

古典語彙の出現頻度の分布

——『日本古典対照分類語彙表』を利用して

石井 久雄（同志社大学文学部）

語彙研究会の定例研究会が百回という節目に至ったことを、お祝い申し上げます。

語彙研究会そのものも昨年既に十周年を迎え、立派なことであると存じます。

毎年の大会が始めて名古屋を離れたのは、第7回、2009年の学習院大学で行われたときであると理解します。その大会では、研究会代表・田島毓堂先生の学術講演「比較語彙論 現状と課題」があり、引く続く研究発表の最初に

宮島達夫「古典対照語い表の増補・改訂作業の中間報告」

がありました。大会は3年後の2012年にも東京の駒沢大学で行われ、そのときにも、研究発表の最後に

宮島達夫「『古典対照語い表』増補改訂の中間報告」

が置かれ、この後に田中章夫先生の学術講演「語彙の類似と偏り」がありました。

今日の発表は、宮島先生その古典対照語い表の増補改訂作業が「中間」を脱して、漸く終結しましたので、ご報告・ご紹介をするものです。本来ならば宮島先生がお話をするところであり、定例百回の記念の人選としても宮島先生がふさわしいと承知していますが、諸般の理由により、共編者陣のなかから私が出ることになりました。

これから、全体を二つに分け、前半では、新しくできた語彙表を、語彙表の言わば思想に触れながら、ご紹介します。後半では、語彙表における数値についての一つの想いを、極私的なものですが、めぐらせることとします。この上に記した標題と異なり、話の順序は逆転します。

1 『日本古典対照分類語彙表』の体裁

新しい語彙表は、『日本古典対照分類語彙表』の名を与えられ、今日から1箇月後の6月末に、笠間書院から発刊されます。

書籍本文の様子を、まずお目にかけます。本文最初の見開きを、この予稿2～3ページに少少縮小して写しました。書籍はB5版ですが、見開きのB4版の幅で1行を構成し、1語についての次の10種類のことがらを26列で記します。

- ・ 見出し 古典かなづかいのかな表記で、用語はこの五十音順に配列されています。
……最初の行では「あ」です。以下、この最初の行を読みます。
- ・ 順序 同音語のなかの順番。同音語がなければ空白です。
……「1」とあります。次の行から5行目まで同音語があります。
- ・ 漢字 おおむね総索引により、典型的な一つを挙げます。……「吾」とあります。
- ・ 語種 空白は和語、「漢」で漢語、「混」で混種語です。外来語はありません。
……最初の行の「あ」は空白で和語です。下に同音語「あ<案>」の漢語が見えます。
- ・ 品詞 空白は名詞、他は品詞名の初めの漢字により、動詞は活用の種類を添えます。

見出し	順	漢字	語種	品詞	注記	作品	合計	徒然	平家	宇治	方丈	新古	大鏡
あ	1	吾				1	43						
あ	2	足				1	3						
あ	3	彼				1	3						
あ	4	案	漢			1	1						
あ	5			感		1	1		1				
ああ		嗚呼		感		2	2						
あい		愛	漢			1	1						1
あいぎやう		愛敬	漢			6	25	1		2			1
あいぎやうづく		愛敬付	混	動四		4	51						1
あいげう		愛樂	漢			1	1	1					
あいしふ		愛執	漢			1	1						
あいす		愛	混	動サ変		5	20	8	5	3	3		1
あいだちなし				形		1	3						
あいたどころ		朝所				1	1						
あいだる			混	動下二		1	2						
あいたんどころ		朝所				1	2		2				
あいちやく		愛着	漢			1	1	1					
あいなさ		愛無				1	1	1					
あいなし		愛無		形		6	128	5					
あいなだのみ						2	4						
あいなだのめ						1	1						
あいべつりく		愛別離苦	漢			1	2		2				
あいみん		哀愍	漢			1	2		2				
あいれん		哀憐	漢			1	2		2				
あう		奥	漢		「あう無し」の	5	17			2			
あうしう		奥州	漢			1	13		13				
あうぜつ		鶯舌	漢			1	1		1				
あうむ		鸚鵡	漢			1	1						
あえか				形動		2	20						
あえもの		肖物				1	5						
あか	1	垢				3	4		2				
あか	2	閑伽	漢			2	12		3				
あが	1	吾				1	3						
あが	2	吾		連体		7	431	1		1			
あかいろ		赤色				5	20						2
あかえ		赤兄				1	1		1				
あかかう		赤香	混			1	1			1			
あかがね		銅				3	7		2	4			
あかがはをどし		赤草威				1	2		2				
あかがり		輝				1	1		1				
あかがりだいた		輝大太	混			1	2		2				
あかぎ		赤木				1	1						
あがき		足掻				2	5	1					
あかぎぬ		赤衣				2	4						
あかぎぬすがた		赤衣姿				1	1						
あがく		足掻		動四		3	4		2	1			
あかくちば		赤朽葉				2	2						
あかげさ		赤袈裟	混			1	1						
あかごま		赤駒				1	11						
あかさ		明・赤				3	4		1				
あかぎ						1	1	1					
あかし	1	灯				7	11		1	1			
あかし	2	証				1	1						1
あかし	3	明石				7	71		9			4	1
あかし	4	明・赤		形		15	152	2	5	28	1		11
あかしう		明得		動下二		1	1						
あかしおほと		明石大門				1	1						
あかしがた		明石湯				2	2					1	
あかしがたし		明難		形		1	2						
あかしかぬ		明不堪		動下二		3	9		4				
あかしくらしわぶ		明暮佗		動上二		1	1						
あかしくらす		明暮		動四		7	37	2	8	1			

作品 合計 徒然 平家 宇治 方丈 新古 大鏡

更級 紫 源氏 枕 蜻蛉 後撰 土左 古今 伊勢 竹取 万葉 意味分類

											43	12010(われ・なれ・かれ)
											3	15603(手足・指)
		3										11010(こそあど・他)
		1										13066(判断・推測・評価)
												43000(感動)
	1	1										43010(間投)
												12390(固有人名)
	1	14	6									13030(表情・態度)/13500(交わり)
	1	42	7									23030(表情・態度)
												13020(好悪・愛憎)
		1										13042(欲望・期待・失望)
												23020(好悪・愛憎)/23680(待遇)
		3										33680(待遇・礼など)
			1									14410(家屋・建物)
		2										23020(好悪・愛憎)
												14410(家屋・建物)
												13042(欲望・期待・失望)
												11332(良不良・適不適)
1	4	101	9	8								31332(良不良・適不適)/31920(程度)/33045(意志)/33070(意味・問題・趣旨など)
1		3										13042(欲望・期待・失望)
		1										13042(欲望・期待・失望)
												13014(苦悩・悲哀)
												13020(好悪・愛憎)
												13020(好悪・愛憎)
	2	10	2	1								11622(年配)/11623(時代)/11771(奥・底・陰)/13061(思考・意見・疑い)
												12590(固有地名)
												13031(声)
			1									15502(鳥類)
	3	17										33300(文化・歴史・風俗)
		5										11100(類・例)
			1									1 15607(体液・分泌物)
		9										15130(水・乾湿)
												3 12590(固有地名)
		16	2	4					1	406		31010(こそあど・他)
	6	5	6	1								15020(色)
												12390(固有人名)
												15020(色)
			1									15110(元素)
												13840(裁縫)
												15720(障害・けが)
		1										12390(固有人名)
												14120(木・石・金)
			3									4 13392(手足の動作)
		1										1 14210(衣服)
												11800(形・型・姿・構え)
		1										1 23392(手足の動作)
		1		1								15020(色)
			1									14210(衣服)
												11 15501(哺乳類)
			1									2 15010(光)/15020(色)
												15402(草本)
1	1	3	1	3								14600(灯火)
												11113(理由・目的・証拠)
		48						1		1		7 12590(固有地名)
15	3	38	33	8	2	2	1	1				2 33420(人柄)/35010(光)/35020(色)
												1 21635(朝晩)
												1 12590(固有地名)
												1 12590(固有地名)
		2										31346(難易・安危)
		4										1 21635(朝晩)
		1										23330(生活・起臥)
	1	21		3						1		23330(生活・起臥)

更級 紫 源氏 枕 蜻蛉 後撰 土左 古今 伊勢 竹取 万葉

- ……空白で名詞です。下に同音語「あ」の感動詞が見えます。その下のほうの「あいぎやうづく<愛敬付>」は、動詞・四段活用です。
- ・ 注記 用語のはば、地名、漢字などにつき、必要に応じて記します。
……この見開きでは、中ほどに、「あう<奥>」が「あうなし」のものであると記したのが見えるにとどまります。
 - ・ 作品 17作品中のいくつかの作品で使われたかを示します。
……最初の行の「あ」は1作品に使われたにとどまります。
 - ・ 合計 17作品で合計何回使われたか。 ……その1作品で43回使われました。
 - ・ 作品ごと それぞれの作品で何回使われたかを示し、使われなければ空白です。
17列に分けて、作品の配列は成立時期の逆順、徒然→万葉とします。
……「あ」が使われた1作品は、万葉集でした。なお、徒然から万葉までで、数値が入っている箇所数が「作品」になり、数値の合計が「合計」になります。
 - ・ 意味分類 分類語彙表（増補改訂版）の意味分類項目の番号および名称です。
……意味分類項目「1.2010(われ・なれ・かれ)」で、番号の小数点は省きます。
なお、いくつかの意味分類にわたる使われかたをしているばあいは、必要なだけ示すこととして、この見開きで行が太さ2倍の「あいなし<愛無>」「あう<奥>」は、それぞれ四つ挙がっています。意味分類二つ・三つのものも見えています。
以上が 1行の構成です。書籍全体で、行数すなわち異なり語数は34,180です。見開き一つに60行程度を収めますが、その見開きは569、ページ数が1138になります。

2 旧版『古典対照語い表』と異なる特徴

日本古典対照分類語彙表は、宮島先生の2度の研究発表の標題に見られるように、宮島達夫『古典対照語い表』（1971年、笠間索引叢刊 4）の増補改訂版に相当します。古典対照語い表と大きく異なる特徴があり、3点を挙げることができます。

新しい特徴の第1として、取り上げた**古典文学作品の数を17**としています。17作品が何であるかは、見開きの見本で、最上および最下にゴシック体で記した軸によって、ご承知いただけるかと思います。古典対照語い表で14作品を取り上げていたところに、3作品を加えました。加えた3作品を、基にした索引で掲げます。

- ・ 近藤政美・武山隆昭・近藤三佐子『平家物語語彙用例総索引』（1996年、勉誠社）
- ・ 増田繁夫・長野照子『宇治拾遺物語総索引』（1975年、清文堂出版）
- ・ 滝沢貞夫『新古今集総索引』（1970年、明治書院）

作品の成立年代から見ると、鎌倉時代に厚くなったことになるでしょう。

このほかに、万葉集も、研究の進展に対応するために、索引を入れ替えました。

- ・ 古典索引刊行会『万葉集総索引』（2003年、塙書房）

なお、万葉集については、宮島先生が巻別に出現頻度を集計していて、近いうちに公刊するように準備を進めています。

古典対照語い表と異なる特徴の第2として、**一語一語に意味分類を記述**しています。用

語の属性として語種・品詞はかねて記していて、それに意味分類が加わることになりました。

ただし、いまここで「意味分類」と言っているのは、

国立国語研究所『国立国語研究所資料集14 分類語彙表 増補改訂版』

(2004年, 大日本図書)

で項目として立てられている意味分類でありまして、すなわち「意味」そのものではありません。例えば、「あく吾」の意味分類「1.2010(われ・なれ・かれ)」もそういうものであって、いま現に話している人であるのみでなく、話し手と対立すると普通は理解される聞き手も含み、話し手・聞き手との関係で第三者も含んでいます。ここに言う意味分類は、つまり、分類語彙表が895の意味分類項目を設定して、あらゆる現代日本語の語句・表現をそこに配属させようとした、その思想を、古典日本語に踏襲し展開しようとする日本古典対照分類語彙表として図るものである、ということになるでしょう。3項目を追加して、「3.9999(枕詞)」, および動詞と判断されながら意味不明である「2.9999」, 他の意味不明「9.9999」を設けて、898項目としたのが、項目についての微細な変更です。

意味分類は、取り上げた古典文学17作品のうちに1回でも確認したものは、示しています。日本国語大辞典第二版などで、17作品いずれかの事例が用例として挙がっていれば、その意味は取り込むように努めたつもりです。しかしながら、示さなかった意味分類が、取り上げた作品のうちに現れていないとは、断言する自信がありません。本文の解釈が定着していないときでも、意味は一つを選びました。一つの意味が分類語彙表で幾つかの意味分類に該当しうることもあり、そのようなばあいには、煩雑になるのを避けて、示すのは1項目のみとしています。

また、複合語では主要部分の意味分類を示し、あるいは品詞によって制限したりしています。例えば、次の複合動詞に示した意味分類は、「さく<咲>」による「2.5701(生)」のみです。漢字「咲」はいま省略しました。

さきあふ<逢> さきいづ<出> さきかかる<懸> さきく<来>
さきこぼる<零> さきすさぶ<荒> さきそむ<初> さきちりすぐ<散過>
さきちる<散> さきづ<出> さきつぐ<継> さきでく<出来>
さきでてる<出照> さきにほふ<匂> さきはじむ<初> さきまがふ<紛>
さきまさる<勝> さきまじる<雑> さきます<増> さきみだる<乱>
さきゆく<行> さきわたる<渡> さきををる<撓> みだれさく<乱咲>

複合形容詞「さきがたし<咲難>」の意味分類は、他の「動詞+がたし<難>」とともに、「がたし<難>」に対応する「3.1346(難易・安危)」のみです。

意味分類を行ったために、用語の認定も影響を受けたところがあります。人名・地名は、古典対照語い表では相互にあるいは普通名詞と合併されていましたが、今回はほとんど分離しています。同語として意味分類幾つかを示すということもありえましたが、分離が難しいものもありましたが、試みてみました。

日本古典対照分類語彙表という一書の標題に「分類」を謳っているのは、特徴を表面に打ち出すとともに、分類語彙表を継承していることを示しています。意味分類が2以上にわたる用語については、意味分類ごとに出現頻度を計数すべきですが、今後の課題とします。

——予稿8ページに飛びます。

	合計	徒然	平家	宇治	方丈	新古	大鏡	更級	紫
全語種異なり	34180	4240	13173	6540	1148	2543	4819	1950	2468
延べ	581970	17110	100091	49183	2527	17165	29253	7243	8736
平均出現頻度	17.02	4.03	7.59	7.52	2.20	6.74	6.07	3.71	3.53
和語異なり	24478	2897	6899	4916	896	2530	3359	1769	2104
延べ	524074	14740	74071	43800	2235	17125	24144	6890	7731
平均出現頻度	21.41	5.08	10.73	8.90	2.49	6.76	7.18	3.89	3.67
漢語異なり	8146	1189	<u>5678</u>	1371	231	12	1230	146	276
延べ	50524	2114	<u>23831</u>	4744	268	39	4347	295	849
平均出現頻度	6.20	1.77	4.19	3.46	1.16	3.25	3.53	2.02	3.07
混種語異なり	1556	154	596	253	21	1	230	35	88
延べ	7372	256	2189	639	24	1	762	58	156
平均出現頻度	4.73	1.66	3.67	2.52	1.14	1.00	3.31	1.65	1.77
異なり比率 (ppm)									
和語	716149	683254	<u>523722</u>	751681	780487	994887	697032	907179	852512
漢語	238326	280424	<u>431033</u>	209633	201219	4718	255239	74871	111831
混種語	45523	36320	45244	38685	18292	393	47727	17948	35656
延べ比率 (ppm)									
和語	900517	861484	<u>740036</u>	890551	884447	997669	825351	951263	884958
漢語	86815	123553	<u>238093</u>	96456	106054	2272	148600	40728	97184
混種語	12667	14962	<u>21870</u>	12992	9497	58	26048	8007	17857
	源氏	枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉
全語種異なり	11416	5246	3599	1923	984	1993	1692	1310	6601
延べ	207788	32904	22400	11955	3496	10013	6931	5119	50056
平均出現頻度	18.20	6.27	6.22	6.21	3.55	5.02	4.09	3.90	7.58
和語異なり	9930	4413	3280	1916	926	1988	1586	1201	6565
延べ	194737	30242	21461	11933	3369	9998	6729	4859	50010
平均出現頻度	19.61	6.85	6.54	6.22	3.63	5.02	4.24	4.04	7.61
漢語異なり	1019	640	235	6	44	4	89	88	28
延べ	10528	2164	774	21	103	13	183	220	31
平均出現頻度	10.33	3.38	3.29	3.50	2.34	3.25	2.05	2.50	1.10
混種語異なり	467	193	84	1	14	1	17	21	8
延べ	2523	498	165	1	24	2	19	40	15
平均出現頻度	5.40	2.58	1.96	1.00	1.71	2.00	1.11	1.90	1.87
異なり比率 (ppm)									
和語	869831	841212	911364	996359	941056	997491	937352	916793	994546
漢語	89260	121997	65295	3120	44715	2007	52600	67175	4241
混種語	40907	36789	23339	520	14227	501	10047	16030	1211
延べ比率 (ppm)									
和語	937190	919097	958080	998159	963672	998501	970855	949208	999081
漢語	50667	65767	34553	1756	29462	1298	26403	42977	619
混種語	12142	15134	7366	83	6864	199	2741	7814	299

主要品詞	合計	徒然	平家	宇治	方丈	新古	大鏡	更級	紫
名詞 異なり	21057	2507	9867	3611	658	1511	3079	957	1250
延べ	279715	8461	58954	22245	1306	9982	15546	3387	4583
平均出現頻度	13.28	3.37	5.97	6.16	1.98	6.60	5.04	3.53	3.66
動詞 異なり	10480	1246	2564	2277	335	793	1244	693	837
延べ	192236	5558	29558	18923	751	5372	8637	2495	2460
平均出現頻度	18.34	4.46	11.52	8.31	2.24	6.77	6.94	3.60	2.93
形容詞異なり	893	209	249	221	61	91	206	129	170
延べ	44379	1430	3478	2662	161	848	1679	601	766
平均出現頻度	49.69	6.84	13.96	12.04	2.63	9.31	8.15	4.65	4.50
形容動詞異なり	898	132	229	191	34	49	143	81	118
延べ	16057	474	1771	1103	60	258	592	216	302
平均出現頻度	17.88	3.59	7.73	5.77	1.76	5.26	4.13	2.66	2.55
副詞 異なり	597	119	207	186	49	78	116	77	78
延べ	35080	790	3876	2337	171	477	1880	385	513
平均出現頻度	58.76	6.63	18.72	12.56	3.48	6.11	16.20	5.00	6.57
異なり比率 (ppm)									
名詞	616062	591273	<u>749032</u>	552140	573170	594180	638929	490769	506482
動詞	306612	293867	<u>194640</u>	348165	291811	311836	258144	355384	339141
形容詞	26126	49292	18902	33792	53135	35784	42747	66153	68881
形容動詞	26272	31132	17384	29204	29616	19268	29674	41538	47811
副詞	17466	28066	15713	28440	42682	30672	24071	39487	31604
延べ 比率 (ppm)									
名詞	480634	494506	589004	452290	516818	581532	531432	467623	524610
動詞	330319	324839	295311	384746	297190	312962	295251	344470	281593
形容詞	76256	83576	34748	54124	63711	49402	57395	82976	87683
形容動詞	27590	27703	17693	22426	23743	15030	20237	29821	34569
副詞	60278	46171	38724	47516	67669	27789	64266	53154	58722
主要品詞	源氏	枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉
名詞 異なり	4847	2807	1689	1023	540	1096	920	579	3928
延べ	86482	14504	9107	6211	1738	5315	3409	2294	26191
平均出現頻度	17.84	5.16	5.39	6.07	3.21	4.84	3.70	3.96	6.66
動詞 異なり	5096	1818	1375	670	301	657	540	526	2046
延べ	68430	10775	8544	4217	1209	3490	2560	2039	17218
平均出現頻度	13.42	5.92	6.21	6.29	4.01	5.31	4.74	3.87	8.41
形容詞異なり	606	233	212	97	62	100	98	84	239
延べ	22534	3559	1777	828	195	585	340	283	2653
平均出現頻度	37.18	15.27	8.38	8.53	3.14	5.85	3.46	3.36	11.10
形容動詞異なり	585	233	157	42	22	39	49	45	37
延べ	8965	988	628	133	43	115	114	76	219
平均出現頻度	15.32	4.24	4.00	3.16	1.95	2.94	2.32	1.68	5.91
副詞 異なり	223	124	135	70	47	82	67	54	217
延べ	17231	2686	1902	327	177	256	310	254	1508
平均出現頻度	77.26	21.66	14.08	4.67	3.76	3.12	4.62	4.70	6.94
異なり比率 (ppm)									
名詞	<u>424579</u>	535074	469297	531981	548780	549924	543735	441984	595061
動詞	<u>446391</u>	346549	382050	348413	305894	329653	319148	401526	309953
形容詞	53083	44414	58905	50442	63008	50175	57919	64122	36206
形容動詞	51243	44414	43623	21840	22357	19568	28959	34351	5605
副詞	19533	23637	37510	36401	47764	41144	39598	41221	32873
延べ 比率 (ppm)									
名詞	416203	440797	406562	519531	497139	530809	491848	448134	523233
動詞	329326	327467	381428	352739	345823	348546	369355	398319	343974
形容詞	108447	108163	79330	69259	55778	58424	49054	55284	53000
形容動詞	43144	30026	28035	11125	12299	11485	16447	14846	4375
副詞	82925	81631	84910	27352	50629	25566	44726	49619	30126

古典対照語い表と異なる第3の特徴として、内容をエクセルファイルで作成したことがあります。いま書籍本文を紹介しましたが、エクセルファイルも、ロックを掛けたりせずに、付録としています。と言うよりも、エクセルファイルを印字したものが書籍であると理解するのが、よいように思います。

書籍で、1行1語となつて、種類の属性や出現頻度が並んでいるのは、実は、エクセルファイルの画面の体裁そのままです。古典対照語い表では1行に3語を詰め込みましたが、今回の新しい語彙表では贅沢をさせていただきました。

古典対照語い表には、データを電子化した経歴があります。

宮島達夫・中野洋・鈴木泰・石井久雄

『フロッピー版古典対照語い表および使用法』（1989年、笠間書院）そこでのデータは、今日で言えば .csv ですが、.dat の名を与えられて、使用者のプログラミングに委ねられました。.csv や .xls , .xlsx というファイル名によってエクセルが起動する今日とは、隔世の感があります。媒体も、直径5インチないし3.5インチのフロッピーディスクでした。このデータに新たなものを加えることによって、今回の日本古典対照分類語彙表が成立したことになります。

以下、エクセルファイルによって、二三のことを述べます。

3 出現頻度および類似度

ある用語が古典文学作品に現れるか、という問いに直ちに答える、言わば索引の索引として、古典対照語い表は広く用いられてきたのではないかと思います。現れるか否かは数字の有無として知ることができ、現れる多寡は数値の大小で読み取ることができます。

古典文学作品における用語の出現頻度を示す、という古典対照語い表の根幹は、新しい日本古典対照分類語彙表においても、変わりません。出現頻度についての基本的な集計を、二つ示します。語種についてのものと、主要な品詞についてのものとです。語種の集計が予稿6ページ、品詞の集計が7ページです。

和歌集で和語が圧倒的であるのは予想できますが、新古今和歌集でも異なり99.4%、延べ99.7%という数値は、予想し難いところでしょう。新しく加えた作品・平家物語の特徴が目立ちます。平家物語の語種は、他の作品に比べて、漢語の比率が異なり・延べでともに大きく、異なりで4割を超え、延べで2割を超えて、大雑把に言えば他の作品の2倍以上になります。漢語の異なり語数という見かたをしても、他の作品の4倍以上です。平家物語の品詞は、名詞が異なりで7割を超え、動詞が2割に届かない、などというのは、他の作品と対比して異常と言ってよいようなものです。源氏物語の品詞は、異なりの比率で見るときに名詞・動詞とも4割を上回り、その名詞は諸作品のうちで最小値となり、動詞は最大値となっています。またまた大雑把ですが、源氏物語の名詞の最小値の倍、動詞の最大値の半分が、平家物語の分量です。

こうした数値は、古典対照語い表の14作品を取り出すならば、かねて集計されていたところに一致するはずですが、必ずしも一致しません。意味分類を記述しながら、出現頻度を全面的に点検することになり、多少の修正は免れませんでした。地名・人名を分離したことは、すぐ上に言ったとおりであり、それだけでも異なり語数を変更することになります。

ところで、宮島先生は、統一した基準のもとで用語というものを調整して、出現頻度をなぜ計数したか。単純な索引の索引であったならば、用語の基準などというものは不要でした。用語を統一的に設定したのは、**作品間の用語の類似度**を算出するためです。古典対照語い表は、類似度を算出するための基礎作業でした。

日本古典対照分類語彙表でも、用語の基準に従って出現頻度を示しています。類似度は次のように算出されます。

	平家	宇治	方丈	新古	大鏡	更級	紫			
徒然	.440582	.535947	.428680	.312719	.466623	.462330	.435328	徒然		
	平家	.486336	.336297	▼268747	.451088	.377858	.354630	平家		
		宇治	.375082	.303263	.507575	.495847	.426384	宇治		
			方丈	.314857	.326667	.378315	.317321	方丈		
				新古	▼255818	.413942	.286286	新古		
					大鏡	.437159	.466398	大鏡		
						更級	.463930	更級		
							紫	紫		

	源氏	枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉	
徒然	.510674	.511751	.494277	.346852	.391718	.338271	.443317	.454215	.292227	徒然
平家	.405488	.399838	.383441	.277331	.309201	▼269322	.347684	.375330	▼271060	平家
宇治	.489242	.545748	.524694	.329349	.426018	.318774	.461403	.522739	.310485	宇治
方丈	.376336	.340913	.372456	.336771	.331794	.326055	.352490	.356161	▼267700	方丈
新古	.354508	.305369	.385068	▲637668	.331276	▲625481	.413282	.288514	.452296	新古
大鏡	.527175	.504855	.457965	▼274005	.342322	▼266026	.388797	.429895	▼253342	大鏡
更級	.523984	.534538	▲560961	.435252	.440149	.410496	.497909	.470323	.345315	更級
紫	▲569244	▲554268	.480596	.308550	.340376	.300440	.391163	.392664	▼252513	紫
	源氏	▲571628	▲590332	.377947	.369344	.353971	.447681	.455774	.293418	源氏
		枕	▲579150	.337144	.408761	.325587	.460558	.468350	.296235	枕
			蜻蛉	.424007	.452675	.404815	.507937	.471343	.337238	蜻蛉
				後撰	.371882	▲735076	.454290	.333727	.492621	後撰
					土左	.365714	.476234	.434419	.339487	土左
						古今	.462329	.318408	.521621	古今
							伊勢	.457624	.393001	伊勢
								竹取	.305506	竹取
									万葉	万葉

数値は全部で136組=17(作品全数)×16(対比作品数) / 2(重複) あって、特に大きいものまた小さいもの、それぞれ9個について、小数点を▲▼に置き換えてみました。大きいほうでは、.625以上の古今和歌集・後撰和歌集・新古今和歌集の一群と、.550以上の蜻蛉日記・枕草子・源氏物語を中心とする一群とが、目立ちます。小さいほうでは、いずれも.275未満の、万葉集の関係と大鏡の関係とが目立ちます。最大値は古今和歌集－後撰和歌集の.735076、最小値は紫式部日記－万葉集の.252513でした。

なお、いま、大小9個ずつを選びましたが、9個という数に特段の意味はありません。表に印を付けてみて、あまり入り組まないところで止めたというに過ぎません。誌面の都合で表を2段に分けてあって、下のほうが大きくなっている一因は、数値が大きいもの一群がよく見えるように、源氏物語の縦の列を下に移したことにあります。大きいほうの数値、小さいほうの数値は、結果的にそうなったというもので、切り出すのに先立って数値を定めたというものではありません。数値を解析して切り出すという方法も、当然、あると思います。

集計ということは、合計だけでも筆算では大変でしょう。宮島先生の元の古典対照語い表は、数表を作るなどの工夫を加えつつ、類似度の算出までを筆算で進める、という大作業の成果です。出現頻度を一覧することができても、そこから新たな数値を算出するのは大変です。今回の新たな日本古典対照分類語彙表では、一語一語について、出現した作品の数を示し、そこにおける出現頻度の合計も示しています。今はコンピュータの時代であり、エクセルを使うならば、それぞれ、最初に覚える関数の使用例となるようなものです。それにもかかわらず、出現作品数・頻度合計を示したのには、大野晋さんの編著『古典基礎語辞典』（2011年、角川学芸出版）「執筆者序」に次のように記されることに関係しています。「先生」は大野さんです。

平成10年頃から先生が病気がちになられたこともあり、辞書の完成のために、先生は計画の変更を決断された。すなわち、収録項目数を大幅に減らし、基礎語の辞書を作成するという方向転換である。項目数は大幅減となるが、個々の項目については必要なだけの分量を確保することは変わらない。具体的な基礎語としては、『古典対照語い表』（笠間書院刊）において41例以上の用例のある1353語を中核とし、隣接する語や同一範疇の語とのバランスなどを考慮して語を増やしていった。

「41例以上」というのは14作品の出現頻度の合計値です。大野さんたちはフロッピー版をコンピュータで処理したと推測しますが、宮島先生は、親しい先輩である大野さんのこの作業を知って、集計値を当初から示すことにも意義があると承知したとのことでした。

4 語彙表の組み換え

日本古典対照分類語彙表は、繰り返しますが、エクセルファイルで、ロックもせずに、提供されます。すなわち、目的によってデータを組み換えることができます。行や列を移動したり追加したりすることもできます。出現頻度を比率などに入れ換えることもできます。この表を土台として、用語を形態素に分割した対照形態素表を作ったり、作品を加えて新たに大きな古典語彙表を作ったりするというのも、よろしいのではないのでしょうか。そうして組み変えた例も併せて幾つか提供しますので、紹介します。

かつて国立国語研究所が続けてきた語彙の計量調査では、結果が、五十音順および出現頻度順の二つの語彙表として示されました。さらに意味分類順語彙表も構想されて、実際、

国立国語研究所『国立国語研究所報告 13 総合雑誌の用語 後編』

(1958年、秀英出版)

にありますし、

国立国語研究所『国立国語研究所報告 21 現代雑誌九十種の用語用字

第一分冊 総記および語彙表』（1962年、秀英出版）

では、五十音順語彙表を第1表として、次のように予告されました。

語を意味によって分類した表を、別の分冊に掲げるが、第1表に見出しとして掲げている語の範囲では、語が意味分類語彙表のどの分類項目の所にあげてあるかを知るために、第1表が利用出来る。すなわち知ろうとする語を見出しに持つ所の、左端の「意味分類」の欄に、その分類項目番号がしるしてある。（この意味分類表は報告

13の§4に掲げた表に相当するが、分類番号の与え方は一部改めた。)

なおこの方法で意味分類語彙表のその分類項目の所を見ることによって、その語の類義語・対義語の大体を知ることが出来る。(p.35)

ただし、ここからイメージされる形では実現されず、他の語彙を併せて

国立国語研究所『国立国語研究所資料集 6 分類語彙表』(1964年、秀英出版)にまとめられることになりました。いずれにせよ、分量が大きくなることは避けられ、例えば出現頻度順語彙表では用語一つ一つの詳細は示さずに五十音順語彙表を参照させる、というようなことになっていました。

さて、日本古典対照分類語彙表でも、書籍としては五十音順語彙表のみであり、これと同じ大きさの書籍がさらに2冊できるというのも、おおかたにとって迷惑なことでしょう。しかし、コンピュータファイルとしてならば、作成も保存も携行も容易にできます。できました。五十音順語彙表がエクセルファイルとしてあり、その行を並べ替えただけの操作で、出現頻度順語彙表も意味分類順語彙表もエクセルファイルとして、CDに収めています。出現頻度順のものは全作品合計により、意味分類順のものは、意味分類が幾つかあるばあいには、最も若い分類項目番号によっています。それぞれの最初の左半分程度を印刷してみます。出現頻度順が予稿12ページ、意味分類順が13ページです。意味分類順は、意味分類の列を印刷していませんが、この範囲ではすべて「1.1000(事柄)」です。

国立国語研究所の語彙計量調査は、現代雑誌九十種を終えたあと、コンピュータ処理の時代に入り、意味分類を記すことはなくなりました。現代雑誌九十種までの語彙表3点の、大袈裟に言うならば悲願は、今回の日本古典対照分類語彙表によって遂げられることになります。

なお、仮名見出しを右から左に反転させて五十音順に配列した語彙表も、できています。語形の反転は、こちらにいらっしゃる会長・田島毓堂先生が夙に実現なさっています。

田島毓堂・丹羽一彌『日本語尾音索引 現代語篇』(1978年、笠間索引叢刊 65)

田島毓堂・丹羽一彌『日本語尾音索引 古語篇』(1979年、笠間索引叢刊 73)

この書の意義については、宮島先生に紹介があり、ここに両先生のお名前を並べるのも縁でしょう。

宮島達夫「田島毓堂・丹羽一弥共編『日本語尾音索引——現代語編』

風間力三編『綴字逆順排列語構成による大言海分類語彙』」

(1979年、国語学 118)

内容は同じでも数値に別の見かたをすることはできて、例えば出現頻度を各作品における順位に置き換えた語彙表もできます。全作品の合計頻度にも順位を付けることができますので、順位で整列したものを予稿14~15ページに示します。数値のほかは、出現頻度順語彙表と全く重なるものですから、列を移動したり追加したりすることも、してみました。見出しの五十音順の配列を保存したり、かな反転用の順・出現頻度順・意味分類順の語彙表を作成したときの配列を保存したりするために、列を追加しました。一語に与えられた意味分類の数も、記しました。用語の属性を左にまとめ、出現頻度の関係を右にまとめることも、しました。動詞「あり」は、全作品合計で首位であると同時に、作品の半数においても首位であったと直ちに理解できます。副詞「いと」は作品による癖が見えます。

見出し	順	漢字	種	品詞	注記	作	合計	徒然	平家	宇治	方丈	新古	大鏡	更級	
あり	4	有		動ラ変	「あら	17	12004	490	1490	1214	51	207	397	169	
す	6	為		動サ変	「せん	17	9778	360	1857	1117	47	202	455	128	
こと	2	事			「こと	17	9749	538	1141	849	45	57	766	98	
ひと						17	8869	572	575	691	46	249	326	134	
なし	3	無		形	「心な	17	7381	305	817	557	44	220	221	104	
いと	2	甚		副		17	6323	69	11	102	2	5	362	95	
いふ	1	言		動四	「いは	17	6267	354	728	1409	17	48	223	120	
おもふ		思		動四		17	5989	129	479	559	4	206	155	74	
もの		物	・者		「もの	17	5612	263	624	768	25	128	195	52	
ころ		心			「心憂	17	5598	168	345	170	29	158	129	49	
みる	2	見		動上一		17	5299	111	421	548	11	218	163	97	
この		此		連体		17	4547	108	594	901	13	27	506	67	
ほど		程				17	3706	80	382	445	12	44	200	56	
なる	1	成	・為	動四		17	3014	77	540	296	13	66	233	49	
よ	1	世			「よに	17	2802	96	267	91	21	172	179	37	
その	3	其		連体		17	2700	132	775	522	27	17	219	52	
きみ	1	君				16	2572	9	194	32		78	135	4	
さり	3	然	有	動ラ変	「さら	16	2378	57	440	217	4	14	217	21	
きこゆ		聞		動下二		17	2373	21	268	17	3	9	96	16	
いま	1	今				17	2284	29	340	286	11	66	155	28	
みゆ		見		動下二		17	2279	60	237	143	4	82	73	45	
まうす		申		動四	「まを	15	2277	104	1073	296	1		405	9	
わが		我		連体		17	2148	35	192	157	7	113	40	18	
おぼす		思		動四		11	2068	3	31	26			92	3	
しる	2	知	・領	動四		17	2066	127	182	180	21	115	58	25	
また	2	又		副		17	2065	80	251	189	21	43	230	25	
これ		此				17	2060	73	783	351	21	11	98	24	
やう		様	漢			13	1978	44	74	466	2		106	57	
いか	3	如何		形動	「いか	17	1955	41	373	257	4	72	90	22	
かく	11	斯		副	「かく	17	1944	37	76	198	9	13	111	9	
さぶらふ		侍	・候	動四		11	1940	45	925	295			155	12	
かた	1	方				17	1938	30	200	141	4	23	58	27	
まゐる		参		動四		12	1923	30	378	223			200	23	
さま	1	様				16	1919	38	48	39	8	1	46	19	
のたまふ		宣		動四		11	1888	8	343	93			74	3	
はべり		侍		動ラ変		11	1774	55		100	9		271	6	
み	3	身			「身の	17	1761	59	280	96	19	152	29	12	
みや		宮			「野の	16	1729	3	152	25		8	228	18	
ところ	1	所				17	1714	83	331	183	14	2	49	52	
きく	3	聞		動四		17	1672	39	187	139	7	60	82	35	
ただ	3	唯	・直	・行	副	17	1646	40	209	177	10	20	78	20	
いみじ		甚		形		12	1620	41	2	164	1		188	82	
おぼゆ		覚		動下二		14	1618	65	197	122	3	4	107	37	
あはれ		哀		形動	名詞	・	17	1606	37	132	55	1	62	66	44
とき	1	時				17	1601	98	384	207	18	25	127	12	
われ	1	我			「われ	17	1533	19	203	173	2	56	49	18	
なほ		尚		副		17	1484	29	106	44	5	57	48	11	
よし	2	良		形		17	1365	109	82	171	2	4	77	10	
うち	1	内	・内裏			17	1342	45	255	81	7	11	73	22	
さ	2	然		副	「さし	15	1333	35	183	71	3	27	94	18	
ひとびと		人	人			13	1306	6	186	95			82	37	
ひ	1	日				17	1293	30	257	84	4	31	83	19	
はな	1	花				17	1282	22	61	21	2	162	19	20	
こち		心	地			15	1225	16	33	33		8	38	23	
ゆく		行		動四		17	1220	22	107	189	2	80	23	16	
いづ	4	出		動下二		17	1205	14	204	124	5	50	60	17	
うへ	1	上			「身の	16	1201	21	223	73		36	41	12	
やま		山				17	1200	12	120	71	6	134	18	36	

見出し	順	漢字	種詞	注	作	合計	徒然	平家	宇治	与丈	新古	大鏡	尾級	紫源氏	枕		
あまりこと		余事				1	1								1		
いちじ	1	一事	漢			3	14	6	7			1					
いちだいじ		一大事	漢			1	1	1									
いなびどころ		否所				1	1							1			
いはひごと		祝事				1	2							2			
うひごと	1	初事				2	2							1			
うみ	1	有為	漢			1	1	1									
えうじ		要事	漢			2	2	1	1								
かうなこと						1	1							1			
かたぎま		方様				6	46		3	6		5	1	30			
きどく		奇特	漢			3	6	1	4	1							
くだん		件				3	38	26	11						1		
くま	1	隅				10	94	3	5		4	3	5	63	2		
ここ		此处				17	615	11	91	76	2	3	9	6	4	251	22
こと	2	事		「こ		17	9749	538	1141	849	45	57	766	98	134	4804	449
ことがら		事柄				2	4	1	3								
ことごと	1	異事				7	44			7		4	2	1	24	5	
ざふじ		雑事	漢			3	5	1	1						3		
しさい	1	子細	漢			3	40	35	4						1		
しようし	1	勝事	漢			1	3	3									
しよじ	1	諸事	漢			2	2	1	1								
しれごと	2	痴事				1	1	1									
すずろごと	2	漫事				1	3								3		
すぢ		筋				10	253	1	1	2	4		3	225	8		
すぢすぢ		筋筋				1	1							1			
せうじ		小事	漢			1	1	1									
そうべつ		惣別	漢			1	1	1									
そこ	1	其処				12	198	32	39	4	8	16	1	24	18		
そぞろごと	2	漫事				2	5	4					1				
だいじ	2	大事	漢			6	89	12	43	15		6		11	2		
だいせうじ		大小事	漢			2	5	4				1					
たじ	2	他事	漢			2	11	6	5								
ただごと	2	只事				4	11	6	3	1							
でう		条	漢			1	14	14									
ところ	1	所				17	1714	83	331	183	14	2	49	52	20	598	148
とをこと		十事				1	1	1									
なにごと		何事				15	503	24	62	49	4	24	3	2	269	29	
なにわぎ		何業				7	22	8	1		1			7	2		
なほあらしごと		猶事				1	1							1			
なほざりごと		等閑事				1	10							10			
にはかごと		俄事				1	1	1									
はかなごと		果無事				1	1							1			
はかなしごと		果無事				1	7							7			
ばんじ		万事	漢			2	6	5	1								
ひとこと	2	一事				5	13	1	6			1		4	1		
ひとごと	2	人事				2	2	1						1			
ひとつこと		一事				1	1							1			
ひとふし		一節				6	20	1			1		2	14	1		
ふしぶし		節節				4	11	1				1	1	8			
ふるごと	2	古事				2	13					1		12			
まさなごと		正無事				1	1	1									
まめごと		忠実事				2	5							4	1		
みごと		見事				1	1	1									
もの		物・者		「も		17	5612	263	624	768	25	128	195	52	80	1942	535
やくなしごと		益無事	混			1	1							1			
やすだいじ		易大事	混			1	1	1									
ようなしごと		用無事	混			1	1								1		
よし	1	由				17	613	19	202	89	1	2	20	9	6	160	10

仮名順	反転順	頻度	分類順	見出し	順	漢字	種	品詞	分	意味分類	注記	作	計
1471	27338	1	21026	あり	4	有		動ラ変	4	21200(存在)/21584(限定)	17	1	
16418	11293	2	21281	す	6	為		動サ変	5	21211(発生・復活)/21220	17	2	
11990	17140	3	15	こと	2	事			2	11000(事柄)/13800(事業)	17	3	
25815	17433	4	3707	ひと		人			2	12000(人間)/12020(自他)	17	4	
22393	10384	5	31614	なし	3	無		形	1	31200(存在)	17	5	
2603	16996	6	32695	いと	2	甚		副	1	31920(程度)	17	6	
3031	20688	7	25912	いふ	1	言		動四	4	23031(声)/23100(言語活)	17	7	
6580	21936	8	25498	おもふ		思		動四	3	23020(好悪・愛憎)/23030	17	8	
30647	19181	9	54	もの		物	・者		4	11000(事柄)/12000(人間)	17	9	
11564	31357	10	532	こころ		心			6	11302(趣・調子)/11742(□)	17	10	
29797	30530	11	26770	みる	2	見		動上一	7	23065(研究・試験・調査)	17	11	
12178	18870	12	31469	この		此		連体	1	31010(こそあど・他)	17	12	
27801	17624	13	540	ほど		程			6	11302(趣・調子)/11611(≠)	17	13	
22942	29820	14	21305	なる	1	成	・為	動四	4	21220(成立)/21500(作用)	17	14	
31993	26478	15	1339	よ	1	世			5	11611(時機・時刻)/11621	17	15	
17774	18891	16	31477	その	3	其		連体	1	31010(こそあど・他)	17	16	
9423	23888	17	3623	きみ	1	君			3	12000(人間)/12010(われ)	16	17	
13811	27568	18	21070	さり	3	然有		動ラ変	5	21200(存在)/31010(こそあど)	16	18	
9144	26269	19	27320	きこゆ		聞		動下二	3	23093(聞く・味わう)/231	17	19	
3107	23047	20	2334	いま	1	今			1	11641(現在)	17	20	
29781	26463	21	22975	みゆ		見		動下二	5	21527(往復)/23061(思考)	17	21	
28138	11405	22	27640	まうす		申		動四	3	23100(言語活動)/23102(≠)	15	22	
33119	3777	23	31533	わが		我		連体	1	31040(本体・代理)	17	23	
5998	12764	24	25470	おぼす		思		動四	3	23020(好悪・愛憎)/23030	11	24	
16131	29215	25	26282	しる	2	知	・領	動四	6	23050(学習・習慣・記憶)	17	25	
28471	14287	26	31501	また	2	又		副	4	31010(こそあど・他)/316	17	26	
12598	31142	27	72	これ		此			4	11010(こそあど・他)/116	17	27	
31040	1910	28	160	やう		様	漢		4	11100(類・例)/11300(様)	13	28	
1746	3234	29	31447	いか	3	如何		形動	1	31010(こそあど・他)	17	29	
7260	5252	30	31462	かく	11	斯		副	1	31010(こそあど・他)	17	30	
13698	22300	31	21068	さぶらふ		侍	・候	動四	2	21200(存在)/21527(往復)	11	31	
7629	13862	32	1303	かた	1	方			6	11611(時機・時刻)/11700	17	32	
28960	30940	33	22973	まゐる		参		動四	6	21527(往復)/23331(食生活)	12	33	
13721	23138	34	505	さま	1	様			2	11300(様相・情勢)/13081	16	34	
23956	21905	35	27596	のたまふ		宣		動四	1	23100(言語活動)	11	35	
25031	28082	36	21078	はべり		侍		動ラ変	1	21200(存在)	11	36	
28991	23736	37	122	み	3	身			5	11040(本体・代理)/12020	17	37	
29682	26150	38	5063	みや		宮			3	12320(君主)/12630(社寺)	16	38	
21279	31417	39	35	ところ	1	所			3	11000(事柄)/11690(場合)	17	39	
8994	5391	40	27304	きく	3	聞		動四	3	23093(聞く・味わう)/231	17	40	
18603	14152	41	31545	ただ	3	唯	・直・行	副	5	31110(関係)/31130(異同)	17	41	
3166	10875	42	31794	いみじ		甚		形	2	31332(良不良・適不適)/3	12	42	
6210	26454	43	21001	おぼゆ		覚		動下二	5	21130(異同・類似)/23001	14	43	
906	31206	44	11609	あはれ		哀		形動	6	13020(好悪・愛憎)名詞	・17	44	
21137	4410	45	1264	とき	1	時			4	11600(時間)/11611(時機)	17	45	
33489	31248	46	3817	われ	1	我			2	12010(われ・なれ・かれ)	17	46	
22635	23014	47	32399	なほ		尚		副	3	31612(毎日・毎度)/31920	17	47	
32139	10971	48	31838	よし	2	良		形	1	31332(良不良・適不適)	17	48	
3701	14416	49	2473	うち	1	内	・内裏		6	11652(途中・盛り)/11720	17	49	
12764	8659	50	31470	さ	2	然		副	1	31010(こそあど・他)	15	50	
25989	17511	51	3710	ひとびと		人人			2	12000(人間)/12010(われ)	13	51	
25305	19817	52	2124	ひ	1	日			2	11633(日)/15210(天体)	17	52	
24814	17953	53	726	はな	1	花			4	11345(美醜)/14360(薬剤)	17	53	
11540	14629	54	531	こち		心地			4	11302(趣・調子)/13000(≠)	15	54	
31742	6951	55	23023	ゆく		行		動四	3	21527(往復)/21600(時間)	17	55	
2364	15199	56	21112	いづ	4	出		動下二	2	21210(出沒)/21531(出・)	17	56	
4431	22539	57	2858	うへ	1	上			6	11741(上下)/11750(面・)	16	57	
31381	23467	58	13622	やま		山			3	13333(住生活)/14700(地)	17	58	

徒然	平家	宇治	方丈	新古	大鏡	更級	紫	源氏	枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉	見出
3	2	2	1	6	6	1	5	2	1	1	1	2	2	1	2	1	あり
4	1	3	2	8	4	3	3	7	5	2	7	4	8	7	4	4	す
2	3	5	4	51	1	6	4	1	7	7	15	11	25	14	3	65	こと
1	12	7	3	2	8	2	1	4	4	5	4	3	1	3	5	9	ひと
6	6	9	5	3	14	5	6	6	10	8	3	15	5	10	11	7	なし
24	1428	62	204	557	7	8	2	3	3	6	246	28	403	12	20	889	いと
5	9	1	18	68	13	4	10	17	2	3	16	1	24	6	1	19	いふ
10	14	8	93	7	27	10	11	8	11	4	5	19	3	8	14	6	おも
7	10	6	8	16	20	15	9	9	6	13	6	10	9	15	7	20	もの
8	23	37	6	12	31	18	8	5	19	14	8	20	10	15	18	15	ここ
12	16	10	24	4	25	7	15	11	13	11	2	12	4	11	9	2	みる
14	11	4	21	119	2	11	38	14	36	25	123	5	132	13	8	25	この
19	19	13	23	73	18	13	11	12	12	9	64	150	286	59	39	2246	ほど
22	13	15	21	41	10	18	42	26	22	10	18	32	19	17	18	46	なる
17	33	74	10	9	22	23	28	15	87	52	29	471	28	48	53	120	よ
9	8	11	7	193	16	15	28	49	35	55	1039	49	666	18	29	78	その
267	52	236		32	30	318	14	24	56	92	13	150	16	32	117	3	きみ
29	15	21	93	227	17	44	26	22	18	42	1039	192	666	35	53		さり
102	32	432	131	325	45	71	79	13	40	27	128	192	252	115	601	341	きこ
64	25	18	24	41	27	31	113	29	59	28	32	35	31	39	65	31	いま
27	42	43	93	26	64	20	17	28	15	21	25	38	26	36	65	29	みゆ
15	4	15	367		5	133	682	170	49	301		268	666	78	10	386	まう
48	53	39	44	20	105	55	60	61	134	48	9	60	7	32	72	4	わが
764	496	286			48	399	120	10	286	166				768	39		おぼ
11	59	32	10	19	76	35	79	48	28	40	14	35	15	26	32	22	しる
19	37	27	10	75	11	35	35	39	14	15	246	23	222	65	140	129	また
23	7	14	10	282	42	38	101	79	30	33	304	24	286	49	17	616	これ
36	187	12	204		37	12	25	27	37	24		29		153	16		やう
38	21	19	93	36	49	41	46	39	32	19	91	192	116	232	140	123	いか
45	183	24	27	245	33	133	79	23	151	23	116	29	95	49	23	38	かく
34	5	17			27	97	44	74	63	156				153	287		さぶ
59	50	44	93	146	76	32	28	18	54	31	92	119	85	65	79	758	かた
59	20	20			18	39	15	32	20	67				271	53	3287	まみ
44	308	190	33	1433	98	50	20	16	65	59	562	268	666	494	140		さま
310	24	72			63	399	94	19	33	63				351	12		のた
30		63	27		9	204	7	20	52	76				768	32		はべ
28	31	68	15	14	149	97	68	44	199	41	10	268	18	39	79	172	み
764	74	298		363	12	55	17	21	72	247	562	471	503	51	601	188	みや
18	26	30	20	1017	94	15	55	47	23	18	562	9	403	28	41	3287	とこ
43	54	46	44	47	56	28	101	55	31	16	26	32	54	39	21	48	きく
40	45	33	26	164	60	47	35	35	21	44	221	78	222	153	95	126	ただ
38	5032	38	367		21	9	50	43	9	44				192	95		いみ
25	51	54	131	674	36	23	55	33	23	22	720			109	72		おぼ
45	93	130	367	45	69	22	85	25	48	30	146	119	85	99	236	889	あは
16	18	22	17	127	32	97	94	133	157	55	44	38	27	23	34	23	とき
111	48	34	204	57	94	55	79	80	128	43	19	78	17	30	86	17	われ
64	127	171	75	51	96	111	137	30	23	35	96	26	132	55	60	326	なほ
13	167	36	204	674	61	117	34	58	17	76	720	47	332	78	29	112	よし
34	36	87	44	282	64	41	31	45	72	66	395	68	222	153	95	215	うち
48	58	107	131	119	46	55	50	46	16	37	720			666	494	117	さ
417	55	69			56	23	21	39	78	92		26	962	130	26		ひと
59	35	83	93	102	54	50	79	124	49	17	128	20	64	109	164	27	ひ
95	228	342	204	11	239	47	372	108	78	145	11	96	6	44	287	21	はな
145	471	232		363	108	39	43	31	62	20	221	150	286	351	65		ここ
95	125	27	204	28	201	71	205	173	259	75	31	22	32	24	164	14	ゆく
173	47	52	75	62	70	61	79	81	59	44	116	42	50	36	86	69	いづ
102	44	101		93	104	97	19	73	26	115	99	78	110	89	86	51	うへ
198	106	107	61	15	254	27	372	219	184	115	17	60	13	56	26	13	やま

出現頻度を順位に置き換えるなどということをしていると、つい、次のようなものも出してしまい、ここに示すことはしませんが、CDには添えてあります。

- ・ 17作品のいずれでも100位以内にある用語。抽出したところ、13語でした。
- ・ 　　　　　　いずれかで20位以内にある。88語。
- ・ 17作品のいずれでも上位10%に入っている。250語。
- ・ 　　　　　　いずれかで上位1%に入っている。21語。

数値を扱ってのエピソードを一つ記します。よく検算をします。何かの拍子で、作品数の合計を出してしまいました。作品数の合計というものの意味は理解しかねましたが、ともかく数値を得てしまいましたので、検算ができないかと習性で考えます。各作品の異なり語数の合計が合致するはずであると直ちに気づき、実行して事無きを得るに至りました。しかし、異なり語数を加算するなどは、相互に相容れない語種のような語群で行って全体の異なり語数とする、といったばあいしか、想像できません。暫くして、出現頻度が入っているセルの総数であると了解しました。その数値、37,818です。全セル581,060個＝17列(作品)×34,180行(語)の6%弱、1行当たり1.1個に過ぎないと言えば、それまでですが、宮島先生がこれだけの数値を書き、あるいはコンピュータで打ち込んだということでありまして、そのことに思い到ったとき、私は感動を覚えました。これだけの分量を、黙々と一人でこなしただのか。その数値の合計581,970の用例を数えて調整したのか。

なお、言い添えます。日本古典対照分類語彙表は、宮島先生に3名の名を並列させて共編の形をとっていますが、実質は宮島先生のものであります。旧版古典対照語彙表は宮島先生の独創ですし、これまでの中間報告は宮島先生が続けてきましたし、言わずもがなではありますが、明言して残したく思います。今回の万葉集の索引の交替についても、用語の出現頻度を索引に一一書き込み、用語の基準によって変更する際には記録し、従来の作品の語彙表に統合することを図りつつ、コンピュータで数字を入力し、点検し、といった一連の作業は、宮島先生の独力で成し遂げられました。意味分類を付ける着想も宮島先生のもので、旧版分類語彙表で作業を開始したときには諸方面から協力していただきましたが、分類語彙表の増補改訂版が公刊されて全面的に分類し直すこととしたときから、宮島先生に作業が集中するようになり、全体にわたる最終的な調整までそうでした。他の3名は、時折、宮島先生の話聞いてきただけのようなもので、それがなぜ共編であるかと人に問われたとしても、宮島先生の意向であるからとしか答えることができません。

5 語彙表の展開

日本古典対照分類語彙表から分類語彙表のようなものを作ることは、本書刊行に先立って、次で公にしています。出現頻度などは記さず、用語を列記したにとどまります。

宮島達夫・鈴木泰・石井久雄・安部清哉・前田直子

『研究報告 古典分類語彙表(稿)』

(2013年、学習院大学計算機センター年報33 pp.40-121)

分類語彙表と異なり、一つの分類項目のうちで語句の意味関係を見て配列することをしていません。この改訂版もCDに入れはしましたが、その点は同じです。

さて、出現頻度を示しつつ分類語彙表の分類項目で配列した語彙表を、嚮の意味分類順

語彙表とは別に作成しました。分類項目の内では配列を意味によらずに五十音順としていますが、その古典の用語に、現代の分類語彙表増補改訂版の語句も並記しました。分類語彙表の語句の配列も五十音順とする暴挙に出ましたが、分類語彙表を写したことによって、一項目に集まった古典の用語がどのような一群であるかということ、感得しやすくなったと思います。現代の語句がどの分類項目に配置されているかは、分類語彙表の五十音順索引から求めることになりますが、現代の語句に対応する古典の語句を求めるという用いかたをすることができます。古典で出現する作品も、同時に知ることができます。

1.1030(真偽・是非)	17
うつつ(現) *	徒 1 平17 字 3	新10 大 5 更 2 源19 枕 4 蜻 1 後15 古11 伊 4 万18
うつつさま(現様)		源 1
しじやう<至誠>	平 1	
じち<実>		大 1 源 7
じつぶ<実否>	平 2	
しんじつ<真実> *	平 3	大 1
しんによ<真如>	平 1	
しんり<真理>	平 1	
ひ<非> *	徒 5 平 1	
ふしぎ<不思議>	徒 3 平37 字10 方 1	大 3
ふしぎ<不思議>	平 4	
ほつしやう<法性>	平 2	
ほんたい<本体>	字 1	大 1
まこと<真・誠> *	徒37 122 字93 方 2 新 5 大59 更 4 紫 6 233 枕78 蜻 5 後 2 土 2 古 1 伊 2 竹 8 万20	
ゆめうつつ<夢現> *	平 1	源 1 後 1 古 1
りひ<理非>	平 1	大 1
をつづ(現)		万 4
[[分類語彙表]] ありのまま あるがまま 一場の夢 イレギュラー オカルト 仮性 仮想現実 可否 可不可 偽(ぎ) 虚(きよ)		
虚偽 曲直 虚実 空中楼阁 黒白(くろしろ) 下(げ) 現実 現実性 現実味 公式 黒白(こくびやく) 虚仮(こけ) 史実 事実		
実(じつ) 実際 実態 実地 実のところ 実否 邪 邪曲 順逆 情実 正真正銘 上(じょう) 正銘 白黒(しろくろ) 真(しん)		
真価 真贋 真偽 真実 真実性 真実味 真性 真如 真否 神秘 真面目(しんめんもく・しんめんぼく) 真理 素(す)[~の顔] 是		
正(せい) 正格 正邪 正則 正否 是非 是非 善悪 尊厳 試し 当 当然 当否 内実 七不思議 生(なま)[~の声]		
バーチャルリアリティー 背理 迫真[~の演技] 非 否 不可 複雑怪奇 不思議 不正 不偏 不良 別格 変格 変則 本格 本式		
本然 本当 ほんま 本来 真(ま)[~に受ける] 誠 まとも ミステリー 夢物語 よしあし リアリティー 理非 略式 良否		
レギュラー		

1.1040(本体・代理)	22
かたしろ<形代> *		源 3
かはり<変・代> *	平 2 字11	大 1 更 1 紫 1 源26 竹 1
きさきがね<后>		大 1 源 2
こともの<異物>	字 1	大 1 更 1 源 4 枕 2 土 1 竹 1
しやうたい<正体>	平 1 字 1	
じよ<自余>	平 4	大 1
しろ<代>		万 1
しんたい<神体>	平 1	
たく<他> *	徒 7 平 4 字 2 方 4	大 1
ためしく<例・試> *	徒 5 平15 字 1 方 2 新 5 大 3 更 2 紫 2 源87 枕 3 蜻 2 後 2 古 1	
どう<胴> *	平 1	大 1
とりかへ<取替>		源 1
ぼうがね<坊>		源 1
ははしろ<母代>	平 1 字 1	大 1
ひとがたく<人形> *		源 9
ほか<外> *	徒30 平91 字44 方 6 新14 大34 更13 紫 6 192 枕13 蜻11 後12 古 4 伊 3 竹 1 万 4	
まめやかもの<実物>	字 1	
み<身> *	徒59 280 字96 方19 152 大29 更12 紫18 685 枕26 蜻77 140 土 2 古80 伊25 竹11 万50 伊 2	
むこがね<婿>		
むね<宗>	徒 3 平 4 字 3	
よ<余>	平 4 字 1	枕 1
れい<例> *	徒 3 平36 字37	大28 更 5 紫35 539 枕39 106 土 4 伊 5 竹 2
[[分類語彙表]] 亜流 得得 親代わり 替え玉 掛け替え 肩代わり 仮 借り物 かわり(代・替) 贗造物 贗物(がんぶつ) 偽(ぎ)		
偽造物 機体 偽物(ぎぶつ) 客体 好材料 後身 候補 個体 材料 サブ サンプル 自 自身 自他 自体 実(じつ) 実体 実物		
車体 主(しゅ) 従 主体 主流 正体 庶流 支流 資料 史料 神体 人体 正(せい) 正副 先蹤 前身 船体 選択肢 先例		
前例 素材 外枠 その他 その外 そのもの それ自体 他(た) 例(たい) 代替(だいたい) 代表 体用(たいよう・たいゆう) 代理		
他事 ダミー ためし 単体 嫡流 データ データベース 度外 二次 偽(にせ) 偽物(にせもの) 偽者(にせもの) 媒体		
バックアップ 張り子のとら 判例 控え 被写体 ひな型 標本 ファイル 不安材料 部外 副 複製 複体 別扱い 別立て 別物		
補外(ほか) 補欠 補助 ボディー 本質 本筋 本体 本人 本物 本流 末流 マテリアル 間に合わせ 見返り 身代わり 見本		
名代(みょうだい) 名実 メーン 模型 模式 世のためし 例 レプリカ		

この表に立ち入るつもりは、すみませんが、今はありません。ただ、見てしまって分からないというのにもなりますので、一言加えます。分類項目の番号・名称のずつと右、行末の数字は、その意味分類で出現したと認められる用語の数です。ここでは17語および22語であり、その用語について、出現した作品と頻度とが下に掲げられます。見出しのところどころに * が付いているのは、出現のすべてがその意味分類のものであるのではなく、頻度には他の意味分類のものも併せられている、という用語です。出現頻度の直前の漢字は作品名であり、縦罫線を代えたようなものです。一般に、表で列の上軸を読むために視線を上下動させるのが結構ストレスになりますので、それを避ける試みです。左右方向を読むために、今回の書籍のほうでは行の網掛けを1行おきよりは3行おき程度にしています、そのような工夫をこの表ではしなかったことを遺憾とします。

この、現代語彙を並記した意味分類語彙表では、意味分類を幾つか与えた用語は、意味分類ごとに繰り返して掲げています。部分部分で印刷されることをイメージして作成し、項目ごとにページを改めるようなこともしたために、膨らんで900ページを超えるものになっています。ちなみに、CDに収めてある種類の語彙表を、書籍と同じ体裁で印刷すると、すべてで15冊分くらいにはなるのではないかと思います。

上に名前だけを挙げた古典分類語彙表や、この現代語彙並記の意味分類語彙表のように、日本古典対照分類語彙表のうちで、あるいは他の語彙表と結合させて、新たな語彙表を構築することができます。そのような展開がなされることを念じて、日本古典対照分類語彙表はデータを提供しています。

なお、日本古典対照分類語彙表には別冊付録があって、解説を二つ載せています。

- ・ 宮島 達夫「古典語の統計と意味」13ページ
- ・ 小木曾智信「Excel による『日本古典対照分類語彙表』の活用」15ページ

宮島先生のは、いまここで粗あら述べたような統計を示し、さらに次のような統計も示しています。

作品数の分布 意味数の分布 語数が多い意味分類項目

全作品に共通に出現する用語

各作品出現頻度上位20語 Kilgarriff方式による各作品の特徴語20語

宮島先生の解説は、また、意味分類の様相について幾つかを述べています。日本古典対照分類語彙表の特に力を注いだ点に関係し、熱意を感じさせられます。

意味分類の方針および認定 分類語彙表の問題点

意味分類項目間の対比 意味分類の作品間の対比 現代語彙との対比

意味分類について考察したものとして宮島先生には次のものがあり、同様の詳細緻密な考察が期待されます。

『国立国語研究所報告89 雑誌用語の変遷』（1987年、秀英出版）

小木曾さんの解説は、特にお願いして、エクセルの入門書として執筆していただきました。日本古典対照分類語彙表でエクセルを使い始めるといった人を念頭に置いて、

データを検索・選択する

ピボットテーブルを用いて集計する

ということを、コンピュータの画面の写真も示しながら、丁寧に説明しています。

6 数値の利用

日本古典対照分類語彙表に見られる数値をめぐって、さまざまな研究ができそうに思います。その一例を試みます。この語彙表の出現頻度は著しい特徴をもって、すなわち、現実の全数の調査の結果です。数値が推計の幅を伴うものでなく、全数を推計させるものでもなく、種々の推計を基礎づける素材となるものであると考えられます。その語彙の出現頻度をめぐって、異なり語数の半数が頻度1であるという問題を聊か見てみようと思います。

異なり語数の半数が出現頻度1であると、いま唐突に言いましたが、経験的にそうです。日本古典対照分類語彙表で、そのことを確かめます。後に考えることとの関係で、出現頻度2・3・4・5も示します。比率は、異なりに対するその頻度の語数の百分率です。例えば、徒然草では、異なり4,240語、そのうちで出現頻度1のものが2,466語、異なりに対するその比率は2466語／4240語＝58.16%、頻度2では679語／4240語＝16.01%、のようになります。

——	異なり	頻度 1	比率	頻度 2	比率	頻度 3	比率	頻度 4	比率	頻度 5	比率
全体	34180	15857	46.39	5028	14.71	2522	7.37	1607	4.70	1146	3.35
徒然	4240	2466	58.16	679	16.01	332	7.83	155	3.65	123	2.90
平家	13173	5990	45.47	2152	16.33	1110	8.42	712	5.40	484	3.67
宇治	6540	3213	49.12	1070	16.36	527	8.05	330	5.04	218	3.33
方丈	1148	782	68.11	163	14.19	73	6.35	38	3.31	18	1.56
新古	2543	1111	43.68	416	16.35	222	8.72	121	4.75	117	4.60
大鏡	4819	2499	51.85	794	16.47	375	7.78	221	4.58	145	3.00
更級	1950	1105	56.66	306	15.69	141	7.23	81	4.15	57	2.92
紫	2468	1407	57.00	380	15.39	193	7.82	117	4.74	79	3.20
源氏	11416	4688	41.06	1649	14.44	827	7.24	599	5.24	418	3.66
枕	5246	2855	54.42	752	14.33	385	7.33	224	4.26	171	3.25
蜻蛉	3599	1886	52.40	547	15.19	292	8.11	172	4.77	113	3.13
後撰	1923	885	46.02	319	16.58	158	8.21	103	5.35	64	3.32
土左	984	514	52.23	203	20.63	76	7.72	42	4.26	31	3.15
古今	1993	1032	51.78	296	14.85	163	8.17	100	5.01	71	3.56
伊勢	1692	925	54.66	274	16.19	143	8.45	80	4.72	39	2.30
竹取	1310	710	54.19	222	16.94	92	7.02	51	3.89	39	2.97
万葉	6601	3315	50.21	1041	15.77	480	7.27	292	4.42	208	3.15

異なり語数に対する出現頻度1の用語の比率は、17作品で、最小は源氏物語41.06%から最大は方丈記68.11%まで幅がありますが、中央値が土左日記52.23%です。それらが混然と入り組んでいる全体でも46.39%であり、頻度1が異なり語数の半数であることは、大体当たっていると行ってよいと思います。

異なり語数の半数が出現頻度1であることは、

ジップ Zipf, George Kingsley の法則 出現頻度×順位＝一定

によって導くことができそうです。いま出現順位1位の頻度が1000であったとすると、単純に見て、1000位で頻度1になり、異なり語数1000ということになります。その順位1から1000までに対して、理論値として小数を用いて出現頻度を計算すると、概略次のようになるでしょう。実際の出現頻度には小数はありませんから、小数を処理する必要があり、切り捨て、あるいは四捨五入をすることもしてみます。

順位	理論頻度	切り捨て	語数	四捨五入	語数
1	1000.000	1000		1000	
2	500.000	500		500	
3	333.333	333		333	
4	250.000	250		250	
5	200.000	200		200	
166	6.024	6	----	6	
167	5.988	5		6	
181	5.524	5		6	----
182	5.494	5		5	
199	5.025	5		5	
200	5.000	5	34	5	
201	4.975	4		5	
222	4.504	4		5	41
223	4.484	4		4	
249	4.016	4		4	
250	4.000	4	50	4	
251	3.984	3		4	
285	3.508	3		4	63
286	3.496	3		3	
333	3.003	3	83	3	
334	2.994	2		3	
400	2.500	2		3	115
401	2.493	2		2	
499	2.040	2		2	
500	2.000	2	167	2	
501	1.996	1		2	
666	1.501	1		2	266
667	1.499	1		1	
999	1.001	1		1	
1000	1.000	1	500	1	
2000	0.500	-		1	1334
延べ	7485.470	7069			7449
	8178.360				8449

小数を切り捨てた、あるいは四捨五入した結果として、出現頻度が同じになるところが大量に現れます。例えば、501位から1000位までの500語は、理論的頻度の整数部が1ですから、小数を切り捨てるとすべて出現頻度1となります。小数を操作した結果の語数を、頻度の右に記しました。中央の1列は、出現頻度6以上については語数を見ることを省略することとして、頻度5が34語あり、頻度4が50語あり、……、頻度1が500語あると言っています。四捨五入したほうは、理論的頻度が0.5.であっても操作結果では1になりますから、当初の前提である異なり1000語ということと齟齬しますが、最大範囲2000位までを見ています。なお、理論頻度でも調整後でも、出現頻度を合計すれば延べ語数が出ますので、併せて示しました。

お気付いただけかもしれません。用語の出現順位と頻度とがジップの法則に従うならば、出現頻度の小数を切り捨てた理論値で、出現頻度1の語数が異なり語数のちょうど半分になる——ということ。四捨五入した出現頻度に対しては、出現頻度1が異なり語数

の66.7%=1,334語/2,000語、頻度2が13.3%=266語/2,000語のようになります。頻度1への傾きが、理論頻度の小数を切り捨てた結果より大きくなり、頻度2以上では、切り捨てた結果より語数が少なくなります。しかし、四捨五入したほうでも、それなりの味がありまして、現実に近いところがあるようにも思えます。

現実の古典17作品に戻ります。問題の契機は、出現頻度1の用語の多さ、異なり語数の半数を占めるという経験でした。いまジップの法則に触れましたので、頻度2・3・4・5のあたりも、理論値と17作品の実際とで比べてみます。小数を切り捨てた理論的語数に近い比率に、19ページの表で下線を施しました。近い比率というのは、具体的には

	頻度 1	頻度 2	頻度 3	頻度 4	頻度 5
範囲の最小	47.50 ~	15.87 ~	7.89 ~	4.75 ~	3.23 ~
最大	52.50	17.53	8.71	5.25	3.57

というもので、理論的な語数の95%~105%の範囲に収まります。宇治拾遺物語が驚くべくよく理論値に合致し、古今和歌集がそれに次ぎます。方丈記・更級日記・紫式部日記には、合致するところがありません。

累積比率で見えます。小数切り捨ての操作をした理論頻度によれば、次のように算出されます。範囲は、理論的語数に対して、その大きさによらずに前後2.5%ずつとします。

	頻度 1	頻度 2	頻度 3	頻度 4	頻度 5
理論的語数	500	667	750	800	834
範囲の最小	47.50 ~	64.20 ~	72.50 ~	77.50 ~	80.90 ~
最大	52.50	69.20	77.50	82.50	85.90

	累積 1 比率	累積 2 比率	累積 3 比率	累積 4 比率	累積 5 比率
全体	34180 15857 46.39	20885 61.10	23407 68.48	25014 73.18	26160 76.53
徒然	4240 2466 58.16	3145 74.17	3477 82.00	3632 85.66	3755 88.56
平家	13173 5990 45.47	8142 61.80	9252 70.23	9964 75.63	10448 79.31
宇治	6540 3213 49.12	4283 65.48	4810 73.54	5140 78.59	5358 81.92
方丈	1148 782 68.11	945 82.31	1018 88.67	1056 91.98	1074 93.55
新古	2543 1111 43.68	1527 60.04	1749 68.77	1870 73.53	1987 78.13
大鏡	4819 2499 51.85	3293 68.33	3668 76.11	3889 80.70	4034 83.71
更級	1950 1105 56.66	1411 72.35	1552 79.58	1633 83.74	1690 86.66
紫	2468 1407 57.00	1787 72.40	1980 80.22	2097 84.96	2176 88.16
源氏	11416 4688 41.06	6337 55.50	7164 62.75	7763 68.00	8181 71.66
枕	5246 2855 54.42	3607 68.75	3992 76.09	4216 80.36	4387 83.62
蜻蛉	3599 1886 52.40	2433 67.60	2725 75.71	2897 80.49	3010 83.63
後撰	1923 885 46.02	1204 62.61	1362 70.82	1465 76.18	1529 79.51
土左	984 514 52.23	717 72.86	793 80.58	835 84.85	866 88.00
古今	1993 1032 51.78	1328 66.63	1491 74.81	1591 79.82	1662 83.39
伊勢	1692 925 54.66	1199 70.86	1342 79.31	1422 84.04	1461 86.34
竹取	1310 710 54.19	932 71.14	1024 78.16	1075 82.06	1114 85.03
万葉	6601 3315 50.21	4356 65.99	4836 73.26	5128 77.68	5336 80.83

範囲に収まるものには、下線を施しました。宇治拾遺物語が理論値に近いことはここでもそうであり、大鏡・蜻蛉日記・古今和歌集もそのようなものであると知られます。

ジップの法則は言語学の語彙論で着想され、経済学など諸分野にも広がりを見せて、また他の法則との関係もよく論じられてきたようです。その一方で、言語現象については、

あまりよい近似を示さないとして、改訂が続けられてきたと見受けます。しかしながら、ジップの法則の思い掛けない一面が見付かったように思います。ジップの法則は、高頻度・高順位の語彙にかかわるものであり、研究の対象としても、一般に、高頻度・好順位の語彙の諸現象の解明に向かうのが正道です。低頻度の語彙に向かうことにどのような意義があるかは、問われなければならないでしょうし、ジップの法則の使いかたも特異です。ただ、言語現象の末端にせよ、ジップの法則から導かれる現実があることを認めて、ジップの法則なりその現実なりからまた新たな研究を切り開くことも考えられてよいと思います。

ジップの法則の一般的な有効性を考えるのには、17作品の上位10位ずつを見るということも、よいでしょう。ここでは、出現比率(ppm)で掲げます。

合計	徒然	平家	宇治	方丈	新古	大鏡	更級	紫
20626	33430	18553	28648	20182	15671	26185	23332	21634
16801	31443	14886	24683	18599	14506	17297	18500	16941
16751	28638	11399	22711	18203	12816	15895	17672	16712
15239	21040	10720	18319	17807	12700	15553	16567	15338
12682	20689	9241	17262	17411	12175	13844	14358	12591
10864	17825	8162	15615	11476	12059	13571	13530	11790
10768	15371	7822	14049	10684	12001	12374	13392	10531
10290	9818	7742	11365	9893	11768	11144	13116	9958
9643	7714	7273	11325	8705	10020	9264	11321	9157
9619	7539	6234	11142	8310	9787	7964	10216	7898
源氏	枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉
23119	23188	45357	18318	35469	23469	31597	28716	16221
21416	22337	22187	17900	29462	17577	28855	23051	16081
20390	21729	20803	17733	22883	16278	24960	21097	14723
17960	20331	19821	17649	18592	15779	21497	18753	12965
16415	19876	19553	16060	17448	15280	19477	18362	12965
16102	16259	19464	14554	16304	14581	19333	15823	11547
14726	13645	17812	14470	16304	13981	19189	14846	10308
11877	12825	14821	14052	16018	13881	17024	14455	10168
9346	10424	14732	12128	12585	12383	13417	13479	9089
8869	8631	11160	11710	11727	11684	10820	13088	8430

出現順位2位の用語の出現比率が1位の1/2であるというのは、蜻蛉日記のみです。動詞「あり」と動詞「す」とが、それを実現しています。しかし、その蜻蛉日記の他のところを含めて、ジップの法則 出現順位×頻度(比率)＝一定 が成り立つところがないことは、一目瞭然です。ジップの法則には、いまはこれ以上には立ち入りません。

なお、低頻度のことを扱おうとするのは、異なり語数の推計をしたいとかねて思っているからです。すなわち、一作品のようなある全体が与えられたとき、その延べ語数は大体推測することができます。現代の普通の文章であるならば、経験的に400字で100文節であり、語彙の計量調査を設計するときなどは、それを基にして規模を見積もります。古典文学作品でも、現代人向けに漢字・仮名を調節して本文が作成されていれば、同様の見込みをもつことができます。それに対して、異なり語数を推定することは、今のところ、できるとは到底言えそうにありません。問題に重みがないのかもしれませんが、気になり続

けています。

私は、語彙の計量調査について、何ほどのことをしたわけでもありません。しかしながら、携わってみると、一作品のものを徐々に処理するときによ、連続していない文章を次つぎに処理するときによ、新たな語句が絶え間なく立ち現れてきます。量をこなせば新たな語句はなくなりそうなものであるのに、なぜかそうはならず、頻度1の語句が常に異なり語数の半分として存在し続ける。例えば、古今和歌集と新古今和歌集とでは、次のようになります。余分な数値も並べているのは、検算をしているつもりです。

	異なり	頻度1	比率%	頻度多	延べ
古今	1993	1032	51.78	961	10013
新古今	2543	1111	43.68	1432	17165
両古今	3313	1448	43.70	1865	27178
両古今の内訳					
二古今共通	1223			1223	
古今のみ	770	620		150	
新古今のみ	1320	828		492	

古今和歌集と新古今和歌集とが一体となった両古今和歌集というものを作成したとすると、異なり語数は3313になり、古今和歌集・新古今和歌集それぞれの異なり語数の合計にはなりません。二つの古今和歌集に元もと共通に出現していた用語が1223語あって、両古今和歌集ではその分が異なり語数としては増えないからです。頻度1のほうは、古今和歌集で頻度1であって新古今集に出現しない620語と、新古今和歌集で頻度1であって古今和歌集に出現しない828語と、それが両古今和歌集の頻度1になります。古今和歌集で頻度1、新古今和歌集で頻度1であれば、それぞれでは頻度1として見られるものの、両古今和歌集では頻度2となって頻度1から外れることとなります。異なり語数や出現頻度1は、作品間でこのように絡み合いながら、全体を構成してゆくことになる。その絡み合いを定量的に捉えてみたいのです。異なり語数の半数を占め続ける出現頻度1の様相が捉えられないと、問題を解決することができないのではないかと、思う次第です。

出現頻度1にこだわらなければ、異なり語数の推定の問題は、解決できるかもしれませぬ。源氏物語から1/10を抽出して、異なり語数を見る。次に2/10の異なり語数を見て、1/10との関係をみる。それを源氏物語で延長するなり、他の作品に展開するなりして、異なり語数の様相を記述し、作品全体との関係を把握すればよい。このばあい、われわれには、記述の最終目標となる作品全体の様相が与えられています。私は今その作業に着手しているわけでもなく、問題提起のみというのは、研究を進めていらっしやるかたがたにご迷惑ですから、今日はおしまいということにします。

日本古典対照分類語彙表のご紹介をご清聴くださいまして、ありがとうございました。

研究会ご出席者（着席順）

田島 毓堂	名古屋大学名誉教授
大島 中正	同志社女子大学表象文化学部日本語日本文学科
丸山健一郎	同志社大学大学院文学研究科博士後期課程在学
吉岡真由美	同志社大学大学院文学研究科博士前期課程在学
成田 徹男	名古屋市立大学人文社会学部人間科学科
入江さやか	同志社大学文学部国文学科
石井 久雄	同志社大学文学部国文学科
前川 武	大阪国際大学短期大学部ライフデザイン総合学科
広瀬 英史	静岡文化芸術大学文化政策学部国際文化学科
広瀬菜桜子	株式会社エディット
ミフタクル・アムリ	愛知学院大学大学院文学研究科博士後期課程在学
木下 朗	株式会社三省堂出版局辞書出版部国語辞書第一編集室