  
『源氏物語語彙索引』(上田・今西・村上他, 1994)
- 宇津保物語  
『宇津保物語 全 改訂版』(室城, 1995)
- 山路の露  
『日本古典全書 源氏物語7』(池田, 1955)
- 雲隠六帖  
『源氏物語の研究』(長谷川, 1957)

<sup>†</sup> 同志社大学大学院文化情報学研究科博士後期課程  
Doshisha University, Graduate School of Culture and Information Science  
<sup>††</sup> 同志社大学文化情報学部  
Doshisha University

これらのうち『源氏物語語彙総索引』は『源氏物語大成』(池田, 1984)の単語分割に従い、『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。本研究では、『宇津保物語』、『山路の露』、『雲隠六帖』についても、『源氏物語語彙総索引』の形態素解析の規格に従い、それぞれ形態素解析が為され、電子データベース化されているものを用いた。

### 3. 手法

#### 3.1 対象

主たる研究対象は『源氏物語』であるが、先述したように『宇津保物語』、『山路の露』、『雲隠六帖』についても検討を加える。

『源氏物語』は全54帖、『宇津保物語』は全20帖によって構成される。そこで、本研究の分析においてはそれぞれの1帖をひとつのケース(case)と称し、分析を行う。ただし、源氏物語において、第27巻「篝火」の総語数が653語であるのに対し、第35巻「若菜下」が20223語であり、このように各巻の総語数における大きなばらつきが認められるため、総語数が2000語未満の巻は分析から除外することにする。よって、「篝火」に加えて、源氏物語の第11巻「花散里」(724語)、第16巻「関屋」(934語)を分析から除いた。

また、『雲隠六帖』は作品名にあるように6帖より構成されるが、1帖の総語数が2000語を下回るためこれを6分割せず『雲隠六帖』を1つのケースとする。『山路の露』も同様に1つのケースとする。

表1 各品詞の出現率

	出現頻度	出現率
名詞	127712	0.1950
動詞	110296	0.1684
形容詞	33202	0.0507
形容動詞	12556	0.0192
副詞	24920	0.0381
助詞	208322	0.3182
助動詞	72478	0.1107
補助動詞	31751	0.0485

次に、分析は品詞別に行う。分析に使用した変数は表1に示すように、名詞、動詞、形容詞、形容動詞、副詞、助詞、助動詞の7品詞である。これらの品詞は品詞別の出現率を基準として選択した。品詞別の出現率が0.01以上となる品詞を本研究における分析に用いる。

#### 3.2 変数選択

計量的に分析するに際し、語を変数とし語の頻度を求め、

これを分析する。本研究では、分析対象となる作品に現れるすべての語について、その出現頻度を集計し、ケース別に相対頻度を求める。つまり、ケースAの語 $t$ において、 $t$ の出現頻度を $f_t$ とし、ケースAの総語数(tokens)を $S_A$ とすると、 $t$ の相対頻度は $f_t/S_A$ となる。

分析における変数の数は異語数(types)であるため、品詞にもよるが、名詞や動詞はきわめて高次元になることがある。このようなことから、変数選択を行い分析に使用する語数を決定する。

変数選択においては、以下のような条件を設ける。

- (1) 出現頻度が高い
- (2) 出現するケースが多い
- (3) 作品(対象)間での変数の数が大きく偏らない

(1)および(2)については、出現頻度が低く、なおかつ少数ケースにのみ用いられている語は分析におけるノイズになり得ることから、これらを分析から除外する為の条件である。また、(3)は変数選択における閾値を決定する為の条件となる。すなわち、作品(対象)間での変数の数がおよそ同様になるように変数を選択し、分析を行う。

表2 主成分分析の変数

	宇津保物語	山路の露	雲隠六帖	宇治十帖
名詞	40	34	106	106
動詞	128	64	86	321
形容詞	45	66	58	195
形容動詞	89	16	11	65
副詞	69	63	20	53
助詞	53	49	44	33
助動詞	25	28	25	22

表3 ランダムフォレストの変数

	宇津保物語	山路の露	雲隠六帖	宇治十帖
名詞	704	507	499	1039
動詞	550	539	538	769
形容詞	207	228	225	272
形容動詞	89	111	109	112
副詞	69	80	79	87
助詞	49	48	49	49
助動詞	22	22	22	22

また、後述するように分析では、主成分分析とランダムフォレストを用いる。各手法を用いることの目的、および分析手法の特徴の応じ、主成分分析とランダムフォレストでは使用する変数の数が相違する。主成分分析では、(3)の条件において作品(対象)間での変数の数が同数になるように閾値を定めた。他方、ランダムフォレストではより広く特徴語を抽出する為に作品(対象)間での変数の数の

比率が 0.90 から 1.10 になるように定めた。分析に使用する変数の数は表 2 および表 3 に示す通りである。

### 3.3 主成分分析

主成分分析は多次元データに対する次元縮約の手法であり、もとのデータの変数より新たに合成変数を求めることで、情報の縮約を行う。分析結果は主に 2 次元の散布図にケースを付置しケース間の関係を視覚化されたものである。本研究では主成分分析を行うことで、分析対象のクラスタリングを行う。これにより、作品（対象）間に語の使用傾向の相違が認められるか検討する。

### 3.4 ランダムフォレスト

ランダムフォレストとは機械学習 (machine learning) の手法の 1 つであり、CART 法による決定木のアンサンブル学習である。ランダムフォレストは判別や回帰の為の手法であるが、分析の際に、変数の重要度を推定する。この重要度が高い変数ほど判別や回帰において有効であると言える。本研究では推定される重要度を利用し、特徴語の抽出を行う。

『源氏物語』と『宇津保物語』を例とすると、これらの 2 つの作品の判別にもっとも有効とされる変数、つまりもっとも重要度が高く推定された変数が 2 つの作品の間において、使用頻度に顕著な相違が認められる語、すなわち特徴語であると言える。

## 4. 作者の相違についての分析

「源氏物語と宇津保物語における語の使用傾向について」(土山・村上, 2011) において、『源氏物語』および『宇津保物語』を対象とし、分析を行った。用いた分析手法は主成分分析およびランダムフォレストである。なお、『宇津保物語』は『源氏物語』とおよそ同時期に成立したと考えられる現存最古の長編物語のひとつである。

前掲の論文において、古典文を分析対象とする場合も現代文と同様に、語の使用頻度に対する主成分分析によって作者の相違を見いだせることを論じた。次に、ランダムフォレストによる特徴語の抽出から、両作品はともに和文体の作品であると考えられるが、『宇津保物語』の特徴語として漢文訓読文体に属する語彙が抽出されることを示し、『宇津保物語』が『源氏物語』に比べ、漢文訓読文体の影響を強く受けている可能性を論じた。

本研究においても、これと同様に主成分分析およびランダムフォレストを使用する。

## 5. 語の重み付けの有効性

### 5.1 重み付け指標について

後述の宇治十帖他作者説の検討の為に、『山路の露』および『雲隠六帖』を対象とし、語に重み付けを行った分析についての検討を加えた。後世の著述家の手による続編の執筆や『源氏物語』に描かれていないエピソードの補作など

が行われてきた。続編および補作として『山路の露』や『雲隠六帖』が特に有名である。『山路の露』は作者不詳であり、成立は南北朝時代であるとされ、『雲隠六帖』も作者不詳で、成立は室町時代であるとされる。『山路の露』や『雲隠六帖』といった作品は、『源氏物語』のストーリーを意識して著された作品であるから、『源氏物語』の文体と親近性を有するものと考えられる。よって、これら続編や補作と言われる作品の文章を『源氏物語』のそれと比較し、計量的観点から文体の相違に検討を加える。

また、詳細は後述するが、語の使用頻度に対する『源氏物語』と『山路の露』、および『源氏物語』と『雲隠六帖』の主成分分析においては、『源氏物語』と『宇津保物語』の分析において見出せたような作者の相違は認められない。そこで、変数となる各語に重み付けを行い、主成分分析を行った。

本研究において使用した語の重み付け指標は *idf* (inverse document frequency), *Gfidf* (global function inverse document frequency), *Normalisation*, *Entropy* である (Dumais, 1992)。これらはそれぞれ下記の式によって求められる。

- *idf*

$$idf = \log_2 \left( \frac{N}{df} \right)$$

*N* はケースの総数

*df* は語 *t* が出現するケース数

- *Gfidf*

$$Gfidf = \frac{gf_i}{df_i}$$

*gf* は語 *t* が出現する全回数

*df* は語 *t* が出現するケース数

- *Normalisation*

$$Normalisation = \frac{1}{\sqrt{\sum_j tf_j^2}}$$

*tf<sub>j</sub>* はケース *j* における語の頻度

- *Entropy*

$$1 - Entropy = 1 - \left( \sum_j p(t, j) \cdot \log(p(t, j)) \right) / \log(N)$$

*N* はケースの総数

*p(t, j)* はケース *j* における語 *t* の出現する確率

これらの指標によって語に重み付けを行い、主成分分析を行ったところ、idfによる重み付けを行ったときの助詞の分析において作者の相違が認められた。

### 5.2 『源氏物語』と『山路の露』

『源氏物語』と『宇津保物語』に対する分析と同様に語の頻度について主成分分析を行った。図1に示すように、分析の結果、分析に用いた7品詞すべてにおいて文体の相違は認められたなかった。ゆえに、『山路の露』は『源氏物語』の語の使用頻度と類似していると言える。

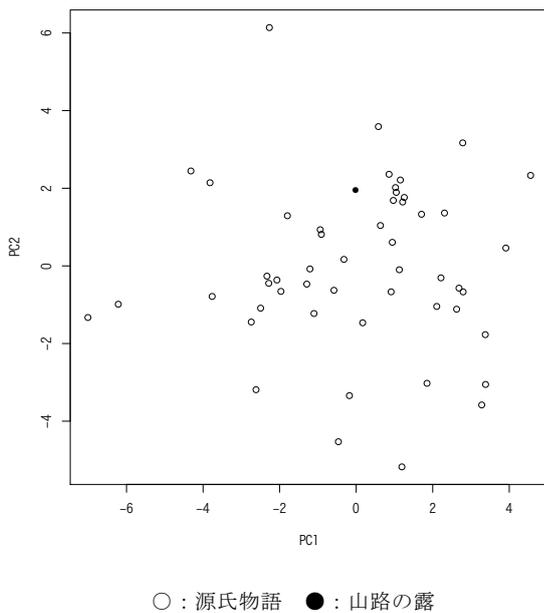


図1 助詞の相対頻度に対する主成分分析の結果

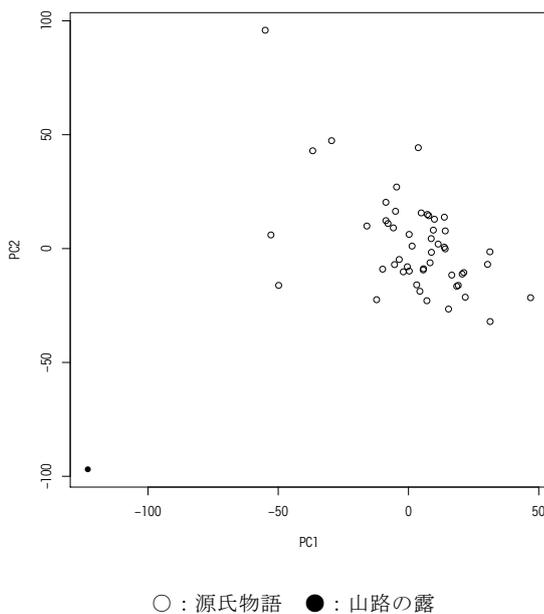


図2 助詞のidfを用いた主成分分析の結果

次いで、上記の重み付け指標を用い主成分分析を行った。図2に示したように、分析の結果、idfを用いた重み付けによる助詞に対する主成分分析において、『源氏物語』と『山路の露』との間における文体の相違を見出せる。なお、作者が相違する場合、助詞の分析に文体の相違が見出せるのは現代文に対する分析と同様である。(金, 2002)

また、主成分分析によって文体の相違が認められた助詞について、ランダムフォレストを用い、特徴語の抽出を行った。分析の結果、「ども」、「だに」、「へ」といった語の重要度が高く推定された為、これが特徴語であると考えられる。これら3語は『源氏物語』に比べ『山路の露』に現れやすい語である。

### 5.3 『源氏物語』と『雲隠六帖』

はじめに語の重み付けを行わず、主成分分析を行ったところ、『源氏物語』と『山路の露』に対する分析と同様、図3に示すように、文体の相違はどの品詞の分析においても認められなかった。すなわち、『山路の露』同様、『雲隠六帖』も『源氏物語』の語の使用頻度と類似していると言える。

『源氏物語』と『雲隠六帖』に対する分析においても、上記の指標による語の重み付けを行い、再度、主成分分析を行った。図4に示す通り、分析の結果、ここにおいてもidfを用いた重み付けによる助詞に対する主成分分析において、『源氏物語』と『雲隠六帖』の間に文体の相違を見出せる。

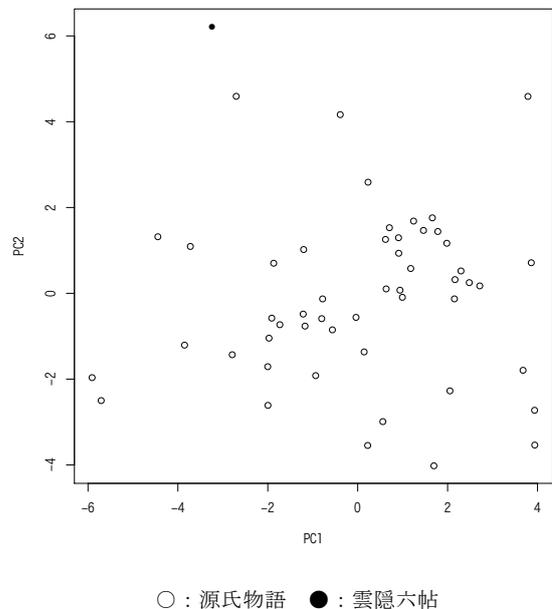
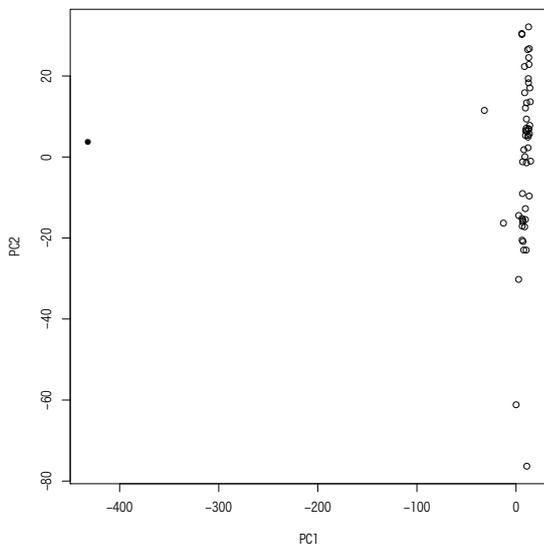


図3 助詞の相対頻度に対する主成分分析の結果

次に、助詞についてランダムフォレストを用い、特徴語の抽出を行った。分析の結果、「ども」、「へ」、「まで」といった語の重要度が高く推定された為、これが特徴語である

と考えられる。これら3語は『源氏物語』に比べ『雲隠六帖』に現れやすい語である。



○：源氏物語 ●：雲隠六帖

図4 助詞のidfを用いた主成分分析の結果

#### 5.4 重み付けを行った分析の考察

相対頻度に対する主成分分析では認められなかった作品間における語の使用傾向の相違を、idfによる重み付けを行うことで、明らかにすることが可能であると考えられる。また、語の使用傾向の相違が顕著に現れたのは助詞であり、これは現代文で記述された文章の計量的な研究の結論と同様である。

また、「ども」は『源氏物語』に比べ、『山路の露』および『雲隠六帖』の双方に頻出しており、これら2作品を特徴づける語であることがランダムフォレストの結果から言える。『平安時代の訓読語につきての研究』（築島, 1963）によれば、「ども」は接続助詞であり、同じく接続助詞である「ど」の2つが平安時代には併用されていたが、「ど」は和文体において、「ども」は漢文訓読文体において用いられる語であったとされる。しかし、「ば・と・とも・ども・ども <ても> <けれども> <ところが> <ところで>」（飛田, 1970）によれば、中世以降になると、「ど」は衰退し、「ども」が優勢になったとされる。

『山路の露』においても、『雲隠六帖』においても、助詞を除く、名詞、動詞、形容詞、形容動詞、副詞、助動詞の6品詞に対する主成分分析より『源氏物語』の語の使用傾向をよく模倣していると言える。その一方で、『源氏物語』では多用されない「ども」という接続助詞については『山路の露』および『雲隠六帖』においては『源氏物語』に比べ相対的に多く使用されていると言える。これは飛田(1970)

の指摘と合致する。

## 6. 「宇治十帖」における他作者説の検討

### 6.1 分析対象

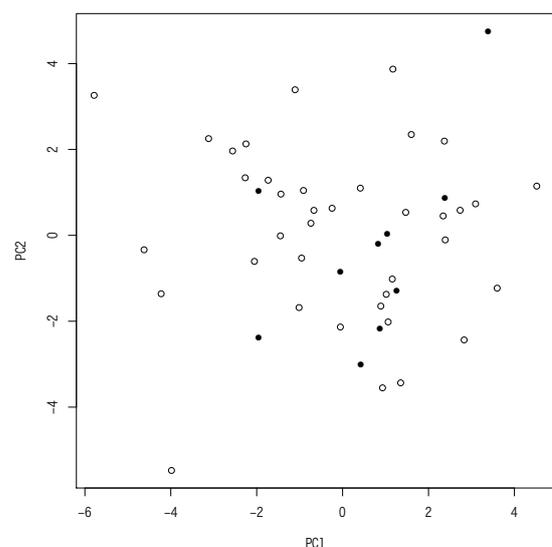
先述のように『源氏物語』の第42巻「匂宮」以降の13巻は光源氏没後の物語である。一般に、これら13巻のうち「匂宮」、第43巻「紅梅」、第44巻「竹河」の3巻は「匂宮三帖」と称され、第45巻「橋姫」以降の10巻は「宇治十帖」と称される。「匂宮三帖」および「宇治十帖」のどちらにおいても、それぞれ別個に他作者説が唱えられている。

「匂宮三帖」については、「竹河」巻末の官位昇進に記述が「宇治十帖」と矛盾することや、これら3巻の文章の質が他の巻の文章の質と相違する（石田, 1961）とされるから「匂宮」、「紅梅」、「竹河」の3巻は紫式部と親しい同時代の別人の作であるという説（石田, 1961）が論じられている。よって、「匂宮三帖」は「宇治十帖」および第41巻「幻」までの諸巻の双方と作者が相違する可能性があるため、本研究から除外した。

本研究においては、「幻」以前の諸巻と比較することで、「宇治十帖」についての他作者説に検討を加える。なお、便宜上、「幻」以前の諸巻を「他38帖」と称する。分析から除外したケースは「匂宮三帖」に加え、第11巻「花散里」、第16巻「閑屋」、第27巻「篝火」を加えた6巻である。

### 6.2 主成分分析によるクラスタリング

『源氏物語』と『宇津保物語』に対する分析と同様に名詞、動詞、形容詞、形容動詞、副詞、助詞、助動詞の7品詞に対し、品詞別に変数となる各語の相対頻度を求め、主成分分析およびクラスター分析を行った。ケースの総数は48となる。

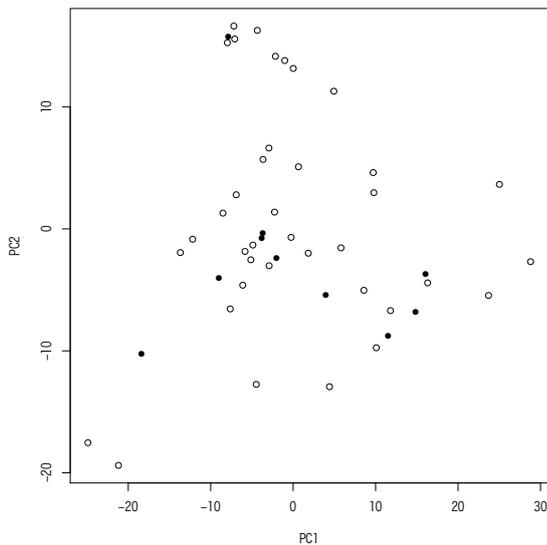


○：他38帖 ●：宇治十帖

図5 助詞の相対頻度に対する主成分分析の結果

主成分分析の結果、図5のように「宇治十帖」と「他38帖」の間において、語の使用傾向からみた文体の相違はすべての品詞において認められない。

次に、『源氏物語』の補作あるいは続編と見なされる『山路の露』および『雲隠六帖』との分析と同様に分析に用いる語に対し重み付けをし、主成分分析を行った。重み付けの指標は上記の4つの指標である。



○ : 他38帖 ● : 宇治十帖

図6 助詞のidfを用いた主成分分析の結果

図6に示すように、分析の結果、どの語の重み付け指標を用いても「宇治十帖」と「他38帖」との間に語の使用傾向の相違は認められなかった。

このように、本研究においては「宇治十帖」の作者が「他38帖」と異なるとする、すなわち宇治十帖他作者説を支持する積極的な分析結果は得られなかった。

### 6.3 特徴語の抽出

「宇治十帖」と「他38帖」の文章のクラスタリングでは、どの品詞の分析においても、対象間における語の使用傾向の相違が認められなかった。しかし、先述のように「宇治十帖」は他の巻に比べ、語の使用法が相違するという指摘があるため、どのような語の使用傾向が相違するのか明らかにするために、特徴語の抽出を行った。なお、特徴語の抽出については、各品詞の相対頻度に基づき特徴語を抽出した。

また、「宇治十帖」と「他38帖」とでは語の使用傾向に強い親近性を有していると言えるから、すべての品詞において抽出された特徴語から意味のある分析結果を得られなかった。

形容詞および形容動詞において抽出される特徴語は図7と図8に示す。図7にあるように、「こころうし」、「こころ

ぼそし」、「こころやすし」は形容詞の特徴語であると言える。一方、図8にあるように「こころよりほかなり」、「こころのどかなり」、「こころひとつなり」は形容動詞の特徴語であると言える。これら6語はすべて「宇治十帖」に多く現れる語である。

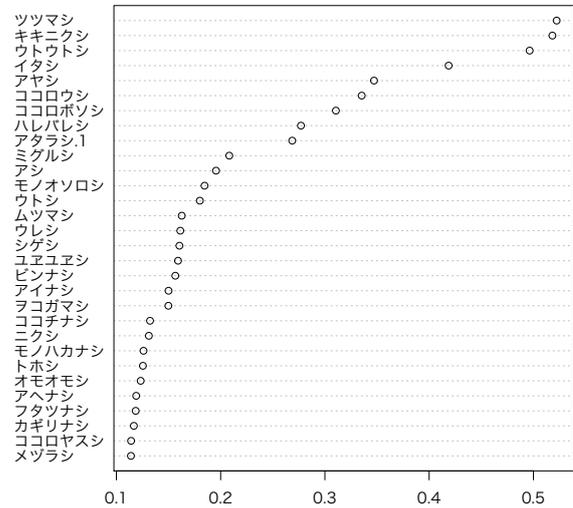


図7 形容詞のランダムフォレストによる重要度の推定

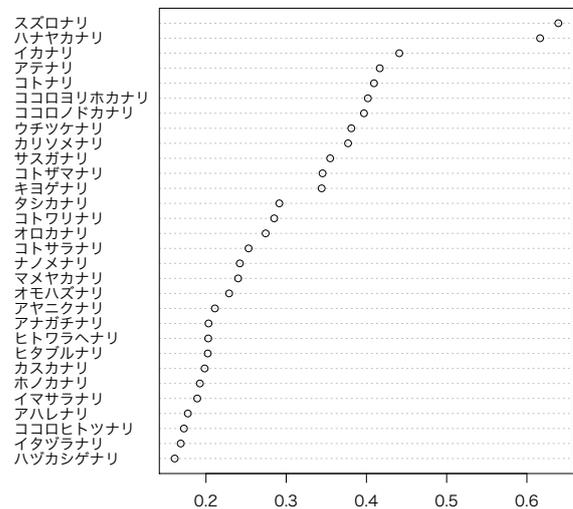


図8 形容動詞のランダムフォレストによる重要度の推定

「うし」、「ほそし」、「やすし」といった形容詞は特徴語として抽出されないことから、「宇治十帖」においても、「他38帖」においても使用傾向に大きな偏りを見いだすことはできない。しかし、「こころ」という接頭語をこれらの語に前接させる複合語の多用は注目される。これは形容動詞の

「こころよりほかなり」、「こころのどかなり」、「こころひとつなり」の3語についても同様である。したがって、「宇治十帖」は「こころ」を前接語としてもつ形容詞・形容動詞を「他38帖」に比べ、多用していると言える。

『源氏物語』(大野, 1984)によれば、『源氏物語』において形容語は、つまり形容詞と形容動詞は、整然と組織的に使い分けられている、と論じられており、接頭語の「もの」や「なま」は状況に応じてそれぞれ使い分けられていると指摘している。この指摘は、本研究における形容詞および形容動詞のランダムフォレストの結果とも合致するものと考えられる。

#### 6.4 考察

すでに述べたように、「宇治十帖」は他作者説が論じられるが、「宇治十帖」と「他38帖」との間に、分析に用いた各品詞の相対頻度に対する主成分分析から、語の使用傾向の相違は認められなかった。くわえて、語の重み付けを行った場合、『源氏物語』と『山路の露』に対する分析、『源氏物語』と『雲隠六帖』に対する分析において、語の使用傾向の相違が認められたが、「宇治十帖」と「他38帖」との間に語の使用傾向の顕著な相違は認められなかった。各品詞の相対頻度に対する分析、語の重み付けを行った分析のどちらにおいても語の使用傾向が認められなかったことから、計量的なアプローチからは「宇治十帖」の作者は「他38帖」と異なるとする積極的な根拠は得られなかったと言える。したがって、本研究の分析からは、「宇治十帖」は「他38帖」と作者が異なるとは断言できない。

次に、「宇治十帖」と「他38帖」との語の用い方の相違について、ランダムフォレストを用い、顕著に使用傾向が相違する語の抽出を行った。特徴語の抽出から、形容詞および形容動詞において「こころ」という接頭語を持つ語が「宇治十帖」に多用されることが認められた。

#### 7. 結語

まず、『源氏物語』とその補作または続編とされる『山路の露』および『雲隠六帖』を対象とした分析では、『源氏物語』と『宇津保物語』についての分析で用いた相対頻度に対する主成分分析では作品間における語の使用傾向の相違を認められなかった。これは『源氏物語』の語の使用傾向と『山路の露』および『雲隠六帖』の語の使用傾向が顕著に相違していないことを意味する。

次に、分析に用いる語について重み付けを行うことで、助詞の語の使用傾向に相違があらわれること明らかにした。本研究においては、特にidfが有効であることを示した。

なお、相対頻度に対する分析から、『山路の露』および『雲隠六帖』は『源氏物語』におけるこれらの語の使用傾向を継承していることが考えられ得る。すなわち、ここに両作品が文学史上『源氏物語』の続編、補作として現在まで伝承されてきた根拠の一つを見出すことが可能である。また、

『源氏物語』においては多用されない「ども」という助詞を『山路の露』、および『雲隠六帖』のどちらもが『源氏物語』に比べ、特徴的に使用していることを指摘した。これは「ども」の使用が後代になると顕著になるという国語学史上の知見と一致する。

最後に、他作者説が論じられる「宇治十帖」を採り上げ、「宇治十帖」とそれ以外の諸巻との間に語の使用傾向の相違が認められるか検討を行った。特徴語の抽出から、語構成上の特徴として形容詞および形容動詞において「こころ」という接頭語を持つ語が「宇治十帖」に多用されることが認められた。しかしながら、主成分分析の結果、各品詞の相対頻度、および語に重み付けを行った分析から、「宇治十帖」とその他の諸巻の間に語の使用傾向の相違が認められなかった。ゆえに、本研究における語の使用傾向に対する分析から「宇治十帖」の作者は「他38帖」の作者と異なるとする積極的な根拠は得られなかった。

#### 参考文献

- 1) 新井皓士: 源氏物語・宇治十帖の作者問題: 一つの計量言語学的アプローチ. 一橋論叢, 117(3), pp397-413 (1997).
- 2) Breiman, L.: Random forests. Machine Learning, 24, pp123-140 (2001).
- 3) Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J.: Classification And Regression Trees. Wadsworth. (1984).
- 4) Dumais, S. Enhancing Performance in Latent Semantic Indexing (LSI) Retrieval. Technical Report, Bellcore. (1992)
- 5) 長谷川和子: 源氏物語の研究. 東宝書房. (1957).
- 6) 飛田良文: ば・と・とも・ども・も <ても> <けれども> <ところが> <ところで>. 国文学 解釈と鑑賞, 35(13), pp54-61 (1970).
- 7) 伊井春樹: 源氏物語注釈書・享受史事典. 東京堂. (2001).
- 8) 池田亀鑑: 源氏物語の構成. 至文堂(『新講源氏物語(上)』所収). (1951).
- 9) 池田亀鑑: 日本古典全書 源氏物語7. 朝日新聞社. (1955).
- 10) 池田亀鑑: 源氏物語大成. 中央公論社. (1984).
- 11) 池田亀鑑: 合本源氏物語事典. 東京堂. (1987).
- 12) 稲賀敬二: 宇津保物語は合作か?. 明治書院(阿部秋生編, 『講座日本文学の争点(二) 中古編』所収). (1968).
- 13) 石田穰二: 匂宮・紅梅・竹河の三帖をめぐって. 国文学: 解釈と鑑賞, 26(12). (1961).
- 14) 金明哲: 読点の打ち方と文章の分類. 計量国語学, 19(7), pp317-330 (1994).
- 15) 金明哲: 助詞の分布における書き手の特徴に関する計量分析. 社会情報, 11(2), pp15-23 (2002).
- 16) 金明哲, 村上征勝: ランダムフォレスト法による文章の書き手の同定. 統計数理, 55(2), pp255-268 (2007).
- 17) 金明哲: 文章の執筆時期の推定-芥川龍之介の作品を例として. 行動計量学, 36(2), pp89-103 (2009).
- 18) 小林栄子: 源氏・伊勢物語新研究. 晋文館. (1935).
- 19) 小林雄一郎, 田中省作, 富浦洋一: ランダムフォレストを用いた英語科学論文の分類と評価. 人文科学とコンピュータ研究会報告, 2011-CH-90(6), pp1-8 (2011).
- 20) 村上征勝: 文化を計る 文化計量学序説. 朝倉書店. (2002).
- 21) 村上征勝, 今西祐一郎: 源氏物語の助動詞の計量分析. 情報処理学会論文誌, 40(3), pp774-782 (1999).
- 22) 室城秀之: うつほ物語 全 改訂版. おうふう. (1995).
- 23) 大野晋: 源氏物語. 岩波書店. (1984).

- 24) 杉本知之・下川敏雄・後藤昌司: 樹木構造接近法と最近の発展. 計算機統計学, 18(2), pp123-164 (2005).
- 25) 武田宗俊: 源氏物語の最初の形態. 文学, 18(6-7). (1952).
- 26) 築島裕: 平安時代の漢文訓読語につきての研究. 東京大学出版会. (1963).
- 27) 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シンポジウム論文集. Vol.2011, No.8, pp125-132 (2011)
- 28) 上田英代, 今西祐一郎, 藤田真理, 村上征勝, 樺島忠夫, 上田裕一: 源氏物語語彙総索引. 勉誠社. (1994).
- 29) 和辻哲郎: 日本精神史研究. 岩波書店. (1992).
- 30) 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, 2(1), pp147-156 (1957).
- 31) 安本美典: 文章心理学の新領域. 創元社. (1960).書式を設定する

### 著者紹介

土山玄 (非会員)  
同志社大学大学院文化情報学研究科博士後期課程

村上征勝 (正会員)  
同志社大学文化情報学部