

## 中古における接続表現の統計的分析：指示詞を構成要素とするものを中心に

著者	小林 雄一郎, 岡? 友子
雑誌名	国立国語研究所論集
号	13
ページ	65-77
発行年	2017-07
URL	<a href="http://doi.org/10.15084/00001372">http://doi.org/10.15084/00001372</a>

## 中古における接続表現の統計的分析 ——指示詞を構成要素とするものを中心に——

小林雄一郎<sup>a</sup> 岡崎友子<sup>b</sup>

<sup>ab</sup>東洋大学／国立国語研究所 共同研究員

### 要旨

本研究の目的は、中古資料における接続表現の使用の違いを明らかにすることである。具体的には、「日本語歴史コーパス（平安時代編）」と統計手法を活用することで、時代、ジャンル、書き手等の要因による接続表現の頻度の変異を分析した。その結果、(a) 紀貫之の筆による『土左日記』と『古今和歌集』（仮名序）の類似性、(b) サテの使用による歌物語の類似性、(c) カカリ系とサテ系の使用法に対する執筆年代の影響、等が見られた\*。

キーワード：中古（平安時代）資料、接続表現、日本語歴史コーパス、多変量解析

### 1. はじめに

近年、国立国語研究所による「日本語歴史コーパス（平安時代編）」（以降「歴史コーパス」と呼ぶ）の開発が進み、2016年現在、中古（平安時代）の主要な16作品が公開されている。この歴史コーパスを利用することにより、従来は不可能であった調査や分析が可能となった。

中古の言語現象については、これまでも多くの精緻な歴史的研究の成果がある。しかし、それらの歴史的研究のいわば土台となる中古の資料群に関する問題（時代、ジャンル、書き手の位相等）に関して、どれほど注意が払われてきたであろうか。もちろん、残されている資料が少なく偏りもあり、また索引等を用いた従来のやり方で中古の主要作品全体を俯瞰するのは、骨が折れる作業でもあった。そのような理由から、いくつかの示唆的な研究（e.g., 山口1984<sup>1</sup>, 安部1996, 村田2001, 金水2006）を除き、本格的に言及するものは少なかったと考えられる。

それにしても、中古における和文の資料群は、『竹取物語』（10世紀前半?）や『土左日記』（935年頃）から『源氏物語』（1001年以降?）、そして『大鏡』（1086～1123年頃）までに約200年の時間幅があること、また日記・和歌・物語・歌物語・随筆といった多様なジャンルを含むこと、書き手も男性・女性という位相も含むこと等、これらを同時代の均質なものとして扱うには問題があると思われる<sup>2</sup>。

\* 本論は国立国語研究所機関拠点型基幹研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」（プロジェクトリーダー：小木曾智信）の研究成果である。

<sup>1</sup> 山口（1984）でも、接続詞等の使用から中古の資料の文体の特徴について考察を行っているが、サンプリング調査（各作品から、40か所を等間隔抽出によって選び、次に1か所につき連続した文と文とのつなぎ目5個を調査している）であり、全体調査ではない。

<sup>2</sup> 例えば、小林・小木曾（2013）は物語文学と日記文学で大きく異なることを明らかにしている。

そこで、本論では歴史コーパスに見られる中古の接続表現（従来は「接続詞」と呼ばれるもの。本論では「接続表現」とする）を分析することにより、中古資料（14 作品<sup>3</sup>）の性質について深く考えることを目的とする。本論の目的は以下の 3 点の research questions (RQ) に答えることである。

RQ1：中古作品における接続表現の使用頻度に作品間の差があるか？

RQ2：どの中古作品でどの接続表現が多く（あるいは少なく）用いられているのか？

RQ3：接続表現の使用傾向の類似した中古作品はどれとどれか？

なお、具体的な作業としては、歴史コーパスから抽出・解釈・検討を行った用例を、各種の統計手法を用いて分析する。さらに、統計解析から得られた結果を日本語学・日本文学的な観点から再解釈していく。

## 2. 接続表現について

これまでの研究では、接続詞の発達は中世以降とされている。そのため、中古における接続詞の研究はあまり多くない。例えば、歴史コーパスで品詞が接続詞とされているものは少なく、7 語のみである。しかし、中古でも、連語の域を超えて複合語化し、接続詞的な働きをしているものがある。

なお、京極・松井（1973: 91-93）は古代語の接続詞について、以下のようなグループ分けを行っている（一部省略した）。

### (1) 京極・松井（1973: 91-93）でのグループ分け

#### A：複合接続詞

##### (a) 指示語を構成要素とするもの

①「か」系（本論注：指示副詞「カク」＋動詞「有り」→「カカリ」）

○「かく」類　かくあるほどに　かくいふほどに　かくして　かくて等

○「かかり」類　かかりければ　かかるあひだに　かかるうちに等

○「かり」類　かれ　かるがゆゑに

②「こ」系（本論注：指示代名詞「これ・ここ」等）

○「こ」類　このあひだに　このゆゑに

○「ここ」類　ここに　ここをもちて

○「これ」類　これによりて　これをもって

③「さ」系（本論注：指示副詞「サ」＋動詞「有り」→「サリ」）

○「さ」類　さいふいふ　さいふとも　さこそいへ　さて　さては等

○「さり」類　さらずは　さらば　さりけるに　さりけれど　さりければ等

<sup>3</sup> 本論では、『日本語歴史コーパス』（バージョン 2015.3, 中納言バージョン 2.0.1）を使用した。分析対象とした 14 作品は稿末の「資料」に示した。

- ④「しか」系（本論注：指示副詞「シカ」+動詞「有り」→「シカリ」）  
 ○「しか」類 しかありとも しかあるに しかあるを しかあれば等  
 ○「しかり」類 しからば しかりといへども しかるに しかるあひだ等
- ⑤「そ」系（本論注：指示代名詞「それ」等）  
 ○「そ」類 そのゆゑに そのゆゑは そも そもそも そゑに  
 ○「それ」類 それ それに それにつけても それも それを

(b) その他の語を構成要素とするもの

- ①動詞系 あるいは あるは ならびに もって よりて  
 ②名詞系 ゆゑは ゆゑをもって  
 ③副詞系 ただし もしは

**B：転成接続詞**

（あて） および かつ すなわち はた また

**C：借用接続詞**

乃至<sup>ないし</sup>

京極・松井（1973: 93）は、各グループについて、以下のような文体的な傾向があると指摘している。（イ）A-(a)「か」系（「かり」類除く）・「さ」系は和文に用いられて訓読文に用いられない、（ロ）A-(a)「こ」系・「しか」系及び「か」系の「かり」類は、訓読文に用いられて和文に用いられない、（ハ）A 複合接続詞の (b)、及び B 転成接続詞・C 借用接続詞は、ほとんどが訓読文に用いられる、（ニ）A-(a)「そ」系は語により異なるが、訓読文に用いられる傾向が強い。

また、和文ではなく漢文訓読文に現れる接続詞について、築島（1963: 328-329）では、訓読系接続詞として「カルガユエニ・カレ・ココニ・ココロモテ・コノユエニ・コレニヨリテ・コレヲモテ・シカノミナラズ・シカウシテ・シカウシテノチニ・シカルニ・シカニハアラズハ・シカモ・シカルヲ・シカラバ（シカレバ）・シカリトイヘドモ・シカレドモ・ソレ・タダシ・モシソレ・ユエニ・ユエヲモテ」、また『源氏物語』にも見えるが用法や用例が限られているものとして「ソモソモ」があげられている。

そこで、本論では、上記の先行研究を参考として、以下のようにグループ分けし、分析を行う（岡崎 2015 を修正した）。なお、高橋（1985）によれば、接続詞の語構成において最も多いのは「指示語等（+助詞）」であり、構成語 60 語のうち 51 語までが指示語系のもので占められていると指摘しており、本論の調査対象も指示詞（特に「カク・サ」といった指示副詞）を含むものを中心に調査・分析を行うこととする。

(2) 本論でのグループ分け

- 「A\_サテ」及び「A\_その他」

歴史コーパスでは、「マタ、アルイハ、タダシ、ソエニ、スナハチ、サハレ、サテ」の 7 語が接続詞とされている。そこで、これらを A 類とし、さらに使用が多い「サテ」を「A\_サテ」、その他 6 語を「A\_その他」に分けて分析する。

## ○「B\_全体」

築島(1963), 京極・松井(1973)において訓読系接続詞とされるものをB類とする。なお, 京極・松井(1973)で指摘する訓読系の接続詞は, 「コノアヒダニ・コノユエニ・ココニ・ココヲモチテ・コレニヨリテ・コレヲモッテ」, 「シカアリトモ・シカアルニ・シカアルヲ・シカアレバ・シカシテ・シカノミナラズ」, 「シカラバ・シカリトイヘドモ・シカルニ・シカルアヒダ・シカルホドニ・シカルモノヲ・シカルヲ・シカレドモ・シカレバ」, 「カレ・カルガユエニ」, 「(A類:アルイハ)・アルハ・ナラビニ・モッテ・ヨリテ・ユエハ・ユエヲモッテ・(A類:タダシ)・モシハ」, 「オヨビ・カツ・(A類:スナワチ)・ハタ・(A類:マタ)」, 「(A類:ソエニ)」(築島(1963)の「ソレ」は対象外とする)である。これらについては一括して扱うため, 「B\_全体」と呼ぶ(A類とあるものは, 歴史コーパスで接続詞とされているためA類として扱う)。

## ○「C\_カカリ系」及び「C\_サリ系」

指示副詞「カク・サ」+助詞等(「カクテ・サテ」等), また「カク・サ」+動詞「有り」である「カカリ・サリ」を構成要素とする「カカレバ・サレバ」等をC類とし, さらに「カク」を含むものを「C\_カカリ系」, 「サ」を含むものを「C\_サリ系」と呼び, 分けて分析する。

大まかに言って, 接続表現は文頭に位置し, 直前の文と接続表現を含む文をつなぐものであると考えられる。そのため本論では, 文頭で用いられているもののみを分析対象としている(ただしA類は全例を対象とする)。

なお, 本論5.3節で行う統計処理に向けて, 各作品から得られた接続表現の使用頻度を文数で正規化する(表1)。文数による正規化については, 既に文数で接続詞の使用頻度を正規化している先行研究がある(山口1984, 福島2008)。また, 中古の和文は, 現代語等に比べて1文がかなり長いので, 語数ではなく文数で正規化した方が他の時代の使用傾向との比較がしやすくなると判断した(接続表現は基本的に文頭に位置するため, 理論上は, 文と文の境界の数だけ接続表現の出現可能な位置が存在する)。

表1 歴史コーパスにおける文数（本文種別）

\* 作品名は以降、括弧内の略称で示す

	歌	会話	手紙	地	詞章（古注）	計
竹取物語（竹取）	15	204	25	330	0	574
古今和歌集（古今）	1069	0	0	125	1334	2528
伊勢物語（伊勢）	235	20	3	637	0	895
土左日記（土左）	61	29	0	471	0	561
大和物語（大和）	297	146	3	972	0	1418
平中物語（平中）	154	92	0	403	0	649
落窪物語（落窪）	72	1658	0	1371	0	3101
枕草子（枕草）	39	787	0	2959	0	3785
和泉式部日記（和泉）	147	218	0	295	0	660
源氏物語（源氏）	794	5640	107	10072	0	16613
紫式部日記（紫式）	18	52	0	739	0	809
堤中納言物語（堤中）	51	293	0	419	0	763
更級日記（更級）	89	87	0	384	0	560
讃岐典侍日記（讃岐）	23	185	0	406	0	614
計	3064	9411	138	19583	1334	33530

### 3. 分析データ

表2は、本研究の分析データをまとめたものである。表中には、中古作品における接続表現の使用頻度だけでなく、各作品の文数で正規化した頻度が括弧内に記されている。

表2 中古作品における接続表現の使用頻度（括弧内は、文数で正規化した頻度）

	A_ サテ	A_ その他	B_ 全体	C_ カカリ系	C_ サリ系
竹取	2 (0.34)	9 (1.56)	2 (0.34)	5 (0.87)	8 (1.39)
古今	0 (0.00)	10 (0.39)	8 (0.31)	1 (0.03)	2 (0.07)
伊勢	10 (1.11)	5 (0.55)	0 (0.00)	2 (0.22)	18 (2.01)
土左	5 (0.89)	12 (2.13)	12 (2.13)	8 (1.42)	3 (0.53)
大和	34 (2.39)	21 (1.48)	3 (0.21)	7 (0.49)	20 (1.41)
平中	41 (6.31)	40 (6.16)	1 (0.15)	3 (0.46)	33 (5.08)
落窪	29 (0.93)	16 (0.51)	1 (0.03)	15 (0.48)	35 (1.12)
枕草	46 (1.21)	68 (1.79)	2 (0.05)	1 (0.02)	61 (1.61)
和泉	2 (0.30)	1 (0.15)	0 (0.00)	4 (0.60)	7 (1.06)
源氏	93 (0.55)	48 (0.28)	4 (0.02)	8 (0.04)	239 (1.43)
紫式	0 (0.00)	8 (0.98)	1 (0.12)	0 (0.00)	16 (1.97)
堤中	8 (1.04)	6 (0.78)	0 (0.00)	0 (0.00)	17 (2.22)
更級	1 (0.17)	2 (0.35)	0 (0.00)	0 (0.00)	3 (0.53)
讃岐	1 (0.16)	4 (0.65)	1 (0.16)	2 (0.32)	8 (1.30)

#### 4. 分析手法

本研究で接続表現の頻度解析に用いる統計手法は、以下のとおりである。まず、中古作品における接続表現の使用頻度に作品間の差があるかという問い (RQ1) に答えるために、カイ2乗検定 (Agresti 2007) を用いる。本研究のカイ2乗検定は、表2における使用頻度を対象とし、「全ての作品における接続表現の使用傾向が同じである」という仮説 (帰無仮説) を検証する。なお、帰無仮説を採択するか否かを判断する有意水準は5%とする。

カイ2乗検定の結果に有意差が見られた場合は、Habermaの残差分析 (Haberma 1973) を行い、どの中古作品でどの接続表現が多く (あるいは少なく) 用いられているか (RQ2) を特定する。カイ2乗検定で有意差が見られたとしても、「全ての作品における接続表現の使用傾向が同じである」という仮説を棄却するだけで、表中にどのような差が見られるのかは分からない。このような場合、残差分析を行うことで、特定の作品で有意に多く (あるいは少なく) 現れている接続表現を明らかにすることができる。残差分析の有意水準は、カイ2乗検定と同様に5%とする。

そして、対応分析と階層型クラスター分析を用いて、接続表現の使用傾向の類似した中古作品はどれとどれか (RQ3) 等を明らかにする。対応分析とは、頻度表に含まれる複雑な情報を2次元の散布図等で分かりやすく可視化するための手法である (Greenacre 2016)。また、階層型クラスター分析とは、個々の作品の (非) 類似度を「距離」として表現し、距離の近い作品同士をまとめてクラスター (グループ) を作っていく手法である (Moisl 2015)。これらの多変量解析手法を用いることで、接続表現の使用傾向に基づく中古作品のグループ化が可能になる。

なお、全ての統計処理には、データ解析環境 R (R Core Team 2015) を用いる。

### 5. 結果と考察

#### 5.1 RQ1の結果と考察

まず、表2の使用頻度を用いてカイ2乗検定を実行した結果、 $p$ 値は0.05未満であり、「全ての作品における接続表現の使用傾向が同じである」という帰無仮説が棄却され、中古作品における接続表現の使用頻度に有意差があることが示された ( $\chi^2 = 429.40$ ,  $df = 52$ ,  $p\text{-value} < 0.001$  \*\*\*)。従って、この結果は、接続表現の使用傾向を精査することで中古作品における文体上の違いが浮き彫りにされることを示唆している。

#### 5.2 RQ2の結果と考察

次に、上記のカイ2乗検定の結果を踏まえ、具体的に特定の作品で有意に多く (あるいは少なく) 用いられている接続表現を特定するために、Habermaの残差分析を行った。表3は、その結果をまとめたものである。表中の値は標準化残差を表し、太字で強調されている値は有意差が見られるものである。また、太字になっている正の値は、他の中古作品と比べて、その接続表現が有意に多く用いられていることを示す。逆に、太字になっている負の値は、その接続表現が有意に少なく用いられていることを示す。そして、標準化残差の絶対値が大きい (あるいは小さい) ほど、その接続表現がその作品の文体を強く特徴づけていることを意味する。

表3 Haberman の残差分析の結果

	A_ サテ	A_ その他	B_ 全体	C_ カカリ系	C_ サリ系
竹取	<b>-2.07</b>	<b>1.41</b>	1.30	<b>3.28</b>	-1.32
古今	<b>-2.68</b>	<b>2.69</b>	<b>9.12</b>	-0.09	<b>-3.16</b>
伊勢	0.48	-1.26	-1.10	0.15	0.97
土左	-1.87	1.06	<b>9.76</b>	4.32	<b>-4.67</b>
大和	<b>3.30</b>	<b>0.37</b>	0.16	1.33	<b>-3.85</b>
平中	<b>2.56</b>	<b>2.95</b>	-1.55	-1.37	<b>-3.58</b>
落窪	1.21	-1.56	-1.27	<b>4.85</b>	-1.44
枕草	0.24	5.24	-1.74	<b>-3.04</b>	<b>-2.69</b>
和泉	-0.94	-1.42	-0.69	<b>3.98</b>	0.50
源氏	-0.79	-6.38	<b>-3.10</b>	<b>-3.50</b>	<b>8.79</b>
紫式	<b>-2.93</b>	<b>1.07</b>	0.22	-1.18	<b>2.10</b>
堤中	0.09	-0.50	-1.03	-1.32	1.30
更級	-0.48	0.60	-0.45	-0.57	0.33
讃岐	-1.75	0.18	0.69	1.33	0.54

表3における作品に注目すると、『古今和歌集』では、5項目中の4項目に有意差が見られ、他の中古作品とは接続表現の使用傾向が大きく異なっていることが分かる。また、『竹取物語』、『大和物語』、『平中物語』、『源氏物語』、『紫式部日記』では、5項目中の3項目に有意差が見られる。その反対に、『堤中納言物語』、『更級日記』、『讃岐典侍日記』の3作品には有意差がまったく見られず、接続表現に関する際立った特徴がないことが示されている。そして、5種類の接続表現の方に注目すると、「C\_ サリ系」が14作品中の7作品で有意差が見られ、中古作品の文体を識別する有効な指標となり得ることが分かる。

### 5.3 RQ3の結果と考察

続いて、対応分析を用いて、接続表現の使用頻度（文数で正規化）に基づく中古作品のグループ化を行った。表4～5は、対応分析の結果として得られた4次元のうち最も寄与率の高い2次元の得点を集計したものである<sup>4</sup>。なお、第2次元までの累積寄与率は82.30%である。

<sup>4</sup> 対応分析の結果として得られる次元の数は、行数と列数のうちで小さい方から1を引いた数となる。



表4 対応分析の次元得点（行得点）

	Dim 1	Dim 2
竹取	-0.69	0.93
古今	-2.71	-0.84
伊勢	0.72	0.70
土左	-2.42	-0.21
大和	0.16	-0.93
平中	0.48	-0.98
落窪	0.20	0.46
枕草	0.50	-0.56
和泉	-0.06	2.41
源氏	0.85	1.12
紫式	0.44	1.68
堤中	0.88	0.57
更級	0.69	0.51
讃岐	-0.15	1.59

表5 対応分析の次元得点（列得点）

	Dim 1	Dim 2
A_ サテ	0.52	-1.21
A_ その他	-0.14	-0.60
B_ 全体	-3.39	-0.26
C_ カカリ系	-1.44	1.33
C_ サリ系	0.62	1.08

そして、図1は、表4～5の次元得点を用いて、14種類の中古作品と5種類の接続表現の関連性を可視化した結果である。因みに、表中の次元得点と図中の座標が若干異なるのは、描画にあたって対称的正規化（Greenacre 2016）という処理がなされているためである。

対応分析の結果を可視化した図1では、接続表現の使用傾向が類似した作品同士が近くに布置され、使用傾向が異なる作品同士が遠くに布置されている。また、原点から引かれた矢印に注目することで、どの作品（群）にどの接続表現が顕著に使われているかが分かる。

例えば、図中の左側に『古今和歌集』と『土左日記』が布置されていて、これらの2作品では「B\_全体」の使用頻度が他の作品よりも極めて高い（この結果は、前掲の残差分析の結果等とも一致する）。これについては、『古今和歌集』（仮名序）と『土左日記』の作者が同一の男性（紀貫之）であるためであると考えられる。和歌の部分では、あまり接続表現が用いられないが、和歌の本質、その起源・歴史や歌人の優劣等といった抽象的な論を仮名で書いた仮名序では、紀貫之が対句（駢儷体の影響）や接続表現等を多用している（渡辺 1981）。その結果、当時の男性の一般的な教養であった漢文訓読の表現である「B\_全体」が用いられたものと予想される。

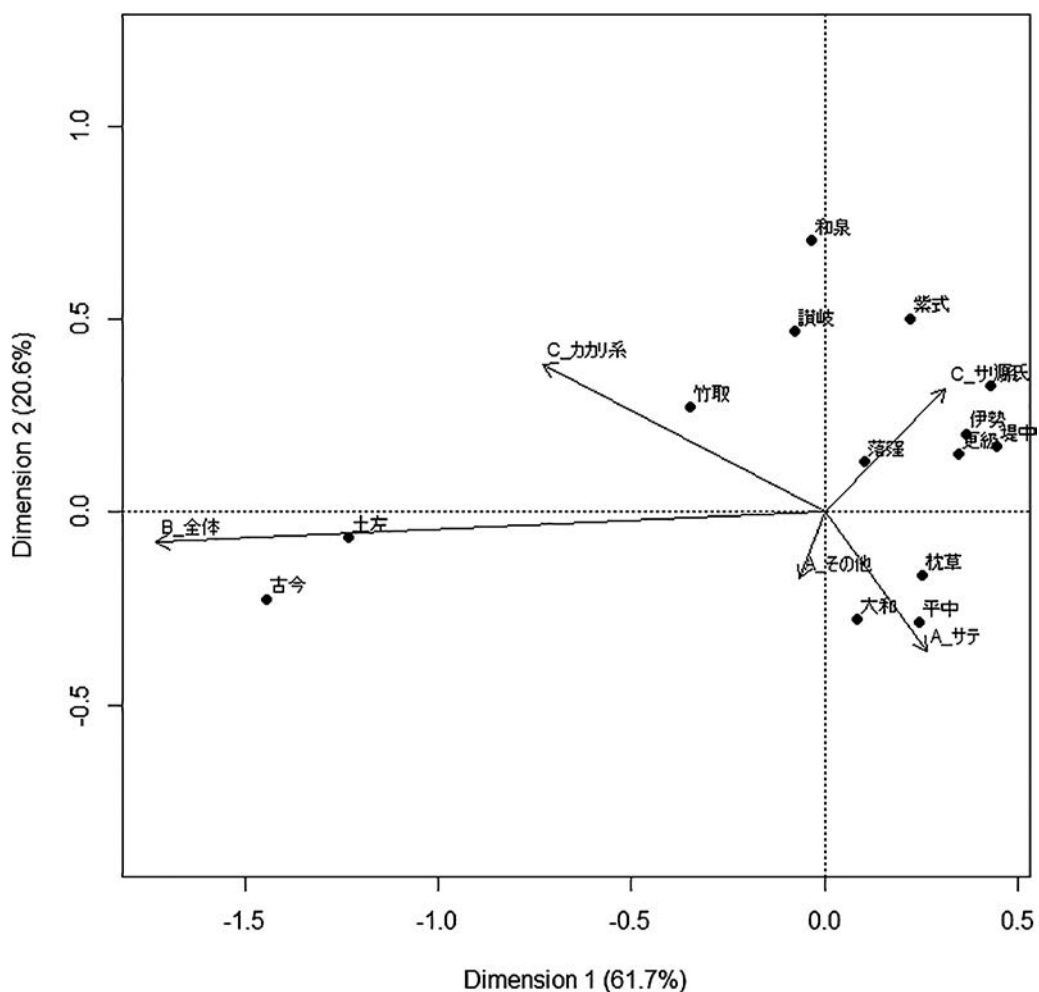


図1 対応分析による中古作品のグループ化

また、図中の右下に注目すると、『大和物語』や『平中物語』等で「A\_サテ」が多く使われていることが分かる。これについては、岡崎（2011）で述べられている歌物語のストーリー構成による影響と考えられる。歌物語は主に各段が短いストーリーで構成されており、また、以下の（3）に示すように、その歌がどのように歌われたかという状況を「歌が詠まれる背景の説明 → 『サテ』歌」という形式で提示している。そのために、このような結果が出たものと予想される。

- (3) 今日明日あひなむとしけるほどに、伊勢の斎宮の御占にあひたまひにけり。「いふかひなくくちをし」と、思ひたまうけり。さてよみて奉りたまひける。  
伊勢の海の千尋の浜にひろふとも今はかひなくおもほゆるかな（大和物語，p. 316）

なお、山口（1984）は、歌物語（『平中物語』、『大和物語』、『伊勢物語』）が文と文との接続に接続詞と指示詞（「ソノ＋名詞」, 「ソレ」等）を重ねて用いる傾向が高いことから、これらは「耳で聞く」文学であった特色を示しているのではないかと予想している。しかし、山口（1984）でも指摘されているように、『伊勢物語』は『平中物語』や『大和物語』に比べて接続表現の使用が少ないことから、同一に扱うのは問題があると思われる。これについては、他の口承文学とされるものの調査も含め、今後、慎重に考えていく必要がある。

因みに、「A\_その他」の矢印が伸びる方向には作品が存在せず、中古作品の文体識別には（対応分析の結果において、他の接続表現ほど）寄与していない。これは、「A\_その他」の用例がほぼ「マタ」であり（マタ 225 例、アルイハ 6 例、タダシ 2 例、ソエニ 1 例、スナハチ 1 例、サハレ 15 例）、この語が『和泉式部日記』を除く全作品に満遍なく用いられているためと考えられる。

そして、「C\_カカリ系」に注目すると、『竹取物語』を特徴づけている。また「C\_サリ系」は、紫式部によって書かれた『源氏物語』と『紫式部日記』等に顕著である。これについて、さらに詳しいデータを見てみると、カカリ系で例数の多かった語「かかるほどに・かかれば」は『竹取物語』、『伊勢物語』、『落窪物語』、『大和物語』に見られ、『古今和歌集』、『紫式部日記』には見られない（『源氏物語』、『枕草子』にも少ない）。それに対して、サリ系は『紫式部日記』や『源氏物語』等に多く、『土左日記』、『古今和歌集』には少ない。このことについては、先行研究（e.g., 安部 1996, 村田 2001, 金水 2006）でも、950 年代後半の『宇津保物語』や『蜻蛉日記』、『落窪物語』以降の作品とそれより前の作品との間に差異が見られるという指摘がある。従って、それと同様に時代の差が現れている可能性がある。ただし、カカリ系を中心とした前半の作品は男性、後半は女性の手によるものであると考えられており、書き手の位相によるものとも考えられる。特に、『竹取物語』の素材・背景が仏典や漢籍等の典籍をもとにしているという指摘が江戸時代からあり、近年も盛んに研究されているように、書き手が男性であること、そして漢文訓読の影響といったことも考えられよう。

最後に、『紫式部日記』、『讃岐典侍日記』、『和泉式部日記』が図中で近くに布置されているが、同じ日記でも『更級日記』のみは遠くに布置されている。これについて、『更級日記』は、日記とされているものの、作者の旧歌稿ないし家集をもとに晩年に編まれたと言われており（孝標女の家集とする考え方があり）、時系列的に書き綴ってゆく他の日記とは文体が異なると予想される。

このように対応分析を用いると、接続表現の使用傾向に基づく中古作品のグループ化ができる。しかしながら、散布図上で近くに布置された作品を手作業でグループ化する場合、解釈が恣意的になる恐れがある。そして、この問題に対処する方策として、対応分析の結果として得られる次元得点に基づく階層型クラスター分析が知られている（Alberti 2013）。そこで本研究でも、対応分析の結果として得られた第 1～2 次元得点を用いて、階層型クラスター分析を行った。その際、距離の計算には標準化ユークリッド距離、クラスターの結合法には群平均法を用いた。図 2 は、その結果である。階層型クラスター分析の結果として得られる樹形図において、任意の 2 作品の距離（非類似度）は、それらの作品を結ぶ線の長さに対応している。つまり、一番左の『古今和歌集』（仮名序）は『土左日記』と類似しており、これらの 2 作品は他の作品全てと異なる性質を持つ

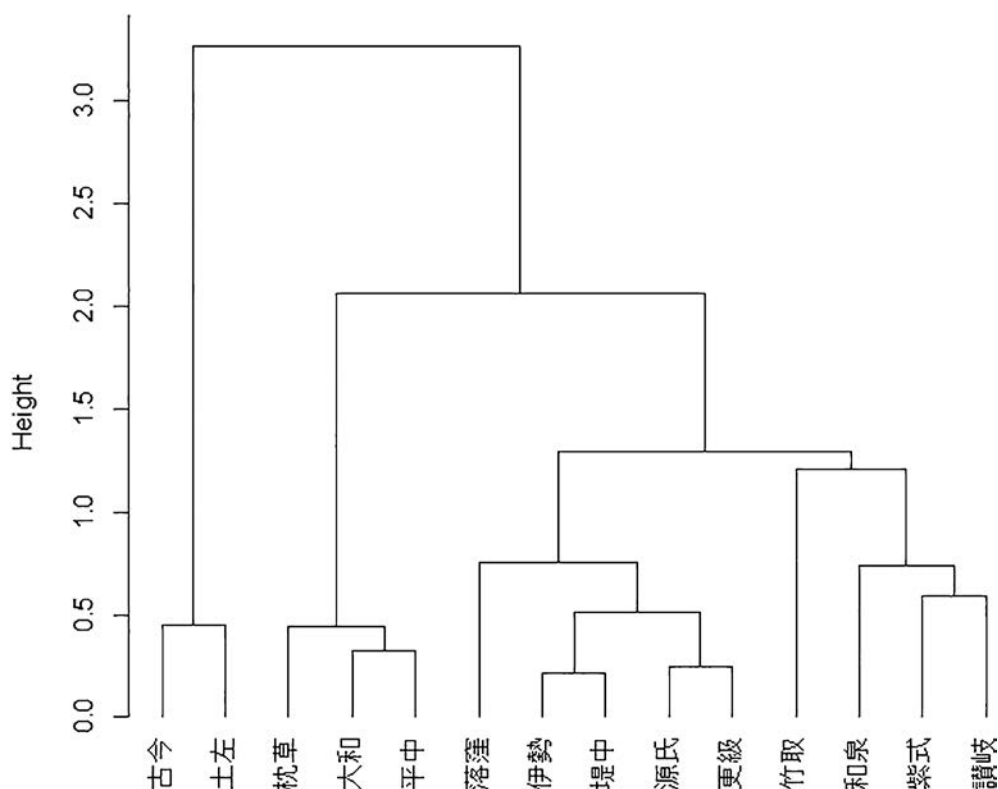


図2 階層型クラスター分析による中古作品のグループ化

ている。また、対応分析の結果と同様に、『大和物語』や『平中物語』の類似性、日記3作品（『紫式部日記』、『讃岐典侍日記』、『和泉式部日記』）の類似性が見てとれる。

## 6. 今後の課題

ここまで、中古14作品の接続表現から、中古の資料性について考察を行ってきた。その結果、特に作者（紀貫之）とジャンル（歌物語、日記）による類似性と、C類による時代性が浮き彫りにされた。今後は、接続表現以外の言語表現にも書き手やジャンルによる類似性や時代性が見られるのかを調査する必要がある。これまでも、助詞・助動詞（小林・小木曾 2013）や「コノ・ソノ・アノ」+ N（名詞句）（岡崎 2014）等が異なるジャンルの文体識別に有効であることが示されている。従って、コーパス研究の特長を活かし、様々な言語表現の網羅的な調査を行っていきたい。また、2016年に日本語歴史コーパス（平安時代編）に『大鏡』、『蜻蛉日記』の2作品が追加されたため、より多くの中古資料が利用可能となった。そして、年代差やジャンル差を考慮した統計解析を行うにあたっては、多重因子分析（小林・小木曾 2013）や様々な回帰モデリング（Gries 2015, Gries and Deshors 2014）が有効になると予想される。なお、回帰モデリングを用いる場合は、

情報量基準 (Konishi and Kitagawa 2007) 等を用いて、異なる時代・ジャンル・書き手等を弁別する言語表現を特定することができるであろう。

## 参考文献

- 安部清哉 (1996) 「語彙・語法史から見る資料—『篁物語』の成立時期をめぐりて」『国語学』184: 14-27.
- Agresti, Alan (2007) *An introduction to categorical data analysis*. Second edition. New York: John Wiley & Sons.
- Alberti, Gianmarco (2013) An R script to facilitate correspondence analysis: A guide to the use and the interpretation of results from an archaeological perspective. *Archeologia e Calcolatori* 24: 25-53.
- 福島直恭 (2008) 『書記言語としての「日本語」の誕生—その存在を問い直す』東京：笠間書院.
- Greenacre, Michael (2016) *Correspondence analysis in practice*. Third edition. Boca Raton: Chapman & Hall.
- Gries, Stefan Th. (2015) The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10: 95-125.
- Gries, Stefan Th. and Sandra C. Deshors (2014) Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9: 109-136.
- Haberman, Shelby J. (1973) The analysis of residuals in cross-classified tables. *Biometrics* 29: 205-220.
- 金水敏 (2006) 『日本語存在表現の歴史』東京：ひつじ書房.
- 小林雄一郎・小木曾智信 (2013) 「中古和文における個人文体とジャンル文体—多変量解析による歴史的資料の文体研究」『国立国語研究所論集』6: 29-43.
- Konishi, Sadanori and Genshiro Kitagawa (2007) *Information criteria and statistical modeling*. New York: Springer.
- 京極興一・松井栄一 (1973) 「接続詞の変遷」鈴木一彦・林巨樹 (編) 『品詞別日本語文法講座 6 接続詞・感動詞』89-136. 東京：明治書院.
- Moisl, Herman (2015) *Cluster analysis for corpus linguistics*. Berlin: Mouton De Gruyter.
- 村田菜穂子 (2001) 「平安時代の形容動詞—ゲナリと〜カナリ」『国語学』52: 16-30.
- 岡崎友子 (2011) 「指示詞系接続表現の歴史的变化—中古の『カクテ・サテ』を中心に」青木博史 (編) 『日本語文法の歴史と変化』67-87. 東京：くろしお出版.
- 岡崎友子 (2014) 「指示詞再考—コロケーション強度からみる中古のコノ・ソノ・カノ+名詞句」『日本語学』11月臨時増刊号, 33(14): 138-150.
- 岡崎友子 (2015) 「中古和文における接続表現について」近藤泰弘・田中牧郎・小木曾智信 (編) 『コーパスと日本語史研究』71-92. 東京：ひつじ書房.
- R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- 高橋尚子 (1985) 「中古語接続詞の機能と変遷—物語文学作品を資料にして」『愛文』21: 8-17.
- 築島裕 (1963) 『平安時代の漢文訓讀語につきての研究』東京：東京大学出版会.
- 山口伸美 (1984) 『平安文学の文体の研究』東京：明治書院.
- 渡辺実 (1981) 『平安朝文章史』東京：東京大学出版会.

## 資料

国立国語研究所コーパス開発センター編 (2015) 『日本語歴史コーパス』(バージョン 2015.3, 中納言バージョン 2.0.1) <https://maro.ninjal.ac.jp> (2015年3月31日確認)

和歌集：『古今和歌集』

日記：『土左日記』, 『和泉式部日記』, 『紫式部日記』, 『更級日記』, 『讃岐典侍日記』

物語：『竹取物語』, 『落窪物語』, 『源氏物語』, 『堤中納言物語』

歌物語：『伊勢物語』, 『大和物語』, 『平中物語』

随筆：『枕草子』

\* 用例には小学館「新編日本古典文学全集」の頁数を付した。

## A Statistical Analysis of Conjunctions in Early Middle Japanese with a Focus on Deictic Expressions

KOBAYASHI Yuichiro<sup>a</sup>

OKAZAKI Tomoko<sup>b</sup>

<sup>ab</sup>Toyo University / Project Collaborator, NINJAL

### Abstract

The purpose of the present study is to investigate the use of conjunctions in Early Middle Japanese. By using the Corpus of Historical Japanese and statistical methods, this study examined variations of the frequencies of connectives across periods, genres, and authors. The results show (a) the similarity of *Kokin Wakashu Kanajo* and *Tosa Nikki* due to the style of Ki no Tsurayuki, (b) the similarity of song stories due to the high frequency of *sate*, (c) the influence of periods upon the use of *kakari*- and *sate*-conjunctions.

**Key words:** Early Middle Japanese, conjunctions, Corpus of Historical Japanese, multivariate analysis