

## 中古散文22作品の類似度の測定

前 川 武\*

### Measuring the Degree of Similarity of 22 Ancient and Medieval Prose Works

Takeshi Maekawa\*

#### Abstract

In research on Japanese documents, as computerization of full-text database and glossary indexes advances, it has become possible to perform empirical analyses from various angles.

Although computerization of full-text databases has progressed greatly, that of glossaries has not advanced as much. This is because what constitutes a compound word is a subjective judgment by the maker of the index and it is necessary for the researcher to examine the text again to add attributes of words to the index.

In such situations Associate Prof. Nahoko Murata makes a glossary index of the adjectives and adverbs of ancient words and performs analyses.

The degree of similarity of 22 ancient and medieval prose works were measured using the methods detailed by Murata.

#### キーワード

類似度、語彙、形容詞、形容動詞、中古散文作品

#### I はじめに

日本語の文献に関する研究においては、本文テキストのデータベース化と作品中に使用されている語の属性、出現頻度などをまとめた語彙索引の電子化が進めばさまざまな角度からの計量的分析が可能になる。

日本語の古典文学の分野では、国文学研究資料館が、利用目的を研究の範囲に限定し、さらに登録制にした上で、岩波書店旧版『日本古典文学大系』全作品（100巻560作品）の全文データベースとその検索システムを試験的に公開している。これを利用すると、複数の作品を対象とした文字列検索や文字頻度分析などが簡単にできる。<sup>1)</sup>

このように、本文テキストのデータベース化は進んでいるが、それに比べて、索引の電子化は、なかなか進んでいないのが現状である。その理由の一つとして、印刷物の索引を

---

\*まえかわ たけし：大阪国際大学短期大学部准教授〈2007.6.18受理〉

ベースにする場合、複合語の認定基準が索引作成者の判断に左右されるところが大きいことがあげられる。そのほかにも、従来の索引にない語の属性を付加したい場合など、再度本文を調査する必要があることなどがあげられる。

このような現状の中で、村田葉穂子氏は、古代語の形容詞と形容動詞について、単なる作品における語の出現度数、用例だけではなく、語構造・単位数・結合タイプ・結合型といった語構造に関わる属性、造語形式・造語型といった造語論に関わる属性、活用の種類などの詳細な情報を付加した語彙表を作成し、古代語形容詞・形容動詞の活用別、出現時代別、結合タイプ別あるいはこれらを組合わせた形での延べ語数・異なり語数の分布状況を分析し、それらの史的変遷について考察を行い、『形容詞・形容動詞の語彙論的研究』<sup>2)</sup>の中で分析・考察内容を明らかにしている。

今回、同書に別表として掲載されているデータに基づき、中古の散文資料22作品<sup>3)</sup>について、その語彙の使用状況から作品間の類似度を機械的に測定する試みを行った。

## Ⅱ 類似度の測定の方法

語彙の使用状況から作品間の類似度を測定する方法については、これまでいくつかの手法が考案されてきた。

宮島達夫氏は、「語いの類似度」<sup>4)</sup>において、見出し語数の共通度に注目した水谷静夫氏の式<sup>5)</sup>と安本美典・本多正久氏の式<sup>6)</sup>について解説しながら、独自の説を展開し、様々な分野への適用を試みている。

以下に、宮島氏の手法を簡単に説明する。

見出し語 $W_1, W_2, W_3, \dots, W_n$ があったとき、見出し語 $W_i$ の作品Aにおける使用率を $P_i(A)$ 、見出し語 $W_i$ の作品Bにおける使用率を $P_i(B)$ で表し、両者の小さい方を $\min(P_i(A), P_i(B))$ で表すと、以下の式で算出される値 $C_{AB}$ が作品AB間の類似度となるというものである。

$$C_{AB} = \sum_i \min(P_i(A), P_i(B))$$

この式は、不一致度の合計を用いた次の式から導き出されている。

$$C_{AB} = 1 - \frac{1}{2} \left( \sum_i (|P_i(A) - P_i(B)|) \right)$$

さらに、宮島氏は、この類似度を基にして、複数作品間の近さを図示する方法を考案している。それは、異なる2作品間の不一致度（1－類似度）をすべて算出し、その中で一番高い値をとったときの2作品の組合せ（例えばAとB）を求め、その不一致度を底辺の長さとする三角形を想定し、その両端に2つの作品を配置し、比較したい作品（例えばCとD）との不一致度を算出し、それぞれの長さを残りの2辺とし、その頂点に比較したい作品を配置するというものである。これを図示すると図1のようになる。

## 中古散文22作品の類似度の測定

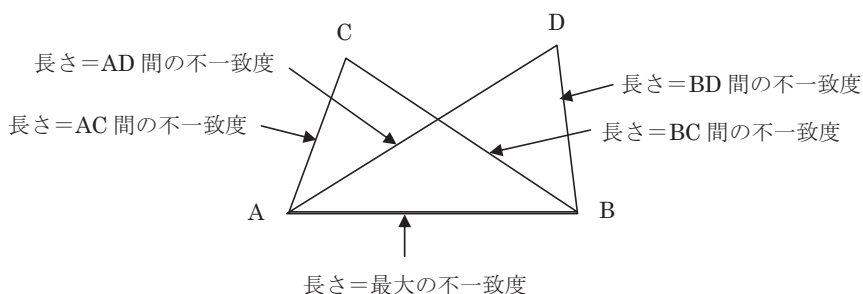


図1 複数作品間の近さを図示する方法

今回の類似度の測定にあたっては、この宮島氏の式および図示の方法を使うこととした。

### Ⅲ 類似度の測定

#### 1 対象となるデータ

今回対象としたデータは、『形容詞・形容動詞の語彙論的研究』<sup>2)</sup>の別表三「中古散文作品の形容詞対照語彙表」収録の形容詞1124語および別表四「中古散文作品の形容動詞対照語彙表」形容動詞1093語とした。

データの内容は、見出し語と中古散文22作品での述べ語数（出現回数）である。参考までに形容詞のデータの例を表1に示す。

#### 2 各作品の使用率の算出

データの件数が多いため、Excel VBAを使って算出した。

以下は、Excel VBAで作成したプロシージャの内容である。

```
Sub syukei1 ( )
    last = 1126          '元データの集計行の番号
    yoko = 22            '作品数22
    For x = 1 To yoko    '作品数分繰り返す
        Cells(1, x) = x  '項目行 (1, 2, 3, ..., 22) の出力
        For y = 2 To last '縦方向の繰り返し (項目行の次からデータ数分)
            '元データの対応セルから使用率を算出
            Cells(y, x) = Sheet2.Cells(y, x + 3) / Sheet2.Cells(last, x + 3)
        Next
    Next
End Sub
```

#### 3 2作品間の類似度の算出

22作品の異なる2作品の組み合わせは ${}_{22}C_2 = 231$ とおりと多いので、これもExcel VBAを使って算出した。

以下は、Excel VBAで作成したプロシージャの内容である。

```
Sub syukei2 ( )
```

表1 測定対象となるデータの例 (中古散文作品の形容詞1124語)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124
1	見出し語	漢字	読み	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			

中古散文22作品の類似度の測定

```

last = 1126          '元データの集計行の番号
kosu = last - 2      '元データのデータ数
yoko = 22            '作品数22
x3 = 0               '横方向の列番号の初期値
syukei = kosu + 1 + 1 '集計行の番号（個数+見出し行+集計行自身）
For x = 1 To yoko - 1 '2作品の組合せの一方を変化させる
    For x2 = x + 1 To yoko '2作品の組合せのもう一方を変化させる
        x3 = x3 + 1 '横方向の列番号を1増やす
        syokei = 0 '列ごとの小計の初期化
        '項目行（(1-1),(1-2),(1-3),..., (21-22)）の出力
        Cells(1, x3) = "(" & LTrim(Str(x)) & "," & LTrim(Str(x2)) & ")"
        For y = 2 To kosu + 1 '縦方向の繰り返し（項目行の次から
            'データ数分）
            '項目に対応する2つの使用率の小さい方のデータを算出
            Cells(y, x3) = _
            Application.WorksheetFunction.Min(Sheet1.Cells
            (y, x), _Sheet1.Cells(y, x2))
            syokei = syokei + Cells(y, x3) '小計の算出
        Next
    Next
    syokei = Round(syokei, 3) '小計の小数点3位以下四捨五入
    Cells(syukei, x3) = syokei '小計の転記
Next
Next
End Sub

```

#### 4 2作品間の類似度および不一致度のまとめ

3で求めた値を見やすい形に整理するとともに、類似度の最小値、類似度が最初になる2作品の組を求め、さらに、その2作品とそれらを除いた20作品との類似度および不一致度を求める。

以下は、Excel VBAで作成したプロシージャの内容である。

```

Sub syukei3 ()
    last = 1126          '類似度データの集計行
    yoko = 22            '作品数22
    x3 = 0               '横方向の列番号の初期値
    oldmin = 1           '仮の最小値
    For x = 1 To yoko - 1 '横方向の列番号を変化させる
        Cells(x, 1) = x + 1 '項目列（2, 3, 4, ..., 22）の出力
        Cells(22, x + 1) = x '項目行（1, 2, 3, ..., 21）の出力
    Next

```

```

For y = x To yoko - 1 '縦方向の行番号を変化させる
    x3 = x3 + 1 '類似度データの横方向の列番号を1増やす
    '類似度データの集計値を転記
    Cells (y, x + 1) = Sheet3.Cells (last, x3)
    '今の類似度<記憶しておいた類似度の最小値ならば
    If Cells (y, x + 1) < oldmin Then
        oldmin = Cells (y, x + 1) '記憶しておいた類似度
        を置き換える
        minx = x 'そのときの列番号を記憶
        miny = y + 1 'そのときの行番号+1を記憶
    End If
Next
Next
minx1 = 5 '類似度最小のときの2作品を書き出す列
miny0 = 25 '20作品の見出しを書き出す行
miny1 = 26 '20作品の不一致度を書き出す行 (片方)
miny2 = 29 '20作品の不一致度を書き出す行 (もう片方)

Cells (miny1, minx1) = minx '類似度最小のときの作品1を書き出す
Cells (miny2, minx1) = miny '類似度最小のときの作品2を書き出す
Cells (miny2 + 2, minx1) = oldmin '類似度最小のときの類似度を書き出す
Cells (miny2 + 3, minx1) = 1 < oldmin '類似度最小のときの不一致度を書き出す
'類似度最小のときの2作品と他の20作品との類似度、不一致度を算出
For i = 1 To yoko '類似度最小のときの片方の場合
    If i <> minx And i <> miny Then '類似度最小の組み合わせ以外なら
        '(i,minx)の組合せが (行<列) になるように入れ替える
        If minx > i Then
            y1 = minx
            x1 = i
        Else
            y1 = i
            x1 = minx
        End If
        Cells (miny0, minx1 + i) = i '20作品の見出しを表示
        ruiji = Cells(y1 - 1, x1 + 1) '類似度を転記
        hiruiji = 1 - ruiji '不一致度の算出
        Cells(miny1, minx1 + i) = ruiji '類似度の転記
        Cells(miny1 + 1, minx1 + i) = hiruiji '不一致度の転記
    End If
Next

```

## 中古散文22作品の類似度の測定

```

Else
    If i <> 1 Then          '一番最初でなければ
        minxl = minxl - 1  '列番号を一つ前に戻す
    End If
End If
Next
minxl = 5
For i = 1 To yoko          '類似度最小のときのもう片方の場合
    If i <> minx And i <> miny Then
        If miny > i Then
            y1 = miny
            x1 = i
        Else
            y1 = i
            x1 = miny
        End If
        ruiji = Cells (y1 - 1, x1 + 1)
        hiruiji = 1 - ruiji
        Cells (miny2, minxl + i) = ruiji
        Cells (miny2 + 1, minxl + i) = hiruiji
    Else
        If i <> 1 Then
            minxl = minxl - 1
        End If
    End If
End If
Next
End Sub

```

### 5 22作品の距離感の図示

不一致度の最大の組合わせの2作品とその2作品と他の20作品との不一致度が求まったので、あとは、20作品のそれぞれについて、(最大の不一致度、不一致度最大の組合わせの一方との不一致度、不一致度最大の組合わせのもう一方との不一致度)を3辺とする三角形の頂点の位置を求めて点をプロットしていけばよい。

しかし、Excelの標準のグラフ作成機能には、このような場合に適合するグラフがなく、変数を変換して適用するにもやはり適切なものがない。

また、Excel VBAのユーザーフォームでは、グラフィックス系のコントロールが使用できない。

別アプリケーションやDLLを呼び出すこともできなくはないが、環境に依存することに

なるため、できれば使用したくない。

シートの倍率を縮小して、セルのサイズを小さくして、ドットの代わりにする方法も考えたが、これだと文字を表示したときに小さすぎて見えなくなってしまう。

そこで、図形のオートシェープの楕円をExcel VBAからコントロールして、楕円をドットの代わりに使って22作品の点をプロットしていくこととした。

以下は、Excel VBAで作成したプロシージャの内容である。

```
Sub syukei4 ( )
    size0 = 800                'グラフ領域の大きさ
    x0 = 150                   'グラフ領域の開始位置 (x0,y0)
    y0 = 450
    tsize = 16                 'テキストボックスのサイズ
    atate = 32                 'a(底辺に相当する不一致度)の入ってるセルの列
    ayoko = 5                 'a(底辺に相当する不一致度)の入ってるセルの行
    btate = 30                 'b(右辺に相当する不一致度)の入ってるセルの列
    byoko = 6                 'b(右辺に相当する不一致度)の入ってるセルの行
    ctate = 27                 'c(左辺に相当する不一致度)の入ってるセルの列
    cyoko = 6                 'c(左辺に相当する不一致度)の入ってるセルの行
    leftx = 5                 '底辺の左端の作品番号の入ってるセルの列
    lefty = 26                '底辺の左端の作品番号の入ってるセルの行
    rightx = 5                '底辺の右端の作品番号の入ってるセルの列
    righty = 29               '底辺の右端の作品番号の入ってるセルの行
    midashix = 6              '見出しの入ってるセルの最初の列
    midashiy = 25             '見出しの入ってるセルの行
    a = Sheet4.Cells(atate, ayoko) 'a(底辺に相当する不一致度)を取得
    x1 = x0 + a * size0        '底辺の右端の座標(x1,y1)を算出
    y1 = y0

    '底辺の左端(x0,y0)に半径2の円を描く
    Shapes.AddShape msoShapeOval, x0 - 2, y0 - 2, 2, 2
    '1文字分のテキストボックスを描く
    Shapes.AddTextbox msoTextOrientationHorizontal, x0 - tsize / 2, y0 - tsize, tsize,
    tsize gcount = Shapes.Count

    'テキストボックスに作品番号を描く
    Shapes(gcount).TextFrame.Characters.Text = Sheet4.Cells(lefty, leftx)
    Shapes(gcount).Line.Visible = False    'テキストボックスの線を非表示に
    Shapes(gcount).Fill.Visible = False    'テキストボックスの内部を非表示に
    Shapes.AddShape msoShapeOval, x1 - 2, y1 - 2, 2, 2
    Shapes.AddTextbox msoTextOrientationHorizontal, x1 - tsize / 2, y1 - tsize, tsize,
    tsize gcount = Shapes.Count
```



```

Shapes(gcount).TextFrame.Characters.Text = Sheet4.Cells (righty, rightx)
Shapes(gcount).Line.Visible = False
Shapes(gcount).Fill.Visible = False

For i = 1 To 20                                '20個の円を描く
    b = Sheet4.Cells (btate, byoko + i - 1) 'b(右辺に相当する不一致度)を取得
    c = Sheet4.Cells (ctate, cyoko + i - 1) 'c(左辺に相当する不一致度)を取得
    cosb = (a ^ 2 + c ^ 2 - b ^ 2) / (2 * a * c) 'cosBとsinBを算出
    sinb = (1 - cosb ^ 2) ^ (1 / 2)
    yl = c * sinb                                '頂点の (x0,y0) からの垂直方向の距離を算出
    xl = c * cosb                                '頂点の (x0,y0) からの水平方向の距離を算出
    x2 = x0 + xl * size0                        '頂点の座標 (x2,y2) を算出
    y2 = y0 - yl * size0

    Shapes.AddShape msoShapeOval, x2 - 2, y2 - 2, 2, 2
    Shapes.AddTextbox msoTextOrientationHorizontal, x2 - tsize / 2, y2 - tsize, _
    tsize, tsize
    gcount = Shapes.Count
    Shapes(gcount).TextFrame.Characters.Text = Sheet4.Cells(midashiy,
    _midashix + i - 1)
    Shapes (gcount) .Line.Visible = False
    Shapes (gcount) .Fill.Visible = False

Next
End Sub

```

#### Ⅳ 結果および考察

##### 1 形容詞の使用率から見た中古散文22作品の類似度

まず、異なる2作品間での類似度および不一致度の算出結果は、表2のようになった。

上部の階段状のものは、横軸の1～21が2作品の組合わせの一方を表し、縦軸の2～22がもう一方を表し、その交点に対応する数値が2作品間の類似度を表す。

例えば、横軸が3、縦軸が5の交点を見ると値は0.554であるが、これは、作品3と作品5の類似度が0.554であることを表す。

下部の左側の2列は、全体の中で最も高い類似度と最も低い類似度をそれぞれ5つずつ示している。

下部の右側部分の内、最左列は、全体の中で最も類似度の低かったものが作品6と作品12の組合せであったこと、そのときの類似度が0.283、不一致度が0.717であることを示している。

表2 中古散文22作品の形容詞の使用率から見た類似度および不一致度

[illegible]

# 中古散文22作品の類似度の測定

下部の右側部分の内、最上部の行は、22作品から最も類似度の低かった組合わせの作品6と作品12を除いた20作品を示している。

その下の2行は、作品6と上記20作品との間の類似度（上段）と不一致度（下段）を示している。

さらに1行空けたあとの2行は、作品12と上記20作品との間の類似度（上段）と不一致度（下段）を示している。

表2では、作品を作品名ではなく1～22の連番で表しているが、これは、成立年代の順を反映しているもので、年代の推移を念頭に置いた場合、この方がわかりやすいこともあって、敢えて連番で表記した。参考までに中古散文22作品の作品名と連番との関係を表3に示す。

表3 中古散文22作品の作品名と連番との関係

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
竹取物語	土佐日記	伊勢物語	平中物語	大和物語	多武峯少将物語	篁物語	宇津保物語	蜻蛉日記	落窪物語	和泉式部日記	枕草子	源氏物語	紫式部日記	堤中納言物語	夜の寝覚	浜松中納言物語	更級日記	狭衣物語	大鏡	讃岐典侍日記	とりかへばや物語

表2の中で類似度の高いものは、次の5つである。

17-22 : 0.718    19-22 : 0.711    13-19 : 0.710    16-17 : 0.707    16-22 : 0.703

逆に低いものは、次の5つである。

6-12 : 0.283    6-14 : 0.298    2-6 : 0.322    7-12 : 0.355    7-18 : 0.356

次に、22作品間の距離感は図2のような結果になった。

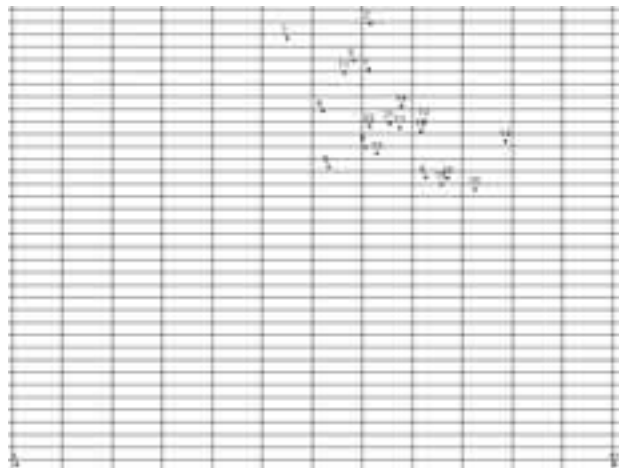


図2 中古散文22作品の形容詞の使用率から見た距離感

図3は、図2の上部を拡大したものである。



図3 図2の上部を拡大したもの

近いものを集めてみると、次のようになる。

グループ1：8,10,15

グループ2：9,13,16,17,18,19,21,22

グループ3：1,3,11

しかし、これらのグループに属する2作品間の類似度を表2で確認すると、図2の位置関係と符号しないものが出てくる。

例えば、17-22の類似度が0.718、17-21の類似度が0.538であることからわかるように、数値的には17から見れば22の方が近いにも関わらず、図2では22の方が遠くなっている。あるいは、1-4の類似度が0.494、1-11の類似度が0.410であるのに対して、図2上では、1-4の距離>1-11の距離となっているなどである。

仮に、図2を3次元的に見たとすれば、11や21はもっと奥の方にあるのかもしれない。

## 2 形容動詞の使用率から見た中古散文22作品の類似度

まず、異なる2作品間での類似度および不一致度の算出結果は、表4のようになった。表4の中で類似度の高いものは、次の5つである。

15-22：0.669 13-16：0.663 17-22：0.655 16-17：0.642 13-17：0.641

逆に低いものは、次の5つである。

2-7：0.116 1-7：0.136 1-6：0.164 2-10：0.174 1-21：0.177

次に、22作品間の距離感とは図4のような結果になった。

図5は、図4の上部を拡大したものである。

近いものを集めてみると、次のようになるが、形容詞の場合と同様の問題がある。

グループ1：5,6,9,10

グループ2：11,13,15,16,17,18,19,21,22

グループ3：3,4,14,20

### 中古散文22作品の類似度の測定

表4 中古散文22作品の形容動詞の使用率から見た類似度および不一致度

[illegible]

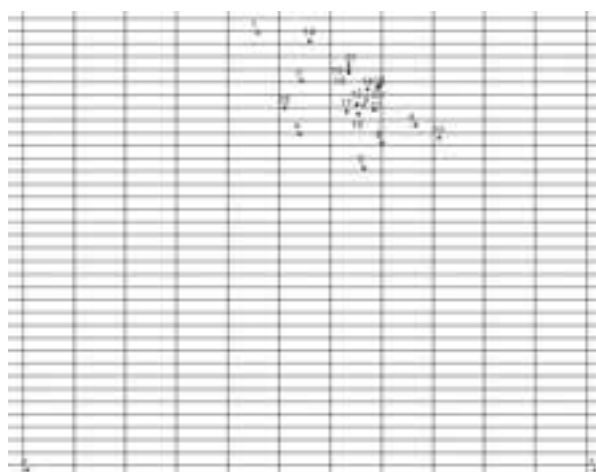


図4 中古散文22作品の形容動詞の使用率から見た距離感

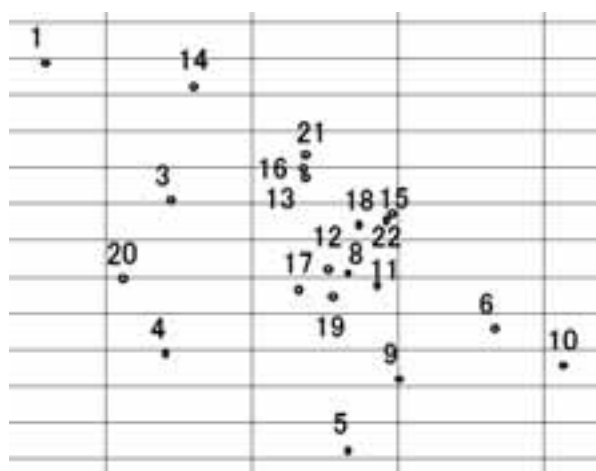


図5 図4の上部を拡大したもの

### 3 形容詞の使用率から見た場合と形容動詞の使用率から見た場合との比較

図2、3と図4、5を見比べると、多くの部分で共通点が見られる。

例えば、6と7は、年代的に近いにも関わらず、形容詞、形容動詞とも距離が大きく隔たっていること、14と20がともに前後の作品から離れた点に位置していること、13,16,17,18,19,21,22が極めて近い位置にあることなどがあげられる。

しかし、ここでも、1で述べた問題点があることを忘れてはいけない。

## V おわりに

今回は、宮島氏の提唱した手法に基づいて、形容詞および形容動詞の見出し語の使用率から中古散文22作品の距離感を測定して図示するプロセスをExcel VBAを使って実現して

みた。

その結果、一つの平面上に図示することはできたが、距離感に矛盾が生じることも明らかになった。

今後は、他の分析手法との比較などを通して、これらの矛盾を解消していきたい。

また、今回は、見出し語の使用率に着目したが、形容詞や形容動詞の個体そのものの持つ属性と用いられている作品との関係、作品の特性や作者の特性などについても着目し、分析を行っていきたい。

### 【付記】

本稿は、日本学術振興会平成19-22年度科学研究費補助金（基盤研究（C）課題番号19520407）による研究成果の一部である。

### 注

- 1) 国文学研究資料館 <http://www.nijl.ac.jp/index.html>
- 2) 村田菜穂子、『形容詞・形容動詞の語彙論的研究』、初版、和泉書院、2005年
- 3) 『竹取物語』、『土佐日記』、『伊勢物語』、『平中物語』、『大和物語』、『多武峯少将物語』、『篁物語』、『宇津保物語』、『蜻蛉日記』、『落窪物語』、『和泉式部日記』、『枕草子』、『源氏物語』、『紫式部日記』、『堤中納言物語』、『夜の寝覚』、『浜松中納言物語』、『更級日記』、『狭衣物語』、『大鏡』、『讃岐典侍日記』、『とりかへばや物語』
- 4) 宮島達夫、「語いの類似度」、『國語學』、82、1970年
- 5) 2作品の見出し語の数をa,b、共通の見出し語の数をとしたとき、類似度 =  $\frac{x}{a+b-x}$
- 6) 2作品の見出し語の数をa,b、共通の見出し語の数をとしたとき、類似度 =  $\frac{x}{\sqrt{ab}}$

### 参考文献

- 1) 宮島達夫、「総索引への注文」、『國語學』、76、1969年
- 2) 村上征勝・金明哲、『講座 人文科学研究のための情報処理〔第5巻 数量的分析編〕』、初版、尚学社、1998年