情報処理学会論文誌 Vol.63 No.2 347–354 (Feb. 2022)

本文表記のNgramを用いた室町時代書写の源氏物語写本の分類

齊藤 鉄也[1,a)]

受付日 2021年5月9日, 採録日 2021年11月2日

概要：本稿では，源氏物語写本の本文表記の Ngram を用いて源氏物語写本の書承関係を推定した．調査方法として，文献学の知見から本文の類似性が指摘されている写本を用いて，写本間関係を表す定量的な尺度を設計し，その尺度に基づいて調査対象とした源氏物語写本本文の親疎を探索した．室町時代書写の写本を中心とした「桐壺」から「松風」18 帖の本文の調査結果からは，転写過程の共有といった，これまで指摘されていない書承関係を持つ可能性がある写本が存在することが明らかになった．

キーワード：表記，Ngram，室町時代源氏物語写本，クラスタリング，cosine 距離

Clustering among Textual-line Origins between Muromachi-period Manuscripts of the *Tale of Genji* as Seen in Ngram Analysis of Orthography

Tetsuya Saito[1,a)]

Received: May 9, 2021, Accepted: November 2, 2021

Abstract: This paper sought to determine the relationships between various manuscripts of the *Tale of Genji* using an Ngram analysis of differences between these manuscripts in orthography. The method of investigation used first posited a quantitative measure for expressing the degree of closeness between manuscripts, one developed based on examples of manuscripts whose closeness had previously been determined using the expertise of traditional philology. This measure was used in turn to estimate the degree of closeness between those particular *Genji* manuscripts selected for the study. Using such a method of investigation, differences in orthography over the first 18 chapters of the *Tale of Genji* were examined in a number of manuscripts, centered around those copied during the Muromachi period. The results showed that by such a method it is possible to discover manuscripts whose mutual closeness and stemmatic affiliations had not been previously recognized.

Keywords: orthography, Ngram, Muromachi-period manuscripts of the *Tale of Genji*, clustering, cosine distance

1. はじめに

本稿では，計量的な手法を用いて，表記の点から源氏物語本文を分類した結果を報告する．以下，本稿では，計 14 の源氏物語写本を調査対象としているため，通称に基づく省略した写本名称を用いる．たとえば，大島本源氏物語であれば，大島本と記す．

調査対象には，現在まで写本が多く残り，出版または画像公開されている室町時代に書写された源氏物語写本を中心に選択した．これら写本の本文から，仮名と漢字の使い分けや送り仮名，仮名遣い，音便といった本文表記の相違を反映した本文データを作成した．調査方法は次の 2 段階で行った．最初に，文献学の分野において，本文の類似性が指摘されている写本の本文データを用いて，写本間の書承関係を表す定量的な尺度を設計した．次に，その尺度を用いて調査対象とした写本群の写本間距離を比較し，書承関係があると考えられる写本関係を調査した．定量的な尺度の設計では，以下の方法を採用した．最初に，写本の本文データから，文字の Ngram と呼ばれる，本調査で検討

1 淑徳大学
Shukutoku University, Iruma-gun, Saitama 354–8510, Japan
a) saimune@u.shukutoku.ac.jp

© 2022 Information Processing Society of Japan

した適切なN文字の長さを持つ文字列を，本文の先頭文字から順に生成した．次に，その文字列の頻度からなる，本文の特徴を表す数値データに対して，本文間の相違を表す写本間距離をcosine距離により計算し，それを写本の書承関係を表す尺度の値に用いた．

計量的な手法を用いた調査結果からは，文献学から指摘されている，大島本と書陵部蔵三条西家本，大正大学本の本文の類似性の知見に関して，本文表記の点からも，その蓋然性を高める結果を得た．加えて，大島本と国文研蔵正徹本に書承関係がある可能性を示す仮説を提案した．

2. 本研究の目的と関連研究

本研究の目的は，源氏物語写本を対象とし，計量的な手法で，それらの関係を調査することである．本調査では，共通の祖本を持つといった書承関係のある写本を発見することを目標とした．調査方法として，文献学の調査により，本文の類似性が指摘されている写本群を対象に，写本本文の表記のNgramを用いた写本間距離の尺度を設計し，その尺度を用いて写本間距離が近い写本を探索した．

2.1 本調査の位置付け

源氏物語の写本は，その伝来を示す奥書が記された写本が少なく，写本間の本文の関係を発見するためには本文異同を比較する [1]．校訂された本文異同の比較結果に基づき，源氏物語写本は，青表紙（定家）本系統，河内本系統と，それら以外の写本を総称した別本群と呼ばれる，2系統1群の本文に分類することが通説となっている．

これに対して，本調査では，校訂されていない写本本文の表記に着目し，本文を分類することを試みた．校訂前の本文は，同系統に分類される写本であっても，本文異同に加えて，本文の仮名と漢字の使い分けや送り仮名，仮名遣い，音便といった表記の相違がある．表記が近い本文は，共通祖本を持つ可能性や転写における世代が近い可能性といった写本の転写過程を共有する可能性が指摘できる [2]．

表記の一致または類似が指摘される写本の報告は少ないことから [3], [4], [5]，表記が似ている本文を持つ写本は，何らかの写本間関係を指摘でき，伝来に関する知見が得られる可能性が高い．しかし，これまで表記に着目し，複数の巻にわたって写本間関係を統一的に調査した事例は少なく [6]，本文に大量に存在する表記の相違の傾向は不明なことが多い．そこで，本研究では，写本が多く現存し公開されている室町時代書写の源氏物語の複数の巻を対象に，計量的な方法を用いて，表記の点から源氏物語本文を分類し，書承関係が指摘できる可能性がある写本を探索する．

2.2 関連研究

奥書を持たず写本間関係が不明な源氏物語に対する，文献学の方法を用いた調査においては，本文異同に着目した分類に加えて，本文表記に用いられている漢字に着目した調査報告がある．たとえば，一揃いの写本に対して，特定の漢字表記に基づいた分類 [3] や漢字の異体字の相違に着目した分類 [7]，複数の一揃いの写本を対象とした，漢字の宛て方に基づいた分類 [1] がある．これらの研究では，特徴的な漢字の使い分けに着目して写本間関係の考察を行い，本文異同に着目した本文の2系統1群といった分類とは別に，表記の特徴に基づいた写本の分類やその共通祖本の考察を行っている．

源氏物語写本に対して計量的方法を適用して，本文異同の分類を試みた研究が存在する [8], [9], [10] が，表記を対象として計量的方法を用いた研究は少ない．計量的方法を適用した本文表記の分類 [11] の調査報告では，少数の写本を元に写本間距離を求め，暫定的な尺度を設計し，写本を分類している．本調査では，この報告で提案した尺度を再検討し，調査対象とする複数の一揃いの写本を拡大した．

3. 調査対象写本と本文データ

本調査では，本文表記の写本間距離を測る尺度を設計するための写本と，室町時代書写の源氏物語写本を中心に表記の近い写本を探索するための写本を用いる．尺度設計に用いた写本と探索対象とした写本とは異なる．

稿者が作成した尺度設計と調査に用いた本文データは，出版または画像公開されている写本を用いて，写本と同一の行数と改行位置を持つ．本文データの入力の際には，1音で読む漢字は仮名と見なし，誤写を修正した文字や挿入された文字といった傍記は調査対象とせず，本行本文のみを採字対象としている．漢字の異体字は通行漢字に統一している．写本本文の調査範囲は，各巻の写本間で比較可能とするために，本文の同一箇所までとしている．

この本文データは，写本の本行本文を反映しているため，本文は校訂されず，表記の相違に加えて，脱落や誤写，虫損といった相違を含む．このため，本文異同の比較や品詞分解といった処理には適していない．その一方で，仮名と漢字の使い分けや送り仮名，仮名遣い，音便を反映していることから，表記の相違を調査することに適している．

写本間距離の尺度を設計するために用いた，これまで文献学において写本本文の関係が指摘されている写本の組は，(1) 字形まで忠実に書写され臨模されている尊経閣文庫本 [12] と明融臨模本 [13] の第36帖「柏木」の1帖，(2) 行詰と字詰といった字配，仮名と漢字の使い分けが同一に書写されている日大本 [14] と中院文庫本 [15] の19帖，(3) 同一の親本を持つとの指摘がある池田本 [16] と明融臨模本の第1帖「桐壺」と第8帖「花宴」の2帖，(4) 同一系統内で異同が少ないとの指摘がある河内本系統を多く含む尾州家本 [17] と高松宮家本 [18] の41帖である．これらの写本組は，この順で写本間関係の親疎がある．

これらの写本組の写本名と巻の序数，その本文データの

© 2022 Information Processing Society of Japan

表 1　尺度設計に用いた写本組

Table 1　Pairs of manuscripts used to develop the measure.

資料名	文字数	漢字率	資料名	文字数	漢字率	資料名	文字数	漢字率	資料名	文字数	漢字率
尊経閣文庫本-36	17,855	6.4%	**明融臨模本-36**	17,875	6.4%	尾州家本-14	6,458	8.7%	高松宮家本-14	6,542	7.2%
日大本-01	6,210	8.7%	中院文庫本-01	6,202	8.6%	尾州家本-15	6,954	3.6%	高松宮家本-15	6,921	4.2%
日大本-02	6,180	6.1%	中院文庫本-02	6,165	6.0%	**尾州家本-16**	2,138	3.3%	**高松宮家本-16**	2,158	4.1%
日大本-03	4,839	5.3%	**中院文庫本-03**	4,841	5.3%	尾州家本-17	6,117	8.9%	高松宮家本-17	6,050	10.9%
日大本-05	5,900	7.1%	中院文庫本-05	5,916	7.0%	尾州家本-18	5,839	6.9%	高松宮家本-18	5,750	8.1%
日大本-11	1,688	5.7%	**中院文庫本-11**	1,686	5.8%	尾州家本-19	6,193	4.0%	高松宮家本-19	6,158	4.4%
日大本-16	2,099	5.2%	**中院文庫本-16**	2,099	5.3%	尾州家本-20	5,970	3.3%	高松宮家本-20	5,508	5.2%
日大本-23	6,451	4.7%	**中院文庫本-23**	6,446	4.8%	尾州家本-21	6,084	7.1%	高松宮家本-21	6,083	7.3%
日大本-27	1,461	7.2%	**中院文庫本-27**	1,462	7.2%	尾州家本-22	5,583	4.3%	高松宮家本-22	5,527	5.1%
日大本-29	12,517	4.7%	**中院文庫本-29**	12,531	4.7%	**尾州家本-23**	6,375	6.1%	**高松宮家本-23**	6,512	4.6%
日大本-30	6,185	7.4%	**中院文庫本-30**	6,189	7.4%	尾州家本-24	6,308	5.2%	高松宮家本-24	6,324	5.2%
日大本-34	6,309	7.3%	中院文庫本-34	6,322	7.3%	尾州家本-25	6,566	5.2%	高松宮家本-25	6,574	5.3%
日大本-35	6,095	8.4%	中院文庫本-35	6,096	8.4%	尾州家本-26	6,574	4.3%	高松宮家本-26	6,569	5.2%
日大本-36	5,670	6.1%	中院文庫本-36	5,664	6.0%	**尾州家本-27**	1,434	9.3%	**高松宮家本-27**	1,447	8.4%
日大本-38	6,472	6.5%	**中院文庫本-38**	6,470	6.5%	尾州家本-28	7,165	6.3%	高松宮家本-28	7,164	6.8%
日大本-42	6,308	7.9%	**中院文庫本-42**	6,312	7.8%	**尾州家本-29**	12,209	8.6%	**高松宮家本-29**	12,160	8.4%
日大本-43	5,901	5.6%	**中院文庫本-43**	5,901	5.6%	尾州家本-31	6,970	6.5%	高松宮家本-31	7,059	5.2%
日大本-45	5,501	3.8%	中院文庫本-45	5,504	3.8%	尾州家本-32	5,977	8.2%	高松宮家本-32	5,977	8.3%
日大本-48	8,116	5.0%	**中院文庫本-48**	8,109	4.9%	尾州家本-33	6,038	6.1%	高松宮家本-33	5,909	8.2%
日大本-51	5,839	6.4%	中院文庫本-51	5,850	6.4%	尾州家本-34	6,431	5.9%	高松宮家本-34	6,396	6.4%
池田本-01	10,980	9.3%	**明融臨模本-01**	10,761	9.4%	尾州家本-35	6,173	7.8%	高松宮家本-35	6,167	8.3%
池田本-08	4,528	7.9%	**明融臨模本-08**	4,522	7.7%	尾州家本-36	5,703	4.7%	高松宮家本-36	5,745	4.2%
尾州家本-02	6,451	4.1%	高松宮家本-02	6,380	5.4%	**尾州家本-38**	6,568	6.3%	**高松宮家本-38**	6,457	7.1%
尾州家本-03	4,886	4.0%	**高松宮家本-03**	4,825	4.7%	尾州家本-39	6,172	5.1%	高松宮家本-39	6,096	6.3%
尾州家本-04	6,202	3.7%	高松宮家本-04	6,155	4.6%	尾州家本-40	6,394	5.5%	高松宮家本-40	6,277	7.7%
尾州家本-05	6,258	3.2%	高松宮家本-05	6,238	3.7%	尾州家本-41	7,030	4.8%	高松宮家本-41	6,900	6.2%
尾州家本-06	6,274	3.2%	高松宮家本-06	6,249	3.9%	**尾州家本-42**	6,265	7.9%	**高松宮家本-42**	6,347	7.9%
尾州家本-07	6,765	6.2%	高松宮家本-07	6,598	9.0%	尾州家本-44	6,144	7.6%	高松宮家本-44	6,152	7.5%
尾州家本-08	5,005	6.2%	**高松宮家本-08**	4,821	8.8%	尾州家本-46	6,076	7.0%	高松宮家本-46	6,035	8.0%
尾州家本-09	7,166	5.6%	高松宮家本-09	7,138	5.8%	尾州家本-47	5,980	5.7%	高松宮家本-47	5,863	6.8%
尾州家本-10	6,722	6.8%	高松宮家本-10	6,642	8.0%	**尾州家本-48**	8,090	5.5%	**高松宮家本-48**	8,073	6.1%
尾州家本-11	1,825	3.6%	**高松宮家本-11**	1,820	3.7%						

文字数と，漢字の出現率を漢字率として**表 1** にまとめた．表 1 で太字で示した巻名は全文を調査対象としている．本文の類似性が指摘されている写本であっても，書写者ごとに漢字の宛て方の相違や誤写があるため，表中の文字数や漢字の出現率は異なっている．日大本と中院文庫本で利用した巻は，本文長が短い巻を中心に源氏物語全体の三分の一程度の巻数を選択した．尾州家本と高松宮家本で利用した巻は，調査対象範囲に落丁が存在しない巻であり，かつ，同系統の本文である巻を選択した．

尺度を用いて探索する対象とした写本を**表 2** にまとめた．調査対象は，源氏物語の最初の三分の一の範囲の巻とし，本方法の有用性を確認することとした．後続の巻は今後調査する予定である．表 2 では，表 1 と同様に，本文の調査範囲の文字数と，その漢字の出現率を示している．

國學院大学本 [19] は室町時代末期書写とされ，一部江戸時代に書写された写本である．今回調査対象とした範囲の巻はすべて室町時代に書写されている．書陵部蔵正徹本 [20] と国文研蔵正徹本 [21] は，ともに室町時代中期の歌僧である正徹が書写した源氏物語の転写本とされている．

表 2　探索対象とした写本

Table 2　Manuscripts surveyed.

巻名	國學院大学本		書陵部蔵正徹本		国文研蔵正徹本	
	文字数	漢字率	文字数	漢字率	文字数	漢字率
01 桐壺	6,279	8.1%	5,988	12.4%	6,411	6.2%
02 帚木	6,249	5.5%	6,147	7.4%	6,349	4.3%
03 空蝉	4,700	8.1%	4,777	6.7%	4,932	3.8%
04 夕顔	5,757	10.4%	5,743	10.3%	6,074	4.9%
05 若紫	5,820	9.0%	5,665	10.9%	5,939	6.7%
06 末摘花	5,968	7.2%	5,844	9.1%	6,065	5.9%
07 紅葉賀	6,376	9.2%	6,140	13.4%	6,389	8.7%
08 花宴	4,605	6.7%	4,367	11.7%	4,616	6.7%
09 葵	6,562	6.7%	6,225	11.6%	6,534	6.6%
10 賢木	6,212	11.4%	6,083	13.7%	6,354	8.9%
11 花散里	1,675	6.8%	1,749	8.4%	1,624	11.4%
12 須磨	5,859	9.3%	5,642	13.0%	5,956	7.5%
13 明石	6,236	7.0%	5,979	11.5%	6,320	5.7%
14 澪標	6,318	9.2%	6,021	12.5%	6,044	13.1%
15 蓬生	6,633	6.7%	6,270	11.5%	6,263	12.2%
16 関屋	2,078	6.6%	2,012	9.2%	1,934	13.3%
17 絵合	6,122	8.9%	5,876	12.2%	5,747	14.9%
18 松風	5,906	6.5%	5,562	11.7%	5,440	14.3%

ⓒ 2022 Information Processing Society of Japan

書陵部蔵正徹本は室町時代末期の書写とされる．国文研蔵正徹本は江戸時代初期の書写とされ，第 8 帖「花宴」を除き，第 1 帖「桐壺」から第 10 帖「賢木」に奥書がある．この表中の写本のほかに，比較対象として，鎌倉時代書写の池田本 [16]，室町時代書写の書陵部蔵三条西家本 [22]，大正大学本 [23]，保坂本 [24]，大島本 [25]，日大本 [14] を用いた．これらの写本は前稿 [11] に書誌情報と文字数，漢字の出現率を掲載しているので，本稿では写本の概要を省略している．なお，池田本は第 11 帖「花散里」を欠き，保坂本は第 17 帖「絵合」までが室町時代に書写されている．この結果，探索対象となった写本数は，計 9 写本 160 冊である．

4.　写本間距離尺度の設計

表記の相違を反映する本文データを用いた尺度の設計するために，文字を対象に Ngram を生成し，写本間距離に cosine 距離を用いる．この方法により，定性的な本文表記の相違は，生成される Ngram の頻度の相違として表され，cosine 距離によって，定量的に比較できる．cosine 距離は，写本間距離を 0 から 1 の範囲で表すため，同一巻内だけではなく，異なる巻間を統一的に比較することに適している．

写本間距離は，調査対象とする本文文字数の増加により，本文異同や表記の相違，誤写が本文中に含まれる可能性が高くなり，結果として増加することが想定できる．また，Ngram の文字数 N の増加に応じて，文字列中に本文の相違が含まれる可能性も高くなることから，写本間距離が増加することも想定できる．そこで，最初に，写本を分類するために適切な本文長と文字数 N を調査する．次に，調査によって求められた適切な N に基づいて，同一の巻の写本を比較する尺度を設計する．

4.1　適切な本文長の調査

写本間で仮名と漢字の使い分けが同一であり，奥書の一致とそこに記されている書写年代から，書写順序が明らかな，日大本と中院文庫本の同一巻を対象に，尺度設計に必要な本文長の調査を行う．日大本と中院文庫本は，写本の各行末の文字まで一致するため，書写者の書き癖と考えられる漢字の利用，日大本の修正の反映，誤写といった相違のほかには本文の相違が発生しない．これら写本は，本文異同による写本間距離の増加を考慮する必要がなく，本文長と表記の相違の関係を考察する対象として適している．

調査対象は，表 1 において，巻名を太字で示した写本である．これらの，本文長が 1,500 字から 12,000 字程度の 11 帖の全文を対象として，N が 1 から 10 までの範囲で，本文長と写本間距離の関係を調査した．**図 1** には，本文長が異なる写本の写本間距離を N ごとに折れ線グラフとして表した．図では，横軸に N を，縦軸に写本間距離を示している．巻名は巻の序数とともに表している．

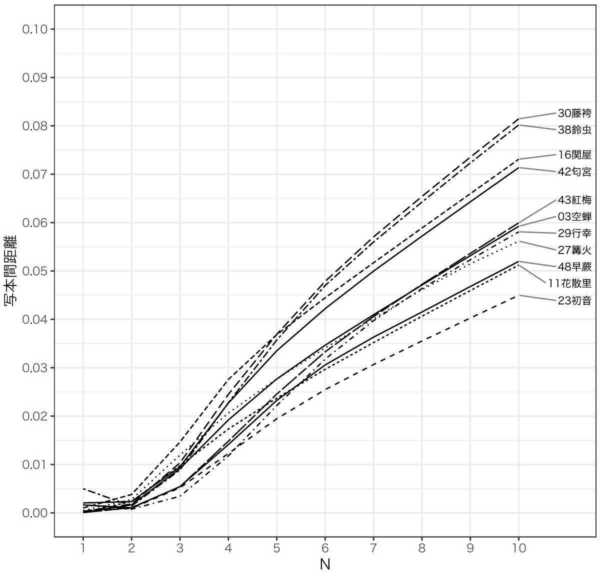

図 1　本文長が異なる写本の写本間距離

Fig. 1　Cosine distances between manuscripts containing texts of different lengths.

図 1 からは，文字数 N の増加に応じて，写本間距離が増加することは明らかではあるが，本文長に応じて写本間距離が増加するという関係は明らかではない．たとえば，第 11 帖「花散里」と第 16 帖「関屋」，第 27 帖「篝火」は本文長が短いため，表記の相違や誤写が発生する可能性が低く，写本間距離が近くなると想定できるが，いずれも最も写本間距離が近い折れ線ではない．同様に，今回の調査対象とした写本のうちで本文長が最も長い，第 29 帖「行幸」の写本間距離は，最も遠い折線ではない．つまり，写本間距離は，本文長ではなく，転写の際に発生する表記の相違やデータ作成の際の採字方針に影響を受けることを示している．たとえば，「ここち」と読む本文を，日大本では「心地」と漢字を用いているが，中院文庫本では「心ち」と仮名を用いることがある．2 文字の Ngram である，2gram の場合であれば，この表記の相違によって，「心地」と「心ち」という異なる文字列が生成され，それぞれの頻度の相違が生じ，cosine 距離を用いて写本間距離として表される．

この結果に基づき，尺度の設計には一定の本文長を対象とすることとした．源氏物語の各巻の本文長は第 11 帖「花散里」，第 27 帖「篝火」が 1,500 字程度，第 16 帖「関屋」が 2,000 字程度，第 3 帖「空蝉」，第 8 帖「花宴」が 5,000 字程度であり，その他の巻は 5,000 字以上である．写本の探索では巻ごとに本文異同も存在することを考慮し，本文が短い巻は全文を，それ以外の本文が長い巻は，より多くの巻が該当する 5,000 字程度の一定長の本文長を用いて，本文データの作成の効率化を行った．

4.2　適切な文字数 N の調査

3 章で，表 1 の写本は，これまでの文献学の調査によって，書承関係が明らかになっていることを述べた．本調査

© 2022 Information Processing Society of Japan

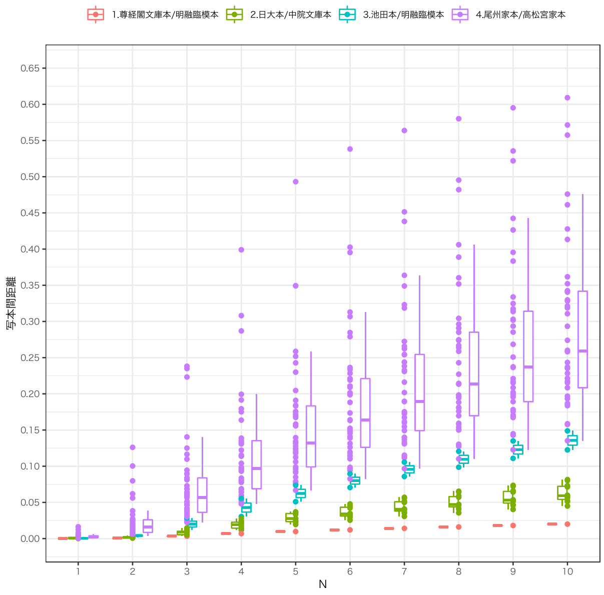

図 2　写本群ごとの箱ひげ図
Fig. 2　Boxplots for each manuscript group.

では，これら写本組の写本間距離を用いて，尺度設計に適切な文字数 N を選択する．書承関係が同じ写本群に対して，それぞれの距離を文字数 N ごとに計算し，箱ひげ図として，**図 2** にまとめた．図 2 では，各写本群が 1 つの箱ひげ図に対応し，各写本間距離が点として表されている．横軸に N を，縦軸に写本間距離を示す．

図 2 は，全体の傾向として，写本間関係の親疎の順で写本間距離が遠くなることを示している．臨模されている尊経閣文庫本と明融臨模本の第 36 帖「柏木」1 帖が最も写本間距離が近く，次に日大本と中院文庫本が写本群を構成している．続いて，共通の親本を持つと考えられる池田本と明融臨模本の第 1 帖「桐壺」と第 8 帖「花宴」の 2 帖が，尾州家本と高松宮家本の写本群と重なりながら位置している．尾州家本と高松宮家本の写本群は，点が広範囲に位置し，外れ値が存在していることから，同一系統内では本文異同が少ないと指摘される写本間関係ながらも，表記は非常に異なっている写本が存在することを示している．

図 2 からは，文字数 N が 3 以下では，各群の写本間距離を示す点の表示が密集し，写本群を分類する写本間距離を決定することが困難であるといえる．これは，N が少なくとも 4 以上であれば写本を分類できることを示している．尺度を設計するためには，書承関係がある写本群ごとに分類が容易であるように，群間の距離がより大きいことが望ましいことから，N が 5 以上であれば分類が容易であるといえる．そこで，尺度の性質としては，写本群の群間に距離があり，各群の写本間距離の重複が少ないことに加えて，尺度を利用する際に簡便であることが望ましいので，共通の親本を持つと考えられている池田本と明融臨模本の写本間距離が 0.1 程度である，文字数 N が 7 の場合を採用することとした．

表 3　写本群の分布の要約統計量
Table 3　Representative statistical values for distribution within each manuscript group.

写本群	中央値	IQR	$Q_{3/4}$	最小値	最大値	母数
1. 尊/明	0.01					
2. 日/中	0.05	0.01	0.05	0.03	0.06	$m = 6.10$, $\eta = 0.05$
3. 池/明	0.10	0.01	0.10	0.09	0.11	
4. 尾/高	0.19	0.11	0.25	0.10	0.56	$\mu = -1.95$, $\sigma = 0.45$

N が 7 である場合に，調査対象写本数が相対的に多い日大本と中院文庫本，尾州家本と高松宮家本の各写本群に対して，写本間距離の分布を最尤推定し，赤池情報量基準を用いて適合した分布を選択した．日大本と中院文庫本はワイブル分布，尾州家本と高松宮家本は対数正規分布が最もあてはまりが良かった．これらの写本群の分布の要約統計量と母数を**表 3** にまとめた．表中では，写本名をそれぞれ先頭の 1 文字に省略し，その他の写本とともに表している．

日大本と中院文庫本の写本間距離の分布を推定し，その分布の 50%と 75%を占める割合の写本間距離は 0.05 と 0.05 であった．同様に，尾州家本と高松宮家本の分布の 50%と 75%を占める割合の写本間距離は 0.20 と 0.27 であった．これらの推定値は，表 3 のそれぞれの分布の中央値と第 3 四分位数に近いことから，調査した写本数が少ない尊経閣文庫本と明融臨模本，池田本と明融臨模本の写本群に関しても，分位数を用いて尺度を設計することとした．

4.3　写本間距離の尺度の決定

これまでの調査結果に基づき，写本間距離の尺度を設計する．尺度の利用目的は，書承関係を持つ可能性がある写本を発見することである．書承関係の調査には，本調査だけでなく，専門家による確認といった複数の視点からの検証を必要とすることから，本尺度を用いて書承関係を指摘した場合であっても，その指摘の訂正は可能である．そこで，本設計では，該当する写本間距離の閾値を小さくすると書承関係のある写本を見落とす可能性を考慮し，書承関係がある写本群で，多くの写本間距離が含まれる第 3 四分位数を尺度の指標として用いることとした．

表 4 は，写本間距離から本文の関係を推定する尺度を示している．表 4 には，設計に用いた写本と，その写本の関係の定性的な記述，写本間距離を用いて写本の関係を操作変数化した，定量的な指標を対応づけてまとめた．これまで，文献学の手法を適用した本文調査に基づく定性的な記述と，定量的な指標である写本間距離の数値との対応は明らかではなかった．ここで提案する，専門家の知見を数値化し，写本の書承関係の相違を測ることができる定量的な指標を持つ尺度によって，定性的な写本間関係を探索することが可能になった．

© 2022 Information Processing Society of Japan

表 4　写本の性質と写本間距離との関係を示す尺度

Table 4　Proposed measure for indicating the relationship between transcription policy and manuscript distance.

対象とした写本	定性的な記述	定量的な指標（N = 7）
尊経閣文庫本/明融臨模本「柏木」1 帖	字形まで忠実に書写されている	cosine 距離 0.01
日大本/中院文庫本 19 帖	字配，仮名漢字が一致する	cosine 距離 0.05
池田本/明融臨模本「桐壺」「花宴」2 帖	共通の親本を持つと考えられる	cosine 距離 0.10
尾州家本/高松宮家本 41 帖	同系統内で本文異同が少ない	cosine 距離 0.25

なお，今後，調査対象となる写本数の増加によって，推定される分布が更新され，その結果，指標の値は変更されうる．このため，写本関係の知見に基づく距離や指標の再計算もありうるが，尺度設計の基本的な方法に変更はない．

5.　室町時代源氏物語写本の調査と考察

4 章で提案した尺度を用いて，室町時代を中心とした源氏物語写本のうち，最初の 18 巻を対象に，本文表記の類似性を持つ写本を探索する．前稿 [11] と尺度を変更し，調査対象写本を増やしたので，ここでは前稿で調査済みの写本も含めて再調査している．

5.1　調査方法

各巻ごとに調査対象写本の写本間距離を計算し，提案した尺度をあてはめ，写本間距離が近い写本を探索する．探索の際には階層的クラスタ分析を用いる．群構成法には最遠距離法を用いた．

階層的クラスタ分析を用いた分類結果の例として，**図 3** に示す．ここでは，提案した尺度にあてはまる写本数が多い第 7 帖「紅葉賀」を例にあげた．図中では尺度の指標を示す距離で点線を引き，それぞれの距離に表 4 の定性的な記述を省略して記している．この「紅葉賀」の分類結果は，「同系統内で本文異同が少ない」と考えられる写本として，2 つの群が分類されたことを示している．1 つは中央に位置する大正大学本と日大本，書陵部蔵三条西家本，保坂本からなる群であり，もう 1 つは下部に位置する大島本と国文研蔵正徹本からなる群である．

5.2　調査結果

調査対象とした第 1 帖「桐壺」から第 18 帖「松風」までの階層的クラスタ分析の結果を**表 5** にまとめた．表では，図 3 と同様に，尺度の指標を示す定性的な記述を省略して記している．

この結果を表でまとめるために，表中の写本名は，基本的には 2 文字の略称で表す．写本名の文字数が多い場合は，写本を区別できる略称を用いた．書陵部蔵正徹本は「書正」，国文研蔵正徹本は「国正」，書陵部蔵三条西家本

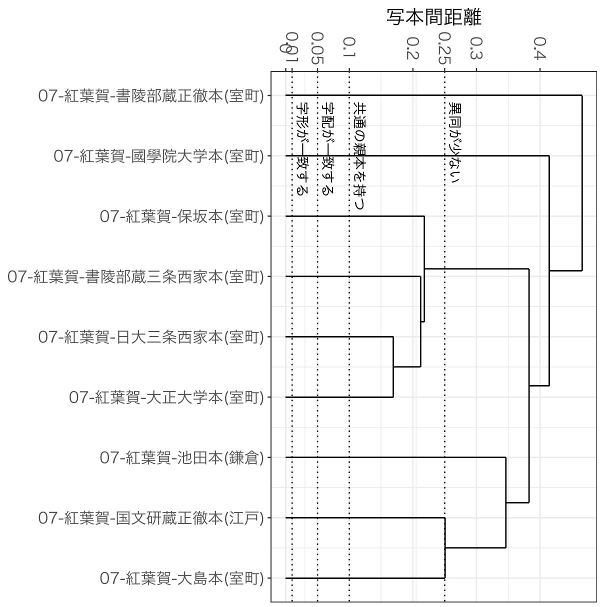

図 3　「紅葉賀」の分類結果

Fig. 3　Hierarchical cluster analysis of all surveyed manuscripts for the chapter *Momiji no ga*.

表 5　尺度を用いた写本の分類結果

Table 5　Results of determining manuscript groupings using the measure.

尺度の記述	巻	写本群
字形が一致する		
字配が一致する		
共通の親本を持つ	08	(大正, 保坂)
異同が少ない	01	(書三, 保坂) (國學院, 国正)
	02	(書三, 日大) (大島, 国正)
	03	(大正, 書正)
	05	((池田, 日大), 国正)
	06	((書三, 日大), 池田) ((大島, 国正), 大正)
	07	(((大正, 日大), 書三), 保坂) (大島, 国正)
	10	(大島, 国正)
	11	(大島, 國學院)
	12	(保坂, 日大) (大島, 国正)
	13	(國學院, 国正) (池田, 日大)
	14	(書三, 大正)
	15	(大島, 國學院)
	16	((大正, 大島), 書三)
	17	(((大正, 大島), 書三), 國學院) (池田, 日大)
	18	((書三, 大正), 大島)

は「書三」，國學院大学本は「國學院」と略している．表では，構成された群を括弧を用いてまとめている．たとえば，図 3 の「紅葉賀」の分類結果は，「(((大正, 日大), 書三), 保坂) (大島, 国正)」と記す．括弧の対応により群を表し，括弧の入れ子が深いほど相対的に写本間距離が近いことを表している．

最初に，表 5 の全体の傾向について概観する．表からは，第 8 帖「花宴」の大正大学本と保坂本だけが「共通の親本を持つ」写本間距離に存在し，それ以外は「同系統内で本文異同が少ない」写本間距離に写本群が存在すること

© 2022 Information Processing Society of Japan

が明らかになった．この結果からは，同時代または近い時代の写本を調査対象としても，表記が一致する写本は少ないこと，特に，「字形が一致する」写本間距離や「字配や仮名と漢字の使い分けが一致する」写本間距離に存在する写本は少ないことが予想できる．

各巻ごとの傾向としては，写本間距離が近い写本群の数や群を構成する写本が巻ごとに異なることが明らかになった．今回の調査対象は第 1 帖から第 18 帖としたが，第 4 帖「夕顔」と第 9 帖「葵」には，表記の点からは，書承関係を持つ可能性がある写本群が存在しない．

出現する一揃いの写本名に着目すると，大島本と近い写本が計 10 巻に存在すること，書陵部蔵三条西家本と近い写本，大正大学本と近い写本，国文研蔵正徹本と近い写本がそれぞれ計 8 巻に存在するといった傾向がある．特に，巻間で共通する写本の組合せとしては，書写年代が近いと考えられる書陵部蔵三条西家本と大正大学本が多く，全体を通して計 5 組存在することが明らかになった．また，組合せが偏在する例として，第 1 帖「桐壺」から第 12 帖「須磨」までは大島本と国文研蔵正徹本の組合せが多く，計 5 組存在すること，第 16 帖「関屋」以降で大島本，書陵部蔵三条西家本，大正大学本の 3 写本の組合せが計 3 組存在すること，といった傾向があることが明らかになった．

5.3　考察

本調査結果から，本文表記の調査で指摘できる多くの写本は，同系統内で本文異同が少ない関係にある可能性が高いと予想できる．加えて，本方法を用いて，写本間距離が近い巻が偏在すること，異同が少ない写本間関係を示す特定の写本の組合せが存在することも明らかになった．

ここで指摘した異同が少ない写本は，写本に共通の祖本が存在する可能性，または，転写過程を共有する可能性を示し，より詳細な調査の対象となりうる．また，巻ごとの偏在や特定の写本の組合せの存在は，一揃いの写本の中で，写本作成の際に取り合わせられた親本選択の状況を示している可能性があり，複数の巻間にわたる調査が必要であることを示している．今回の調査では室町時代の写本を対象として調査したことから，同時代の写本の流布状況を示している可能性もあり，今回未調査の後続の巻においても，この傾向の存在を確認する必要がある．

このように，本文に関する書誌情報が少ない写本に対して，本方法を用いることで，書承関係が存在する可能性を持つ写本を示すことができ，写本の探索の効率化が可能となる．同系統内で異同が少ない写本の発見には，大量に存在する表記の異同の比較が必要であり，計量的な方法を用いた調査が有効であろう．また，複数の巻を調査対象とする場合や，比較対象となる調査対象写本の組合せが増加した場合であっても，同様の方法を用いて統一的かつ網羅的な調査が可能であり，調査の大規模化に対応できる．

本方法の欠点は，仮名と漢字の使い分け，送り仮名，仮名遣い，音便といった膨大な表記の相違によって，写本を比較する際に重要視される脱落や誤写の共有といった，少数しか発生しない特徴的な相違が指摘できないことである．また，特定の漢字の出現といった特徴に着目した分類方法が，調査目的に合致した分類結果を得ることもある．

6.　文献学の知見との照応

表記から分類した結果を用いて考えると，大島本と書陵部蔵三条西家本，大正大学本は，複数の巻にわたって互いに類似した表記を持つ可能性があると指摘できる．これは青表紙（定家）本系統内の写本群の分類の細分化だけでなく，今回提案した尺度を用いることで，これらに書承関係がある可能性と，その親疎も指摘できる．

これまでに，漢字の宛て方の分類が室町時代に流布していた大島本と肖柏本，書陵部蔵正徹本，書陵部蔵三条西家本，大正大学本の間で類似性を持つとの指摘がある [1], [6]．本調査では肖柏本を対象としていないため不明であるが，第 16 帖「関屋」，第 17 帖「絵合」，第 18 帖「松風」に関しては，指摘のある写本の一部が群を構成していることから，計量的な手法を用いた本調査によって，表記の点からもその指摘を支持し，蓋然性を高めた結果となった．

その一方で，今回の調査では，書陵部蔵正徹本が，大島本と書陵部蔵三条西家本，大正大学本の 3 写本と類似性を示す結果とはならなかった．表記が似ている本文を持つ写本であることは，結果として，本文が似た写本である可能性が高い．しかし，特定の漢字の使用や本文異同が似ることと，本調査が対象としている送り仮名や仮名遣い，音便といった表記が似ることは異なる．本調査の表記の類似性を示す結果は，これら 3 写本が書承関係において，より密接な関係にある可能性を示し，その詳細な調査の必要があるといえる．

さらに，本調査結果からは，奥書や書写年代といった書誌情報が限られている写本に対して，これまで指摘されなかった書承関係に関する仮説も提案することができる．その例として，国文研蔵正徹本をあげる．国文研蔵正徹本と異同が少ない写本が存在する巻は第 1 帖から第 13 帖までに偏っている．これらは奥書がある巻と重なることが多い．また，表 2 より，第 11 帖「花散里」を除き，漢字の出現率が低い巻と奥書の有無の傾向が一致することから，国文研蔵正徹本は第 13 帖「明石」以前と第 14 帖「澪標」以後で写本の性質が異なる可能性がある．これは，同じ一揃いの写本の中で異なる書承関係を持つ写本が存在する可能性を示しているといえ，正徹の書状にある，焼失した写本を補写して再建する内容 [26] の傍証の可能性もある．

このうち，大島本と国文研蔵正徹本の第 2 帖「帚木」，第 6 帖「末摘花」，第 7 帖「紅葉賀」，第 10 帖「賢木」，第 12 帖「須磨」は，同系統で異同が少ない写本間関係にあること

© 2022 Information Processing Society of Japan

を示している．表記の類似性から，転写過程の共有といった書承関係を持つ可能性があり，大島本と国文研蔵正徹本に共通の祖本がある可能性を指摘できる．国文研蔵正徹本「箒木」と「末摘花」の奥書には，冷泉為相書写の写本と校合したとの記述 [26] があるので，この奥書の内容の検討と，これら 2 写本との関係の考察も可能であろう．

7. まとめ

本稿では，文献学から関係が指摘されている源氏物語写本の本文表記を用いて，写本間距離の尺度を設計し，その尺度に基づき，写本が多く現存する室町時代の源氏物語写本の本文の類似性を持つ写本を探索した．計量的な手法を用いた本文表記の類似性の調査の結果，初めて明らかになった写本間の書承関係を指摘し，写本間関係に新しい仮説を提案したといえる．さらに，書承関係が指摘できる写本が奥書といった書誌情報を持つ場合には，その書誌情報との関係を考察できる可能性があることも示した．

本方法の利点は，計量的な調査に基づいて定性的な写本間の書承関係の推定が可能であり，大量の写本を対象とした調査が容易であることである．本方法の欠点は，表記が一致する写本間関係という相対的に少数の写本の考察に限定され，従来の本文異同の比較に基づいた確認も必要であることである．これらのことから，本方法を用いた調査は，大量の写本を調査対象として検討することに適しているが，該当する分野の専門家による，従来の方法を用いた精密な調査の必要性も同時に示しているといえる．

今回未調査である後続の巻を含め，公開されている写本に限っても未調査の写本は多く，本方法を用いて書承関係が発見できる写本が存在する可能性がある．今後も同様の調査を続け，写本間関係に対する指摘の確認と仮説の提案を試みたいと考えている．

謝辞 本研究は JSPS 科研費 JP19K00349 の支援により実施された．

参考文献

[1] 中古文学会関西部会（編），加藤洋介：大島本源氏物語の本文成立事情—若菜下巻の場合，大島本源氏物語の再検討，和泉書院，pp.135–165 (2009).

[2] 助川幸逸郎，立石和弘，土方洋一，松岡智之（編），中川照将：『源氏物語大成』校異篇の歴史的意義，新時代の源氏学 7 複線化する源氏物語，竹林舎，pp.121–143 (2015).

[3] 大内英範：源氏物語　鎌倉期本文の研究，おうふう (2010).

[4] 中古文学会関西部会（編），新美哲彦：大島本『源氏物語』と東海大学蔵伝明融筆『源氏物語』の比較から見えるもの，源氏物語　本文研究の可能性，和泉書院，pp.89–129 (2020).

[5] 加藤洋介：中京大学図書館蔵 河内本源氏物語について，中京大学図書館学紀要，Vol.18, No.1, pp.10–19 (1997).

[6] 加藤洋介：本文系統の認定をめぐる諸問題—書陵部蔵三条西家本源氏物語について，詞林，Vol.52, No.1, pp.12–24 (2012).

[7] 田村　隆：「涙」の表記情報，国語国文，Vol.81, No.2, pp.16–30 (2012).

[8] 伊藤鉄也：『源氏物語』の異本を読む「鈴虫の場合」，臨川書店 (2001).

[9] 新美哲彦：源氏物語の受容と生成，武蔵野書院 (2008).

[10] 越野優子：国冬本源氏物語論，武蔵野書院 (2016).

[11] 齊藤鉄也：Ngram を用いた表記から見た紅梅文庫旧蔵本『源氏物語』の調査 (1)，実践女子大学文芸資料研究所年報，Vol.39, No.1, pp.91–119 (2020).

[12] 古典籍覆製叢刊刊行会（編）：源氏物語—花ちるさと・かしは木青表紙原本，前田育徳會尊經閣文庫，雄松堂書店 (1978).

[13] 東海大学桃園文庫影印刊行委員会（編）：源氏物語（明融本）I，II，東海大学出版会 (1990).

[14] 日本大学総合学術情報センター：日本大学蔵源氏物語 三条西家証本，八木書店 (1994，1995，1996).

[15] 京都大学附属図書館：源氏物語，京都大学附属図書館（オンライン），入手先 ⟨https://rmda.kulib.kyoto-u.ac.jp/item/rb00007030⟩（参照 2020-07-28）.

[16] 天理図書館附属天理図書館（編）：源氏物語池田本，八木書店 (2016，2017，2018).

[17] 名古屋市蓬左文庫（監修）：尾州家河内本源氏物語，八木書店 (2010，2011，2012，2013).

[18] 山岸徳平（解説）：高松宮御蔵河内本源氏物語，臨川書店 (1973，1974).

[19] 國學院大学図書館：源氏物語（寄合書），國學院大学図書館（オンライン），入手先 ⟨https://opac.kokugakuin.ac.jp/digital/menus/index09.html⟩（参照 2021-03-28）.

[20] 宮内庁書陵部：源氏物語/正徹本/初音首缺，宮内庁書陵部（オンライン），入手先 ⟨http://dbrec.nijl.ac.jp/KTG_B_100218545⟩（参照 2020-07-28）.

[21] 国文学研究資料館：源氏物語 国文研蔵正徹本，国文学研究資料館（オンライン），入手先 ⟨http://dbrec.nijl.ac.jp/KTG_B_200010454⟩（参照 2020-07-29）.

[22] 宮内庁書陵部：書陵部蔵三条西家本，宮内庁書陵部（オンライン），入手先 ⟨https://shoryobu.kunaicho.go.jp/Toshoryo/Detail/1000629260000⟩（参照 2020-02-17）.

[23] 大正大学図書館・研究所：源氏物語写本，大正大学図書館・研究所（オンライン），入手先 ⟨https://tais.ac.jp/library_labo/library/genji/⟩（参照 2019-12-21）.

[24] 伊井春樹（編）：保坂本源氏物語，おうふう (1995，1996，1997).

[25] （財）古代学協会古代学研究所（編）：大島本源氏物語，角川書店 (1996，1997).

[26] 菅原郁子：源氏物語の伝来と享受の研究，武蔵野書院 (2016).

齊藤 鉄也（正会員）

1971 年生．1996 年慶應義塾大学大学院政策・メディア研究科修士課程修了．2002 年同博士課程単位取得退学．2018 年淑徳大学経営学部教授．古典籍の計量文献学の研究に従事．日本ソフトウェア科学会，中古文学会各会員．

© 2022 Information Processing Society of Japan