

十二ヶ月の別名による古辞書・類書の統計分析

島野 達雄

1. はじめに

1.1 調査と研究のあらまし

一月、二月、三月…の別名として睦月、如月、弥生…があるように、わが国では古くから数多くの十二ヶ月の別名(十二月異名、月の異名)が用いられてきた。たとえば陰暦九月の別名のうち、「菊」が語頭についているものだけでも、菊開月、菊咲月、菊時、菊秋、菊節、菊月、菊天、菊見月がある。そこで、和漢の古辞書、類書(百科事典)、近現代の便覧、辞典など84種の文献から1405個の十二ヶ月の別名を抽出し、出典や解説を添えて、表計算ソフトの「エクセル」でデータベース化した。なお2010年7月現在、データ数(別名の個数)は約1500個に増加している。

ここでは、統計ソフトのSPSSを利用して、主成分分析などの手法により、84種の文献のうち48種に限定して、相互にどのような関係があるかを明らかにしたい。

2. データベースの構築

2.1 文献の選び方

調査した別名や文献は、もともと集めていた約600個の別名を、ひとつひとつ大漢和辞典と日本国語大辞典(第二版)で出典を調べた際、それまで知らなかった別名と文献を「芋づる式」にたぐりよせたものが中心になっている。

84種の文献のうち、統計の対象とした48種の文献は、

(1) 漢籍では和刻があるもの

(2) 版本や影印で原文が確認できるもの

を原則として選んだ。ただし、この選択は「十二ヶ月の別名による分析にふさわしいと思われる」という筆者の主観的な判断にもとづいている。

番号	〔文献名〕	〔略称〕	〔種別〕	〔成立〕	〔別名〕
1	詩経・幽風・七月	詩幽		殷代－春秋戦国	30
2	礼記・月令	月令		周末－秦・漢代	56
3	史記・天官書	史記		前漢初期	20
4	爾雅注疏・釋天	爾雅	辞書	周代－漢代	71
5	梁元帝纂要	纂要		漢魏六朝末期	80
6	倭名類聚抄	倭名	辞書	承平四年(934)頃	12
7	天治本新撰字鏡	字鏡	辞書	天治年間(1124－26)	28
8	色葉字類抄	色葉	辞書	治承年間(1177－81)	43
9	蔵玉和歌集	蔵玉	和歌集	延元三年(1338)頃	37
10	莫伝抄	莫伝	和歌集	室町前期	24
11	古今打聞(秘蔵抄)	打聞	和歌集	永享十年(1438)頃	24
12	壺囊抄	壺囊	類書	文安三年(1446)	72
13	明応五年本節用集	明応	辞書	明応五年(1496)	47
14	因縁抄	因縁	類書	室町時代か	68
15	月庵醉醒記	醉醒	類書	室町時代か	214
16	穂久邇文庫本女訓抄	女訓	類書	室町時代か	41
17	文明本節用集	文明	辞書	室町中期	80
18	藻塩草	藻塩	連歌辞典	天文二年(1533)	102
19	伊京集	伊京	辞書	室町末期	78
20	翰墨全書	翰墨	書札礼	明代	110
21	黒本本節用集	黒本	辞書	室町末期	140
22	饅頭屋本節用集	饅頭	辞書	室町末期	32
23	天正十八年本節用集	天正	辞書	天正十八年(1590)	68
24	易林本節用集	易林	辞書	慶長二年(1597)	126
25	日葡辞書	日葡	和ポ辞書	慶長八－九年(1603－04)	80
26	古今類書纂要	類纂	類書	明・天啓元年(1620)序	222
27	下学集	下学	辞書	元和年間(1615－23)	73
28	国花集	国花	漢文辞典	寛永五年(1628)	126
29	増山の井	増山	歳時記	寛文三年(1663)	135
30	字林拾葉(合類節用集)	字林	辞書	延宝八年(1680)	241
31	邇言便蒙抄	邇言	辞書	天和二年(1682)	89
32	書札調法記	書札	書札礼	元禄八年(1695)	181
33	俳諧通俗志	通俗	歳時記	享保元年(1716)自跋	107
34	書言字考節用集	書言	辞書	享保二年(1717)	42
35	手挑灯	手挑	歳時記	延享二年(1745)	161
36	静嘉堂文庫本莫伝抄	静莫	和歌集	近世か	12
37	事物異名録	事物	類書	清・乾隆四十一年(1776)	100
38	俳諧四季部類	四季	歳時記	安永九年(1780)	125
39	類聚名物考	名物	類書	明和－安永年間(1764－80)	28
40	大広益字尽重宝記綱目	字尽	辞書	天明元年(1781)	239
41	華実年浪草(三餘抄)	年浪	歳時記	天明三年(1783)	130
42	俳諧歳時記	俳歳	歳時記	享和三年(1803)	157
43	式礼用文章	式礼	書札礼	文化六年(1809)	40
44	季引席用集	季引	歳時記	文政元年(1818)	191
45	古今要覧稿	要覧	類書	天保十三年(1842)	442
46	増補俳諧歳時記菜草	増歳	歳時記	嘉永三年(1850)	158
47	和英語林集成	ヘボン	英和・和英辞書	慶応三年(1867)	46
48	言海	言海	辞書	明治二十四年(1891)	87

表 2.1 統計の対象とした 48 種の文献

2.2 データの構成

別名は、各文献の表記にもとづいて採録した。すなわち、明らかな誤記であっても表記を優先した。「しわす」と読む「師走」「師趨」「師走月」「除」「臘」などは、異なる別名としている。また、同じ表記で「月」が異なる場合は、異なる別名として扱った。

一つの別名に対しては、〔番号〕〔名前〕〔代表的読み方〕〔月〕〔注釈〕〔その他の読み方〕に始まる総計約 150 の入力欄（データの見出し）をもうけた。主なものは次の通り。

まず、1.詩幽から 48.言海まで、48 種の文献にその別名があれば 1、なければ 0 を入力する欄がある。1 または 0 を要素にもつ、この〈1405 個の別名〉×〈48 種の文献〉の行列が今回の統計分析の基礎データとなっている。

次に、日本国語大辞典に示されている出典、大漢和辞典の巻号とページおよび出典などを入力する欄がある。

さらに蔵玉和歌集、古今打聞（秘蔵抄）、静嘉堂文庫本莫伝抄、莫伝抄の 4 種の和歌集に掲載されている和歌およびその作者等の入力欄があり、最後に明治新撰俳諧季寄鑑、天下作文必携、和歌俳諧節用集、明治節用無尽蔵の 4 種の明治混乱期の文献における「月」を書きこむ欄がある。明治混乱期とは明治 5 年の太陽暦採用後、「如月」を「3 月」や「4 月」などとした時期を指す。この欄は、各別名が明治混乱期のどの文献に何月として記されているかを示すためのもの。

以上のほか、印刷の指定をするための各種の数式の欄、印刷用一括表示の欄がある。なお、このような印刷のための指定は、ほとんどをエクセルの関数（コマンド）でおこなっており、マクロ（VBA）はごく一部しか使用していない。

3. 統計分析の結果と解釈

3.1 相関係数による近接文献の判定

上に述べたように、48 種の文献は、各要素が 1 か 0 のどちらかの 1405 次元のベクトルになっている。相関係数は、二つのベクトルどうし、つまり二つの文献どうしの関係をあらわす数値で、自分自身とは常に 1、関係が深いほど 1 に近い値、関係が浅いほどマイナスの値が大きくなる。

以下に、48 種の文献のなかから、1.詩幽、2.月令、4.爾雅、9.蔵玉、12.塏囊、13.明応、23.天正、25.日葡、29.増山、30.字林、41.年浪、48.言海、の 12 種を抜粋し、それらと関係が深い上位文献 5 種と関係が浅い下位文献 5 種を相関係数とともに示す。

	1.詩幽	2.月令	4.爾雅	9.蔵玉	12.壺囊	13.明応
上位	翰墨 0.104	女訓 0.462	字鏡 0.316	藻塩 0.448	女訓 0.535	伊京 0.646
	要覧 0.059	翰墨 0.374	字林 0.188	増山 0.429	因縁 0.504	天正 0.622
	字林 0.037	壺囊 0.332	類纂 0.185	俳歳 0.393	邇言 0.470	黒本 0.520
	黒本 0.033	邇言 0.320	黒本 0.184	年浪 0.392	倭名 0.399	名物 0.483
	字尽 0.025	因縁 0.310	要覧 0.152	増歳 0.392	易林 0.390	饅頭 0.449
下位	年浪 -0.047	静莫 -0.019	月令 -0.030	国花 -0.052	静莫 -0.022	莫伝 0.006
	増山 -0.048	莫伝 -0.027	莫伝 -0.030	字林 -0.063	史記 -0.028	蔵玉 -0.006
	俳歳 -0.052	打聞 -0.027	詩幽 -0.034	書札 -0.063	莫伝 -0.031	静莫 -0.017
	増歳 -0.053	爾雅 -0.030	蔵玉 -0.038	酔醒 -0.070	詩幽 -0.034	史記 -0.022
	手挑 -0.053	蔵玉 -0.034	四季 -0.049	類纂 -0.071	蔵玉 -0.038	詩幽 -0.027

	23.天正	25.日葡	29.増山	30.字林	41.年浪	48.言海
上位	伊京 0.742	言海 0.612	増歳 0.633	字尽 0.724	俳歳 0.760	へボン 0.617
	明応 0.622	へボン 0.576	俳歳 0.628	翰墨 0.479	増歳 0.718	日葡 0.612
	黒本 0.611	文明 0.430	年浪 0.613	書札 0.327	増山 0.613	易林 0.467
	饅頭 0.588	易林 0.428	通俗 0.589	国花 0.260	通俗 0.603	邇言 0.430
	文明 0.574	伊京 0.423	手挑 0.520	字鏡 0.219	季引 0.511	文明 0.421
下位	莫伝 -0.004	事物 0.016	爾雅 0.090	史記 0.009	爾雅 0.050	事物 0.032
	蔵玉 -0.016	莫伝 0.015	纂要 0.087	倭名 -0.001	莫伝 0.034	莫伝 -0.011
	静莫 -0.021	静莫 0.011	莫伝 0.013	静莫 -0.042	文明 0.006	静莫 -0.024
	史記 -0.027	史記 -0.030	史記 -0.039	莫伝 -0.045	史記 -0.018	史記 -0.031
	詩幽 -0.033	詩幽 -0.036	詩幽 -0.048	蔵玉 -0.063	詩幽 -0.047	詩幽 -0.038

表 3.1 12 種の文献の近接上位 5 種・下位 5 種文献

表 3.1 からは、こと十二ヶ月の別名の記載だけに限定されるが、各文献の関係について、相関係数を根拠とする、いくつかの結論を導くことができる。

表 3.1 の 1.詩幽から「詩経・豳風・七月の詩はどの文献とも関係が浅い」、2.月令から「礼記・月令は中世の古辞書、類書との関係が若干ある」、13.明応、23.天正から「いわゆる古本節用集どうしの関係は深い」、29.増山、41.年浪から「近世の歳時記どうしの関係は深い」、30.字林から「漢文や書札札に関する文献どうしは関係が深い」などと主張できる。

3.2 主成分分析

次に、もとの 1 または 0 を要素にもつ、〈1405 個の別名〉×〈48 種の文献〉の行列を対象として主成分分析をおこない、12 個の成分を抽出した。

	成分	1	2	3	4	5	6	7	8	9	10	11	12
抽出後の 負荷量 平方和	合計	10.84	4.02	3.03	2.10	1.66	1.54	1.37	1.27	1.16	1.05	1.05	1.03
	分散%	22.57	8.38	6.31	4.38	3.46	3.21	2.86	2.65	2.41	2.20	2.18	2.14
	累積%	22.57	30.95	37.27	41.64	45.10	48.31	51.17	53.82	56.23	58.43	60.60	62.74

表 3.2-1 12 個の成分抽出により説明された分散の合計

12 個の成分を合わせても累積寄与率が 62.74% であり、顕著な因子の抽出とは言いがたいが、成分 1~6 における共通性の高い文献を示し、いちおうの解釈を述べておく。

成分1	成分2	成分3	成分4	成分5	成分6
通俗 0.674	年浪 0.563	書札 0.449	字林 0.505	事物 0.499	字林 0.371
邇言 0.672	俳歳 0.499	翰墨 0.436	字尽 0.447	纂要 0.430	字尽 0.353
増山 0.641	増歳 0.494	字林 0.423	爾雅 0.432	類纂 0.406	へボン 0.341
言海 0.641	蔵玉 0.474	女訓 0.416	字鏡 0.409	爾雅 0.237	書言 0.322
季引 0.633	増山 0.388	類纂 0.413	黒本 0.286	天正 0.222	事物 0.264
明応 0.632	通俗 0.343	堪囊 0.372	天正 0.249	伊京 0.206	言海 0.233
俳歳 0.618	手挑 0.336	纂要 0.354	伊京 0.248	俳歳 0.181	纂要 0.189
増歳 0.611	季引 0.330	月令 0.325	下学 0.217	文明 0.163	日葡 0.173

表 3.2-2 成分 1～6 における共通性の高い文献 8 種

成分 1 は、中世、近世の多くの文献で共通性が高く、「わが国には字書、節用集、歳時記などの分野をこえて十二月の別名に関する知識が存在しており、享保元年（1716）の俳諧通俗志や天和二年（1682）の邇言便蒙抄がその代表格」と言ってもよいであろう。

成分 2 は、歳時記と蔵玉和歌集の融合形。

成分 3 は、書札礼つまり「手紙の書き方」に影響を与えている成分と解釈できる。

3.3 階層クラスター分析

統計ソフトの SPSS では、二つの 1405 次元のベクトル間の距離を定義し、階層クラスター分析によりデンドログラムとよぶ図を描くことができる。（次ページの図 3.3）

4. 課題と展望

4.1 データ入力の諸問題

別名の採録方針については、異論もあるであろうが、「上冬」を「十一とう」と読んだ女訓抄の例や、「晷短」を「匙」と「短」の二つに読んだ藻塩草の例などは、「各文献の表記にもとづいて採録する」という方針がなければ誤読が明示できない、と考える。

また、近世後期の節用集などはビジュアル化が進み、楷書の本字と行書の異体字を並置したり、熟語の音読みを右側に、訓読みを左側に振り仮名として添えたりしている。このようなケースは、データベースの〔備考〕の欄に注記したが、検討の余地は残る。

4.2 統計解釈と手法について

統計でつねに起きる「解釈」の問題は言うまでもなく、このような題材では、主成分分析より数量化理論Ⅲ類のほうが適切かもしれないが、筆者の力不足で適用できなかった。

4.3 総括および今後の展望

幸か不幸か 2007 年の日本国語大辞典のオンライン化以前に、手作業で別名と参照文献の入力作業をおこなったため、「このようなデータベースの構築では、別名の個数と文献の

数の双方が同時に増えてゆく」，すなわち「データベースは対角線方向に増大してゆく」ことが実感できた。

今後，十二ヶ月の別名に限らず，たとえば日本国語大辞典の見出し語 50 万語と大漢和辞典の親字 5 万字・熟語 53 万語について，古辞書や類書ごとにその見出し語の有無を 1 または 0 と入力すれば，今回おこなったものと同様の，古辞書や類書のあいだの関係の統計分析や，成立年代不明の文献の年代推定などに利用できる可能性がある。

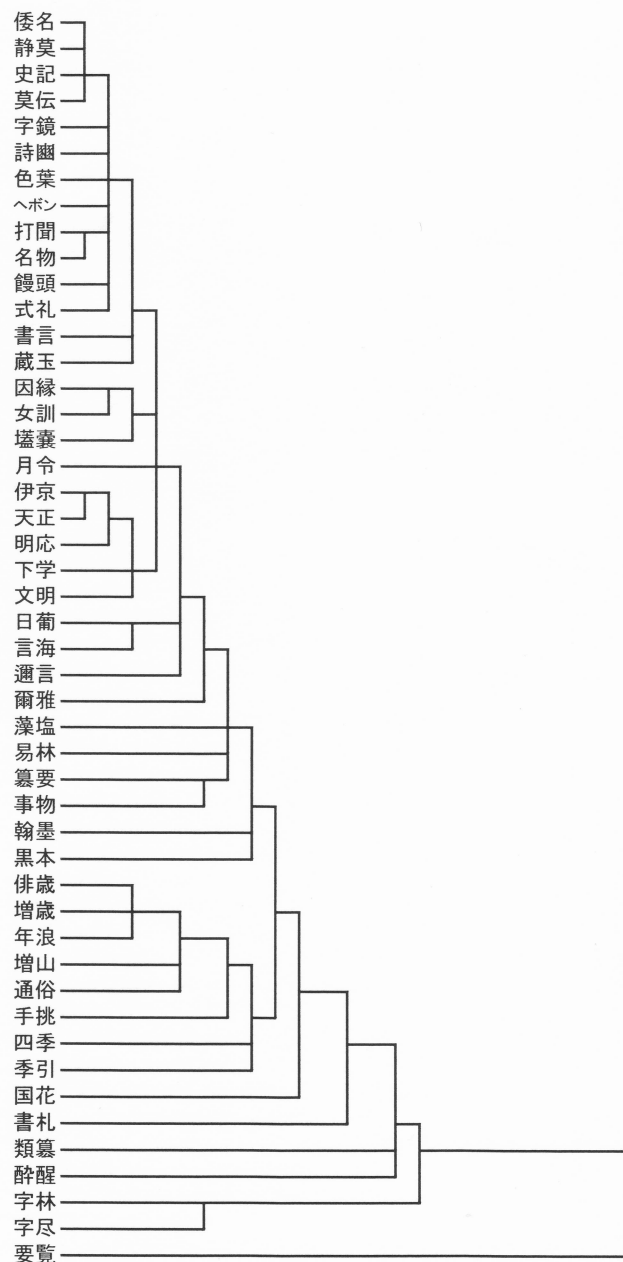


図 3.3 階層クラスター分析による 48 種の文献のデンドログラム