

論文

## 統計的手法による『山海経』編著者の識別

下西紀子\*

## 00. はじめに

『山海経』は概括的に中国古代の地理書とされる。現行の『山海経』は全十八編あり、山経と海経に大別されるが、山経は五蔵山経とも称され五編から成り、海経は海外四経が四編、海内四経が四編、大荒四経が四編、海内経が一編で、十三編から成る<sup>1</sup>。

この五蔵山経と海経では記載内容が大きく異なる。五蔵山経の記述には山岳の系統的つながりがあり、その個々の山々において、山名や方角、距離が示され、そこから流出する河川や産出される鉱物、生息する動植物などの記述がある。このような情報の他に、異形の動物や神、その祭祀方法、また神話伝承の記述が含まれている。海経の海外四経、海内四経では、周辺諸国と思われる国々の人や神、動植物、山岳、河川などの記述があり、主にそれらの奇怪な形状の特徴が記述されている。大荒・海内経では神話伝承の様相が濃くなり、また人々の姓や食が記述され、帝王に結びつく系譜が示されている。その他、海経には古くから絵画説があり、絵図を解説する図解書としての見方もある<sup>2</sup>。

このような記載内容や文体などの違いから、それぞれの由来が異なり、成立に関する時期、地域、編著者も異なると考えられている。しかし、この成立に関する見解には様々な議論があり、現在でも定説がない<sup>3</sup>。『山海経』研究史については（岡本 1960）、（小南 1987）に整理され、成立に関する見解の相違をうかがえる<sup>4</sup>。

近代以降の中国では、（陸侃如 1928）、（玄珠（茅盾） 1929）、（顧頡剛 1934）、（蒙文通 1962）、（袁珂 1978）などの注目される研究があり<sup>5</sup>、例えば袁珂は、大荒・海内経が最も早く、戦国初中期。五蔵山経と海外四経が戦国中期以降。海内四経が最も遅く、漢代初期とし、また戦国の楚の地域で成立したとする<sup>6</sup>。

日本では、（小川 1928）、（伊藤 1969）、（高馬 1969）、（前野 1975）などがあり<sup>7</sup>、例えば小川は、五蔵山経が最も古く、東周（春秋）の洛陽で成立し、海外四経が戦国、海内四経が前漢で加わり、大荒・海内経は『山海経』の最初の注釈者である晋代の郭璞（276-324）までに付加されたとする<sup>8</sup>。

このように成立に関して議論があるが、共通見解としては『山海経』が異なる編著者によって段階的に形成されたことである。本稿では、この編著者が異なるという想定を踏まえて、五蔵山経と海経の編著者、また各編における編著者の識別を試みたい。本研究は『山海経』の成立に焦点を絞るものではないが、コンピュータなどのデジタル技術を用いて文献の情報を数値化、あるいは視覚化、つまり情報を変換することで別の視点をもたらす可能性を意図している。よって積極的な手法の活用を重視しており、本稿では統計的手法を用いて編著者識別を試みる。

この統計的手法による著者推定や識別は、計量言語学における計量文献学、計量文体学などで功績がある。1851年、イギリスの数学者、ド・モルガン（Augustus de Morgan 1806-1871）が『新約聖書』の「パウロ書簡」の著者推定における統計的手法を用いた提案が始まりとされ<sup>9</sup>、統計的手法による著者推定は早くから検証されてきた。近年では、情報システムの普及により、コンピュータを用いた統計的テキスト解析による著者推定などへの展開が見られる。

---

キーワード：山海経、海経、N-gram モデル、統計的手法、編著者識別

\* 立命館大学大学院先端総合学術研究科 2006年度入学 表象領域

すでに前節で挙げた『山海経』の成立に関する諸研究では、これまで『山海経』を校定した前漢末の劉秀（劉歆）（前53-23頃）による序録「上山海経表」や署名、経文の記載内容や叙述の形式、郭璞をはじめとした諸注釈<sup>10</sup>、他文献との比較などによって考察されてきた。『山海経』編著者識別に関して、統計的手法を用いた事例は見当たらない。統計は科学的な分析手法であり、客観的なデータが得られる。しかしデータの最終的な判断には個人の主観が関わり、完全な客観ではないという限界もある。とはいえ、その個人に妥当な判断を下す手掛かりを与えるものであり、『山海経』編著者識別にも統計的手法は有効であると考えられる。

尚、近年では海経各編における成立の多層構造、時代を経た段階的な書き換えが指摘されており<sup>11</sup>、ひとつの編に複数の編著者が推定される。本稿では段階的な書き換えがあったにせよ、主要な編著者の癖のような痕跡が残ると考え、編著者識別の初期的段階として、まずは各編における主要な編著者の相違を検証することを目的とする。

## 01. 研究手法について

### 01.01. テキストデータの前処理

『山海経』のテキストデータは「諸子百家 中国哲学書電子化計画」サイトに公開される、(晋) 郭璞伝『山海経』四部叢刊初編を用いた<sup>12</sup>。以下はテキストデータの前処理である。

【漢字の修正】：(晋) 郭璞伝『山海経』四部叢刊初編の写真画像データに基づき、テキストデータの誤字脱字などを修正。但し、異体字は未処理。

【句読点、記号などの削除】：白文（原文）に復元。

【(前漢末) 劉秀の署名の削除】：海外東経、海内東経の末尾に『山海経』を校定した劉秀などの署名があり、その署名を削除。

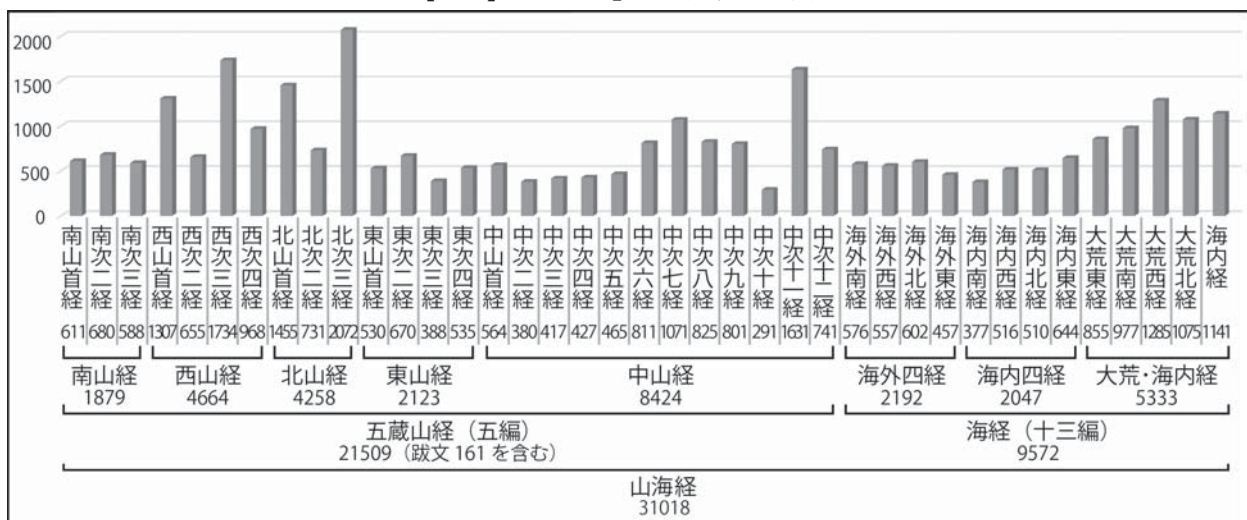
【文字コード】：UTF-8 (Unicode : CJK 統合漢字拡張 B)

### 01.02. 『山海経』の構成と文字数について

(図1)のように『山海経』十八編は五蔵山経の五編と海経の十三編からなるが、五蔵山経の五編には細目二十六編があり、南山経：三編、西山経：四編、北山経：三編、東山経：四編、中山経：十二編で構成される。海経の海外四経、海内四経、大荒・海内経は便宜的な総称で、海外四経は海外南経、海外西経、海外北経、海外東経の四編、海内四経は海内南経、海内西経、海内北経、海内東経の四編、大荒・海内経は大荒東経、大荒南経、大荒西経、大荒北経の四編と海内経の一編を加えた五編で構成される。

編数では海経が多いが、文字数で比較すると全体の約70%が五蔵山経(21509字)で、約30%が海経(9572字)

【図1】『山海経』の構成と文字数



となり、明らかに五蔵山経の文字数が多い。文字数的な比較対象としては、五蔵山経の五編と海経の総称三編である海外四経、海内四経、大荒・海内経、また五蔵山経の細目二十六編と海経の十三編が適合しており、本研究では文字数を優先した組合せで検証を行う。

### 01.03. テキストデータの計量化

テキストの統計的処理のために、テキストから計量可能なデータ抽出が必要となるが、コンピュータによる自然言語処理技術のひとつに形態素解析ツールがある。中国語では現代中国語に対応する、ICTCLAS (NLPIR) という中国語形態素解析ツールが開発され活用されているが<sup>13</sup>、一般利用できる中国古典に対応する形態素解析ツール、また形態素解析辞書は見あたらない<sup>14</sup>。本研究で扱うテキストは中国古典であるため、これらの形態素解析ツールがストレートに使用できない。これにより言語的な特性に依存せず、テキストを単語に分割し、計量化する N-gram モデルを用いることとした。

#### 01.03.01. N-gram モデル

N-gram モデルは自然言語処理における、単語の結びつきを統計的に処理する手法で、テキストを指定した文字数ごとに、一文字ずつずらしながら順番に分割し、文字の共起（出現）、共起頻度を求める。技術的に N-gram 単位でテキストが分割されるため、日本語や中国語といった言語的な特性に依存しない特徴があり、古典などのテキスト解析にも用いられる。

N-gram モデルを用いた日本の古典分析では（近藤 2000）などがあり、中国古典では（山田 2001）など、その他、多数の論稿がある<sup>15</sup>。

#### 01.03.02. N-gram モデルのノイズ処理

N-gram モデルの問題として、テキストの全ての文字を対象として文字が組み合わされるため、ノイズと呼ばれる意味を持たない文字の共起が発生する。このノイズ処理には、分析者の判断において情報を絞り込み、有効とみなされる文字の共起を抽出する方策があげられる<sup>16</sup>。この場合、分析者の視点が重要となるが、計量文体学などでは、文の長さ、単語の長さ、単語や品詞の共起頻度、読点などに着目することで厳選された情報、つまり著者の特徴情報として分析に用いられてきた。著者の特徴情報として日本語では、文章の内容と最も依存性が低い助詞が有効との指摘がある<sup>17</sup>。また英語の機能語、つまり意味を持たない非自立語で文法的な役割を有する語として、前置詞、接続詞、助動詞、冠詞などがあるが、日本語の助詞、助動詞などは、この機能語に相当するとされる。さらに中国語では実詞（実字）と虚詞（虚字、助字）があり、虚詞は機能語と考えられる<sup>18</sup>。日本語で有効とされる助詞を機能語として見ると、中国語の機能語となる虚詞が文章の内容との依存性が低く、著者の特徴情報として有効であると考えられる。実際に（李 1987）は、『紅樓夢』の虚詞として 47 字を抽出し、虚詞の共起頻度によるクラスタ分析や主成分分析などの統計的手法を用いて著者推定を行っている<sup>19</sup>。これにより N-gram モデルのノイズ処理、つまり著者の特徴情報の抽出として虚詞を用いることにした。

要するに本研究で用いるテキストは中国古典であり、形態素解析ツールによる単語の区切り、品詞の抽出には問題を含むため、N-gram モデルをベースとして単語に区切り、文字の共起、共起頻度を集計する。また N-gram モデルで生じるノイズ処理、つまり著者の特徴情報として虚詞を抽出することで、統計的処理に用いるデータを取得する。

### 01.04. 統計的手法と手順

- 1) テキストを 1gram ~ 6gram 単位で単語に区切り、それぞれの平均値からの誤差である標準偏差の比較により、どの単位が相関分析の分析データとして有効かを検証する。
- 2) 分析データ (2gram) を用い、変数間の関連性を検証する相関分析を行う。変数を五蔵山経と海経、また五蔵山経の五編、海経の総称三編を対象とし、関連性を検証する。
- 3) 1gram をベースに虚詞を抽出し、分析データとする。漢字は表語文字と呼ばれ、形態素の多くが 1 文字で構成

される。そのため他言語なら意味のない 1gram 単位の共起頻度でも有効な結果が得られるとされ<sup>20</sup>、ノイズ処理、つまり著者の特徴情報として虚詞を抽出するが、そのベースとして 1gram 単位の共起頻度を用いる。

- 4) 虚詞を対象とし、階層クラスタ分析を行う。クラスタ分析は類似したデータをもつ変数をグループに分類する統計的手法で、データの客観的な分類が可能となる。五蔵山経の五編と海経の総称三編、また海経十三編を対象にクラスタ分析による分類を行う。
- 5) 虚詞を対象とし、判別分析を行う。判別分析は既存の変数に基づき、所属するグループを判定する統計的手法で、海経十三編、また五蔵山経の細目二十六編を加え、クラスタ分析で分類した結果を統計的に評価する。

尚、N-gram モデルのツールとして「MTMineR」を使用する<sup>21</sup>。また虚詞の抽出には、中国語形態素解析ツール「ICTCLAS (NLPPIR)」、中国語哲学書電子化計画サイトの辞書機能、その他、虚詞辞典などを複合的に用いる<sup>22</sup>。統計的分析には「Microsoft Excel」および「Seagull-Stat」を利用する<sup>23</sup>。

## 02. テキスト全体の 2gram 分析結果

### 02.01. N-gram 単位の検証

山海経のテキスト全体、および五蔵山経と海経に分離したテキストを対象とし、それぞれ 1gram ~ 6gram 単位で単語を区切り、その共起頻度を集計した。(表 1) は上位 5 位までの共起頻度である。また (表 2) (表 3) (表 4) は共起数、共起頻度合計、平均、分散、標準偏差などの基礎統計量を示したものである。

共起頻度の平均値を比較すると、1gram の山海経では 15.4、五蔵山経では 14.2、海経では 8.42 であり、15 ~ 8 単語と少なく、2gram では山海経 2.73、五蔵山経 3.25、海経 1.75 で、2 ~ 3 単語となり、3gram 以降では、1 ~ 2 単語と殆ど共起がない。

次に標準偏差を確認すると、1gram では山海経 72.5、五蔵山経 67.3、海経 26.3 と比較的大きく、2gram では山海経 11.6、五蔵山経 14.0、海経 3.45 となり、3gram では山海経 5.47、五蔵山経 6.93 となるものの、海経は 1.25 と小さく、4gram 以降では全体的に小さい値になる。

標準偏差をプロットした折れ線グラフ (図 2) (図 3) (図 4) で、この標準偏差の変化を見ると、1gram が突出し、4gram あたりから、5gram、6gram とグラフが平坦になっている。標準偏差は数値が大きいほど、ばらつきが大きく、それぞれの差が大きいことを示している。つまり差が大きいほど特徴があると言える。最も差があり特徴があるのは 1gram となるが、N-gram モデルの場合、1gram は文字の出現頻度と重なるため、差が突出しているのは当然であり、データの偏りには偏ったものと見られる。また N-gram モデルの標準偏差は、1 という値が最低値になるため、標準偏差の 2 以下の値は差が少なく、特徴がないと言える。

【表 1】 共起頻度 (a: 山海経 / b: 五蔵山経 / c: 海経)

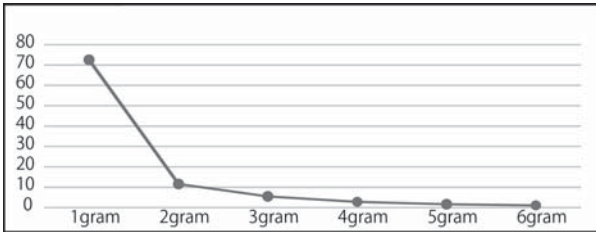
上位 5 位

	1gram	共起頻度	2gram	共起頻度	3gram	共起頻度	4gram	共起頻度	5gram	共起頻度	6gram	共起頻度
a	其	1599	里曰	409	水出焉	242	焉其狀如	139	有獸焉其狀	80	有獸焉其狀如	79
2	之	1101	之山	378	流注于	236	水出焉而	137	獸焉其狀如	79	之山無草木多	40
3	多	1033	名曰	311	十里曰	212	山其上多	122	之山其上多	76	有鳥焉其狀如	37
4	曰	868	水出	299	其狀如	209	之山其上	92	水出焉而東	53	出焉而東流注	34
5	山	867	出焉	242	百里曰	173	有獸焉其	82	之山無草木	52	水出焉而東流	34
b	其	1426	里曰	409	流注于	236	焉其狀如	139	有獸焉其狀	80	有獸焉其狀如	79
2	多	1030	之山	334	水出焉	227	水出焉而	136	獸焉其狀如	79	之山無草木多	40
3	之	713	水出	257	十里曰	212	山其上多	122	之山其上多	76	有鳥焉其狀如	37
4	山	639	注于	242	其狀如	199	之山其上	92	水出焉而東	53	出焉而東流注	34
5	曰	636	山其	238	百里曰	173	有獸焉其	81	之山無草木	52	水出焉而東流	34
c	有	440	名曰	126	大荒之	32	大荒之中	29	大荒之中有	27	荒之中有山名	22
2	之	388	之國	77	有山名	32	有山名曰	28	中有山名曰	23	之中有山名曰	21
3	曰	232	一曰	68	之中有	30	荒之中有	28	之中有山名	22	大荒之中有山	21
4	山	228	國在	48	荒之中	30	中有山名	24	荒之中有山	22	在其東其為人	6
5	在	190	之山	44	山名曰	29	之中有山	22	在其東其為	7	國在其東其為	5

【表 2】『山海経』基礎統計量 有効数字3桁

山海経	1gram	2gram	3gram	4gram	5gram	6gram
共起数	2017	11375	18625	23081	25967	27825
共起頻度合計	31081	31080	31079	31078	31077	31076
平均	15.4	2.73	1.67	1.35	1.20	1.12
分散	5261	134	29.9	7.79	2.71	1.20
標準偏差	72.5	11.6	5.47	2.79	1.65	1.20

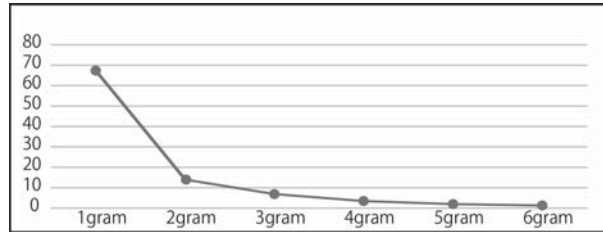
【図 2】『山海経』標準偏差グラフ



【表 3】『五蔵山経』基礎統計量 有効数字3桁

五蔵山経	1gram	2gram	3gram	4gram	5gram	6gram
共起数	1519	6614	10974	14349	16825	18509
共起頻度合計	21509	21508	21507	21506	21505	21504
平均	14.2	3.25	2.00	1.50	1.28	1.16
分散	4531	195	48.0	12.1	4.00	1.72
標準偏差	67.3	14.0	6.93	3.47	2.00	1.31

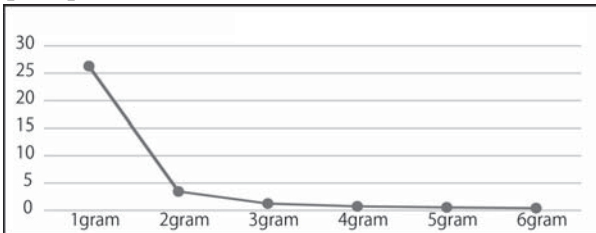
【図 3】『五蔵山経』標準偏差グラフ



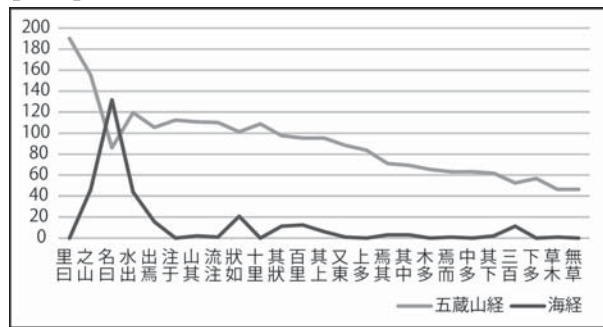
【表 4】『海経』基礎統計量 有効数字3桁

海経	1gram	2gram	3gram	4gram	5gram	6gram
共起数	1137	5457	7861	8787	9150	9314
共起頻度合計	9572	9571	9570	9569	9568	9567
平均	8.42	1.75	1.21	1.09	1.05	1.03
分散	691	11.9	1.55	0.552	0.291	0.168
標準偏差	26.3	3.45	1.25	0.740	0.540	0.410

【図 4】『海経』標準偏差グラフ



【図 5】五蔵山経と海経の相関グラフ 上位 25 位



r=0.1998

このように見ると、1gram、5gram、6gramは分析対象から外れることになる。値が2以上となるのは2gram、3gram、4gramであるが、海経では3gram、4gramも2以下の値となり、これも分析対象から外れる。結果的に相関分析の分析データとして2gramが最も有効となる。

## 02.02. 相関分析

ここでは、2gramで抽出した単語の共起頻度から、五蔵山経と海経の相関分析を行い、その関連性を検証する。また五蔵山経を構成する五編、海経を構成する総称三編でも相関分析を行い、各編の相関性を確認し、編著者識別の予備分析とする。

### 02.02.01. 五蔵山経と海経の相関分析

五蔵山経と海経の相関分析にあたって、まず両者の共起頻度の値を1万語あたりの出現数に正規化した。(図5)は五蔵山経と海経の上位25位までの相関性を折れ線グラフにプロットしたもので、このグラフでは「名曰」のみ五蔵山経と海経に接点があり、その他は殆ど関連性が無いと確認できる。

ピアソンの積率相関係数rでは五蔵山経と海経はr=0.1998と計測された。rは $-1 \leq r \leq 1$ であり、1に近い場合は正の相関、-1に近い場合は負の相関、0に近い場合は無相関となる。ひとつの判断基準として0.2以下は無相関とされ、五蔵山経と海経の相関係数は0.1998であり、無相関と認められる。つまり五蔵山経と海経は別のテキストであり、編著者が異なる可能性がある判断できる。

02.02.02. 五蔵山経：五編の相関分析

五蔵山経と海経では無相関が示され、編著者が異なる可能性が示唆された。よって分析対象を五蔵山経、海経に分離し、五蔵山経、および海経を構成する各編を対象として、相関性を検証する。

まず五蔵山経の五編の相関分析を確認する。南山経、西山経、北山経、東山経、中山経の五編を2つの変数として組み合わせ、相関分析を行った。組み合わせは10パターンで、それぞれの相関係数を(表5)、散布図を(図6)に示す。(図6)の散布図では、図の回帰直線が示すように全体的に右上がりになっている<sup>24</sup>。(表5)に示される相関係数では、すべての値が0.6以上であり、五蔵山経の五編は互いに強い正の相関があると考えられる。

相関係数により相関があると判断された場合、無相関検定でその相関係数が信頼できる値か、その有意性を問う検定を行う必要があるため、以下の無相関検定を行った。

無相関検定に次の仮説を設定する。帰無仮説：2つの変数間に相関がない。対立仮説：2つの変数間に相関がある。また有意水準  $\alpha=0.05$  に定めた<sup>25</sup>。

結果は(表6)の無相関検定に示されるが、10パターンすべてのp値は  $p=0.000$  であった。p値は帰無仮説が正しいとする確率であるが、その確率は0.05未満であり、5%の有意水準で有意差が認められた。よって帰無仮説を棄却し、10パターンすべての2つの変数間に相関があることが補強された。つまり五蔵山経の五編は同一の編著者である可能性が示唆されたことになる。

■相関分析結果

【表5】相関行列

\*:5%有意(>0.024) \*\*:1%有意(>0.032)

五蔵山経	南山経	西山経	北山経	東山経	中山経
南山経	1	0.735**	0.764**	0.776**	0.758**
西山経	0.735**	1	0.820**	0.709**	0.796**
北山経	0.764**	0.820**	1	0.770**	0.762**
東山経	0.776**	0.709**	0.770**	1	0.675**
中山経	0.758**	0.796**	0.762**	0.675**	1

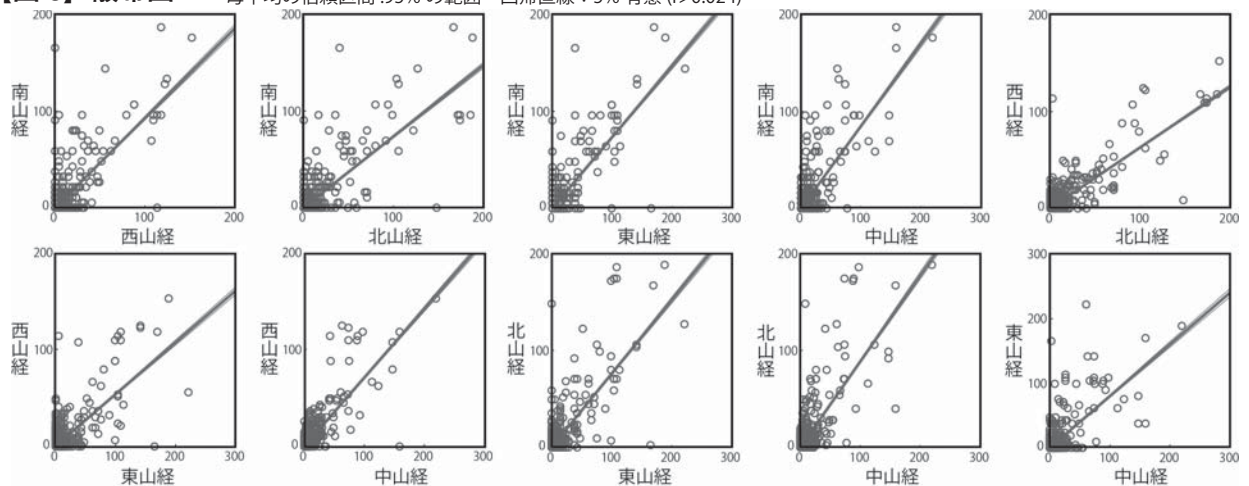
【表6】無相関検定

上p値/下判定(\*:5%有意 \*\*:1%有意)

五蔵山経	南山経	西山経	北山経	東山経	中山経
南山経	—	0.0000	0.0000	0.0000	0.0000
西山経	**	—	0.0000	0.0000	0.0000
北山経	**	**	—	0.0000	0.0000
東山経	**	**	**	—	0.0000
中山経					—

【図6】散布図

母平均の信頼区間:95%の範囲 回帰直線:5%有意( $r>0.024$ )



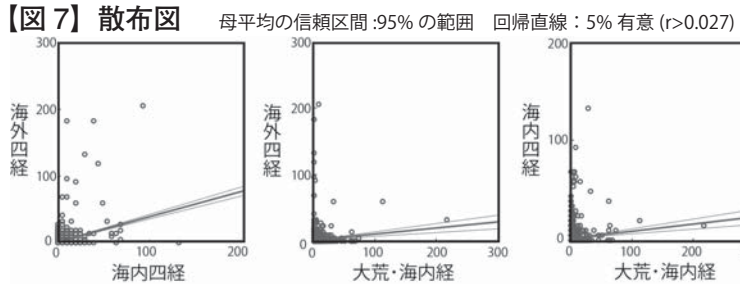
02.02.03. 海経：総称三編の相関分析

海経を構成する総称三編は、海外四経、海内四経、大荒・海内経で、この三編を2つの変数として組み合わせ相関分析を行った。組み合わせは3パターンで、それぞれの相関係数を(表7)、散布図を(図7)に示す。(表7)に示される海外四経と海内四経の相関係数は0.281であり、相関の判断基準では弱い正の相関があると考えられる。海外四経と大荒・海内経は0.0700、海内四経と大荒・海内経は0.075で、相関の判断基準では無相関となる。

海外四経と海内四経について弱い正の相関があると判断されたため、五蔵山経と同様の仮説を設定し、無相関検定を行った。(表8)の無相関検定のp値はいずれも  $p=0.000$  で、帰無仮説は棄却され、5%の有意水準で有意差が認められた。

**【表7】 相関行列** \*5%有意 (>0.027)  
\*\*1%有意 (>0.035)

海経	海外四経	海内四経	大荒・海内経
海外四経	1	0.281**	0.070**
海内四経	0.281**	1	0.0750**
大荒・海内経	0.070**	0.0750**	1



**【表8】 無相関検定** 上 p 値 / 下判定  
(\*5%有意 \*\*1%有意)

海経	海外四経	海内四経	大荒・海内経
海外四経	—	0.0000	0.0000
海内四経	**	—	0.0000
大荒・海内経	**	**	—

海経の総称三編の相関分析では、海外四経、および海内四経と大荒・海内経が無相関であり、編著者が異なることが示唆され、また海外四経と海内四経では弱い正の相関が認められたが、異なるか否かは判断ができない。弱い相関であるため編著者が異なる可能性も考えられる。

相関分析によって、五蔵山経と海経、また五蔵山経の五編、海経の総称三編について編著者の分析を行ったが、五蔵山経と海経では無相関、つまり編著者が異なる可能性、五蔵山経の五編では、すべての組み合わせで強い正の相関が認められ、同一の編著者である可能性がある。海経の総称三編では海外四経、および海内四経と大荒・海内経が無相関で編著者が異なる可能性、海外四経と海内四経では弱い正の相関であるため、判断保留という結果となった。

### 03. 虚詞の 1gram 分析結果

相関分析では、海外四経と海内四経が判断保留となり、海経の総称三編の編著者がすべて異なる可能性も示唆された。これは相関分析による結果であり、さらに分析対象を機能語である虚詞に絞り、階層クラスタ分析、判別分析を用い、編著者識別の検証を行う。

#### 03.01. 階層クラスタ分析

階層クラスタ分析は集団の中から似通ったデータを持つ変数をグループに分類する統計的手法である。この手法により、1gram の共起頻度から抽出した 57 字の虚詞を対象とし (表 9)、五蔵山経の五編と海経の総称三編、また海経十三編の分類を行う。分析結果から各編の編著者の識別を考察する。

##### 03.01.01. 五蔵山経：五編と海経：総称三編の階層クラスタ分析

クラスタ分析では分類する対象の距離を数量的に定義し、距離が最も近い変数を順番にグループに分類しながら、階層構造が構築されるが、その最初の変数間の非類似度の計算 (距離計算) にユークリッド平方距離<sup>26</sup>、またクラスタ間の非類似度の計算にはワード法<sup>27</sup>を用いた。

クラスタ分析の結果としてプロットされたデンドログラム (樹形図) (図 8)、あるいは要約表 (表 10) からクラスタが結合される過程を確認すると、最初に南山経と東山経がクラスタを形成し、そのクラスタに西山経、次に北

**【表9】 虚詞と共起頻度 (カッコ内は海経の共起頻度)**

虚詞	共起頻度	虚詞	共起頻度	虚詞	共起頻度	虚詞	共起頻度	虚詞	共起頻度	虚詞	共起頻度
1 其	1596(173)	11 也	81(18)	21 於	23(18)	31 盡	7(4)	41 因	3(2)	51 厥	1(1)
2 之	1093(388)	12 所	72(60)	22 始	18(18)	32 亦	7(3)	42 未	3(2)	52 向	1(1)
3 焉	470(42)	13 者	69(41)	23 方	15(15)	33 復	6(6)	43 甚	3(0)	53 卒	1(1)
4 而	463(37)	14 則	69(2)	24 與	11(10)	34 各	6(5)	44 已	2(2)	54 唯	1(1)
5 又	433(36)	15 皆	67(21)	25 惟	11(2)	35 故	5(4)	45 先	2(2)	55 既	1(1)
6 于	367(27)	16 凡	62(1)	26 相	10(9)	36 必	5(1)	46 維	2(2)	56 勿	1(0)
7 是	157(55)	17 至	51(11)	27 常	8(0)	37 嚮	4(4)	47 輒	2(1)	57 彼	1(0)
8 以	134(43)	18 愛	26(24)	28 從	7(7)	38 或	4(4)	48 當	1(1)		
9 不	122(47)	19 乃	25(18)	29 即	7(6)	39 莫	4(2)	49 非	1(1)		
10 自	88(19)	20 此	23(21)	30 及	7(4)	40 請	3(3)	50 遂	1(1)		

※「漢字を読むための助字小字典」の品詞分類に基づき、副詞、代名詞を含める。 ※跋文、名詞に含まれる文字は省く。

## ■階層クラスタ分析結果

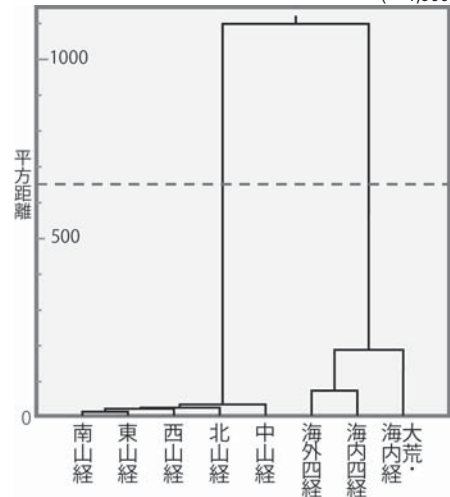
【表 10】 要約表 (ワード法)

(×1,000)

回	番号	名称	番号	名称	平方距離
1	1	南山経	4	東山経	11.261
2	[1]	南山経	2	西山経	19.806
3	[1]	南山経	3	北山経	22.581
4	[1]	南山経	5	中山経	31.629
5	6	海外四経	7	海内四経	69.62
6	[6]	海外四経	8	大荒・海内経	183.711
7	[1]	南山経	[6]	海外四経	1086.808

※ [] は既に結合した代表クラスタの名称を示す

【図 8】 デンドログラム (×1,000)



山経、さらに中山経が融合し、第1クラスタを形成した。その後に海外四経と海内四経がクラスタを形成、そのクラスタに大荒・海内経が融合し、第2クラスタを形成した。そして最後に五蔵山経となる第1クラスタと海経となる第2クラスタが結合し、五蔵山経と海経のクラスタは分離していると確認できる。

非類似度 (ユークリッド平方距離) では、まず五蔵山経クラスタでは最初に結合した南山経と東山経の 11.3 が最も近く、最後に結合した北山経と中山経は 36.9 で、五蔵山経ではこの距離が最も離れている。海経クラスタでは、海外四経と海内四経の非類似度は 69.6 であるが、五蔵山経で最も離れている 36.9 と比較すると、海外四経と海内四経はその約 2 倍離れている。また大荒・海内経と海外四経は 205 で、海内四経とは 105 であり、大荒・海内経は海外四経より、海内四経と距離が近く、大荒・海内経と両者との距離は約 2 倍の差がある。そして最後に結合した五蔵山経と海経クラスタの結合距離 (ワード法) は 1087 で、大きく離れていることが確認できる。

この結果から五蔵山経と海経はクラスタを分かれ、距離的にも大きく離れており、この距離を編著者の特徴とすると、五蔵山経と海経は編著者が異なると考えられる。また五蔵山経は海経に比べて距離が全体的に近く、五蔵山経は編著者の特徴が近いものと見る。

海経では海外四経と海内四経は少し距離があるが、これによって編著者が異なるかを決定できるものではなく、編著者の特徴として少し距離があるものとする。大荒・海内経の場合は、海外四経より海内四経との距離が近いと示されるが、基本的に両者とは別のクラスタを形成しており、編著者の特徴が異なると考えられる。

### 03.01.02. 海経：十三編の階層クラスタ分析

海経の総称三編では編著者が異なる可能性があるため、さらに海経を構成する十三編のクラスタ分析を行った。分析結果であるデンドログラム (図 9)、要約表 (表 11) を確認すると、まず第1クラスタを海外・海内経クラスタとして、第2クラスタを大荒・海内経クラスタとして、大別することができる。

形成された大荒・海内経クラスタで注目されるのは、本来、海外・海内経クラスタに属すべき、海内西経が含まれていることである。(表 11) では、最初に、つまり最も近い距離 29.4 で海内西経は大荒東経と結びついている。その他の海内西経の非類似度 (ユークリッド平方距離) を確認すると、大荒東経 29.4、大荒北経 62.6、大荒南経 74.8、海外西経 92.5、海外北経 96.7 となり、さらに他の編とは値が 100 以上離れ、最大は海外南経 380 である。海外四経や海内四経より、大荒・海内経の各編との距離が近いことが示されている。この結果から海内西経と大荒・海内経の編著者が同一とは判断できないが、海外四経と海内四経の編著者が異なるとする根拠になると考えられる。

クラスタ分析結果では五蔵山経と海経は分離し、さらに海経は海外四経、および海内四経と大荒・海内経が分離してきたが、海外四経と海内四経の分離は曖昧である。両者の各編は混在し、近い距離で結びつくものもあれば、海外南経や海内東経のように少し距離があるが、最終的にひとつのクラスタを形成するものもある。編著者の相違

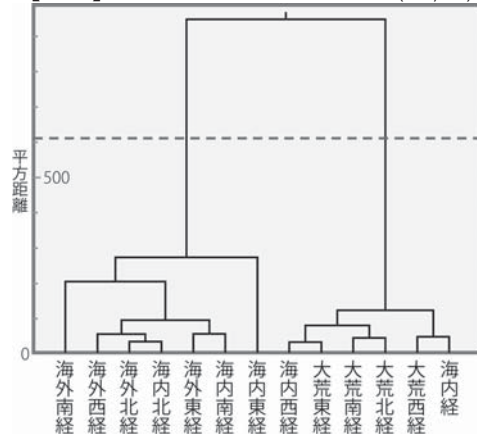


【表 11】 要約表 (ワード法) (×1,000)

回	番号	名称	番号	名称	平方距離
1	6	海内西経	9	大荒東経	29.434
2	3	海外北経	7	海内北経	30.901
3	10	大荒南経	12	大荒北経	41.952
4	11	大荒西経	13	海内経	44.792
5	2	海外西経	[3]	海外北経	52.676
6	4	海外東経	5	海内南経	52.826
7	[6]	海内西経	[10]	大荒南経	76.723
8	[2]	海外西経	[4]	海外東経	91.789
9	[6]	海内西経	[11]	大荒西経	119.557
10	1	海外南経	[2]	海外西経	201.212
11	[1]	海外南経	8	海内東経	269.963
12	[1]	海外南経	[6]	海内西経	944.245

※□は既に結合した代表クラスタの名称を示す

【図 9】 デンドログラム (×1,000)



をこの結果から判断できないが、海内西経が大荒・海内経クラスタにクラスタリングされた結果がさらに強固となれば、海外四経と海内四経の編著者が異なる可能性を補強するものとなる。よって、さらに他の分析手法として判別分析を用い、海内西経の所属を検証する。

### 03.02. 判別分析

海経十三編の階層クラスタ分析によって、本来は海内四経に属すべき海内西経が、大荒・海内経クラスタにクラスタリングされた。海内西経が実際に大荒・海内経に属すれば、海外四経と海内四経の編著者が異なる指標となる。よってクラスタ分析で得た結果を判別分析でも検証する。判別分析は、外的基準がある分類方法で、幾つかのグループで得られる変数に基づき、サンプルがどのグループに所属するかを判別する手法であり、統計的な評価をもった検討が可能となる。

判別分析の設定として、海経十三編を目的変数のサンプルとし、判別前のグループを海外四経、海内四経、大荒・海内経とする3グループを定めた。目的変数を説明する説明変数は虚詞の共起頻度となるが、その際の目的変数のサンプル数は「説明変数の数+1」以上の関係にする必要がある。目的変数のサンプル数は13であるため、説明変数は虚詞の共起頻度、上位11位までを用いた。

また分析結果を補強する目的で、目的変数に五蔵山経の細目二十六編、判別前のグループに五蔵山経を加え、説明変数を虚詞の共起頻度、上位37位までを用いた分析も行った。

#### 03.02.01. 海経：十三編の判別分析

判別手法を変数増減法とし、偏F値：F=2を統計量の基準とした。変数増減法はステップワイズ法とも呼ばれ、基準となる偏F値に従って説明変数の追加、除去を判断し、判別に有効な説明変数の組み合わせを決定する方法である。

海経十三編を対象とした判別結果として、変数増減法による変数選択は最終的に「ステップ4」となり、「其、是、不、于」という4つの虚詞が判別に有効な説明変数として導入された(表12)。個々の説明変数の有意性検定では、偏F値の基準統計量は $F(2,5,0.01)=13.27$ であり、自由度2、5で有意水準1%の場合、偏F値の基準統計量は13.27であるが、この値以上に該当する変数はなかった。有意水準5%にすると、偏F値の基準統計量は5.79となり、この値以上の変数は「是」のみで、検定では導入された説明変数が判別にあまり寄与していないと示された。その他の検定統計量は、 $Wilks-\Lambda=0.042$ <sup>28</sup>、 $F_0=6.32$ 、 $F(8,13,0.01)=4.30$ 、 $p=0.001873$ であり、有意水準1%で、説明変数の有意性が示されている(表12)。

変数によって生成された判別関数から判別得点が計算され、それが最大値を与える群に所属が判別される(表13)。その判定結果として(表14)に判別前と判別後の所属が示される。第1群(海外四経グループ)、第2群(海内四経グループ)、第3群(大荒・海内経グループ)すべて所属の移動がなく、総合100%の正判別率となった。

クラスタ分析により大荒・海内経にクラスタリングされた海内西経は判別分析では第2群(海内四経グループ)

## ■判別分析結果

【表 12】判別関数 (最終 Step4)

説明変数	偏F値	第1群	第2群	第3群
其	5.330261	0.043257	0.013484	0.023391
是	12.98445	0.088092	0.011151	0.243717
不	5.186203	0.133378	0.035046	0.201457
于	2.691031	-0.10227	0.012347	-0.15714
定数項		-11.4308	-1.93976	-16.6117

Wilks- $\Lambda$ =0.042  
 $F_0=6.32$   
 $F(8,13,0.01)=4.30$   
 $p=0.001873$   
 $F(2,5,0.01)=13.27$   $F(2,5,0.05)=5.79$

【表 13】判別得点

目的変数	グループ	第1群	第2群	第3群
海外西経	1	7.337336	3.676023	-1.81895
海外東経	1	8.215409	4.652172	-1.09582
海外南経	1	19.75176	8.97569	3.450173
海外北経	1	10.41885	6.42419	3.018744
海内西経	2	-3.21318	1.225782	-9.12873
海内東経	2	-10.5207	-0.38399	-17.274
海内南経	2	0.75029	5.684388	-10.4278
海内北経	2	-1.25265	1.232871	-11.1079
海内経	3	5.657959	2.352036	24.80347
大荒西経	3	2.759359	2.249623	12.97496
大荒東経	3	2.079639	2.123337	4.316627
大荒南経	3	8.922258	2.801572	18.35007
大荒北経	3	10.92764	3.909991	22.61323

【表 14】判別結果

前/後	第1群	第2群	第3群	正判別率
第1群	4	0	0	100.0%
第2群	0	4	0	100.0%
第3群	0	0	5	100.0%
総合				100.0%

と判別され、所属は変わらなかった。これにより海経の総称三編の編著者がすべて異なる可能性が示唆されたが、個々の説明変数の有意性検定で、有意性が示されなかったという問題が残る。

## 03.02.02. 五蔵山経：細目二十六編を加えた判別分析

最後にこの結果をさらに検証するために、説明変数を追加し、検証データを補強する。まず目的変数として五蔵山経の細目二十六編を加えサンプル数を39とし、判別前のグループを五蔵山経、海外四経、海内四経、大荒・海内経の4グループを設定した。目的変数の追加によって、説明変数も追加可能となり、説明変数を共起頻度、上位37位までの虚詞を対象として、同じ条件で判別分析を試みた。

変数増減法による変数選択は最終的に「ステップ17」となり、「其、焉、而、又、不、也、皆、至、爰、與、常、及、復、各、故」という15の虚詞が判別に有効な説明変数として導入された(表15)。個々の説明変数の有意性検定では、偏F値の基準統計量は $F(3,19,0.01)=5.01$ であり、自由度3、19で有意水準1%の場合、偏F値が5.01以上は「其、又、皆、至、爰、與、常、及、復、故」の変数で認められ、有意水準5%の場合は偏F値が3.13で「焉」が加わり、計11の変数が認められることになり、これらの変数の有意性が確認できる。またその他の検定統計量は、Wilks- $\Lambda=0.000$ 、 $F_0=33.65$ 、 $F(45,62,0.01)=1.89$ 、 $p=1.765E-17$ であり、有意水準1%で説明変数の有意性が示されている(表15)。

(表16)の判別得点、(表17)の判別結果から第1群(五蔵山経グループ)、第2群(海外四経グループ)、第3群(海内四経グループ)、第4群(大荒・海内経グループ)、すべて所属の移動がなく、総合でも100%の高い正判別率となった。変数を追加し、データを補強して分析した結果では、個々の説明変数の有意性が示されたうえで、所属の移動がないことが確認された。

クラスタ分析で海内西経が大荒・海内経クラスタにクラスタリングされたことから、海外四経と海内四経の編著者が異なる可能性を求めて、判別分析による検証を行った。判別分析の結果では、所属の移動は発生せず、最終的に海外四経と海内四経は分離し、編著者が異なる可能性があるということが示唆された。

## 04. 結語

『山海経』編著者の識別に関して、これまでの検証結果を踏まえると、五蔵山経と海経は編著者が異なり、五蔵山

【表 15】判別関数 (最終 Step17)

説明変数	偏F値	第1群	第2群	第3群	第4群
其	15.55972	0.197002	0.120025	0.048645	0.05126
焉	3.762367	0.204511	0.052374	0.000603	0.097753
而	2.932305	0.155183	-0.0421	0.080891	-0.08701
又	12.65253	0.37525	-0.00811	0.066283	0.055617
不	3.000525	0.144042	0.262644	-0.00384	0.283791
也	2.449386	0.390677	0.003645	0.069079	0.190946
皆	9.220356	-0.27516	-1.01569	0.221611	-0.91681
至	13.95251	1.051996	1.942636	0.053117	1.224907
爰	29.05091	-0.69326	2.095218	-0.38018	3.224865
與	20.2327	-0.9159	3.567725	-0.29163	3.231102
常	5.549218	2.34458	1.569295	0.186533	1.134993
及	10.4116	2.254722	-2.20981	0.928327	-4.23421
復	24.15356	2.305598	3.731209	-0.0716	13.48935
各	3.093551	0.049845	1.142995	0.083268	0.629222
故	17.3605	-0.43817	3.804845	-0.98337	14.46975
定数項		-157.558	-115.864	-15.5971	-203.876

F(3,19,0.01)=5.01 F(3,19,0.05)=3.13 Wilks-Λ=0.000  
 F<sub>0</sub>=33.65  
 F(45,62,0.01)=1.89  
 p=1.765E-17

【表 17】判別結果

前/後	第1群	第2群	第3群	第4群	正判別率
第1群	26	0	0	0	100.0%
第2群	0	4	0	0	100.0%
第3群	0	0	4	0	100.0%
第4群	0	0	0	5	100.0%
総合					100.0%

【表 16】判別得点

目的変数	グループ	第1群	第2群	第3群	第4群
西山首経	1	129.8546	-16.0305	45.94631	-139.524
西次三経	1	160.0595	19.87133	56.55691	-116.214
西次四経	1	130.3184	-32.8025	52.08127	-183.668
西次二経	1	154.2409	-6.42621	59.16982	-146.688
中山首経	1	171.0522	-0.5507	53.5579	-115.703
中次九経	1	157.8708	-2.77625	56.68944	-138.426
中次五経	1	117.7691	-18.63	41.64477	-130.428
中次三経	1	171.4945	8.770929	57.68311	-127.065
中次四経	1	179.585	6.157953	61.81911	-134.043
中次七経	1	144.9975	-25.462	54.60748	-132.289
中次十一経	1	188.4537	8.562761	51.5107	-115.137
中次十経	1	178.6744	3.379649	61.44806	-126.832
中次十二経	1	133.3357	-23.4234	53.71165	-158.01
中次二経	1	184.4846	-1.71579	68.59709	-145.529
中次八経	1	148.0465	1.818841	54.58965	-138.789
東山首経	1	155.3672	18.44019	54.327	-130.54
東次三経	1	173.0479	7.860704	54.57108	-129.635
東次四経	1	164.052	-0.86725	55.82182	-121.519
東次二経	1	168.5523	7.910987	57.31668	-125.791
南山首経	1	133.8133	-2.0083	45.62211	-133.355
南山首経	1	154.9951	33.39211	52.48415	-106.386
南次三経	1	158.3469	29.47739	53.17383	-109.587
南次二経	1	153.8305	-20.6213	54.70728	-143.848
北山首経	1	155.257	-19.3517	57.83492	-147.214
北次三経	1	161.493	-6.35143	52.38607	-78.9566
北次二経	1	167.5116	-7.27977	59.56928	-136.843
海外西経	2	-55.7044	110.8468	3.487837	-26.2272
海外東経	2	-36.7472	119.988	11.08158	-29.7851
海外南経	2	-53.9068	127.9083	7.249613	12.25852
海外北経	2	-26.3643	104.7124	4.488531	-29.1527
海内西経	3	-77.71	-77.6662	27.24291	-206.935
海内東経	3	-116.667	-113.456	5.621062	-218.694
海内南経	3	-54.4205	-96.2362	21.34126	-210.607
海内北経	3	-99.4412	-87.4014	8.183072	-202.361
海内経	4	-80.1413	68.07133	-22.0361	198.7891
大荒西経	4	-96.3862	76.03149	-22.103	242.832
大荒東経	4	-78.1826	75.29939	-19.1519	223.0245
大荒南経	4	-86.6841	73.89323	-21.8812	189.8248
大荒北経	4	-83.1748	55.63016	-21.6811	164.9077

経の五編は同一の編著者である可能性がある。海経では海外四経、および海内四経と大荒・海内経では編著者が異なり、また海外四経と海内四経も編著者が異なる可能性が示唆された。

検証の諸段階において、五蔵山経と海経、また海外四経、および海内四経と大荒・海内経の編著者の相違については比較的安定的な結果がもたらされたが、海外四経と海内四経は検証結果を得るため、複数の分析を重ねる必要があった。

この海外四経、海内四経に改めて着目すると、両者には幾つかの共通項があることがわかる。まず海外四経の最終編である海外東経、および海内四経の最終編である海内東経の末尾に「建平元年四月丙戌、待詔太常屬臣望校治、侍中光祿勳臣龔、侍中奉車都尉光祿大夫臣秀領主省。」という劉秀らの署名がある。現行の『山海経』には前漢末の劉秀による「上山海経表」が付載され、そこに彼が『山海経』を校定した旨の記載があり、さらに海外東経、海内東経の末尾にも建平元年（前6）に劉秀らが校定したとの署名がある。この署名は海外東経と海内東経の末尾にのみ記載され、五蔵山経および海経の他編には見られない。また海外四経、海内四経には「一曰」が頻繁に記載されているという特徴がある。清代の畢沅（1730-1797）は、この「一曰」について「凡一曰云云者、是劉秀校此経時附著所見他本異文也。舊亂入経文、當由郭注此経時升為大字。」と注釈を加え<sup>29</sup>、劉秀による校定時に他のテキストの異文が加えられ、晋代の郭璞が注釈を施した際に小文字の異文を大文字に書き直したため、経文に乱入したとしている。郭璞の注釈が乱入の要因であるかは定かではないが、海外東経と海内東経の末尾にのみ劉秀らの署名があるのは、海外四経と海内四経における校定の意義を示すものと推測されている<sup>30</sup>。

また海経には絵画説があり、海経は絵画を説明するテキストとの指摘がある。郭璞による注釈に「畫亦似仙人也（画

は仙人に似たるなり)：海外南経]、「畫似獼猴而黑色也。(画は獼猴に似て黑色なり)：海外南経]、「畫四面各乘靈車、駕二龍(四面に各々靈車に乗り、二龍に駕するを画く)：海内北経]」などが示され、その他にも絵画的痕跡を示す記述や「山海図」「山海経図賛」などの存在を記載する文献があることなどが、その根拠となっている<sup>31</sup>。絵画的痕跡を示す記述のなかで、海外四経の冒頭に「海外自西南陬至東南陬至者(海外の西南陬より東南陬に至る者)：海外南経]」、「海外自西南陬至西北陬者(海外の西南陬より西北陬に至る者)：海外西経]」、「海外自東北陬至西北陬者(海外の東北陬より西北陬に至る者)：海外北経]」、「海外自東南陬至東北陬者(海外の東南陬より東北陬に至る者)：海外東経]」とある。これらは記述の範囲を提示したもので、方形を成す絵画の痕跡とされる。海内四経の冒頭にも「海内東南陬以西者(海内の東南陬以西の者)：海内南経]」、「海内西南陬以北者(海内の西南陬以北の者)：海内西経]」、「海内西北陬以東者(海内の西北陬以東の者)：海内北経]」、「海内東北陬以南者(海内の東北陬以南の者)：海内東経]」のような範囲を示す記述があり、海内四経にも方形を成す絵画の痕跡がある。このように海外四経と海内四経には、末尾に劉秀らの署名、「一曰く」の記載、また記述の範囲を示す形式などの共通項があり、同一の編著者と捉えられるような特徴が際立っている。

しかし近年の日本の研究には両者の相違を明示したのものもある。(竹内 1991)<sup>32</sup>は、海外四経と海内四経における叙述形式や「一曰」の分布を検証し、海外四経にはある種の規則性があり、海内四経は規則性を見出しがたいとしている。また海内四経の特徴として、昆侖之虚のような神話的記述が含まれる一方で、現実実在した国名、地名、河川が多数あり、海内四経は現実地理に近い記載を中心としており、海外四経と海内四経では、その内を含める項目について区別が設けられているとしている。また(大野 1998)<sup>33</sup>は、海内四経に記述される実在の地名、(大野 2000)<sup>34</sup>では、海外四経の絵画性に着目し、それぞれ一括して成立したものではなく、海外四経と海内四経のなかでも新古の層が存在することを指摘している。このように海外四経、海内四経の成立が異なるとの見解があり、本研究の編著者識別の検証結果が補強したことになる。また既述のように海外四経と海内四経は共通項が際立つが、統計的手法を用いたことで、海外四経と海内四経の編著者が異なる可能性が顕在化し、その相違に注視する視点がもたらされたことは大きな成果といえる。

本稿は特定の統計的手法による結果であり、今後は他の統計的手法、および検証対象となるテキストから著者の特徴情報の抽出方法についても検討の必要があるだろう。

## 註

- 1 海外南経～海外東経の四編を「海外四経」、海内南経～海内東経の四編を「海内四経」、大荒東経～大荒北経の四編を「大荒四経」、また海外四経と海内四経を複合する場合「海外・海内経」、大荒四経と海内経を複合する場合「大荒・海内経」と表記する。
- 2 松田稔 1995「海経の絵画的要素」『『山海経』の基礎的研究』pp.20-35、笠間書院
- 3 前野直彬 1975「解説」『山海経・列仙伝(全釈漢文大系 33)』pp.9-26、集英社  
大野圭介「朴齋主頁(『山海経』解説)」<<http://www.hmt.u-toyama.ac.jp/chubun/ohno/>> (2014/08/31 アクセス)
- 4 岡本正 1960「山海経について」『中国古代史研究』pp.383-397、吉川弘文館  
小南一郎 1987「『山海経』研究の現況と課題」『中国 - 社会と文化』2、pp.220-226、東大中国学会
- 5 陸侃如 1928「論『山海経』著作年代」『新月』1(5)、pp.1-3  
玄珠(茅盾) 1929『中国神話研究 ABC』世界書局 / 玄珠・伊藤弥太郎訳 1943『支那の神話』地平社  
顧頡剛 1934「五藏山経試探」『史学論叢』1、北京大学  
蒙文通 1962「略論『山海経』的写作時代及其產生地域」『中華文史論叢』1、pp.43-70、中華書局  
袁珂 1978「『山海経』写作的時地及篇目考」『中華文史論叢』復刊1号、上海古籍出版社
- 6 袁珂 1982「『山海経』写作的時地及篇目考」『神話論文集』p.2、上海古籍出版社
- 7 小川琢治 1928『支那歴史地理研究』弘文堂書房  
伊藤清司 1969「山川の神々(1)」『史学』41(4)、pp.529-559、慶應義塾大学三田史学会 / 「山川の神々(2)」『史学』42(1)、pp.73-106、慶應義塾大学三田史学会 / 「山川の神々(3)」『史学』42(2)、pp.163-212、慶應義塾大学三田史学会  
高馬三良 1969『中国古典文学大系 8』平凡社 / 高馬三良 1994『山海経—中国古代の神話世界』平凡社  
前野直彬 1975: 前掲書
- 8 小川琢治 1928: 前掲書、pp.7-8、p.210

- 9 村上征勝 1994「真贋分析の歴史」『真贋の科学－統計文献学入門』pp.52-55、朝倉書店
- 10 (清) 畢沅『山海経新校正』18巻 / (清) 郝懿行『山海経箋疏』18巻 / 袁珂 1985『山海経校注』上海古籍出版社など
- 11 吉本道雅 2007「山海経研究序説」『京都大学文学部研究紀要』46、pp.36-53、京都大学大学院文学研究科
- 12 Donald Sturgeon「諸子百家 中国哲学書電子化計画」  
テキストデータ <<http://ctext.org/shan-hai-jing/zh>> (2014/08/31 アクセス)  
原本画像データ <<http://ctext.org/library.pl?if=gb&res=77442>> (2014/08/31 アクセス)
- 13 中国科学院計算技術研究所「ICTCLAS (NLPIR) 2014」(中国語形態素解析ツール) 開発:張華平 <<http://ictclas.nlpir.org/>> (2014/08/31 アクセス)
- 14 漢文解析システムなどの開発の動きはある。東アジア古典文献 コーパスの研究 <<http://www.zinbun.kyoto-u.ac.jp/kyodokenkyu/research16.htm>> (2014/11/30 アクセス)
- 15 近藤みゆき 2000「n グラム統計処理を用いた文字列分析による日本古典文学の研究－『古今和歌集』の「ことば」の型と性差－」『人文研究』29、pp.187-238、千葉大学  
山田崇仁 2001「『國語』 韋昭注引系譜資料について－ N-gram 統計解析法による分析－」『立命館史学』22、pp.38-75、立命館史学会
- 16 山田崇仁 2007「N-gram 方式を利用した漢字文献の分析」『立命館白川静記念東洋文字文化研究紀要』1、p.7、立命館大学白川静記念東洋文字文化研究所
- 17 金明哲 2000「自然言語における統計手法を用いた情報処理」『統計数理』48 (2)、p.279、統計数理研究所
- 18 石田基広・金明哲 2012「文章の書き手の特徴情報と書き手の識別」『コーパスとテキストマイニング』pp.59-60、共立出版
- 19 李賢平 1987「『紅樓夢』成書新説」『復旦学報 (社会科学版)』5、pp.3-16
- 20 山田崇仁 2007: 前掲論文、p.5
- 21 金明哲「MTMineR (Multilingual Text Miner) Ver5.2」開発: 金明哲・張信鵬 <<http://mjn.doshisha.ac.jp/MTMineR/index.html>> (2014/08/31 アクセス)
- 22 張華平「ICTCLAS (NLPIR) 2014」: 前掲サイト  
Donald Sturgeon「諸子百家 中国哲学書電子化計画」: 前掲サイト (字典機能)  
中国社会科学院語言研究所古代漢語研究室編 1999『古代漢語虚詞詞典』商務印書館  
三浦勝利 1996『漢文を読むための助字小字典』内山書店
- 23 早狩進「Seagull-Stat10」開発: 早狩進 <<http://www7b.biglobe.ne.jp/~hayakari/>> (2014/08/31 アクセス)
- 24 回帰直線: 2つの変数の分布する個々のデータの中心的な分布傾向を示す線。
- 25 有意水準  $\alpha$ : 帰無仮説の棄却、採択を決定する基準。 $\alpha=0.05$  は、5%の有意水準で検定を行うことであり、100回、同じ検定をした場合に5回は誤った結果を得る危険性があることを意味する。厳密さが求められる場合、 $\alpha=0.01$  (1%)、 $\alpha=0.001$  (0.1%) が用いられる。
- 26 ユークリッド平方距離 (ユークリッド距離の2乗): データ各群の相関の情報と各群の分散、平均値の差の情報を同時に評価する非類似度の定義法。個体  $i$  と  $j$  との非類似度  $d_{ij}$  を以下のように定義する。
- $$d_{ij} = \sum_{k=1}^m (x_{ki} - x_{kj})^2$$
- 27 ウォード法: 2つのクラスタを結合させる方法として、クラスタの重心までの距離に注目し、その距離の平方和の増加量 (距離の二乗の総和) が最小になるクラスタを順次結合。クラスタ  $C1$  と  $C2$  の距離関数  $d(C1, C2)$  の場合、 $d(C1, C2) = E(C1 \cup C2) - E(C1) - E(C2)$  ただし、 $E(Ci)$  は、 $Ci$  のすべての点から  $Ci$  の質量中心までの距離の二乗の総和である。
- 28 Wilks- $\Lambda$ : 検定統計量。グループ間での説明変数の平均値の差を検定。 $0 \leq \Lambda \leq 1$  の値をとり、 $\Lambda$  が0に近いほどグループの差が大きく、有意性を示す。
- 29 (晋) 郭璞伝: (清) 畢沅校 (出版書写 1877)『山海経一八巻首一卷』3、pp.2-3、浙江書局
- 30 玄珠 (茅盾)・伊藤弥太郎訳 1943: 前掲書、p.58
- 31 松田稔 1995: 前掲書、pp.20-22、前野直彬 1975: 前掲書、pp.16-17
- 32 竹内康浩 1991「海外諸経の成立『山海経』現行本の成立の問題について (2)」『史流』31、p.46-48、北海道教育大学史学会
- 33 大野圭介 1998「『山海経』海内四経の成立」『富山大学人文学部紀要』28、pp.180-183、富山大学
- 34 大野圭介 2000「『山海経』海外四経原始」『富山大学人文学部紀要』33、pp.134-136、p.158、富山大学

## 参考文献

石川慎一郎・前田忠彦・山崎誠編 2010『言語研究のための統計入門』くろしお出版

伊藤雅光 2002 『計量言語学入門』大修館書店

大野圭介 1999 『『山海経』大荒・海内経原始』『富山大学人文学部紀要』30、pp.228-258、富山大学

金明哲 2001 「助詞の N-gram 分布にもとづいた書き手の識別」『日本行動計量学会大会発表論文抄録集』29、pp.298-299、日本行動計量学会

金明哲 2009 『テキストデータの統計科学入門』岩波書店

小林立 1994 「中国語における実詞と虚詞について」『香川大学一般教育研究』27、pp.57-60、香川大学一般教育部

竹内康浩 1987 「後漢時代に於ける『山海経』現行本の成立の問題について」『道教と宗教文化 / 秋月観映編』pp.61-80、河出出版社

東京大学教養学部統計学教室 1994 『人文・社会科学の統計学（基礎統計学 II）』東京大学出版会

村上征勝 2002 『文化を計る 文化計量学序説（データの科学 5）』朝倉書店

山田崇仁 2004 「歴史記録としての『春秋』— N-gram モデルと統計解析法による分析—」『中國古代史論叢（立命館東洋史学会叢書 2）』pp. 横 13-42、立命館東洋史学会

涌井良幸・涌井貞美 2011 『実習多変量解析入門 Excel 演習で మరిなくわかる』技術評論社

## Distinguishing the Author-editors of the *Shanhaijing* [The Classic of Mountains and Seas]: A Statistical Study

SHIMONISHI Noriko

Abstract:

The *Shanhaijing* [The Classic of Mountains and Seas], is one of the most important documents in the study of ancient China, but it is unclear when or where it was created, or by whom it was compiled. An  $n$ -gram model was used to list up the co-occurrence frequency of the words that appear in each of the work's eighteen chapters, then correlation analysis, hierarchical clustering analysis and discriminant analysis were applied to compare the sections and chapters of the work. These analyses led to the following results. The same author-editor compiled all of the Classic of the Mountains (chapters 1-5) but not the rest of the *Shanhaijing*. In addition, the Classic of Regions Beyond the Seas (chapters 6-9) as well as the Classic of Regions Within the Seas (chapters 10-13) were compiled by a different author-editor than that of the Classic of the Great Wilderness (chapters 14-17) or the Classic of Regions Within the Seas (chapter 18). Moreover, the Classic of Regions Beyond the Seas and the Classic of Regions Within the Seas (chapters 10-13) were compiled by different author-editors, a finding first revealed through the statistical analysis of this research.

Keywords: *Shanhaijing* [The Classic of Mountains and Seas], The Classic of the Seas,  $n$ -gram model, statistics, classification of authors and editors

## 統計的手法による『山海経』編著者の識別

下西紀子

要旨:

本論文は、中国古代の地理書とされる『山海経』に注目する。現行の『山海経』は十八編あり、五蔵山経が五編、海経は海外四経が四編、海内四経が四編、大荒四経が四編、海内経が一編の十三編から成る。この五蔵山経と海経、また海経各編は、記載内容や文体などに相違があり、中国古典文学研究において、『山海経』は異なる編著者により段階的に形成されたとみなされている。本稿の目的は、相関分析、階層クラスタ分析、判別分析などの統計的手法を用いて、五蔵山経と海経および各編における主要な編著者の識別を明らかにすることにある。検証の結果、五蔵山経と海経の編著者の相違が示され、各編では五蔵山経が同一の編著者、海外四経および海内四経と大荒・海内経では編著者が相違し、また海外四経と海内四経も相違が示された。これにより、同一の編著者と捉えられるような記載が目立つ海外四経と海内四経の間に、編著者の相違が見られることが明らかになった。

