コレスポンデンス分析と英文分析について

コレスポンデンス分析とは、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ンデンス分析によって求める と図 2 のようになる。







ここで述べたコレスポンデ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。



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。

P	🚔 データ編集 3textbooks.txt 📃 💷 💌									
Ŀ		► ► <u>%</u> ■	a 🛍 💺							
		Crown		▲						
•	a	340								
	ability	5								
	able	14								
	about	62								
	above	1								
	accept	2								
	accepted	1								
	according	1								
	achieve	1								
	achieved	1		-						
1/3	(1,1) 5	術:	備考:	1.						

図4 各ページの貼り付け

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

/₽ ツール	
単語数比較ツール	
ページ指定	出力順 ④ 下一句山略細
○ 呈へ ジ ○ 指定(カンマ区切り)	〇 アルファベット順
	実行

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
а	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

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。

🐹 🛃 🍠 = (°! - =				3textbook	s.xlsx - Microso	oft Excel			-	- 🗆 X
ファイル ホーム 吉	挿入 ページレイス	Pウト 数式 :	データ 校開	表示 開発	¥−Δ					a 🕜 — 🗗 X
🖺 👗 МS Р	ゴシック・	12 × A A	= = = >	言:折り返して:	全体を表示する	農準	- 👪 🛛	1	■挿入▼ Σ ▼ ●● 削除 ▼ ■ ▼ …	27 🕅
BAGNIGI7 ▼	<u>n</u> - 🖽 - 🧟	• <u>A</u> · <u>∡</u> ·		譚 国 セルを結合	して中央揃え、	😗 - % • 🕇 👬 🍦	28 朱作何き テーノ 書式 * 書式	ルとして セルの 設定 = スタイル =	□書式 - ② - プ	へ替えと 検索と (ルター▼ 選択 ▼
クリップボード ち	フォント	5		配置	5	数値	5 ZŚ	ษณ	セル	編集
A1	• (°)	f _×								*
A	В	С	D	E	F	G	Н	I	J	K
1	Crown	Discovery	Prominence	Total			Crown	Discovery	Prominence	
2 the	570	154	492	1216		the	435.45	421.46	505.34	
3 to	371	97	272	740		to	283.42	265.46	279.38	
4 of	290	50	232	572		of	221.54	136.84	238.29	
5 and	326	44	197	567		and	249.05	120.42	202.34	
6 in	234	99	202	535		in	178.76	270.94	207.48	
7 i	263	28	115	406		i	200.92	76.63	118.12	
8 was	146	48	128	322		was	111.54	131.36	131.47	
9 that	167	30	112	309		that	127.58	82.10	115.04	
10 you	129	46	128	303		you	98.55	125.89	131.47	
11 he	128	44	116	288		he	97.78	120.42	119.15	
12 is	155	48	84	287		is	118.41	131.36	86.28	-
I€ € ▶ ₽ / crown /	discovery / promit	nence / Sheet1	上位、分析結果	〈分析結果5000 /	/ 🞾 🚺 🖌 📃					▶
コマンド 🎦									Ⅲ 100% —	- · · · ;;

図7 選別されたデータ

<u>ここから実際に演習をしてみましょう。</u>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-コレスポンデンス分析]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。

🔀 多変量解析									
コレスポンデンス分析	変数選択								
○ データから(2変数選択)									
	44(12) 201 (0-1) 201 (3-67								
分割表の作成	」レスポンデンス分析								
×軸 🔽	□ 反転								
уфа	□ 反転								
軸設定	□ 相関重み								
▶ 行 ▶ 列	散布図								

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 スポンデンス分析」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。

☆ コレスポンデンス分 4	折結果				- • •	×		I			
	君羊	第1成分	第2成分	重み1成分	重み2成分		2 -				
▶ 固有値		0.057	0.028			\square					
相関係数		0.240	0.166			1					
寄与率		0.675	0.325				Prominence				
累積寄与率		0.675	1.000			1	1-				
the	1	-0.207	0.384	-0.050	0.064	1					
to	1	-0.114	-0.033	-0.027	-0.005	1					
of	1	-0.917	0.233	-0.220	0.039	1					_
and	1	-1.094	-0.571	-0.262	-0.095	1	·-1	0	1	Discovery	2
in	1	0.703	0.302	0.168	0.050	1					
i	1	-1.257	-1.507	-0.301	-0.250	1					
was	1	0.161	0.388	0.039	0.064	1	-1-				
that	1	-0.716	-0.264	-0.172	-0.044	1	Grown				
you	1	0.190	0.677	0.045	0.112	1	CIOMI				
he	1	0.215	0.461	0.052	0.077	1					
is	1	0.500	-0.719	0.120	-0.119	1	-2 -				
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