

仮名字母の出現頻度率に基づく藤原定家書写資料の年代推定

齊藤 鉄也

淑徳大学経営学部

本稿では、藤原定家が書写した写本について、変体仮名の字母の出現頻度率を計算し、それらに対してクラスター分析と主成分分析を行った結果を報告する。その結果、仮名字母の出現頻度率を用いた場合、天福年間に近い時代に書写された写本は互いに似ていることが明らかになった。また、天福年間に近い時代に書写された写本とそれ以外の写本との間で、出現頻度率に差がある字母についても調査し、差があると考えられる字母を指摘した。

Dating Fujiwara no Teika Autograph Manuscripts by Relative Frequency Rates of *Kana*-character Variants

Tetsuya Saito

College of Business Administration, Shukutoku University

I calculated relative frequency rates for *kana*-character variants in several manuscripts by Fujiwara no Teika, and performed hierarchical cluster analysis and principal component analysis. The results showed that in manuscripts from around the Tenpuku Era, these relative frequency rates were similar to one another. I also compared manuscripts from this period with those from other years, to see if between the two groups there were any *kana* variants whose relative frequency rates showed a difference, finding several for which this was the case.

1. はじめに

これまでに、多くの古典籍の影印本が出版されている。加えて、最近では、資料を所蔵する図書館より、インターネット上の古典籍の画像データの公開が進みつつあり、その本文を参照することが容易になっている[注1]。

本研究では、公開されている古典籍の写本を対象とし、その特徴を表す情報として、現在使われている平仮名とは異なる字体の仮名である変体仮名に着目し、統計的手法を用いて写本を調査した。具体的には、古典籍の写本に使われている変体仮名の元になった漢字である字母の出現頻度率を用いた、藤原定家(1162-1241)が書写したとされる写本の年代推定の可能性を調査した。

調査対象には、書写年代が明らかな写本、または、推定書写年代が指摘されている写本を用いた。調査方法には、階層的クラスター分析と主成分分析を用いて、それらの写本の分類を試みた。

その結果、定家が書写した写本のうち、天福年間(1233-1234)に近い時期に書写したとされる写本が分類できることと、それらの写本で特徴的な使い方をする字母が存在することが明らかになった。

2. 本研究の目的と関連研究

古典籍の写本は、様々な変体仮名を用いて書写されている。このことから、ある人物の書写した写本には、何らかの特徴的な変体仮名の使い方(書き癖)があると仮定すると、同一人物が書写し

た写本であれば、それらの写本の変体仮名の出現傾向は似ていることが予想できる。また、同一人物であっても書写時期によって変体仮名の使い方が異なれば、その使い方の相違によって写本の書写年代の推定ができる可能性がある。

本研究の目的は、古典籍の写本に特有の変体仮名の出現の多様性を利用し、写本に用いられている変体仮名の出現傾向を、文章の特徴を表す情報として用いて統計分析を行い、異なる時代や異なる作品の写本の分類を試みることである。具体的には、変体仮名の字母の出現頻度率に基づく写本間の類似性の有無や、使い方が特徴的な仮名字母の存在に関して調査を試みることを目的としている。そのため、写本からの変体仮名のデータの収集を必要とする。写本の変体仮名を現代の仮名に翻刻して出版された本文は利用できない。

本論文では、調査の対象を定家が書写した写本とし、その中でも、推定書写年代が明らかな写本を中心を選択した。特徴情報として、本文の変体仮名の字母の出現頻度率に着目し、それらに対してクラスター分析と主成分分析を行い、写本の書写年代に対する分類と年代推定を試みた。

定家が書写した写本を調査対象とした理由は、多くの写本が利用できるからである。定家は生涯に渡って多くの古典籍を書写した。その後、それらの写本は長期間に渡り尊重されてきたことから、定家が書写した、または関わった写本は現在まで多く残されている。そのうち、幾つかの写本に関してはその書写年代も明らかである。これは多くの古典籍の書写者が不明であることや、

写本の書写時代も「鎌倉時代中期」といった相対的な時代と時期の指摘にとどまっていることと比較して、調査対象として非常に恵まれた条件にある。加えて、国文学の研究において、定家が残した写本に関して多くの研究が行われているために、それらの写本は影印本として出版され、定家が書写した写本は利用し易い利点がある。

近年の古典籍に対して統計分析を用いた事例としては、源氏物語の助動詞の計量分析[1]や第三部の成立順序に関する研究[2]がある。これらの研究は、『源氏物語大成』[3]の源氏物語本文のデータベースを用いて、単語や品詞の出現率を分析し、成立順序を論じている。似た方法を用いてt検定を行い、井原西鶴と弟子の北条団水の文体比較を行った研究[4]もある。

また、源氏物語の「空蝉」と「夕霧」の本文に対して、系統学的手法を用いて分析した研究[5]がある。この研究では、源氏物語の諸写本の異同を数値に置き換えて、写本間の距離を測り、それらを可視化し分類している。

現代の仮名に活字化された本文ではなく、古典籍そのものを対象とし、仮名に対する計量的な分析を行なった例としては、以下の研究がある。

藤原定家の書写した資料に関して分類と書写年代の推定を行なった研究[6]がある。この研究では、定家監督書写本を中心とした写本を対象とし、出現する仮名の字体(字形)の頻度率を用いて、写本間の関係の考察を行なっている。

また、世阿弥の自筆資料を対象とした研究[7]や上田秋成自筆資料を対象とした研究[8]もある。これらの研究では、それぞれの資料に対して、仮名の字母の用字法を分析し、単語中の字母の位置とその出現の特徴を論じている。

源氏物語の「花散里」の写本の漢字と仮名字母に着目し、その本文における漢字の出現率や仮名字母の種類数を用いて、その写本の特徴や書写された時代の傾向を指摘している研究[9]や、似た方法を用いて源氏物語「篝火」写本や版本に着目した研究[10]もある。

3. 調査対象と生成したデータ

ここでは、本調査の対象とした写本と、対象とするデータ、その処理方法について述べる。

表1：調査対象とした藤原定家が書写した写本

書名	(推定)書写年代	対象本文範囲	仮名文字数	仮名字母数
古今集-伊達本	1223/7以降 1226/4以前	巻三夏まで	10642	91
古今集-嘉禄二年四月本	1226/4/9	巻三夏まで	10108	90
拾遺愚草	1232/7以降 1233/10以前か	38丁裏まで	8041	98
拾遺集-天福元年本	1233/8	巻第五賀まで	9285	104
後撰集-天福二年本	1234/3/2	巻第四夏まで	8144	95
更級日記	1230/6/17以降	全文	23977	105
土左日記	1235/5/13	全文	10771	111
源氏物語-柏木-定家筆		全文(11丁裏5行まで)	3247	88
源氏物語-柏木-他筆		全文(11丁裏6行から)	13459	83
伊勢物語-学習院大学蔵本	1234/1/20(1507/5頃か)	全文	20905	113
伊勢物語-御所本	1234/1/20(1547/1/下旬)	全文	20894	112

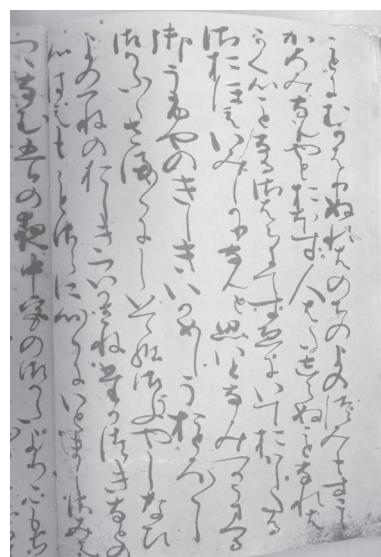


図1：尊経閣文庫蔵源氏物語「柏木」の11丁裏

図1に写本の例として、調査対象とした源氏物語の「柏木」を挙げる。図1は、「柏木」の定家筆部と他筆部に書写者が変更された丁とその箇所を示している。5行目までが定家筆部とされ、6行目からが他筆部とされている。

3.2 写本からの採字方法と生成したデータ

写本から変体仮名の字母を収集し、分析対象とするデータを生成するため、対象とする本文と対象とする仮名字母を以下の様に定めた。

対象とした本文は、原則として写本の本行本文としている。そのため、本文に修正後の文字が右傍に記載され、本行の文字が残されている「見せ消ち」に関しては、修正前の文字を採用している。重ね書きされている場合は、修正前と後の両方の文字を採用している。異文注記といった傍記は対象としない。本文へ挿入されている文字も採用していない。また、踊り字(繰り返し文字)は対象としていない。

この本文を、作成した字母表に基づいて分類する。字母表には『くずし字辞典』[11]の「かな編」と、『常用源氏物語要覧』[12]から字母を選択している。字母表の一部を表2に抜粋した。表2は「あ」行の仮名の字母を表している。

字母の中には、「こ」や「ね」と読む「子」といった複数の読みを持つ字母が存在する。これらを区別するために、字母表の記号を用いて区別している。表2では、「あ」の字母である「安」は“a1”という記号を持つことを表している。

表2：字母表の例(一部抜粋)

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

字母表には計369字母が存在する。漢字と仮名の区別もこの字母表に基づいて行う。熟語や複数の仮名(字音)で読む漢字、作成した字母表に存在しない一音の漢字は、漢字として分類する。読みが一音で、かつ、この字母表に掲載されている漢字は仮名と見做し、漢字と仮名を区別する。この分類に基づいて仮名を集計した結果が、表1の対象本文範囲の仮名文字数と仮名字母数である。

集計した字母は四十八種類の仮名ごとに出現頻度率を計算する。これは、本研究では、異なる作品や、同一の作品であっても異なる巻を調査対象としているため、単純な出現頻度数では、その出現傾向の比較ができないからである。例として、

表3に仮名「あ」に関しての表1の一部の写本の出現頻度率を挙げる。表3中の「柏木」とは源氏物語の「柏木」の巻を示している。

表3：出現頻度率の例(一部抜粋)

	安	阿	愛	亞	惡
更級日記	0.99	0.01	0.00	0.00	0.00
土佐日記	0.99	0.01	0.00	0.00	0.00
柏木-定家筆	0.98	0.02	0.00	0.00	0.00
柏木-他筆	0.99	0.01	0.00	0.00	0.00

表3は、仮名「あ」の字母としては「安」が用いられることが極めて多く、「阿」が用いられることは少ないことを示している。

4. 階層的クラスター分析と考察

表1の写本に対して、3.2で述べた仮名ごとに計算した字母の出現頻度率を変数として用い、階層的クラスター分析を行った。写本間の距離として、ユークリッド距離を用いた。クラスター間の距離の計算方法には、ウォード法を用いた。

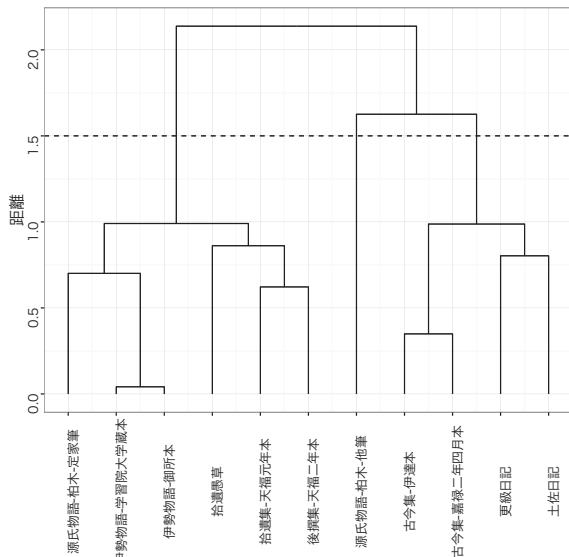


図2：定家書写資料間のクラスター樹形図

点線の位置で樹形図を切断すると、図2からは、主に三つのクラスターを構成することが明らかになった。全体を概観すると、定家筆の写本と他筆の写本に分類されていること、さらに、対象とした定家書写の資料は大きくふたつに分類できること、が考えられる。それらは、左側に位置する6冊の写本が所属するクラスターと、中央の定家が書写していない源氏物語の「柏木」の他筆部だけが所属するクラスター、右側の古今集と日記の計4冊の写本が所属するクラスターである。

次に、外的な根拠として表1の推定書写年代を用いて、クラスター分析の結果を考察する。

左側のクラスターには、拾遺集と後撰集、伊勢物語が分類されている。これらの資料は、表1より、天福年間(1233-1234)に書写されていることが明らかである。加えて、拾遺愚草も天福年間に近い時期に書写されたかと推定されている。このことから、天福年間に近い時期に書写されたと推定される写本は共に似た仮名字母の出現頻度率を持つことが推定できる。

右側のクラスターには、1226 年に書写された古今集と、1235 年に書写された土左日記が分類されている。このことから、これらの異なる年代に書写された資料が、仮名字母の出現頻度率の点からは、共に似ていることを示している。

今回は書写年代が不明の写本は源氏物語の「柏木」の定家筆部しか用いていない。この写本は、左側のクラスターに所属していることから、天福年間に近い時期に書写された可能性がある、と考えることができる。

また、今回の調査では、他筆の写本がひとつしかなく、この結果だけで判断することは危険ではあるが、よく似た字体を持つ定家筆と他筆の写本を分類するために、特徴情報として仮名の字母に着目し、古典籍の書写者の推定ができる可能性がある。この点に関しては、対象とする資料を増やして、今後も調査を続けたい。

5. 主成分分析と考察

4. のクラスター分析の分類結果を異なる手法を用いた結果と比較するため、主成分分析を行った。この分析では、調査対象から源氏物語の「柏木」の他筆部を除き、定家が書写した資料だけを用いている。

分析の前に、字母表にある字母のうち、出現しない字母や、多くの写本で出現率が 100%に近い字母を分析対象から除外した。また、表 3 の「あ」の字母の「安」と「阿」の様に互いに相関が高いふたつの字母はどちらか一方を削除した。その結果、67 字母のデータを変数として用いた。

次に、この 67 字母を対象に、相関行列を用いた主成分分析を行い、次元の縮約をした。表 4 に、その固有値と寄与率、累積寄与率をまとめた。表 4 より、第 6 主成分まで、累積寄与率が 80% を超えることが明らかになった。

表 4 : 固有値と寄与率, 累積寄与率

	PC1	PC2	PC3	PC4	PC5	PC6
固有値	14.66	12.57	9.99	7.81	7.00	6.11
寄与率(%)	21.87	18.76	14.91	11.65	10.40	9.11
累積寄与率	21.87	40.64	55.54	67.19	77.59	86.71

次に、全体を概観するために、主成分分析の結

果と、主成分負荷量を図3の散布図に表した。図3では、横軸を第1主成分の主成分得点、縦軸を第2主成分の主成分得点とした。

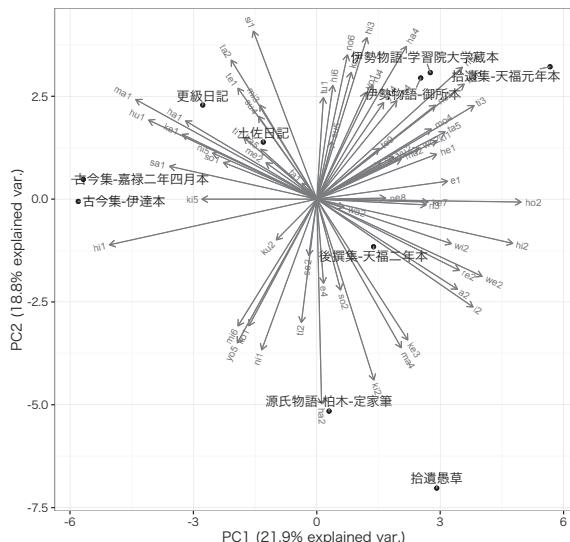


図3：主成分分析(相関行列)の散布図

この図3においては、第1主成分に関して正の値と負の値に分けられる位置に、各写本が存在している。この第1主成分の状況を把握するために、第1主成分得点を図4の棒グラフで表した。

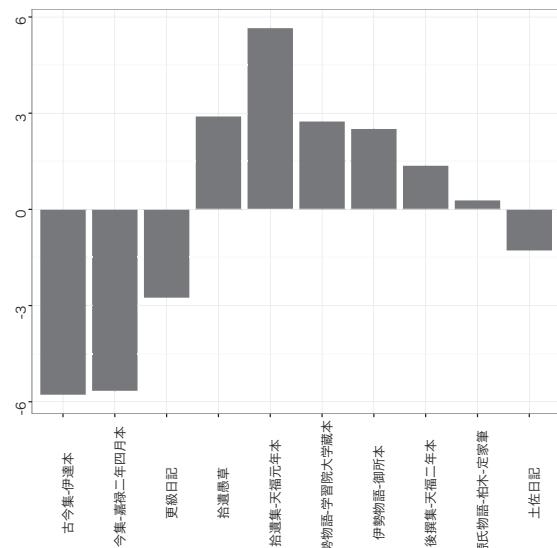


図4：第1主成分得点の棒グラフ

図4の横軸には、推定書写年代が古い写本の順に左から右に配置している。但し、表1より、更級日記の推定書写年代は、上限である1230年以後としかわからぬいため、便宜的に1230年として扱っているが、更級日記の解説[注2f]によれば、定家晩年の書写ではないかとの指摘がある。また、書写年代が不明の源氏物語「柏木」定家筆

部は、図3において天福年間に書写された写本と近い位置に付置されているため、後撰集の次に配置している。

主成分分析の結果からも、4. のクラスター分析の結果と同様に、天福年間に近い時期に書写されたと推定される写本は、第1主成分得点が正の値を示し、分類ができることが明らかになった。尚、第1主成分と同様に、表4に記載している第2主成分から第6主成分までグラフを用いて調査をしたが、傾向を見つけることはできなかった。

また、図3の矢印には、表2で述べた字母表の記号が記されている。この矢印の大きさと向きからは、第1主成分が大きな値を示す拾遺集では、字母「本」を表す“ho2”や、字母「飛」を表す“hi2”といった字母の出現頻度率が大きな値を持っていること、反対に、第1主成分が小さな値を示すふたつの古今集では、字母「比」を表す“hi1”や、字母「末」を表す“ma1”的出現頻度率が大きな値を持っていることを読み取ることができる。

のことから、第1主成分が正の値を示す天福年間に近い時期に書写されたと推定される写本では、共通した字母の出現傾向の存在が考えられる。そこで、それぞれの仮名字母に関して調査を行い、字母の出現頻度率を用いて、年代を区別できる可能性を検討する。

6. 年代を区別する字母と考察

4. と 5. で明らかになった天福年間に近い時期に書写されたと推定される6冊の写本とそれ以外の4冊の写本に対して、字母の出現頻度率から見た特徴を明らかにすることを試みた。5. で分析対象とした67の字母に関して、6冊から成る群とそれ以外の4冊から成る群に対して、ウェルチのt検定を用いて、「ふたつの群の字母の出現頻度率に差がない」とする仮説検定を、有意水準5%で行った。

5%有意で帰無仮説が棄却された字母は計12存在した。それらを、五十音順に、仮名と字母、表2と図3で用いた記号と共に表5にまとめた。表5では、p値と6冊から成る群と4冊から成る群の出現頻度率の平均も示している。

これらの字母は、ふたつの群の間で出現率に差があると言え、その使い方に特徴があると考えられる。例えば、仮名「い」の字母「伊」の場合について述べると、天福年間に書写されたと推定される写本を中心とした6冊から成る群では、本文に出現した仮名「い」の字母のうち「伊」は5.8%出現することに対して、それ以外の4冊から成る群では1.5%しか出現しない。仮名「ひ」の場合については、6冊から成る群では、平均すると、字母「比」が59.3%、字母「飛」が22.1%出現し

ている。これに対し、4冊から成る群では、平均すると字母「比」が76.7%，字母「飛」が5.7%出現している。仮名「ひ」の場合、どちらの群も字母「比」が主に使われているが、6冊から成る群の方が、その出現率が相対的に低い。これに対し、字母「飛」については、6冊から成る群の方が、その出現率が相対的に高い。

表5：5%有意で出現頻度率に差がある字母

仮名	字母	記号	p値	6冊の平均	4冊の平均
い	伊	i2	0.0092	0.0582	0.0150
け	計	ke1	0.0494	0.0653	0.2755
け	遣	ke3	0.0098	0.1611	0.0757
さ	左	sa1	0.0081	0.5644	0.7675
は	波	ha1	0.0029	0.0614	0.1171
ひ	比	hi1	0.0030	0.5933	0.7669
ひ	飛	hi2	0.0072	0.2214	0.0572
ふ	不	hu1	0.0080	0.5893	0.7121
ほ	本	ho2	0.0185	0.5279	0.2631
ま	末	ma1	0.0042	0.1477	0.3602
り	里	ri3	0.0152	0.1178	0.0535
ゑ	衛	we2	0.0063	0.2333	0.0379

この調査結果から、天福年間に書写されたと推定される写本と、それ以外の写本で、出現頻度率に差がある字母が明らかになった。これらの字母に着目して調査を行うことで、天福年間に書写されたと推定される写本とそれ以外の写本を分類できる可能性がある。

本調査方法の利点としては、ある字母の出現頻度率の高低に着目して、年代を区別する調査の「手がかり」を指摘できる可能性がある。例えば、表5からは、4冊から成る群において「伊」や「衛」といった字母の出現率が低いので、これらの字母に着目して、調査することができる。また、ある字母の出現の有無だけに着目して、写本間の年代を区別することができるならば、その区別は容易である。しかし、今回の調査対象には、そのような字母は存在しなかった。

一方で、本調査方法の欠点としては、今回得た出現頻度率の知見を得るために、多くの文字数の調査を必要とすることである。ある字母の出現頻度率を計算するためには、その字母が示す仮名に対応する全ての字母を収集し、さらに、特定の字母に関して出現頻度率を計算する必要がある。特定の字母を対象とした調査だけでは、出現頻度率の差を知ることはできないため、探索的な調査方法として、本調査方法を用いることは容易ではない。

加えて、今後、さらに多くの写本を対象として調査を行うとすると、調査に必要なデータサイズの検討も必要である。表1の仮名文字数から明らかな様に、今回は10000文字程度のデータを中心

として、出現頻度率に基づいた調査を行なっている。源氏物語といった大規模な作品の場合を調査対象として想定すると、その全文を対象として調査することは容易ではない。対象とするデータサイズに関する知見があれば、効率的な調査できる可能性がある。

7. まとめと今後の課題

本研究では、古典籍の写本に用いられた変体仮名の字母を特徴量として着目し、その出現頻度率を用いて、藤原定家が書写した写本の書写年代の推定を行った。調査の結果、定家書写とされる写本の中には、仮名字母の出現頻度率の点から、互いに近い出現傾向を持つ写本があることが明らかになった。それらの写本は、天福年間に近い時期に書写されていることと、出現頻度率に差がある字母が存在することが明らかになった。

一方で、書写年代が明らかに 1226 年に書写された古今集と、1235 年に書写された土左日記は、クラスター分析においては同じクラスターに所属し、また、主成分分析においては、第一主成分得点が共に負の値を示している。このことから、今回の調査対象においては、仮名字母の出現頻度率の点から、これらの写本に関しては、書写年代に 10 年ほど相違があるにもかかわらず、その差は明らかではない。

今回対象としていない書写年代が明らかでない定家筆とされる写本も多く、加えて、定家書写の写本を忠実に書写したとされる写本も存在する。今後の課題は、調査対象とする写本を増やし、今回と同様の方法を用いて調査を継続することである。

注

- 1) 古典籍の影印本は多数出版されているため、例として、近年出版された尾州家河内本源氏物語 (<https://catalogue.books-yagi.co.jp/books/view/471>) を挙げる。同様にインターネット上に公開されている写本の画像データとして、国立国会図書館デジタルコレクションの定家本土佐日記 (<http://dl.ndl.go.jp/info:ndljp/pid/1191108>) を挙げる。
- 2) ここで選択した写本の出典は次の通りである。
 - a) 『古今集・伊達本』、久曾神昇(解題)：伊達本古今和歌集藤原定家筆、笠間書院、(1995).
 - b) 『古今集・嘉禄二年四月本』、財団法人冷泉家時雨亭文庫(編)、片桐洋一(解題)：古今和歌集嘉禄二年本貞応二年本、冷泉家時雨亭叢書 2、(1994).
 - c) 『拾遺愚草』、財団法人冷泉家時雨亭文庫(編)、久保田淳(解題)：拾遺愚草上中、冷泉家時雨亭叢書 8、(1993).

- d) 『拾遺集・天福元年本』、久曾神昇(編)：藤原定家筆拾遺和歌集、汲古書院、(2005).
- e) 『後撰集・天福二年本』、財団法人冷泉家時雨亭文庫(編)、片桐洋一(解題)：後撰和歌集天福二年本、冷泉家時雨亭叢書 3、(2004).
- f) 『更級日記』、松尾聰(解説)：御物本更級日記、武蔵野書院、(1995).
- g) 『土左日記』、育徳財団(編)：定家本土佐日記、尊經閣叢刊、(1928).
- h) 『源氏物語・柏木』、太田晶二郎(解題)：源氏物語・花ちるさと・かしは木 青表紙原本、雄松堂書店、(1979).
- i) 『伊勢物語・学習院大学蔵本』、鈴木知太郎(校注)：天福本伊勢物語、武蔵野書院、(2000,1963).
- j) 『伊勢物語・御所本』、鈴木知太郎(編)：御所本伊勢物語、笠間書院、(1971).

参考文献

- 1) 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、Vol.40, No.3, pp.774-782 (1999).
- 2) 土山玄、村上征勝：『源氏物語』第三部の成立に関する計量的な考察、人文科学とコンピュータシンポジウム論文集、Vol.2014, No.3, pp.213-220 (2014).
- 3) 池田亀鑑(編著)：源氏物語大成 索引編、中央公論社、(1985).
- 4) 上阪彩香、村上征勝：西鶴遺稿集の著者に関する統計分析、人文科学とコンピュータシンポジウム論文集、Vol.2014, No.3, pp.113-118 (2014).
- 5) 新美哲彦：源氏物語の受容と生成、pp.33-62 武蔵野書院、(2008).
- 6) 井浦美幸：中近世における平仮名の字体－藤原定家の平仮名－、東京女子大学『日本文學』Vol.37, pp.48-65 (1971).
- 7) 谷川淳子：世阿弥自筆『花伝第六花修』における用字法について、東京女子大学紀要論集、Vol.58, No.2, pp.51-94 (2008).
- 8) 木越治：上田秋成自筆本「春雨物語」における仮名字母の用法について、金沢大学教養部論集 人文科学篇、Vol.26, No.2, pp.244-168 (1989).
- 9) 斎藤達哉：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, pp.47-63 (2015).
- 10) 沼尻利通：『篝火』卷の表記と字母(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, pp.123-146 (2015).
- 11) 児玉幸多(編)：くずし字用例辞典 普及版、東京堂出版、(1993).
- 12) 中野幸一(編)：新装版常用源氏物語要覧、武蔵野書院、(2012).
- 13) 村上征勝：真贋の科学－計量文献学入門、朝倉書店、(1994).
- 14) 金明哲：テキストデータの統計科学入門、岩波書店、(2011).