

(論文)

データマイニング手法の古典籍への応用

—源氏物語短編写本の分析—

齊藤 鉄也

キーワード

源氏物語 写本 仮名字母 データマイニング

1. はじめに

本論文では、前稿^[1]で実施した古典籍の仮名字母(字源)の出現頻度率に関する調査を、さらに対象とする資料を増やし、より詳細に調査した結果を報告する。具体的には、源氏物語54帖のうち、短編を中心とした第3帖「空蟬」、第8帖「花宴」、第11帖「花散里」、第16帖「関屋」、第27帖「篝火」、第38帖「鈴虫」の6帖、計115写本を対象としたことと、写本本文の仮名字母の出現頻度率を用いて、写本の書写年代の特徴を把握するために、統計的手法を用いて分析したこと、を報告する。仮名字母の出現頻度率に基づいて、推定書写年代ごとにその出現傾向を分析した結果、鎌倉時代を中心とする写本に特徴的に出現する仮名字母を指摘することができた。

以下、第2章では、調査目的と対象とした源氏物語の写本の概要、調査方法を述べる。第3章では、関連研究を踏まえ、前稿で調査した本文の漢字の出現割合と仮名の総字母数の関係を調査する。第4章は、写本の仮名字母と書写年代の関係を、統計的な手法を用いて分析する。第5章では、今後の課題を述べ、第6章はまとめである。

2. 調査目的と方法

ここでは、本調査の目的と対象としたデータ、その処理方法について述べる。具体的には、源氏物語の本文に関する基本的な情報と、調査対象とした本文の概要について述べる。本文からの採字方針や文字の集計方法といった調査方法に関しては、前稿^[1]と同様であるため、その詳細に関しては説明せず、その概要だけを述べている。

2-1. 調査目的

計量文献学の目的は、文章の持つ量的な特徴量に基づき、数理的な処理を行って、著者の推定や書写年代の推定を行うことである。本調査においても同様に、古典籍に出現する文字情報を特徴量として用いて、古典籍の書写年代の推定を行うことが目的である。

そのため、長期間に渡って書写され続け、現在までに大量の写本が存在している源氏物語を対象

さいとう てつや：淑徳大学 経営学部 准教授

とする。具体的には、源氏物語は54帖（巻）の長大な小説であるため、その中の短編の巻を調査対象とし、その仮名字母を特徴量として用いて、その書写年代の推定の可能性を調査する。

2-2. 調査対象とした写本の概要

対象とした源氏物語の写本は、源氏物語のうち短編を中心に、第3帖「空蟬」、第8帖「花宴」、第11帖「花散里」、第16帖「関屋」、第27帖「篝火」、第38帖「鈴虫」の6帖である。これらの短編の複数の巻を選択した理由は、大量の写本を対象とした調査であるために、調査の効率性を考慮したからである。加えて、今後、中編や長編の巻を対象とする際にも、これらの短編の調査結果はその基礎を成すと考えられるからである。

源氏物語54帖のうち、第11帖「花散里」、第16帖「関屋」、第27帖「篝火」は最も短い本文を持ち、写本ごとにその長さは異なるが、およそ10ページ（5丁）、漢字と仮名を合わせ1500文字から2000文字程度の写本である。第3帖「空蟬」、第8帖「花宴」はそれらに続いて短い本文を持ち、およそ18ページ（9丁）、4500文字から5000文字程度の本文を持つ。これらの本は源氏物語54帖のうち、短い巻である。第38帖「鈴虫」は、およそ40ページ（20丁）、6000文字から7000文字程度の本文を持つ。「鈴虫」は短編ではないが、鎌倉時代の写本が多く残る巻であるため、調査対象として選択した。

対象とした写本の内訳は次の通りである。第3帖「空蟬」は18冊、第8帖「花宴」は19冊、第11帖「花散里」は20冊、第16帖「関屋」は20冊、第27帖「篝火」は19冊、第38帖「鈴虫」は19冊、6帖の写本の合計は計115冊、全文字数として約40万文字を対象とした。これらの写本は、影印本として出版されている写本及びインターネット上で公開されている写本である。これらの対象とした写本の出典に関しては、基本的には前稿の参考文献と同様であること、すべての写本が出版または公開され、本文を対象に研究を行う研究者にとっては既知であり、ほぼ統一された写本名を採用していることから、115冊分の参考文献を記載することは省略する。写本の書写年代は鎌倉時代前期から江戸時代前期までである。これら写本の概要に関しては付録1に各帖ごとに表としてまとめている。

2-3. 調査方法

2-1の「調査目的」に述べた通り、本調査を行うにあたり、それぞれの写本や写本の書写年代を特徴づける特徴量を選択する必要がある。ここでは、特徴量の候補として、本文における漢字の割合と仮名の総字母数、仮名字母の出現頻度率を取り上げる。源氏物語に限らず、古典籍の物語の写本においては、漢字の利用は現代の文章ほど多くはない。そのため、漢字の出現率は書写年代を推測する手がかりとして考えられている。また、仮名字母も、古い年代ほどその種類が多く、時代が下るほどその種類が少なくなると考えられている。これらの推定に基づいて、書写年代や写本の持つ特徴を把握する研究が行われている^[4]。

2

加えて、本調査では、前稿に引き続き、仮名字母の出現頻度率に着目する。仮名は、漢字に比べて本文の内容の影響を受けにくいこと、同一作品の同一巻の写本であっても書写者ごとに仮名字母が異なり、字母が同一の写本の存在は極めて珍しいことが、その理由である。さらに、このことから、仮名字母の種類数は多く、仮名を書く際の使い方には書写者個人の書き癖があるとも考えられ、特徴量としてふさわしいと考えた。また、より多くの作品を視野に入れて研究することを考えると、単純な仮名字母の出現頻度数では、異なる作品や異なる巻を比較することはできないために、仮名字母の出現頻度率を求めている。

次に、前稿に述べた写本からの採字方針と写本の文字数の集計方法の要点を述べる。集計対象とした仮名に関しては、原則として本行本文に存在する仮名を採用している。そのため、異文注記といった傍記は採用していない。加えて、本文に書写した文字とその修正として、多くの場合、本行本文の右脇に文字が記入されている「見せ消し」に関しては、本行本文に存在する「消された」文字だけを採用している。この理由は、異文注記や修正後の文字は、本文が書写された時期と同時期であるとは限らないことと、同時期である場合もその判断が困難であることからである。

採字した漢字や仮名の区別は、前稿の付録1の字母表を修正した、付録2の字母表に基づいている。基本的には、この字母表に掲載され、かつ、一音で読む文字の漢字は仮名として採字している。採字した仮名は、字母表に掲載した369種類の仮名の字母の単位で分類し、その出現頻度数を単純集計する。それを通行仮名48種類ごとに、出現頻度率を計算している。

対象とした写本の集計結果は大量であるため、付録3に各巻ごとに表としてまとめた。表には、本文の総文字数、漢字と仮名の割合、本文の総文字数の内訳として、漢字の文字数とその種類数、仮名の文字数と字母数をまとめた。第27帖「篝火」に関しては、前稿にも同様の内容を述べているが、掲載した写本に関する内容を更新しているため、改めて掲載している。

このようにして作成したデータを用いて分析を行う。本調査では、分析ツールとしてR^[6]を用いている。

3. 関連研究とその検証

これまで古典籍の特徴を把握するための変数として、本文中の漢字の割合と仮名の総字母数があることを述べた。ここでは、このふたつの変数を特徴量として用いた場合の書写年代の推定に関して調査及び検証する。

3-1. 写本の推定書写年代と漢字の割合の可視化

ここでは、漢字の割合と仮名の総字母数を用いて、対象とした写本の位置付けを可視化し、写本全体の傾向を調査するために、付録3の本文の文字情報を図1にまとめた。図1では、写本の書写年代の相違を明らかにするために、鎌倉時代、室町時代、江戸時代に区別して、写本の位置を点描している。横軸に漢字の割合、縦軸に仮名の字母の総字母数をそれぞれ、平均を0、標準偏差を1とした標準化得点（z得点とも言う）を用いて表している。

この傾向の可視化に関しては、次の指摘を検証するために行った。参考文献^[4]では、漢字と仮名の字母の出現傾向に関して先行する文献の指摘をまとめている。その指摘のうち、書写年代と、漢字の割合や仮名の総字母数に関する指摘は、「古い本は仮名が中心、新しい本は漢字をより使おうとする傾向がある」と「鎌倉期に書写された伝本の漢字含有率は平均値を下回る」の二点である。これらの指摘に関して、対象とした写本を用いて検証する。

3-2. 書写年代と漢字の割合、仮名の総字母数の傾向の検証

上記の最初の指摘である「古い本は仮名が中心、新しい本は漢字をより使おうとする傾向がある」という点に関して、図1からは、鎌倉時代と室町時代の写本の位置において重複が多く、時代が下るほど漢字の割合が増えるという傾向は明らかではない。そこで、この傾向の有無を明らかにするために、写本の推定書写年代と漢字の割合の関係に対して、写本の推定書写年代を目的変数、漢字の割合を説明変数として回帰分析を行った。漢字に関しては、異なる巻を対象としているため、漢字の総文字数や総種類数は比較できず、漢字の割合のみを比較している。その結果は、推定書写年

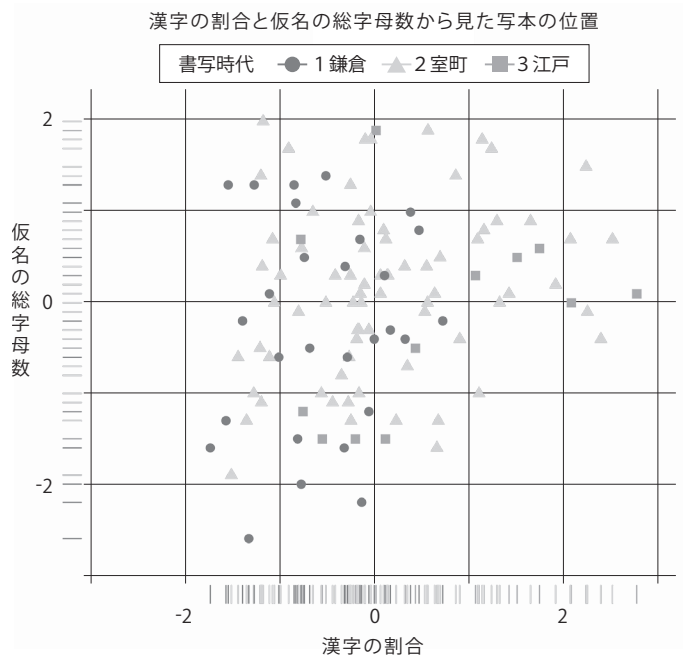


図1 漢字の割合と仮名の総字母数に基づく調査対象写本115本の位置づけ

代と漢字の割合の相関係数は0.399 ($p=0.00^{***}$)、決定係数は0.159であった。この結果は、非常に弱い相関があることを表している。

加えて、推定書写年代と仮名の総字母数との関係に対して、同様に回帰分析を行ったところ、推定書写年代と仮名の総字母数の相関係数は0.186 ($p=0.05^*$)、決定係数は0.035であった。この結果は、ほぼ無相関であることを表している。最後に、推定書写年代を目的変数として、漢字の割合と仮名の総字母数を説明変数とした場合の重回帰分析の結果、自由度調整済み決定係数は0.168であった。

このことは、推定書写年代と、漢字の割合や仮名の総字母数に関する変数との関係からは相関が低い結果を示し、少なくとも対象とした写本に関しては、書写年代と漢字の割合の関係はほとんどわからないと言える。漢字の割合や仮名の総字母数と言った単純な特徴量から書写年代を推定することは難しいだろう。

3-3. 写年代ごとの漢字の割合の平均の差の検証

次の指摘である「鎌倉期に書写された伝本の漢字含有率は平均値を下回る」に関して検討する。図1では、おおよそ、鎌倉時代の写本は、漢字の割合が少ない左側に多く点描されているが、その他の時代の写本の点描と重なっている。そこで、鎌倉時代と、室町時代、江戸時代といった時代ごとの写本の漢字の割合の平均の差の有無を明らかにするために、一要因分散分析を行った。その結果は、0.1%で有意であり ($p=0.000^{***}$)、ボンフェローニ検定を用いた多重比較の結果は、鎌倉時代と室町時代の写本間は1%で有意 ($p=0.006^{**}$)、鎌倉時代と江戸時代の写本間も1%で有意 ($p=0.002^{**}$) であった。このことから、鎌倉時代に書写された写本の漢字の割合は室町時代や江戸時代と比べて低いと言える。

但し、論理的に「逆は真」ではないため、このことから「漢字の割合が少ない写本は書写年代が古い（鎌倉時代の写本である）」と主張することはできない。この「鎌倉期に書写された伝本の漢字含有率は他の時代に書写された伝本の漢字含有率の平均値を下回る」という指摘が正しいとしても、「漢字の割合が少ない写本は書写年代が古い可能性がある（がわからない）」と言えるだけであり、実際に写本の書写年代を推定するために用いることには注意が必要である。

4. 仮名字母に基づく特徴量の調査結果と考察

ここでは、仮名字母の出現頻度率を特徴量として用いた場合の書写年代の推定に関して調査及び検証する。

4-1. 仮名字母に基づく分類

仮名字母についてはその出現傾向に関しての考察^[4]があるため、「2-1. 対象とした写本の概要」で述べた115本を対象に、仮名字母に関して、さらに調査する。調査目的は、特定の時代に出現する仮名の字母の有無の確認である。図1では、各時代に書写された写本の位置が重なってしまっている。3-2において明らかにしたように、単純に仮名の総字母数の情報だけからは、書写年代との関係はわからない。仮名の総字母数ではない特徴量を対象にして、仮名字母に関して異なる処理方法が必要である。

ここでは、対象とするデータとしては仮名字母の出現頻度率に着目して調査し、書写年代と字母の関係性を明らかにする。付録3では、仮名の単純集計を行っていた。上記に述べたように、本調査においては、異なる帖の写本を対象としているため、単純な文字の出現頻度の集計では、同じ巻同士と比較と、異なる巻同士の比較の際に、同じ巻同士の類似性が高くなることや、長編と短編の比較ができないこと、といった問題がある。そのため、ここでは、仮名字母の出現頻度率に着目し、数値化して分析している。前述した字母の表に基づいて、48種類の仮名ごとに、各仮名（字音）の字母の出現頻度率を計算する。出現頻度の集計と出現頻度率の計算結果から、写本ごとの仮名字母の出現頻度率の分布が明らかにする。

分析方法は、以下の手順で行った。最初に、仮名字母の出現頻度率に基づき、主成分分析と非階層的クラスタ分析を用いて、写本を幾つかの群に分類する。次に、その分類結果である群と書写年代の関係を確認した後、その分類に影響を与えている字母を、複数の群があることを想定して、分散分析と多重比較を用いて調査する。多重比較の結果、分類された写本群の間に5%の有意差がある字母をその出現に特徴がある字母と考えている。

4-1-1. 主成分分析と非階層的クラスタ分析を用いた調査

各写本に出現する字母数は、付録3の写本本文の仮名の字母数より70から120程度であるため、集計表にある字母の369種類うち2/3程度はどの写本にも出現していない。そのため、最初に、分類に影響しないと考えられる出現しない字母や出現率が100%に近い字母を対象としたデータの削除を行った。その結果、計242字母を分析対象から削除し、計127字母を対象とした。

加えて、全ての写本に出現する字母のうち、あるふたつの字母の相関係数が高い場合も同様に削除している。具体的には、次のような場合に該当した時、ふたつある字母のうち、一方を削除している。例えば、多くの写本において、仮名「あ」の字母は、「安」または「阿」を利用する。そのため、「安」と「阿」の相関は極めて高く、どちらか一方が高ければもう一方が低いという関係にある。このような関係を持つ字母の出現傾向の特徴を考える場合には、そのどちらか一方を用いれ

ば良い。その結果、計95字母を対象としたデータとなった。

次に、この95字母を対象に主成分分析を行った。この95字母をそれぞれ変数として写本を分類することは困難であるため、多くの変数をまとめる必要がある。主成分分析では、複数の変数をまとめた合成変数を求め、その合成変数に基づいて分析を行う。主成分分析の際には、変数の標準化を行っていない。これは、頻度率の単位が同じであることが理由である。

分析結果の主成分は95あるため、寄与率は第1主成分であっても、17%程度であった。累積寄与率が50%を超える第7主成分までを、表1としてまとめた。

表1 仮名の95字母に対する主成分分析の寄与率

主成分	1	2	3	4	5	6	7
固有値	0.378	0.196	0.175	0.127	0.119	0.108	0.098
寄与率	17.1%	8.9%	7.9%	5.7%	5.4%	4.9%	4.4%
累積寄与率	17.1%	26.0%	33.9%	39.7%	45.0%	49.9%	54.3%

主成分負荷量を解釈するために、95の字母と第1主成分との数値を確認した。この結果、43の字母が第1主成分へ正の影響を与えていることが明らかになった。しかし、それぞれの字母が影響を与えている合成変数の意味を解釈することはできなかった。次に、主成分得点を解釈するために、第1主成分と第2主成分の散布図を用いて確認した。主成分負荷量の解釈はできなかったが、散布図によると、鎌倉時代の写本が多く第1主成分の正の方向に集まっていることが明らかになった。

そこで、主成分分析で求めた第1主成分と第2主成分に対して、k平均法を用いて、非階層的クラスタ分析を実行した。クラスタ分析をすることで、主成分分析をした結果に対して、類似度が近い写本群に分類できる。主成分得点の解釈結果から、第1主成分の正の方向が時代の古さを表している可能性があるため、時代区分に基づいてクラスタ数を決定した。具体的には、k平均法に与えるクラスタ数の初期値を写本の時代区分に従い、鎌倉時代、室町時代、江戸時代に対応する最も少ない3から、鎌倉時代の前期、中期、後期、室町時代の前期、中期、後期、江戸時代前期に対応する最も多い7までとした。

k平均法の結果は、6群に分類された。写本の書写年代と組み合わせ、この分類結果を表2としてまとめた。

表2 非階層的クラスタ分析の分類結果と写本の書写年代の対応

群	鎌倉前期	鎌倉中期	鎌倉後期	室町前期	室町中期	室町後期	江戸前期
1							6
2				2	2	33	2
3						3	
4			4	3		11	1
5				3	4	5	3
6	3	13	8	6		3	

4-1-2. 調査結果の検証

表2からは、非階層的クラスタ分析の分類結果と写本の書写年代の組み合わせは、ばらつきを含みながらも、おおよそ次のように分類できることが明らかになった。1群は東久邇宮家旧蔵本6

冊全てに対応している。2群と3群はおおよそ室町時代後期を中心とした写本に対応し、4群と5群はおおよそ室町時代の写本に対応し、6群はおおよそ鎌倉時代の写本に対応している。特に、1群は東久邇宮家旧蔵本6冊全てが分類され、しかも他の写本は分類されていないことから、東久邇宮家旧蔵本が非常に特徴的な仮名字母の出現傾向があることが予想できる。

また、6群は鎌倉時代の写本の多くが分類されている。6群の中に存在する室町時代の写本は、室町時代前期に書写されたとされる御物本「空蟬」「花散里」「篝火」「鈴虫」、陽明文庫本「花宴」、尾州家本「篝火」と、室町時代後期に書写された日本大学蔵三条西家本「花宴」「花散里」「関屋」が含まれている。本調査では、南北朝時代は室町時代前期と重複する時代であることから、室町時代前期に分類している。これらの写本に関しては、御物本と陽明文庫本の写本は鎌倉時代または南北朝時代に書写されたとの指摘がある。写本の書写年代に関して、幅広い年代が指摘されている写本に関しては、より下の時代を採用していることから、これらの写本はより時代が遡る可能性があるだろう。日本大学蔵三条西家本の3冊はいずれも三条西実隆が書写した本であるという共通点がある。これは三条西実隆の書写態度、具体的には鎌倉時代の写本を字母まで忠実に書写していることを表している可能性も考えられ、非常に興味深い結果となった。

この他の鎌倉時代の写本は、4群に含まれる玉里文庫本古筆源氏の「空蟬」「花宴」「関屋」と榊原家本「関屋」であった。これも、いずれも伝承筆者を世尊寺行房とする玉里文庫本古筆源氏の3冊がひとつの集合を示すことが興味深い。これらの写本に関しては、鎌倉時代に書写された写本が多い6群とは異なり、室町時代に書写された写本が多い4群に含まれていることから、室町時代の写本との共通点がある可能性が高い。

この分類結果からは、仮名字母の出現頻度率の視点から写本の書写年代を分類できる可能性があると言えるだろう。特に、鎌倉時代の写本と室町時代の写本を分類できる仮名の字母の存在や傾向を明らかにすることができれば、仮名字母の出現頻度率の傾向に基づいてより古態を示す写本を明らかにできる可能性がある。そこで、室町時代後期の写本が中心である2群と3群、室町時代の写本が中心である4群と5群、鎌倉時代の写本が中心である6群を対象に、仮名字母の出現頻度率から見た鎌倉時代の写本と室町時代の写本の特徴を調査する。

4-2. 特徴のある仮名字母の調査

鎌倉時代の写本と室町時代の写本の字母から見た特徴を明らかにするために、対象とする群に所属する写本の数（サンプルサイズ）を揃える。ここでは江戸時代前期に書写されたとされる東久邇宮家旧蔵本だけが所属する1群は除く。写本の数を揃えるため、室町時代の写本が多い2群と3群をまとめたA群（サンプルサイズ42）、同様に室町時代の写本が多い4群と5群をまとめたB群（サンプルサイズ34）、鎌倉時代の写本が多い6群（サンプルサイズ33）の三つの群とする。これら三群の合計写本数（サンプルサイズ）は109となった。

4-2-1. クラスカル・ウォリス検定を用いた調査

最初に、分類を集約した三つの群に対して、3群以上のノンパラメトリック検定の手法である、クラスカル・ウォリス検定を行った。これは、それぞれの字母において、ルービン検定を用いて、各群が正規分布に従っていないことを確認した結果、多くの字母が正規分布に従っていないことが明らかになったからである。このことに基づき、ノンパラメトリック検定を用いている。この結果を考慮して、検定結果が5%で有意であれば、さらにシェッフェの方法を用いて多重比較を行った。クラスカル・ウォリス検定と多重比較のソフトウェアは参考文献^[7]に公開されている統計処理ソ

ソフトウェアであるRのプログラムを用いた。

調査の目的は、鎌倉時代に書写された写本を中心とした群（C群）と、ふたつの室町時代に書写された写本を中心とした群（A群とB群）を比較し、鎌倉時代に書写された写本を中心とした群を区別できる字母を見つけることである。そのため、鎌倉時代に書写された写本を中心とした群とその他の群を比較し、多重比較の結果も5%で有意の場合、その字母の出現の有無や傾向が特徴的である、と判断している。

4-2-2. 調査結果の検証

多重比較まで実行した結果、各字母の中で、鎌倉時代の写本を中心としたC群と室町時代の写本を中心としたA群とB群で、その出現に差があり、特徴的であると考えられる字母は、計31存在する。これらの字母を表3としてまとめた。

表3 鎌倉時代と室町時代で出現に差がある字母

仮名	字母	仮名	字母	仮名	字母	仮名	字母
え	江	す	寸、春、壽/寿	は	者	め	女
き	起	せ	勢	ふ	婦	よ	与、世
け	遣、希、氣/気	た	多、堂	へ	邊/辺	り	里
さ	左	に	仁、丹、耳、二	ま	満	る	流
し	之、新	ね	祢/禰、年	み	三	れ	連

これらの字母は、出現の有無に特徴がある字母と、出現率の傾向に特徴がある字母に分けることができる。出現の有無に特徴がある字母とは、鎌倉時代に特徴的に出現するか、または出現しない字母のことである。出現率に特徴がある字母とは、鎌倉時代にも室町時代にも出現するが、その出現率に差があり、増加または減少傾向にある字母のことである。

出現の有無に特徴がある字母のうち、鎌倉時代に特徴的に出現する字母には、「江」「新」「二」「邊/辺」がある。これらに続いて出現率が高い字母には「壽/寿」「勢」がある。これに対して、鎌倉時代に特徴的に出現しない字母には「起」「希」「丹」「年」がある。これらに続いて出現率が低い字母は「流」である。これらの字母は出現の有無によって判断し易いため、書写年代を考える際の直接的な「手がかり」として有効であろう。このうち、字母「新」は参考文献^[6]にあるように鎌倉時代に出現する字母として指摘がある。本調査の結果は、参考文献^[6]と調査方法が異なるにも拘らず同じ結果が明らかになったことは興味深い。

表4 鎌倉時代の出現の有無に特徴を持つ字母

仮名	字母	鎌倉時代の出現率	出現傾向
え	江	78.6% (11/14)	出現自体が少ないが、鎌倉時代に多い
き	起	10.5% (8/76)	鎌倉時代に少ない
け	希	3.8% (2/52)	鎌倉時代に少ない
し	新	79.3% (23/29)	鎌倉時代に多い
に	丹	0.0% (0/44)	鎌倉時代に出現しない
	二	65.6% (21/32)	鎌倉時代に多い
ね	年	15.1% (8/53)	鎌倉時代に少ない
へ	邊/辺	79.6% (11/14)	出現自体が少ないが、鎌倉時代に多い

このうち、特に出現の有無に特徴がある字母を表4にまとめた。表4では、仮名と字母に続いて、対象とした109写本の中で出現した写本数を分母に、そのうち鎌倉時代の写本が多いC群に出現した写本数を分子にして、その字母の鎌倉時代の出現率を記載し、その出現傾向をまとめた。例えば、字母「江」は109写本のうち14写本にしか出現せず、しかも出現した写本のうち11写本が鎌倉時代の写本が多いC群に属している。このことから、出現傾向としては、写本に出現すること自体が少ないが、出現した場合は鎌倉時代の写本である可能性が高いと言える。同様に、字母「起」は76写本に出現するが、鎌倉時代の写本が多いC群には8写本にしか出現しない。字母「起」は室町時代の写本を中心としたA群とB群にも出現するが、鎌倉時代の写本が多いC群には33写本のうち、8写本にしか出現しないので、鎌倉時代には出現が少ないと言える。

出現率に特徴がある字母のうち、鎌倉時代により多く利用されている字母は「遣」「左」「之」「寸」「堂」「仁」「衿/襷」「三」「与」「里」が、室町時代により多く利用されている字母は「氣/気」「春」「多」「耳」「年」「者」「婦」「満」「女」「世」「連」がある。

これらの字母は、同音の仮名や翻字、採字に関係する場合があるため、個別に考察を行う。同音の仮名の事例として、仮名「け」を挙げる。「け」では、鎌倉時代には「遣」が41.6%出現し、「希」は0.2%出現しているのほぼ出現せず、「氣/気」が9.4%出現する。これに対し、室町時代のA群では、「遣」が13.9%、「希」が10.5%、「氣/気」が26.8%出現し、室町時代のB群では、「遣」が20.3%、「希」が3.5%、「氣/気」が17.3%出現する。100%に足りない残りの出現率の差は、仮名「け」の場合「个」が占めている。全体の傾向としては、「个」が出現率の半数を占め、「遣」が鎌倉時代から室町時代にかけて減る傾向にあり、「希」と「氣/気」が鎌倉時代から室町時代にかけて増加する傾向がある。このように書写年代によって、字母の出現率が変化する仮名の例として他に「た」「ね」がある。

ここでは、出現率の傾向に特徴のある字母を表5にまとめた。表5において、鎌倉時代の出現傾向とは、鎌倉時代のC群に出現する割合を表している。括弧の中のふたつの割合は、室町時代のA群とB群に出現する割合を表している。ここに挙げた字母は、ほぼ全ての写本に出現するが、鎌倉時代と室町時代の群の間で出現率の差が20ポイント程度の差がある字母を表している。

表5 鎌倉時代の出現の傾向に特徴を持つ字母

仮名	字母	鎌倉時代の出現傾向	出現傾向
さ	左	90.5% (66.3%, 77.4%)	鎌倉時代の出現率が相対的に高い
す	寸	59.8% (29.2%, 39.1%)	鎌倉時代の出現率が相対的に高い
た	多	55.7% (75.4%, 73.1%)	鎌倉時代の出現率が相対的に低い
	堂	38.0% (11.8%, 17.7%)	鎌倉時代の出現率が相対的に高い
ね	衿/襷	96.9% (79.2%, 71.3%)	鎌倉時代の出現率が相対的に高い
め	女	49.1% (88.1%, 84.5%)	鎌倉時代の出現率が相対的に低い
り	里	41.4% (15.7%, 15.9%)	鎌倉時代の出現率が相対的に高い
れ	連	9.3% (41.2%, 20.5%)	鎌倉時代の出現率が相対的に低い

9

翻字に注意を要する事例として、仮名「す」を挙げる。「す」の場合、字母「春」と「壽/寿」の区別は難しく、翻字の際の間違いも発生し易いため、その出現率の精度は低いことが予想される。採字に関係する事例として、仮名「よ」をあげる。「よ」の字母「世」の場合、既に2-2の採字方針の際に述べた通り、一文字の読みを持ち、かつ、付録2の字母表に掲載されている漢字は、仮名として採用している。一方で「世中」という単語が本文中に出現した場合は熟語であるため、漢

字として採用している。このように採字方針によって、影響を受ける字母は注意して扱うべきであろう。

これらの出現率に差がある字母は、同音の仮名に存在する字母ごとに出現率を計算し、さらに時代ごとに比較することが必要であるため、書写年代を考える際の「手がかり」としては、気付くことが難しいと考えられる。

5. 考察と評価

今回行った調査により、鎌倉時代に特徴的に出現する字母が明らかになった。これは、これまで指摘のあった「新」の字母以外にも多数の特徴のある字母があることを示した点で、新しい知見を得たと言える。しかし、「新」や「江」といった特徴的に出現する字母を持つ写本は多くなく、多くの写本に適用できる特徴としては利用することが難しい。

これに対して、多くの写本に出現する字母の特徴としては、その出現率の相違を用いることが考えられる。但し、この場合は、写本の中で同じ仮名となる字母を全て採字し、さらに特定の字母の出現率を計算することが必要である。特定の字母だけを翻字しただけでは、その出現率を計算することはできないため、利用することが難しい。

このことから、この調査結果が妥当であるとしても、分野の異なる研究者、特にこれらの知見を利用する文学や国語学の研究者が容易に利用できなければ、その結果を用いて研究を進めることは難しい、と言えよう。今回のような、分野が融合する領域の方法や結果は、利用者にとって、その考え方のプロセスや調査結果の使い易さやわかり易さを重視した手法が必要ではないだろうか。

今回は漢字に関しての詳細な調査を行っていない。漢字に関しては、本文が異なれば出現する漢字も異なることと、異なる巻や長編と短編の巻を比較する場合には、単純な出現頻度では比較できないことから、その扱いの検討が必要だろう。書写年代と漢字の割合に関しても、単純な数値計算ではなく、傾向を把握するための方法が必要であろう。

今後、写本からの正確な翻字が自動化できれば、字母の出現率の相違に着目することも容易になるだろう。もし翻字の自動化が可能になった場合には、調査対象とした全ての鎌倉時代の写本に当てはまる傾向ではないが、鎌倉時代の多くの写本の仮名の字母に当てはまる傾向があることがわかったことから、今後はこれらの字母に着目して写本の字母を考察することで、古態を伺える写本の検討ができるだろう。特に、統計的な手法を用いる場合は、特定の仮名の字母の出現の有無だけでなく、出現率の傾向の変化も対象とすることができるため、本調査から得た知見を利用して、判別分析と言った分類を目的とした統計的な手法を用いて、時代ごとに写本を分類できる可能性がある。

6. まとめ

10 本論文では、古典籍に出現する文字情報を特徴量として用いて、古典籍の書写年代の推定の可能性について検討してきた。具体的には、源氏物語の短編を中心に115本の写本の漢字と仮名字母の出現傾向の分析を行った。

漢字の割合と仮名の総字母数に着目した調査からは、写本の書写年代と漢字の割合や仮名の字母数との相関が低く、書写年代との関係は明らかではないこと、鎌倉時代の写本の漢字の割合は他の時代と比較して低いこと、が明らかになった。

仮名の字母の出現頻度率の調査からは、鎌倉時代の写本に特徴的な出現傾向を持つ字母が31存在することが明らかになった。このことから、仮名字母の出現頻度率に着目することで、鎌倉時代

の写本を推定できる可能性があると言える。今後は仮名字母に着目して、さらに調査を進め、書写年代との関係を分析する手法を検討していきたい。

謝辞

調査と研究を進める際に様々な助言と示唆を与えていただいた、尾州家源氏物語研究会と早稲田大学文学学術院陣野研究室の皆様には感謝いたします。

参考文献

- [1] 齊藤鉄也、「源氏物語『篝火』本文の文字の分布」、国際コミュニケーション学会 国際経営・文化研究、Vol.20, No1, 2015
- [2] 児玉幸多編、「くずし字用例辞典普及版」、東京堂出版、2010（平成二十二年）
- [3] 中野幸一、「変体仮名の手引」、武蔵野書院、1991（平成三年）
- [4] 中村一夫、「別本研究の現在／今後」『新時代への源氏学7 複線化する源氏物語』、竹林舎、2015
- [5] 齋藤達哉、「文字表記による伝本分類の試み」（科学研究費補助金基盤研究(A)2014年度研究成果報告書「日本古典籍における【表記情報学】の基盤構築に関する研究IV」）、国文学研究資料館、2015（平成二十七年）
- [6] 酒巻隆治、里洋平、「ビジネス活用事例で学ぶデータサイエンス入門」、SBクリエイティブ、2014
- [7] 青木繁伸、「クラスカル・ウォリス検定（plus多重比較）」、<http://aoki2.si.gunma-u.ac.jp/R/kruskal-wallis.html>, 2009

付録1. 調査対象とした源氏物語の帖と写本の一覧

ここには調査対象とした第3帖「空蟬」、第8帖「花宴」、第11帖「花散里」、第16帖「閑屋」、第27帖「篝火」、第38帖「鈴虫」の計115本の書名、時代区分、本文系統、伝承筆者をまとめている。

時代区分は前稿^[1]と同様に、次のように分けた。鎌倉時代の前期、中期、後期の区分は、3代執権泰時（在職1224－1242）までを前期、8代執権時宗（在職1268－1284）までを中期、それ以降を後期とする。南北朝時代は元弘3年（正慶2年1333）の建武の親政から明德3年（元中9年1392）年の明德の和約までとし、室町前期を含む、とする。室町時代の前期、中期、後期の区分は、4代將軍義持（在職1394－1423）までを前期、9代將軍義尚（在職1473－1489）までを中期、それ以降を後期、とする。安土桃山時代は、足利義昭が京都から追放された元亀4年から徳川家康が征夷大將軍に就任した慶長8年（1573－1603）とする。江戸時代の前期、中期、後期の区分は、4代將軍家綱（在職1651－1680）までを前期、10代將軍家治（在職1760－1786）までを中期、それ以降を後期、とする。

時代のみ記述され、その時代の区分が記述されていない写本や、時代区分の範囲が広い記述のある写本は、より後の時代の書写として扱い、時代区分に記述している。解説に伝承筆者の記述がある場合は、伝承筆者を記している。解説の推定書写時期と伝承筆者の存命期間が異なる場合は、推定書写時期を優先して採用している。

第27帖「篝火」に関しては、前稿の参考文献^[1]と同様の写本の情報を記述していることに加え、伝承筆者と大島本の情報を修正したため、改めて掲載している。

第3帖「空蟬」

	書名（略称）	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	陽明文庫本	鎌倉中期	別	後鳥羽院
2	尾州家本	鎌倉中期	河	正嘉2年（1258）、二条為氏
3	伏見天皇本	鎌倉後期	青	永仁（1293－1298）
4	池田本	鎌倉末期	青	甲筆
5	玉里文庫古筆源氏本	鎌倉	別	世尊寺行房
6	御物本	南北朝	別	鎌倉時代もしくは南北朝時代、兼好法師
7	穂久邇文庫本	南北朝	青	鎌倉末期吉野朝時代、後光厳院
8	高松宮家本	室町中期	河	長享2年（1488）か、冷泉為広
9	書陵部蔵三条西家本	室町後期	青	延徳元年～永正3年（1489－1506）
10	大正大学本	室町後期	青	近衛政家（1444－1505）
11	保坂本	室町後期	青	永正年間（1510（1504－1521））頃、飛鳥井雅綱
12	日本大学蔵三条西家本	室町後期	青	大永5年（1525）三条西公条（1487－1563）
13	飯島本	室町後期	河	冷泉為和（1486－1549）一部書写か
14	大島本	室町後期	青	文明13年～永禄6年（1481－1563）か、宮河印
15	覚勝院抄本	室町後期	青	元亀2年（1571）
16	國學院大學本	室町末期	青	岩山尚宗入道（道堅）（－1532）
17	中院文庫本	江戸前期	青	
18	東久邇宮家旧蔵本	江戸前期	青	

第8帖「花宴」

	書名（略称）	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	尾州家本	鎌倉中期	河	正嘉2年（1258）、六条有忠
2	伏見天皇本	鎌倉後期	青	永仁（1293－1298）
3	玉里文庫古筆源氏本	鎌倉	河	世尊寺行房
4	國學院大學伝為家本	鎌倉	河	藤原為家
5	陽明文庫本	南北朝	別	鎌倉後期吉野時代、冷泉為相
6	御物本	南北朝	別	鎌倉時代もしくは南北朝時代、伝後伏見院
7	穂久邇文庫本	南北朝	青	鎌倉末期吉野時代、二条為明

8	高松宮家本	室町中期	河	長享2年(1488)か、増運
9	書陵部蔵三条西家本	室町後期	青	延徳元年～永正3年(1489-1506)
10	保坂本	室町後期	青	永正年間(1510(1504-1521))頃、万里小路秀房
11	日本大学蔵三条西家本	室町後期	青	享禄3年(1530)、三条西実隆(1455-1537)
12	飯島本	室町後期	河	冷泉為和(1486-1549)須磨巻一部書写か
13	大島本	室町後期	青	文明13年～永禄6年(1481-1563)か
14	覚勝院抄本	室町後期	青	元亀2年(1571)
15	桃園文庫明融本	室町後期	青	冷泉明融(1513?-1581)
16	大正大学本	室町後期	青	
17	國學院大學本	室町末期	青	四条大納言隆永卿(1478-1538)
18	書陵部蔵正徹本	室町末期	青	
19	東久邇宮家旧蔵本	江戸前期	青	

第11帖「花散里」

	書名(略称)	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	尊経閣文庫定家本	鎌倉前期	青	
2	陽明文庫本	鎌倉中期	別	甘露寺資経
3	尾州家本	鎌倉中期	河	正嘉2年(1258)、藤原為家
4	伏見天皇本	鎌倉後期	青	永仁(1293-1298)
5	穂久邇文庫本	南北朝	青	鎌倉末期吉野時代、二条為明
6	御物本	南北朝	別	鎌倉時代もしくは南北朝時代、三条公忠
7	高松宮家本	室町中期	河	長享2年(1488)か、隆旬
8	大正大学本	室町後期	青	延徳2年～明応2年(1490-1493)
9	國學院大學本	室町後期	青	慈照院御台所
10	書陵部蔵三条西家本	室町後期	青	延徳元年～永正3年(1489-1506)
11	保坂本	室町後期	青	永正年間(1510(1504-1521))頃、三条西実隆
12	日本大学蔵三条西家本	室町後期	青	享禄3年(1530)、三条西実隆
13	飯島本	室町後期	河	冷泉為和(1486-1549)須磨巻一部書写か
14	大島本	室町後期	青	文明13年～永禄6年(1481-1563)か
15	覚勝院抄本	室町後期	青	元亀2年(1571)
16	明融本	室町後期	青	冷泉明融(1513?-1581)
17	書陵部蔵正徹本	室町末期	青	
18	書陵部蔵河内本(44冊)	室町	河	
19	中院文庫本	江戸前期	青	正保4年(1647)
20	東久邇宮家旧蔵本	江戸前期	青	

第16帖「関屋」

	書名(略称)	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	陽明文庫本	鎌倉中期	別	慈鎮
2	尾州家本	鎌倉中期	河	正嘉2年(1258)、阿仏
3	伏見天皇本	鎌倉後期	青	永仁(1293-1298)
4	榊原本	鎌倉	青	
5	玉里文庫古筆源氏	鎌倉	別	世尊寺行房
6	御物本	南北朝	別	鎌倉時代もしくは南北朝時代、後伏見院

データマイニング手法の古典籍への応用

7	穂久邇文庫本	南北朝	青	鎌倉末期吉野時代、二条為定
8	高松宮家本	室町中期	河	長享2年(1488)か、隆旬
9	大正大学本	室町後期	青	延徳2年～明応2年(1490-1493)
10	書陵部蔵三条西家本	室町後期	青	延徳元年～永正3年(1489-1506)
11	保坂本	室町後期	青	永正年間(1510(1504-1521)頃、飛鳥井雅俊
12	日本大学蔵三条西家本	室町後期	青	享禄3年(1530)、三条西実隆
13	飯島本	室町後期	河	冷泉為和(1486-1549)須磨巻一部書写か
14	大島本	室町後期	青	文明13年～永禄6年(1481-1563)か、飛鳥井雅康、宮河印
15	覚勝院抄本	室町後期	青	元亀2年(1571)
16	國學院大學本	室町後期	青	賀田武光
17	書陵部蔵正徹本	室町末期	青	
18	書陵部蔵河内本(44冊)	室町	河	
19	中院文庫本	江戸前期	青	正保4年(1647)
20	東久邇宮家旧蔵本	江戸前期	青	

第27帖「篝火」

	書名(略称)	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	保坂本	鎌倉中期	別	
2	陽明文庫本	鎌倉中期	別	後深草院
3	伏見天皇本	鎌倉後期	青	永仁(1293-1298)
4	穂久邇文庫本	南北朝	青	鎌倉末期吉野時代、二条為明
5	御物本	南北朝	別	鎌倉時代もしくは南北朝時代、慶雲法師
6	尾州家本	室町前期	河	応永2-3年(1395-96)、清水谷実秋
7	高松宮家本	室町中期	河	長享2年(1488)か、隆旬
8	大正大学本	室町後期	青	延徳2年～明応2年(1490-1493)
9	書陵部蔵三条西家本	室町後期	青	延徳元年～永正3年(1489-1506)、三条西実隆
10	國學院大學本	室町後期	青	正親町公兼
11	日本大学蔵三条西家本	室町後期	青	享禄3年(1530)、三条西実隆
12	飯島本	室町後期	別	冷泉為和(1486-1549)須磨巻一部書写か
13	大島本	室町後期	青	文明13年～永禄6年(1481-1563)か
14	覚勝院抄本	室町後期	青	元亀2年(1571)
15	河野美術館本	室町後期	青	青蓮院尊朝親王
16	書陵部蔵正徹本	室町末期	青	
17	中院文庫本	江戸前期	青	正保4年(1647)
18	東久邇宮家旧蔵本	江戸前期	青	
19	書陵部蔵河内本(44冊)	江戸初期	河	

14

第38帖「鈴虫」

	書名(略称)	時代区分	系	解説に基づく推定される書写時期と伝承筆者
1	天理図書館本	鎌倉前期	河	藤原俊成、青表紙本か
2	中山本	鎌倉前期	別	藤原為家
3	保坂本	鎌倉中期	別	
4	陽明文庫本	鎌倉中期	別	九条教家
5	日大鎌倉諸本	鎌倉中期	青	二条為氏

6	尾州家本	鎌倉中期	河	冷泉為相
7	伏見天皇本	鎌倉後期	青	永仁(1293-1298)、伏見天皇
8	穂久邇文庫本	南北朝	別	鎌倉末期吉野時代、二条為定
9	御物本	南北朝	河	鎌倉時代もしくは南北朝時代、後二条院または後伏見院
10	高松宮家本	室町中期	河	長享2年(1488)か
11	大正大学本	室町後期	青	延徳2年~明応2年(1490-1493)
12	書陵部蔵三条西家本	室町後期	青	延徳元年~永正3年(1489-1506)
13	國學院大學本	室町後期	青	正親町公兼
14	大島本	室町後期	青	文明13年~永禄6年(1481-1563)か
15	日本大学蔵三条西家本	室町後期	青	享禄4年(1531)、三条西公条
16	飯島本	室町後期	別	冷泉為和(1486-1549)須磨巻一部書写か
17	寛勝院抄本	室町後期	青	元亀2年(1571)
18	中院文庫本	江戸前期	青	寛永19年(1642)
19	東久邇宮家旧蔵本	江戸前期	青	

付録2. 字母表

参考文献 [2][3] に基づいて作成した字母の表である。本文中での漢字と仮名の字母は下記表に基づいて分類した。また、字母の出現頻度分布の図と表も下記表に基づいている。全ての現代仮名遣いの仮名の字母を表の一番目に配置している。下記表の読み方は、『『あ』の字母である『安』はa1と表す』と読む。「き」の16番目 (k116) のようにコンピュータ上で文字として表現できない漢字は偏や旁を組み合わせて表している。

	1	2	3	4	5	6	7
あ	安	阿	愛	亜	惡		
い	以	伊	移	意	異	夷	射
う	宇	有	雲	右	鶴	羽	憂
え	衣	江	要	盈	延	得	縁
お	於	隱	意				

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
か	加	可	閑	賀	駕	家	歌	哥	譌	荷	嘉	佳	香	我	敷	霞	鹿	下
き	幾	起	支	費	木	喜	期	記	季	祈	伎	藝	宜	吉	(七/(七+七))			
く	久	具	九	求	供	俱												
け	計	介	遣	希	氣/気	稀	个	家										
こ	己	古	故	許	胡	期	興	子	兒	乞								

	1	2	3	4	5	6	7	8	9	10
さ	左	佐	散	斜	狹	差	沙	乍	作	障
し	之	志	事	新	斯	四	師	使	思	指
す	寸	春	壽/寿	須	敷/敷	受	爪			
せ	世	勢	聲	瀬						
そ	曾	楚	所	處	蘇	序	其			

た	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ta	太	多	堂	當	田	唾										
ち	知	地	千	遲	致	智	馳	池								
つ	川	徒	津	都	頭	豆	追									
て	天	帝	亭	轉	傳/伝	弓	(イ+弓)	而	手	提	氏					
と	止	登	東	度	等	斗	刀	徒	土	砥	戸	兔	音	十	常	跡

な	1	2	3	4	5	6	7	8	9
na	奈	那	難	南	名	菜			
に	仁	尔/爾		耳	二	兒	尼	而	(尔+,)
ぬ	奴	怒	努	驚	沼				
ね	祢/禰	年	念	根	音	熱	寢	子	
の	乃	能	農	濃	迺	野			

は	1	2	3	4	5	6	7	8	9	10	11	12
ha	波	者	盤	八	半	葉	破	芳	婆	羽	頗	齒
ひ	比	飛	日	悲	非	火	妣	避				
ふ	不	布	婦	風	敷	夫	經					
へ	β	部	遍	篇	偏	邊/辺	反	敝	倍	幣	辨	弊
ほ	保	本	報	寶	奉	穗						

ま	1	2	3	4	5	6	7	8	9	10
ma	未	万	萬	滿	真	間	馬	麻	摩	漫
み	美	見	三	微	薇	身	民			
む	武	無	舞	无	牟	務	夢	六		
め	女	免	面	馬	目	妻	米			
も	毛	无	茂	裳	母	藻	蒙	聞	方	物

10

	1	2	3	4	5	6	7	8	9
や	也	屋	夜	耶	哉	野	弥	八	
ゆ	由	遊	游	喻	湧	柚	湯		
よ	与	余	餘	夜	世	代	容	欲	慾

	1	2	3	4	5	6	7	8
ら	良	羅	蘭	等	落			
り	利	梨	里	理	李	離		
る	留	流	類	累				
れ	禮/礼	連	麗					
ろ	呂	侶	論	婁	樓	路	露	盧

	1	2	3	4	5	6	7
わ	和	王	倭	輪	○		
み	為	井	遣	居	委	威	渭
系	惠	衛	慧				
を	遠	越	乎	惡	緒	尾	

	1
ん	无

付録3.

ここでは、本論文中の「2-2. 写本からの採字方針」に基づいて、調査対象とした写本の本文の文字数を、本文の総文字数、総文字数に占める漢字と仮名の割合、漢字は使用総数とその種類数、仮名は使用総数とその字母数についてまとめた。写本の順番は、付録1. の「調査対象とした源氏物語の帖と写本の一覧」の時代順に従って掲載している。

第3帖「空蟬」

書名(略称)	本文総文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
陽明文庫本	4834	4.80	95.20	232	36	4602	93
尾州家本	4886	3.99	96.01	195	32	4691	97
伏見天皇本	4768	5.14	94.86	245	52	4523	112
池田本	4842	5.18	94.82	251	34	4591	110
玉里文庫古筆源氏本	4926	3.67	96.33	181	50	4745	112
御物本	4883	4.24	95.76	207	28	4676	89
穂久邇文庫本	4832	4.84	95.16	234	34	4598	102
高松宮家本	4825	4.66	95.34	225	31	4600	106
書陵部蔵三条西家本	4844	5.02	94.98	243	42	4601	116
大正大学本	4834	5.56	94.44	269	58	4565	109
保坂本	4786	6.73	93.27	322	63	4464	117
日大蔵三条西家本	4839	5.31	94.69	257	30	4582	105
飯島本	4941	4.45	95.55	220	42	4721	119
大島本	4894	4.43	95.57	217	43	4677	103
覚勝院抄本	4783	7.07	92.93	338	71	4445	100
國學院大學本	4700	8.13	91.87	382	88	4318	118
中院文庫本	4840	5.29	94.71	256	30	4584	106
東久邇宮家旧蔵本	4794	6.97	93.03	334	61	4460	118

第8帖「花宴」

書名(略称)	本文総文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
尾州家本	5005	5.49	94.51	275	63	4730	94
伏見天皇本	4589	6.93	93.07	318	76	4271	95
玉里文庫古筆源氏本	4954	5.85	94.15	290	65	4664	113
國學院大學伝為家本	5019	5.22	94.78	262	59	4757	84
陽明文庫本	4504	8.33	91.67	375	64	4129	83
御物本	4578	6.64	93.36	304	60	4274	99
穂久邇文庫本	4419	9.73	90.27	430	88	3989	99
高松宮家本	4821	8.84	91.16	426	92	4395	95
書陵部蔵三条西家本	4463	9.34	90.66	417	114	4046	117
保坂本	4586	6.85	93.15	314	82	4272	109
日大蔵三条西家本	4623	6.58	93.42	304	95	4319	108
飯島本	4958	4.70	95.30	233	56	4725	99
大島本	4547	8.12	91.88	369	82	4178	99
覚勝院抄本	4392	12.25	87.75	538	134	3854	106

データマイニング手法の古典籍への応用

桃園文庫蔵明融本	4549	7.67	92.33	349	76	4200	92
大正大学本	4575	6.86	93.14	314	77	4261	117
國學院大學本	4605	6.71	93.29	309	78	4296	101
書陵部蔵正徹本	4367	11.66	88.34	509	114	3858	114
東久邇宮家旧蔵本	4448	10.12	89.88	450	106	3998	104

第11帖「花散里」

書名(略称)	本文総文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
尊経閣文庫定家本	1693	4.13	95.87	70	22	1623	73
陽明文庫本	1668	6.65	93.35	111	32	1557	77
尾州家本	1825	3.62	96.38	66	19	1759	86
伏見天皇本	1660	7.29	92.71	121	49	1539	96
穂久邇文庫本	1671	6.40	93.60	107	42	1564	102
御物本	1693	6.20	93.80	105	30	1588	91
高松宮家本	1820	3.74	96.26	68	19	1752	80
大正大学本	1796	6.40	93.60	115	41	1681	112
國學院大學本	1675	6.81	93.19	114	37	1561	96
書陵部蔵三条西家本	1662	7.22	92.78	120	43	1542	102
保坂本	1699	6.71	93.29	114	32	1585	105
日大蔵三条西家本	1688	5.75	94.25	97	26	1591	89
飯島本	1790	4.41	95.59	79	20	1711	88
大島本	1678	6.56	93.44	110	38	1568	96
覚勝院抄本	1602	11.99	88.01	192	66	1410	95
明融本	1667	6.00	94.00	100	35	1567	88
書陵部蔵正徹本	1749	8.40	91.60	147	44	1602	104
書陵部蔵河内本(44)	1803	4.38	95.62	79	31	1724	94
中院文庫本	1682	5.77	94.23	97	26	1585	84
東久邇宮家旧蔵本	1668	7.85	92.15	131	48	1537	94

第16帖「関屋」

書名(略称)	本文文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
陽明文庫本	2112	5.30	94.70	112	36	2000	79
尾州家本	2138	3.27	96.73	70	22	2068	83
伏見天皇本	2009	7.62	92.38	153	55	1856	95
榑原本	2113	4.59	95.41	97	33	2016	100
玉里文庫古筆源氏本	2024	4.25	95.75	86	28	1938	112
御物本	2046	6.35	93.65	130	35	1916	88
穂久邇文庫本	2067	5.85	94.15	121	37	1946	99
高松宮家本	2158	4.08	95.92	88	23	2070	86
大正大学本	2094	6.06	93.94	127	38	1967	102
書陵部蔵三条西家本	2080	6.59	93.41	137	44	1943	96
保坂本	2053	6.62	93.38	136	47	1917	100

日大蔵三条西家本	2099	5.24	94.76	110	36	1989	98
飯島本	2078	6.54	93.46	136	51	1942	95
大島本	2089	6.37	93.63	133	38	1956	93
覚勝院抄本	1995	10.98	89.02	219	75	1776	101
國學院大學本	2078	6.59	93.41	137	39	1941	89
書陵部蔵正徹本	2012	9.24	90.76	186	57	1826	106
書陵部蔵河内本(44)	2134	3.89	96.11	83	24	2051	93
中院文庫本	2099	5.34	94.66	112	38	1987	87
東久邇宮家旧蔵本	2023	9.19	90.81	186	64	1837	102

第27帖「篝火」

書名(略称)	本文文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
保坂本	1490	5.37	94.63	80	22	1410	104
陽明文庫本	1469	6.81	93.19	100	28	1369	87
伏見天皇本	1451	7.93	92.07	115	41	1336	107
穂久邇文庫本	1472	7.07	92.93	104	35	1368	102
御物本	1524	4.59	95.41	70	21	1454	93
尾州家本	1434	9.27	90.73	133	42	1301	89
高松宮家本	1447	8.36	91.64	121	30	1326	86
大正大学本	1477	7.24	92.76	107	30	1370	102
書陵部蔵三条西家本	1472	7.61	92.39	112	33	1360	103
國學院大学本	1498	6.41	93.59	96	30	1402	86
日大蔵三条西家本	1461	7.19	92.81	105	31	1356	106
飯島本	1439	8.06	91.94	116	29	1323	98
大島本	1469	7.42	92.58	109	35	1360	86
覚勝院抄本	1403	11.69	88.31	164	56	1239	98
河野美術館本	1439	9.94	90.06	143	52	1296	100
書陵部蔵正徹本	1437	9.67	90.33	139	47	1298	108
中院文庫本	1462	7.18	92.82	105	31	1357	84
東久邇宮家旧蔵本	1421	11.33	88.67	161	52	1260	99
書陵部蔵河内本(44)	1399	12.79	87.21	179	69	1220	100

第38帖「鈴虫」

書名(略称)	本文文字数	割合		漢字		仮名	
		漢字(%)	仮名(%)	文字数	種類数	文字数	字母数
天理図書館本	6549	6.28	93.72	411	75	6138	103
中山本	5705	6.61	93.39	377	69	5328	106
保坂本	6825	6.33	93.67	432	65	6393	93
陽明文庫本	6289	8.46	91.54	532	82	5757	97
日大鎌倉諸本	6442	7.16	92.84	461	101	5981	102
尾州家本	6568	6.26	93.74	411	69	6157	83
伏見天皇本	6447	7.74	92.26	499	139	5948	109
穂久邇文庫本	6540	8.10	91.90	530	80	6010	103

データマイニング手法の古典籍への応用

御物本	6639	4.40	95.60	292	47	6347	113
高松宮家本	6457	7.14	92.86	461	86	5996	107
大正大学本	6368	8.75	91.25	557	95	5811	113
書陵部蔵三条西家本	6291	9.55	90.45	601	120	5690	116
國學院大学本	6380	8.28	91.72	528	75	5852	100
大島本	6292	9.38	90.62	590	106	5702	107
日大蔵三条西家本	6472	6.46	93.54	418	91	6054	99
飯島本	6191	11.31	88.69	700	141	5491	106
覚勝院抄本	6221	10.42	89.58	648	119	5573	108
中院文庫本	6470	6.51	93.49	421	91	6049	84
東久邇宮家旧蔵本	6205	10.62	89.38	659	109	5546	105

(受理 平成28年9月6日)