

コレスポネンス分析と英文分析について

コレスポネンス分析とは、分割表の行成分と列成分の相関をできるだけ大きくするように行成分と列成分に係数を付け、それを各成分の第 1 軸の値とする。また、その係数とは独立に次に相関を大きくするように行成分と列成分に係数を付け、各成分の第 2 軸の値とする。これを繰り返すことにより、各成分の空間配置を決めて行き、どの成分が近いかを調べる手法である。

例えば図 1 のデータから、行と列の成分の関係をコレスポネンス分析によって求めると図 2 のようになる。

	A	B	C	D
中学生	10	19	13	5
高校生	13	8	15	16
大学生	18	11	14	8

図 1 データ

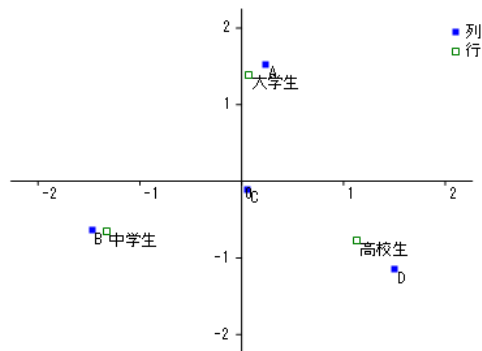


図 2 分析結果

ここで述べたコレスポネンス分析の手法を英文分析に活用してみよう。まず、各テキスト（教科書や本）のデータから、テキスト毎に図 3 のように言語頻度を取得する。

	A	B	C
1		Crown	340
2		a	340
3		ability	5
4		able	14
5		about	62
6		above	1
7		accept	2
8		accepted	1
9		according	1
10		achieve	1
11		achieved	1
12		acorns	1
13		across	3
14		act	1
15		acted	1
16		activities	6

図 3 テキスト言語頻度

このデータを College Analysis のページ毎に図 4 のように貼り付ける。この例では 3 つの教科書のデータを取り込んでいる。

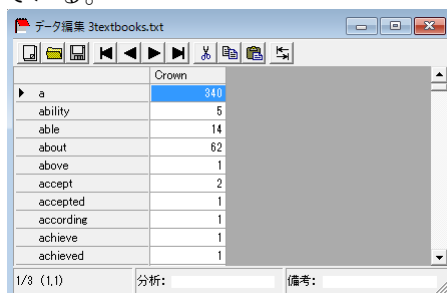


図 4 各ページの貼り付け

次にエディタメニュー [ツール>単語数比較ツール] を選択し、図 5 の単語数比較ツールメニューを表示する。

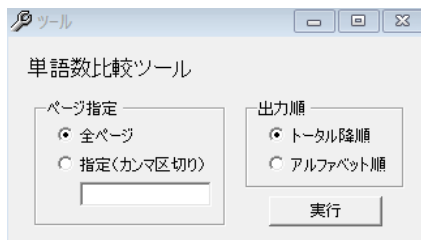


図 5 単語数比較ツール

必要に応じて設定を変えて、実行ボタンをクリックすると、各テキストのデータをまとめて図 6 のような結果を出力する。ここで結果は単語数の合計の多い順に出力されている。

	Crown	Discovery	Prominence	Total
the	570	154	492	1216
to	371	97	272	740
a	340	39	206	585
of	290	50	232	572
and	326	44	197	567
in	234	99	202	535
i	263	28	115	406
was	146	48	128	322
that	167	30	112	309
you	129	46	128	303

図 6 単語数比較表

この表を Excel に移し、専門家が不要な単語を除き、必要なデータまでを選別する。テキストにより語数が異なるので合わせることもできる。結果を図 7 に示す。

	Crown	Discovery	Prominence	Total	
1					
2	the	570	154	492	1216
3	to	371	97	272	740
4	a	340	39	206	585
5	of	290	50	232	572
6	and	326	44	197	567
7	in	234	99	202	535
8	i	263	28	115	406
9	was	146	48	128	322
10	that	167	30	112	309
11	you	129	46	128	303
12	is	128	44	116	288
13	he	128	44	116	288
14	is	155	48	84	287

図 7 選別されたデータ

ここから実際に演習をしてみましょう。このデータについて College Analysis にコピーする。メニュー [分析→多変量解析他→分類手法→コレスポネンス分析] を選択すると図 8 のような分析実行メニューが表示される。

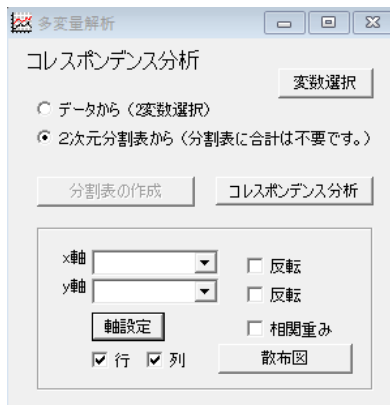


図 8 分析実行メニュー

変数選択ですべて選択し、軸を設定し、その下の「行」チェックボックスを外して、「コレスポネンス分析」及び「散布図」ボタンをクリックすると、図 9 のような結果が表示される。

群	第1成分	第2成分	重み1成分	重み2成分	
固有値	0.057	0.028			
相関係数	0.240	0.166			
寄与率	0.675	0.325			
累積寄与率	0.675	1.000			
the	1	-0.207	0.384	-0.050	0.064
to	1	-0.114	-0.033	-0.027	-0.005
of	1	-0.917	0.233	-0.220	0.039
and	1	-1.094	-0.571	-0.262	-0.095
in	1	0.703	0.302	0.168	0.050
i	1	-1.257	-1.507	-0.301	-0.250
was	1	0.161	0.388	0.039	0.064
that	1	-0.716	-0.264	-0.172	-0.044
you	1	0.190	0.677	0.045	0.112
he	1	0.215	0.461	0.052	0.077
is	1	0.500	-0.719	0.120	-0.119
it	1	1.201	-0.383	0.288	-0.064

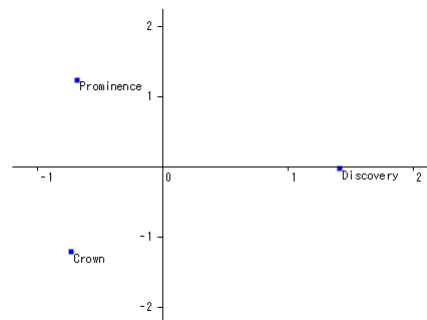


図 9 分析実行結果

最後に、この結果を他の分析手法である主成分分析、因子分析、クラスター分析と比較してみよう。図 10 にそれぞれの手法で表した分類を示す。但し、因子分析は因子数 2 の主成分法でバリマックス回転をした結果、クラスター分析は距離を 1-相関係数、最長距離法を使用した結果である。

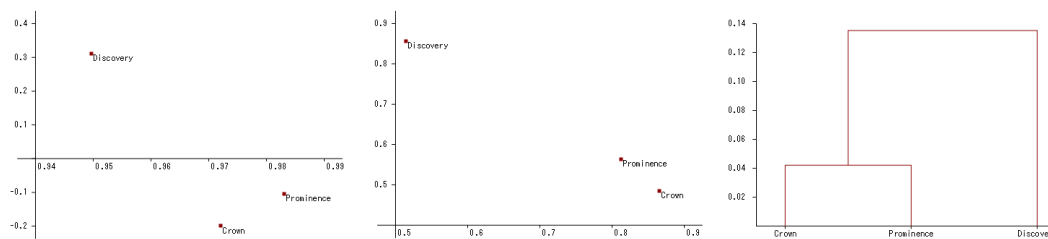


図 10 主成分分析、因子分析、クラスター分析結果