

統計数理研究所

研究レポート53

統計的決定理論の立場からの文献学的 判別問題に対する研究

一日蓮の“三大秘法稟承事”の真偽判別解析—

研究資料

1981年3月

統計数理研究所

統計的決定理論の立場からの文献学的
判別問題に対する研究

一日蓮の“三大秘法稟承事”の真偽判別解析—

この研究レポートは、昭和53・54・55年度科学研究費一般研究（B）の研究
成果報告書として作成されたものである。

当研究所では、現在、

Annals of the Institute of Statistical Mathematics

統計数理研究所彙報

Computer Science Monographs

を発行している。

このレポートは研究調査のデータの発表等を目的とし、必要に応じて発行する。

統 計 数 理 研 究 所

東京都港区南麻布4-6-7

電話(03)446-1501

目次

1. まえがき		
2. 研究目的	藤本 照	1
3. 従来の研究について		
3.1 著者推定における統計的手法	村上 征勝	2
3.2 著者推定の統計学	藤本 照	13
	村上 征勝	
4. “三大秘法抄”の著者推定		
4.1 “三大秘法抄”の真偽問題の由来と経緯	伊藤 瑞叡	27
4.2 統計解析		
4.2.1 解析に用いた文献について	伊藤 瑞叡	32
4.2.2 解析に用いる情報	村上 征勝	35
4.2.3 語の分割と品詞分類	春日 正三	37
4.2.4. 統計解析	村上 征勝	42
4.2.4.1 文の構造に関する情報		42
4.2.4.2 単語に関する情報		60
4.2.4.3 経年変化のある情報		72
4.2.4.4 論理の展開の仕方に関する情報		73
4.3 今後の研究計画について	村上 征勝	73
5 計算プログラムと情報の入出力	村上 征勝	75
5.1 データ処理用サブプログラムとその主な機能		75
5.2 各サブプログラムの使用法		78

1. まえがき

文献の真偽判別に対して有効な統計解析の方法を一般的に述べることは、無理な話のようにも思える。真偽判別問題では、偽なるものの不特定性の故に、通常の統計的判別分析の適用で問題が解決できるかどうかは、利用可能な資料の特徴によることになると思う。

例えば、判別を真作群と偽作群の2群にわけて考えるとしよう。このとき、真作群と対比されている偽作群の特徴を特定化できず、偽作群を構成する偽作資料の数を増加すると、また異なった特徴が現れるように偽作群の構成が不安定であれば、2群判別は成功しなからう。

これに対して、偽作群を構成する偽作資料に、真作に似せようとする作為があって、それが特徴として括れるときには、2群判別で成功する可能性があることになるだろう。

ところで、判別のための基礎資料を得るのに、日本語の文献、特に古典的なものになると、文章を単語に分解する作業が大変である。このことは、この研究に取掛ってみて、あらためて認識させられたことであるが、このような資料を電子計算機に読み込めるようにするのに、研究の初年度の殆どを使ってしまった。

3年を経過し、一応成果報告書の作成が可能な程に纏まったとはいうものの、問題の結着までには、まだまだ検証すべき事柄が残っているし、真偽判別問題に対する統計的方法には、更に検討されるべくものを含むように思う。

このようなことから、この研究成果報告書では、統計解析における方法論や、それからの結論よりは、必要があれば、もとに戻って検討し直せるように、なるべく経過の記録をとどめておくように配慮した。

代表者 藤本 照

2. 研究目的

文献の真偽判別に対する統計的方法を実証的に研究する。そこで現下の文献学的レベルでは真偽未決とされている“三大秘法稟承事”をとりあげた。

目的とするところは、この種の問題は、通常の統計的判別分析の対象にはなりにくい問題であろうという認識から、これを統計的判別分析の枠内のものとして取扱えるようにするには、どうしたらよいかを検討し、出来得べくは、そのための方法への足掛りを得る為である。

問題を総合的に検討するために、統計以外の領域からも専門家の参加を得て、研究を継続してゆく計画でいる。

3. 従来の研究について

3. 1. 著者推定問題における統計的手法

統計数理研究所 村 上 征 勝

(1978年12月 受付)

1. はじめに

洋の東西を問わず、著名な思想家、作家の著書として愛読され続けてきた作品の中にも、その成立について疑問がもたれているものが少なからず存在する。勿論、これらの疑問を解決すべく、それぞれの分野固有の伝統的な方法によって、内容、成立等に関して多くの研究がなされてきた。しかし、これとは別に作品の数量的な性質に注目して、この種の問題の解決を試みた研究者達もいる。この稿では、彼等が用いた研究方法やその成果を統計学の観点から要約、整理し、内在する問題点を明確にする。第2節では著者推定問題と密接に関連する執筆順序推定問題を、第3節では代表的な著者推定問題を紹介し、最後の第4節で統計的パターン認識の立場から問題点を考察する。

2. 執筆順序推定問題

西暦500年～1000年頃、パレスチナの地において旧約聖書の研究に従事していたマソラ学者(Masoretes)達は、ヘブライ語の旧約聖書の索引作りの仕事の一部として、聖書各巻の文字数、語数、宗教的な意味を持つ語の頻度などを調べていた。文献の数量的な性質に対するこのような興味は19世紀に入り、執筆順序の推定や著者の推定に関する研究へと発展する。そして、前者における主な対象はプラトンとシェクスピアであった。

1867年、スコットランドのL. Campbellはプラトンの著作の執筆順序の推定に関する研究を発表した[1]。彼はプラトンの絶筆とされている「法律」を手がかりとして、主に哲学的用語について晩年の6つの対話篇の語彙を「法律」の語彙と比較して、その著作年代を推定した。この研究を契機に、その後、W. Dittenberger (1896年) [2], W. Lutosławski (1897年) [3], C. Ritter (1935年) [4], CoxとBrandwood (1959年) [5]等によってプラトンの執筆順序の推定が、またC. Bathurst (1857年) [6], J. K. Ingram (1874年) [7], F. G. Fleay (1874年) [8], F. J. Furnival (1887年, 1907年) [9], [10], G. König (1888年) [11], H. Conrad (1897年) [12], E. K. Chambers (1930年) [13], M. R. Yardi (1946年) [14]等によってシェクスピアの著作の執筆順序の推定がおこなわれた。シェクスピアについては、生年月日を始めとして、彼の生涯や作品について謎の部分が多く、また当時、合作や劇団所有の古い脚本や他人の作品の修正加筆が盛んであったこともあって、著者推定問題としても興味あるものが少なくない*が、さらに一部の人はシェクスピアの存在までも疑いを抱いていた(第3節のA参照)。ところで、これらの研究の多くは語彙の比較がその中心であり必ずしも統計が重要な役割をはたしてはいない。統計的な意味で、また著者推定問題との関連で興味を引くのはYardiの研究である。彼はシェクスピアの各作品中の“full split line”, “redundant final

* 疑問がもたれている作品としては「ヘンリー六世」、「タイタス・アンドロニカス」、合作の疑いのある作品としては「じゃじゃ馬ならし」、「トロイダスとクレラダ」、「アテネのタイモン」、「ペリクリーズ」、「ヘンリー八世」があげられている

syllable” “unsplit line with pause” の割合を基に、各作品について彼のいうところの識別関数値 u を計算し、 u とシェクスピアの処女作品が執筆された年から、ある作品が書かれた年までの経過年数 v との間には近似的に

$$u = 0.7204 + 0.07645v$$

なる関係があると報告している。

3. 著者推定問題

A. モードによる推定

著者推定問題における統計学の有用性を最初に示唆したのは Augustes de Morgan である。1851年、彼はケンブリッジ大学時代の友人にあてた手紙の中で、単語の長さの平均値に作家の特徴が表われるであろうと記している [15] [16]。この de Morgan の考えにヒントを得たオハイオ州立大学の地球物理学者 T. C. Mendenhall は、1887年に Science 誌に発表した論文 [17] の中で Thackeray, Dickens, Mill 等の文章に現われる単語の長さの度数分布——彼はこの度数分布を word spectrum と呼んだ——を示し、著者推定問題に対する word spectrum の有効性を説いた。1901年、彼はシェクスピアとフランシス・ベーコンの word spectrum を実に40万語、20万語という膨大なデータから求め [18]、シェクスピアのモードは4文字から成る単語であり、ベーコンのモードは3文字から成る単語であることを示すことによって (図1)、当時一部の人々によって信じられていた、シェクスピアなる人物は歴史上に存在せず、ベーコンが王政抗議のため一連の風刺劇を書いたという説を否定した。これが、著者推定問題に統計が主役となって登場した最初の研究である。

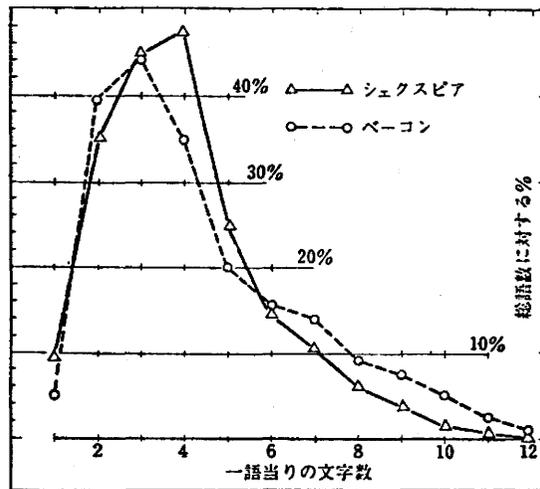


図1 シェクスピア、ベーコンの語の長さの分布

B. 平均値、中央値等による推定

1939年、統計学者 Undy Yule は、Bacon, Coleridge, Lamb, Macauley の文の長さを調査し、その平均値、中央値、四分位範囲等は作家間で異なるが、同一作家の作品では安定した値となるという結果を得た。彼はこの結果をもとに “The Imitation of Christ (原題名 De Imitatione Christi キリストにならいて)” と “Natural and Political Observations...made upon the Bills of Mortality (死亡表に関する…… 自然のおよび政治的諸観察)” の著者推定をこれらの

表 1. The Imitation of Christ と à Kempis, Gerson の比較

	The Imitation of Christ	à Kempis	Gerson 1	Gerson 2
平均値	16.2 語	17.9 語	23.4 語	22.7 語
中央値	13.8	15.1	19.6	18.9
四分位範囲	9.2	11.8	18.8	17.8

表 2. Graunt と Petty の比較

	Graunt				Petty		
	標本A	B	C	全体	標本A	B	C
平均値	50.1 語	45.5 語	46.9 語	47.5 語	66.1 語	60.2 語	56.3 語
中央値	45.2	38.0	37.4	40.1	56.9	51.3	44.0
四分位範囲	32.1	31.7	39.2	35.5	47.1	44.3	44.7

統計量を用いておこなっている [19].

“The Imitation of Christ” に関しては、その著者の可能性のある人物として Thomas à Kempis と Jean Charlier de Gerson の二人をあげ、彼等の作品と “The Imitation of Christ” の文の長さの平均値、中央値、四分位範囲等を計算し、表 1 のような結果を得ている。これらの数値の比較から、彼は、à Kempis が著者であると結論している。

一方 “Natural and Political Observations…” は、統計学者 John Graunt の著作とされているが、Yule はこの作品を Graunt の友人で “A Treatise of Taxes and Contributions (租税負担論)” 等で知られる統計学者 Sir William Petty の著作と比較し、表 2 で示されるようにこの 2 人の文の長さの分布は異なっていることを明らかにした。

この Yule の研究と同じ様な研究に 1957 年、W. C. Wake がおこなった研究がある [20].

彼は偽作説があるプラトンの “第七書簡” の文の長さをプラトンの著作「法律」と比べ (表 3)、その偽作説を否定している。彼はまたこの論文の中で、アリストテレスの著作についても言及している。

表 3. 「第七書簡」と「法律」の比較

	「第七書簡」	「法律」の平均
平均値	28.1 語	27.7 語
中央値	22.3	23.1
第一四分位数	13.1	13.5
第三四分位数	36.0	36.7

C. 特別な指標による推定

ところで、文の長さを用いた推定に満足できなかった Yule は 1944 年、作家の語彙量を測る指標として、 K -characteristic と呼ばれる指標を提案し、再度 “The Imitation of Christ” の著者推定をおこなっている [21]. いま、ある作品の中に n_i 回使用された単語が f_i 個あるとすると、この時 $S_1 = \sum f_i n_i$, $S_2 = \sum f_i n_i^2$ として K は次のように定義される。

$$K = 10^4 \left(\frac{S_2}{S_1^2} - \frac{1}{S_1} \right)^*$$

彼はある単語が出現するプロセスのモデルとしてポアソン分布を想定した。ポアソン分布では 平均=分散 という関係が成立するが、現実にポアソン分布に従うと考えられる事象を観察すると多くの場合 分散 \geq 平均 となっていること、また、分散を平均で除した変動係数が標本の大きさに影響されないこと等の考察にもとづき、この指標を導出している**。問題の書

* 10^4 は K 値が小さくなりすぎるのを防止するため他に意味はない。

** この指標の性質については [22], [23], [24], [25] 等で議論されている。

表 4. K 値の比較

作 品	K 値
The Imitation of Christ	84.2
à Kempis 1	113.7
" 2	110.9
" 3	100.1
" 4	66.9
" 5	59.7
Gerson	35.9

“The Imitation of Christ” と à Kempis, Gerson の作品の K 値は表 4 の様になり、この結果から彼は à Kempis の方が著者の可能性が高いとした。

また A. Ellegård は 1962 年に発表した論文 [26] の中で、1768 年から 1772 年にかけて英国内閣の政策を批判した一連の公開状を “Junius” という偽名で発表した人物が誰であるかを、Distinctiveness ratio という指標を用いて推定している。彼は Junius letter のような 18 世紀中

頃の型にはまった文章に対しては、Yule の K-Characteristic は効力が無いとして、次のような Distinctiveness ratio という指標を提案した。

$$\text{Distinctiveness ratio} = \frac{\text{Junius letter 中でのある単語の相対度数}^*}{\text{Junius letter 以外の百万語中のある単語の相対度数}}$$

彼はこの指標を用いて、100 名にもあまる候補者の中から、Junius として Sir Philip Francis の名をあげている。

この 2 つの指標以外に、語彙量に注目した指標として E. H. Simpson の Index of diversity [27], K-Characteristic の変形である G. Herdan の K*-Characteristic [28] などが、また品詞の相対度数に注目した指標として、V. Neubaver & A. Schlisman の動詞一形容詞比 [29], C. B. Williams の名詞一形容詞比 [30] などが提案され議論されているが、具体的な著者推定問題には用いられていないようであるのでここでは割愛する。

D. χ^2 検定, t 検定による推定

1861 年にニューオリンズ・ディリー・クレセント誌に Quintus Curtius Snodgrass という偽名で南北戦争に関する手紙 (Q. C. S. letter) が掲載された。この手紙の著者は「トム・ソーヤの冒険」等の作者 Mark Twain であると一般に信じられてきたが、C. S. Brinegar は 1963 年 J. A. S. A. に発表した論文 [31] の中で、単語の長さとその頻度に注目して χ^2 検定, t 検定をおこない、この説を否定した。彼は Q. C. S. letter の前の時期、同じ時期、後の時期の Twain の文章から、それぞれ 1885 語、6106 語、2974 語の単語を調べ (表 5)、それをコントロールとして χ^2 検定をおこない、0.5% 水準でも有意な差があることを示した。また、彼は 2, 3, 4 文字からなる単語の頻度の各々の平均値に関しても t 検定をおこない、それらが 5%, 1%, 5% で有意であることも示している。

1960 年に安本は源氏物語の宇治十帖と呼ばれる後半の 10 巻が前半の 44 巻と同様紫式部の手になるものであるかを調べた研究を発表している [32], [33]。彼は、源氏物語各巻の (1) 長編度 (頁数) (2) 和歌の使用度 (3) 直喩の使用度 (4) 声喩の使用度 (5) 心理描写の数 (6) 文の長さ (7) 色彩語の使用度 (8) 名詞の使用度 (9) 用言の使用度 (10) 助詞の使用度 (11) 助動詞の使用度 (12) 品詞数の 12 項目について各巻に順位をつけ、U 検定法によって後半の 10 巻と前半の 44 巻の差が偶然によって生じる確率を計算し、さらに χ^2 検定をおこなって、宇治十帖と前半の 44 巻の著者が異なっていると主張している。

また、A. Q. Morton は 1965 年に新約聖書の中の 14 通のパウロの書簡**の著者推定に關す

* たとえば uniform (形容詞) の場合には $\text{distinctiveness ratio} = 0.000280 / 0.000065 = 4.3$ であった。

** ローマ人への手紙、コリント人への第 1 の手紙、第 2 の手紙、ガラテヤ人への手紙、エペソ人への手紙、ピリピ人への手紙、コロサイ人への手紙、テサロニケ人への第 1 の手紙、第 2 の手紙、テモテへの第 1 の手紙、第 2 の手紙、テトスへの手紙、ピレモンへの手紙、ヘブル人への手紙の 14 通で、いずれも「使徒パウロから…」というような書き出しになっているのでパウロの書簡とされている。

表 5. Mark Twain の著作と Q. C. S. letter の比較

語の長さ	1858年 1861年	の著作	1863年の著作	1867年の著作	Q. C. S. letter
1	74		312	116	424
2	349		1146	496	2635
3	456		1394	673	2752
4	374		1177	565	2302
5	212		661	381	1431
6	127		442	249	992
7	107		367	185	896
8	84		231	125	638
9	45		181	94	465
10	27		109	51	276
11	13		50	23	152
12	8		24	8	101
13以上	9		12	8	61
	1885		6106	2974	13175

る研究を発表している [34]. 彼はパウロの書簡の中に出てくる kai (=and), de (=but), en (=in), autos (代名詞) の頻度を調べ χ^2 検定をおこなって、次のように結論している。

「パウロの書簡といわれているものの中で、もしガラテヤ人への手紙を書いたのがパウロであるならば、ローマ人への手紙、コリント人への第1の手紙、第2の手紙もまた彼の手によるものであるといえる。しかしながら、残りの10通は少なくとも6人の手によるものと考えられる」

E. 判別関数による推定

ハーバード大学の F. Mosteller とシカゴ大学の D. L. Wallace は、1963年の J. A. S. A. に Federalist paper の著者推定に判別関数を用いた論文を載せている [35]. Federalist paper というのは、米国憲法の批准を勧めるために 1787年から1788年の間に、Alexander Hamilton, John Jay, James Madison の3人がニューヨーク州の新聞に載せた77編のキャンペーン記事のことで、Hamilton 43編, Jay 5編, Madison 14編, Hamilton & Madison 3編というように12編を除いてはその帰属は明らかにされており、著者不明の12編も Jay の可能性はなく、Hamilton か Madison のいずれかが著者であるとされていた。

Mosteller はすでに1941年に文の長さによってこの12編の帰属を明らかにしようと試みていたが、その時は表6のような結果を得て失敗に終わっていた。そこで1963年の論文では、彼等は、記事の内容と無関係で、かつ短い記事の中でも比較的多く用いられている言葉——たとえば、of, all, also のような言葉——に注目した。その結果、たとえば upon は Hamilton の文章では 3/1000 の割合で使用されているが、Madison の文章では 1/6000 の割合であり、また、to は Hamilton の方が使用率が高く、反対に by は Madison の方が高い等のことがわかった。そこでこのような識別力の高い31語を用いて

表 6. Hamilton と Madison の文の長さの比較

文の長さ	Hamilton	Madison
平均	34.55 語	34.59 語
標準偏差	19.2	20.3

$$y = \sum_{i=1}^{31} W_i x_i$$

という線型判別関数を作り、これによる判別を試みている。ここで、 x_i は帰属不明の記事の中の i 番目の語の出現率で、 W_i は i 番目の語の重みであり、この値は Hamilton の記事の場合には γ が大きく、Madison の記事の場合には γ が小さくなるように次の式で求めている。

$$W_i = \frac{\bar{x}_{i(H)} - \bar{x}_{i(M)}}{R_{i(H)} - R_{i(M)}}$$

ただし、 $\bar{x}_{i(H)}$, $\bar{x}_{i(M)}$, $R_{i(H)}$, $R_{i(M)}$ は各々 Hamilton, Madison の文章における i 番目の語の頻度の平均値とレンジである。この判別関数を用いて帰属不明の 12 編の記事の γ 値を計算した結果が表 7 である。次に、彼等 2 人の著作 25 編から求めた γ 値の平均 $\bar{\gamma}_H$, $\bar{\gamma}_M$, 標準偏差 S_H , S_M を用いて t 検定をおこない、No. 55 の記事は疑問が残るものそれ以外は Madison が著者であると推定している。またこの論文の中で Mosteller 達は確率比を用いての判別も試みている。いま Hamilton, Madison の文章において i 番目の語が出現する確率が、それぞれパラメータ $\theta_{i(H)}$, $\theta_{i(M)}$ のある分布 (Mosteller 等はこの分布のモデルとして、ポアソン分布と負の二項分布の 2 つを想定している) に従っている時に、 i 番目の語が x_i 回観測される確率を $f(x_i|\theta_{i(H)})$, $f(x_i|\theta_{i(M)})$ とする。この時

表 7. Hamilton, Madison の記事および著者不明の記事における判別関数値

Hamilton		Madison		著者不明	
記事	γ	記事	γ	記事	γ
13	0.88	134	0.05	49	-0.46
15	0.85	135	-0.50	50	-0.26
16	1.11	201	-0.22	51	-0.76
17	1.26	202	-0.28	52	-0.23
21	1.11	203	-0.27	53	-0.48
22	0.96	204	-0.79	54	-0.46
23	0.62	205	-0.29	55	0.32
25	0.31	206	-0.64	56	-0.01
27	0.80	207	-0.86	57	-0.50
28	0.70	208	-0.38	58	-0.33
29	1.39	209	-0.65	62	-0.41
30	1.32	210	-0.65	63	-0.13
31	1.38	216	-0.16		
32	0.49	217	-0.42		
33	0.75	218	-0.42		
34	1.06	219	-0.53		
35	0.80	220	-0.19		
60	0.73	301	-0.07		
61	0.64	302	-0.22		
67	0.63	311	-0.44		
68	0.34	312	-0.08		
69	1.18	313	0.15		
73	1.19	314	-0.38		
75	1.31	315	-0.59		
76	1.17	316	-0.57		
$\bar{\gamma}$	0.92	$\bar{\gamma}$	-0.38	$\bar{\gamma}$	-0.31
S	0.32	S	0.25		

$$\Pi \frac{f(x_i|\theta_i(H))}{f(x_i|\theta_i(M))}$$

が大ならば Hamilton が、小ならば Madison が著者の可能性が高いと考えられる。この式を用いた結果やはり Madison が著者であると推定している。

この Mosteller & Wallace の方法を用いて、葦沢は由良物語の著者として加茂真淵説よりも、建部綾足説を支持する研究を発表している [36], [37]。また、確率比を用いて無量寿経類漢訳者の推定をした研究も報告されている [38]。

F. 因子分析法による推定

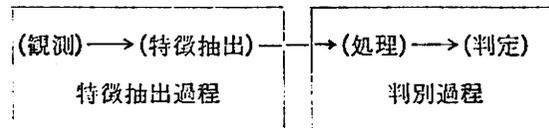
安本は、源氏物語宇治十帖の著者推定を因子分析法によってもおこなっている [39]。彼は前述の 12 項目間の相関係数を計算し、因子分析によって (1)「比喩多用型—比喩節用型」の因子 (2)「歌物語型—作り物語型」の因子 (3)「体言型—用言型」の因子とも名づけられる 3 つの因子を取り出した。彼はこの因子分析の結果をもとに χ^2 検定をおこない、前の研究と同様、源氏物語の宇治十帖と呼ばれる後半の 10 巻の文体は、前半の 44 巻の文体とやや異なっていると報告している。

G. その他

ここであげた以外に、具体的に用いられてはいないが、シラブル間の距離やエントロピー [40], [41], 系列相関 [42] 等も検討されている。

4. 著者推定における問題点

この節では、文体もまた一つのパターンであると考えることによって、著者推定問題を統計的パターン認識の問題に還元し、この観点から前節で紹介した著者推定問題を考察し、問題点を整理する。パターン認識は通常、観測、特徴抽出、処理、判定の 4 つの過程をへるとされているが、ここでは、下記のように特徴抽出と判別の二つに大別して議論を進めることにする。



さて、前節の著者推定問題をこの観点から整理すると表 8 のようになる。

次にこれらの研究で明らかにされた問題点を、特徴抽出、判別の二つの観点から整理してみる。

A. 特徴抽出における問題点

Le style c'est l'homme même (文は人なり) と P. Guiraud は述べているが、文体のいかなる面に著者の特徴が現われるかを知ることが、著者推定問題解決の鍵である。そのためには、少なくとも次の 5 つの側面、イ)文章の構成要素の性質 ロ)文法構造 ハ)音韻 ニ)リズム ホ)問題特有の性質 (例えば、源氏物語における和歌の数のような) の側面から調べるべきであろう。ところで表 8 が示すように、これまでの研究でとりあげられてきたのは、単語の長さ、文の長さ、単語の使用頻度、語彙量など主に イ)に関するものである。これらについては、次のような問題点が指摘されている。まず単語の長さに関しては、単語の定義が困難であり、その上、文の長さとは違い単語の長さは著者の自由にならないため特徴が出にくい等のことが、また文の長さに関しては、単語と同様に文の定義の困難さや、文の長さの測定単位を何にすべきかという問題。さらにはまた、記述体か対話体か等の文の種類によって長さが変わる可能性がある

表 8. 主な著者推定問題

問題となった文献、著者	抽出した特徴	用いられた統計量・統計的手法	研究者
シェクスピア、ベーコン	語の長さの分布	モード	T. C. Mendenhall (1887年)
The Imitation of Christ	文の長さの分布	平均値, 中央値, 四分位範囲, etc.	Undy Yule (1939年)
プラトンの第七書簡	・	・	W. C. Wake (1957年)
The Imitation of Christ	語彙量	K-Characteristic	Undy Yule (1944年)
Jónius letter	語の使用頻度	Distinctiveness ratio	A. Ellegard (1962年)
Quintus Curtius Snodgrass letter	語の長さの分布	χ^2 検定, t 検定	C. S. Brinegar (1963年)
新約聖書中のパウロの書簡	語の使用頻度	χ^2 検定	A. Q. Morton (1965年)
源氏物語の宇治十帖	頁数, 和歌数等の12項目	U 検定, χ^2 検定	安本 美典 (1960年)
Federalist paper	語の使用頻度	線型判別関数, 確率比	Mosteller & Wallace (1963年)
由良物語	・	・	葦沢 正 (1965年)
源氏物語の宇治十帖	頁数, 和歌数等の12項目	因子分析法	安本 美典 (1977年)

ることなどが指摘されている。単語の使用頻度に関しては、内容によっては同じ語が多数用いられる可能性があり、したがってどの単語について調べるべきかの決定の困難さが指摘されている。([20], [35] では、文章の内容と無関係で、しかも使用頻度の多い語を用いているが、その選択のために予備調査をおこなっている。)

さらに、語彙量の測度である K -characteristic については [21], [30], [43] 等の結果をみると、(データが少なく、明言できないが) 同一作家の作品でも K 値のパラッキが大きいような感じを受け、この測度の使用に関してはさらに研究が必要であろう。

このようにイ) に関しては、問題点はかなり明らかにされているが、ロ), ハ), ニ), ホ) に関しては今後の研究を待たねばならない。これとは別に、Yardi の研究 [14] からわかるように、文体が著者の一生を通じて不変のものではなく、時とともに変化する可能性があることも常に考慮に入れておかねばならない。この他に、サンプリングおよび測定法の問題点についても [21], [44], [45] 等で議論されている。また具体的な作業をおこなう場合に、解析に用いる資料として何を用いるべきか、また古典の著者推定問題で真筆が存在しない場合にどの写本を用いるべきかというような問題もある (日本の古典の場合に、文の長さが校訂者によって異なる場合があるという報告がある [46])。

B. 判別における問題点

ところで現在までになされた研究は、主にパラメトリックな統計手法によって解析をおこなっている。パラメトリックな方法を用いる際には当然分布型など初めとして種々の統計上の仮定が必要であり、したがって常にそれらの仮定の妥当性が問題となる。著者推定の場合においても、このことは例外ではない。たとえば、語の長さの分布として、対数正規分布 [47], ワーリング分布 [48], 複合ポアソン分布 [21] などが、文の長さの分布として対数正規分布 [50], ガンマ分布 [51] が、単語の使用頻度の分布として、ポアソン分布, 負の二項分布 [35], 対数指数分布 [49] などが提案されているが、やはりこれらの分布の妥当性についてはまだ検討の余地がありそうである。またこれ以外に、各変数の独立性の仮定など、個々の問題で前提とした種々

の統計上の仮定についてもより詳細な吟味が必要である。勿論このような問題点を避けるために、ノンパラメトリックな方法で判別を試みた例もないではない [32]。しかし、この場合にも各変数の独立性の仮定はおいている。

次に問題になるのは、解析(判別)結果から導びいた結論についてであるが、この点における問題点は実は著者推定問題のタイプと関係させて考えねばならない。一般に著者推定問題は次の2つのタイプに大別できると考えられる。第1のタイプは、問題となっている作品 x が A の著書であるか B の著書であるか、つまり、 $x \in A$ か $x \in B$ かの判別問題であり、第2のタイプは x が A の著書であるか否か、つまり、 $x \notin A$ かどうかの判別問題である。第1のタイプの著者推定問題としては [21], [35] などがあげられるが、この場合には、判別関数や、 t 検定やあるいは何らかの意味で A や B からの近さを求めてそれらの値を基にして判別がおこなわれている。この場合にも、たとえば線型の判別関数で充分であるか等の統計的方法に関する問題は勿論あるが、しかしより本質的な問題点は x の作者の可能性のある人物をどのようにして A , B の2人に絞るかという統計以前の点にある。たとえば、Yule は [19], [21] の “The Imitation of Christ” の著者推定において、著者の可能性のある人物として à Kempis と Gerson を考えているが、この2人以外に à Kempis が属していた教団の創立者、ヘーラルト・ホロートが書いたものを à Kempis が修正加筆しながら編集したという説もある*。前述の2人と比較した場合に、ヘーラルト・ホロートの方の可能性が高くなることも考えられないではない。Federalist paper の著者推定 [35] のように著者の可能性のある人物が2人でそれ以外に考えられないというような問題はむしろ著者推定問題としては特殊であって、一般には可能性のある人物を何らかの形で2人に絞って問題を解いている。したがって、この場合には、 A と B ではどちらの方がより可能性があるかという結論しか得られない。

一方、2番目のタイプの著者推定問題としては [20], [32], [33] などがあり、主に t 検定や χ^2 検定等の統計的仮説検定法によって結論を導出している。この場合に、仮説が棄却されたなら $x \notin A$ と結論でき問題はないが、仮説が棄却されなかった場合は注意しなければならない。仮説が棄却されなかったことが即 $x \in A$ とはならないからである。この意味で [20] の結論導出には問題がある。

5. おわりに

この稿でとり上げた研究例の多くは欧米のものである。したがって、由良物語の著者推定のように、Mosteller 達が Federalist paper に対して用いたのと同じ手法である程度の成果をあげた例はあるにしろ、しかし一般的に、言語体系が異なる日本文の場合に、これらの研究成果がどの程度役立つかは、今後の研究を待たねばならない。また、前節で指摘したように、統計的手法による判定は客観性を持つという反面、多くの仮定のもとでの結論であることに留意する必要がある。したがって、最終的な結論は、統計的な研究も含め多くの側面からの研究を通じてなされるべきである。

以上、著者推定問題と呼ばれるこの種の問題に対する今後の研究の一つの道標となればと願い、この分野の研究の展望をおこなったが、著者の知識の不足から、重要な研究を欠落していたり、引用論文の解釈に多くの誤謬があるかもしれないこの点に関し読者の御叱正を乞う次第である。

なお、本稿をまとめるにあたり、レフリーの方々より多くの有益な助言をいただきました。ここに感謝の意を表します。またこの研究は文部省科学研究費一般研究 B「統計的決定理論の立場からの文献学的判別問題に対する研究」の一部としておこなわれたことを付記しておきます。

* 安本 [39] 参照

参 考 文 献

- [1] Campbell, L. (1867). *The Sophistes and Politicus of Plato*, Oxford, Clarendon Press.
- [2] Dittenberger, W. (1896). *Sprachliche Kriterien für die Chronologie der Platonischen Dialoge*, Hermes, 16.
- [3] Lutosławski, W. (1897). *The Origin and Growth of Plato's Logic, with an Account of Plato's style and the Chronology of his Writings*.
- [4] Ritter, C. (1935). *Unterabteilungen innerhalb der zeitlich ersten Gruppe platonischen Schriften*, Hermes, 70.
- [5] Cox, D.R. and Brandwood, L. (1959). *On a Discriminatory Problem Connected with the Works of Plato*, J.R. Statist. Soc., B. 21, 195-200.
- [6] Bathurst, C. (1857). *Remarks on the Difference in Shakespeare's Versification in different Periods of his Life*, London, Parker and Son.
- [7] Ingram, J.K. (1874). *On the weak endings of Shakespeare*, New Shakes. Soc., Trans., Ser. 1, Part 3, 442-446.
- [8] Fleay, F.G. (1874). *On metrical tests as applied to dramatic poetry. I. Shakespeare*, New Shakes. Soc., Trans., Ser. 1, Part 1, 1-16; Postscript, 38-39; also Discussion, 17-37.
- [9] Furnival, F.J. (1887). *Introduction to the Leopold Shakspeare*, London, Cassell & Co.
- [10] Furnival, F.J. (1907). *"Loues Labors Lost": the Old-Spelling Shakespeare*. London, Chatto and Windus.
- [11] König, G. (1888). *Der Vers in Shakespeares Dramen*.
- [12] Conrad, H. von (formerly Isaac) (1897). *Shakespeares Selbstbekenntnis*, Stuttgart.
- [13] Chambers, E.K. (1930). *William Shakespeare: A Study of the Facts and Problems*, Oxford, Clarendon Press, 2 vols.
- [14] Yardi, M.R. (1946). *A statistical approach to the problem of the chronology of Shakespeare's plays*, Sankhyā, 7(3), 263-268.
- [15] Lord, R.D. (1958). *De Morgan and the statistical study of literary style*, Biometrika, 45, 282.
- [16] de Morgan, Sophia (1882). *Memoir of Augustus de Morgan, by his wife Sophia Elisabeth de Morgan, with Selections from his Letter*, London.
- [17] Mendenhall, T.C. (1887). *The characteristics curves of composition*, Science, 11, 237-249; Supplement, March 1887.
- [18] Mendenhall, T.C. (1901). *A mechanical solution of a literary problem*, Popular Science Monthly, 60(7), 97-105.
- [19] Yule, G.U. (1939). *On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to two Cases of Disputed Authorship*, Biometrika, 30.
- [20] Wake, W.C. (1957). *Sentence-Length Distributions of Greek Authors*, J.R. Statist. Soc., A, 20, 331-346.
- [21] Yule, G. Undy (1944). *The Statistical Study of Literary Vocabulary*, Cambridge University Press.
- [22] Williams, C.B. (1946). *Yule's "Characteristic" and the Index of Diversity*, Nature, 157.
- [23] Good, I.J. (1953). *The population frequencies of species and the estimation of population parameters*, Biometrika, 40.
- [24] Herdan, G. (1955). *A new derivation and interpretation of Yule's "Characteristic K"*, J. of Applied Mathematics and Physics, VI, 332-334.
- [25] Herdan, G. (1958). *The mathematical relation between Greenberg's index of linguistic diversity and Yule's characteristic*, Biometrika, 45.
- [26] Ellegård, A. (1962). *A Statistical Method for Determining Authorship: The Junius Letter, 1769~1772*, Gothenburg Studies in English, no. 13.
- [27] Simpson, E.H. (1949). *The measurement of diversity*, Nature, 163, 688.
- [28] Herdan, G. (1966). *The Advanced Theory of Language as Choice and Chance*, New York.
- [29] Antosch, F. (1969). *The diagnosis of literary style with the verb-adjective ratio*, Statistics and Style, 57-65.
- [30] Williams, C.B. (1970). *Style and Vocabulary*, Hafner Publishing Co., New York.
- [31] Brinegar, C. (1963). *Mark Twain and the Quintus Curtius Snodgrass Letters: a statistical test of authorship*, J. Amer. Statist. Ass., 58, 85-96.
- [32] 安本美典 (1957). *宇治十帖の作者—文章心理学による作者推定—*, 文学・語学, 1957, 4.

- [33] 安本美典 (1958). 文体統計による筆者推定 一源氏物語, 宇治十帖の作者について一, 心理学評論 Vol. 2, No. 1.
- [34] Morton, A.Q. (1965). The authorship of Greek prose, J.R. Statist. Soc., A, 128 (2), 169-233.
- [35] Mosteller, F. and Wallace, D.L. (1964). Inference in an Authorship Problem, J.A.S.A., 58, 275-309.
- [36] 葦沢 正 (1965). 由良物語の作者の統計的判別, 計量国語学 No. 33.
- [37] 葦沢 正 (1973). ある統計的判別の試み, 数理科学 No. 117, 3月号.
- [38] 後藤義乗 (1978). 数理文献学的方法による無量寿経類漢訳者の推定, 印度学仏教学研究第26巻第2号.
- [39] 安本美典 (1977). 現代の文体研究, 岩波講座日本語, 397-423.
- [40] Fucks, W. (1954). On mathematical analysis of style, Biometrika, 39.
- [41] Fucks, W. (1954). On nahordnung and fernordnung in samples of literary texts, Biometrika, 41.
- [42] 米田正人 (1974). 文の長さの統計学的一考察, 電子計算機による国語研究 VII, 国立国語研究所.
- [43] Bennett, P.E. (1957). The statistical measurement of a stylistic trait in Julius Caesar and As You Like It, Shakespeare Quarterly, VIII, 33-50.
- [44] 安本美典 (1957). 文の長さの分布型, 計量国語学 No. 1.
- [45] 林知己夫 (1957). 安本氏の「文の長さの分布型」について, 計量国語学 No. 2.
- [46] 三上悠紀夫 (1959). 古典校訂本の句点は信用できるか, 計量国語学 No. 9.
- [47] Williams, C.B. (1956). A note on an early statistical study of literary style, Biometrika, 45, 248-256.
- [48] Herdan, G. (1961). Quantitative Linguistics, London, Butlerworth.
- [49] 安本美典 (1977). 語彙の量的構造, 数理科学 No. 168, 6月号.
- [50] Williams, C.B. (1940). A note on the statistical analysis of sentence length as a criterion of literary style, Biometrika, 31, 356-361.
- [51] 佐々木和枝 (1976). 文の長さの分布型, 計量国語学 No. 78.

3. 2. 著者推定の統計学

10.1 書き手のクセ

世に偽作と噂される書物、著作は数多く存在する。偉大な思想家、著名な作家の著書として親しまれ愛読され続けてきた作品とて例外ではない。古くはプラトンの「第七書簡」、新約聖書のなかの「パウロの書簡」から現代の作家の作品に至るまで、このような疑問がもたれている多くの作品に関して、真相を「藪の中」から引き出すべく多くの研究がなされてきた。ここでは、この種の謎の解明に統計学がいかなる手がかりを与えてきたかを、代表的ないくつかの研究を通して振り返ってみることにする。

Le style c'est l'homme même (文は人なり) とフランスの文体研究家、P.ギローは述べているが、人それぞれ他の人と区別できる特徴があるように、文章にも他の人の文章と区別できる特徴があるはずである。著者推定の問題に統計的手法を用いるのは、このような書き手の文章における特徴、つまり書き手の「クセ」を統計という道具を用いてつかみとろうとするわけである。

著者推定問題と統計学とのかかわりあいには数学者ド・モルガンの手紙が始まる。1851年、彼はケンブリッジ大学時代の友人に宛てた手紙のなかで、新約聖書の中の「ヘブル人への手紙」(パウロの書簡と呼ばれる14通の手紙の1

つ) がパウロの手によるものかどうかという問題に触れ、単語の長さ(単語を構成する文字数)の平均値を調べたかぎりでは、ヘブル人への手紙はパウロが書いたものではないと推測している。さらに、ヘロドトス等の著作の単語の長さの平均値の例をあげ、最後にこう結んでいる。「ふたりの人間が同じ主題について書いた場合より、ひとりの人間が異なる主題について書いたほうが単語の長さの平均値は近い値となると考えられる。このような考え方によって、近い将来偽書を見出すことが可能になるであろう」と。彼の予測どおり、この手紙を契機に19世紀の後半から種々の統計的手法を用いた著者推定の研究が次々に発表された。たとえば、前出のプラトンの「第七書簡」については、1957年にW. C. ウェイクが、「第七書簡」とプラトンの「法律」との文の長さ(文を構成する単語の数)の比較から、その偽作説を否定している。また、「パウロの書簡」については、1965年にA. Q. モルトンが、書簡中に現われる特定の単語の使用頻度に基づき、次のような説を発表している。「パウロの書簡といわれているもののなかで、もしガラテヤ人への手紙を書いたのがパウロであるならば、ローマ人への手紙、コリント人への第1の手紙、第2の手紙もまた彼の手によるものであるといえる。しかしながら、残り10通の手紙は少なくとも6人の手によるものと考えられる。」

10.2 古典の著者推定

(1) シェークスピアは実在したのか

文藝シェークスピアは、「ハムレット」、「リア王」、「オセロ」、「マクベス」の四大悲劇を始めとして、約20年の間に36もの戯曲を書いたといわれる。しかし、彼が活躍した当時(16世紀)のイギリスにおいては、合作や劇団所有の古い脚本や他人の作品の修正加筆が盛んに行なわれていたために、シェークスピアの作品にも、「ヘンリー六世」、「タイタス・アンドロニカス」など、彼の著作かどうか疑問がもたれている作品がいくつか存在する。このようなことに加え、シェークスピアに関しては生年月日を始めとして、彼の生涯について不明の部分が多くある。このことが後世一部の人々をして次のような疑問を抱

かせることになる。すなわち、「シェークスピアなる人間は歴史上に存在せず、誰かがシェークスピアという偽名で戯曲を発表したのではないか」と。それでは、その誰かとは誰であろうか。彼らは、それが随筆家でありまた哲学者でもあったフランシス・ベーコンではないかと考えた。この問題に対して統計的な立場からの解答を与えたのがT. C. メンデンホールである。彼は1887年に『サイエンス』という雑誌に次のように書いている。「作家が文章のなかでどのような長さの単語をどの程度用いるかという頻度分布を、サッカレー、ディケンズ、ミル等について調べたところ、頻度分布は作家によって異なることがわかった。したがって、このようにもし単語の長さの頻度分布に作家の特徴が現われるのであれば、著者推定の問題にこのことが利用できる。」

そして1901年に彼は、シェークスピアとベーコンの単語の長さの頻度分布を実に40万語、20万語という膨大なデータから求め(図10.1)、シェークスピアの頻度分布のモード(最も使用回数の多い単語の長さ)は、4文字からなる単語であるのに対して、ベーコンの頻度分布のモードは3文字からなる単語であることを示すことによって、シェークスピアなる人間は歴史上に存在せ

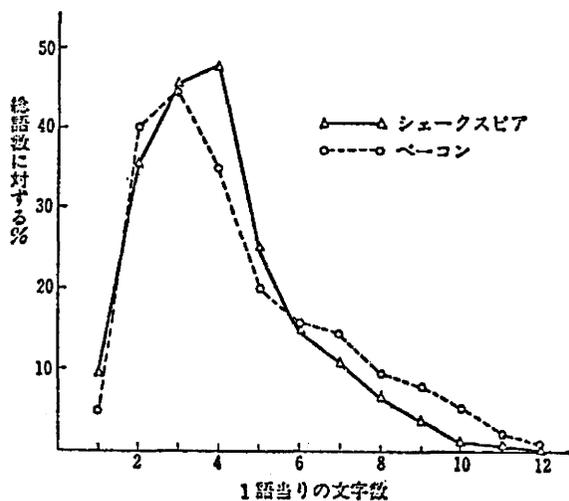


図10.1 シェークスピア、ベーコンの単語の長さの分布

ず。ベーコンが圧制抗議のために一連の風刺劇を書いたという説を否定した。これが著者推定問題に統計が主役となって登場した最初の研究である。

(2) 『キリストにまなびて』

若き日、数学者コーシーはパスカルの「天体力学」、ラグランジュの「解析関数論」とともに、修道士たちの精神生活の完成をめざして書かれたという書物『キリストにまなびて』を受読したという。人間が書いた書物のなかで最も美しいものと評せられ、またカトリック信者必読の書ともいわれるこの15世紀に書かれた書物の著者が誰であるかは、実は明らかではない。

1939年、統計学者ユールは、ベーコン、コールリッジ、ラム、マコーリの文の長さを調べ、その平均値、中央値、四分位範囲等は作家間では異なるが、同一作家の作品では安定した値となるという結果を得た。彼はこの結果をもとに、『キリストにまなびて』の著者をこれらの統計量を用いて推定している。彼はまず著者の可能性のある人物として、聖アグネス修道院の修道士トマス・ア・ケンピスト、最もキリスト教的な学者といわれるジャン・ジェルソンのふたりをあげ、彼らの著作と『キリストにまなびて』の文の長さの平均値、中央値、四分位範囲を求めて、これらの統計量の比較から、ケンピスが著者であろうと考えた。しかし、文の長さに関するこれらの統計量を用いた推定に満足できなかったユールは、1944年、作家の語彙量を示すと考えられる K 統計値という指標を提案し、この指標を用いて再度『キリストにまなびて』の著者推定を行なっている。いま、ある作品のなかに n_i 回出てきた単語が f_i 個あるとすると、この時 $S_1 = \sum f_i n_i$ 、 $S_2 = \sum f_i n_i^2$ として K 統計値は次のように定義される。

$$K = 10^4 \left(\frac{S_2}{S_1^2} - \frac{1}{S_1} \right)$$

さて、問題の書『キリストにまなびて』の K 値は 84.2、それに対してケンピスの著作の K 値は平均 90.26、ジェルソンの著作の K 値は 35.9 であった。この結果から、やはりケンピスのほうがジェルソンより『キリストにまなびて』の著者である可能性が高いとしている。しかし、ケンピス、ジェルソン以

外に、ケンピスが属していた教団の創立者ヘーラルト・ホロートが書いたものをケンピスが修正加筆しながら編集したという説もあり、この問題ははまだ解決をみていない。

(3) マーク・トゥェーンと『Q. C. S. レター』

ここで、ヨーロッパからアメリカへ目を移そう。南北戦争が始まった年1861年に、ニューオリンズの新聞『ニューオリンズ・デェリー・クレセント』誌に「クイントゥス・クルティウス・スノドグラス」という偽名で、南北戦争に関する冒険談が10回ほど掲載された。後の人々が『Q. C. S. レター』と呼ぶこの冒険談の著者は、1934年にある研究者が『トム・ソーヤの冒険』等で知られるユーモア作家マーク・トゥェーンであるという説を発表して以来、一般にそのように信じられ、そしてこの『Q. C. S. レター』が、トゥェーンが南北戦争において果たした役割を推定する重要な文献となった。

ところが、C. S. プリンガーは1963年にアメリカの統計学の雑誌に載せた論文でこの説を否定している。彼はメンデンホールの研究と同じように、使用された単語の長さの頻度分布に注目し、最初に『Q. C. S. レター』とトゥェーンの著作の単語の長さの頻度分布を求めてみた(図10.2)。ところがこの図からもわかるように、この2つの頻度分布のモードは、いずれも3文字からなる

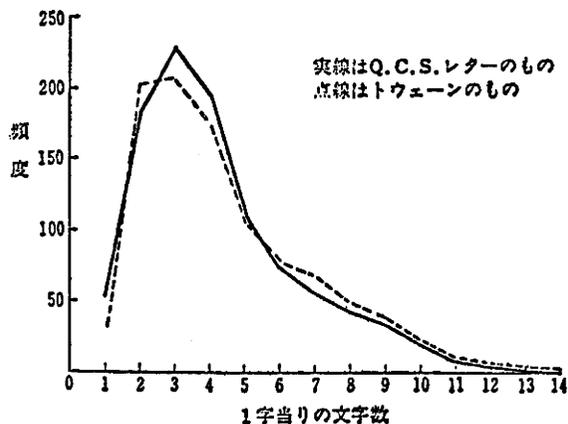


図10.2 『Q. C. S. レター』とトゥェーンの違い

単語であり、したがって、シェークスピア=ペーコン論争のときのように、単にモードだけでは判断できないことがわかった。しかし、この図をよく見ると分布の型が多少異なっていることがわかる。そこで、カイ2乗検定と呼ばれる統計的手法によって、同じ人の文章の単語の長さの頻度分布にこのような違いが生じる可能性を調べてみた結果、その可能性は0.005以下であることがわかった。つまり、この検定法によれば、トウェーンが『Q. C. S. レター』の作者である確率は0.005よりも小さい。さらに、使用頻度の大きい2文字、3文字、4文字からなる単語の使用率に注目して、 χ^2 検定法と呼ばれる方法で、トウェーンが『Q. C. S. レター』の作者である可能性をおおのこの長さの単語について調べたところ、それぞれ0.028、0.002、0.022よりも小さいことがわかった。これらの結果から、彼は『Q. C. S. レター』の作者はトウェーンではないとしている。

ところで、日本でもこの種の研究はなされている。次にそれを簡単に紹介しよう。

(4)『源氏物語・宇治十帖』

平安時代の貴族生活を背景に、光源氏を中心とする華麗な人間関係を描いた紫式部の『源氏物語』は54巻からなる。そのなかで、橋姫の巻以後の宇治を舞台とする10巻は、早くから「宇治十帖」と呼ばれ特別に扱われていたといわれる。安本美典氏は1960年に、この「宇治十帖」と呼ばれる後半の10巻が、前半の44巻と同様紫式部の手によるものであるかどうかを統計的な観点から調べた研究を発表している。

彼は『源氏物語』各巻の(1)長編度(ページ数)、(2)和歌の使用度、(3)直喩の使用度、(4)声喩の使用度、(5)心理描写の数、(6)文の長さ、(7)色彩語の使用度、(8)名詞の使用度、(9)用言の使用度、(10)助詞の使用度、(11)助動詞の使用度、(12)品詞数の12項目に注目して、各項目ごとに54巻の各巻に順位をつけた。たとえば、長編度についてみると、一般的に「宇治十帖」の各巻はページ数が多いので、「宇治十帖」に属する10巻の順位は全体的に小さくなり、反対に前半の44巻の各巻の順位は全体的に大きくなった。しかし、もし「宇治十帖」も前

半の44巻と同じように紫式部の手によるものであれば、このように順位の小さい（ページ数の多い）巻が「宇治十帖」のほうに集中する可能性は少ないであろう。そこで、紫式部が「宇治十帖」の作者であったときに、偶然に「宇治十帖」のほうにこのように小さな順位の巻が集まる可能性を順位検定法によって調べてみたところ、そのような可能性は0.01以下であることがわかった。同じように、他の11項目についても、順位検定法で調べ、さらにこれらの結果をもとにカイ2乗検定を行ない、「宇治十帖」の作者は紫式部以外の人である可能性が強いと主張している。安本氏はまたこの問題に対して、因子分析法という手法を用いての解析も行なっており、やはり同じような結論を得ている。

10.3 「連邦主義者」の著者推定

前節では、文献学的問題に対する統計的方法の寄与の一端を、歴史的順序をおって簡単に触れた。この節では、統計の側からみると、きわめて顕著なものと思える最新の研究を紹介しておこう。それは米国憲法の権威ある解説書として知られる『連邦主義者』という表題の論説集に対するF.モステラーとD.L.ワレスの、著者判別の方法についてである。

おそらく典型的にも思える統計的判別分析法を用いたこの研究を、ここではやや詳細にみてゆくことにしよう。

A.ハミルトンと、J.ジェイおよびJ.マディソン（第4代米大統領）は、米国憲法の批准をすすめるために、1787年から88年にかけて、ニューヨークの新聞に匿名で論説を發表した。その77編と、さらに書き加えられた8編の計85編を含む論説集が、1788年に『連邦主義者』という表題で出版された。これは政治哲学における重要な仕事を残したものとして、アメリカにおいて歴史的に著名なものである。

しかし、誰がこの著者であるかは一般によく知られているものの、論説のそれぞれが誰の手になったかについては、1807年まで議論されたことはなかった。それは決闘によるハミルトンの死後3年のことである。マディソンは大統領の地位を去ってから、そのためのリストをつくったが、そのうち互いに衝突

しあうような多様なリストが現われるようになって、以後この問題は、1世紀半の間、決着のつかない論争を継続させることになった。

ジェイによる5編と、マディソンによる14編、ハミルトンによる51編の、計70編の著者については、一般的意見の一致があり、残り15編のうちの3編は共著とされている。さらにジェイの著作はそれ以外にないとされるので、結局は残り12編のマディソンかハミルトンかへの帰属が問題として残った。

このような紛議のもと、マディソンとハミルトンの政治的立場にあったようで、この著書の出版後数年ならずして、彼らは激しい政敵となって、ときには自分の著述のあるものに対してさえ、反対の立場を表明するようになった。彼らは新聞記者スタイルの文章の熟達者であって、彼らの間の文章の識別はきわめて困難であるといわれる。

文章の長さは、内容の複雑さに関する1つの測度を表わすといわれるが、論争中の12編以外の著述について、マディソンは34.6語、ハミルトンは34.5語がその平均であるという、F.モステラーとF.ウィリアムス(1941年)の報告がある。つまり文章の長さの点ではふたりに差はなく、おそらく文章の形からの判別は、きわめてむずかしからうということを、これは示唆するもののようにみえる。

おそらく、この節のこれまでの叙述を読まれた読者が、統計家であるとしたら、この問題を解析するための統計的方法として、すぐに判別関数の使用を思い付くであろう。F.モステラーとD. L.ワレス(1964年)は、単語の出現頻度だけから、この問題を統計的に解析し、論争中の12編は、マディソンの著作とすべき証拠が圧倒的であると結論した。以降にその方法を平易に解説するが、その狙いは判別関数と同じものである。

(1) 判別子としての単語

まず統計的な方法によらず、使用単語だけから、著者判別をなそうとするなら、その著者独得の使用単語を発見することであろう。いっそう正確には、同一意味内容をもつ異なる単語の対を見つけることである。たとえば、『連邦主義者』中のマディソンによるものと知られる14編では、whileは決して現われ

ず、whilstはその8編で現われるが、ハミルトンでは、15編でwhileが現われ、whilstは決して現われない。これを論争中の12編についてみると、その5編に現われるwhilstの存在は、この5編がマディソンのものであるとする1つの証拠になろう。

しかしこれだけでは、残り7編についてはなんともできない。そこで引続きこれと同じような類の単語の発見ができれば、それらを組み合わせていってその目的が達成されることになろうが、必ずしも可能であるとはかぎらないし、また限られた文献中で発見できたこのような法則が、絶対に確実であるとはかぎらない。事実、この例の場合にも、『連邦主義者』以外の2編にそうでないものがあるのである。

さてこうなると、使用単語を統計的に扱う方法を考えることのほうが、よほど安全確実そうにみえる。ところで、これを統計的に扱うには、まず使用単語の出現頻度分布を比較することであろう。つまり、一方の著者には比較的一般にしばしば使用されているのに、他方の著者にはまれにしか使用されない単語が発見できれば、それは判別子として、判別に強い寄与を与えるだろうと考え、発見されたそのようなもののいくつかを結合して、統計的な判別へ持ち込もうとするのである。もちろん、このときには、単語の対を考える必要はない。

しかし、通常著作に使用されている単語には、その著述の目的および内容に沿って使用されるものと、比較的内容には無関係に使用されるものがある。たとえば、『連邦主義者』についてみると、戦争(war)、法律(law)、自由(liberty)などは、論じている課題に強く関係しよう。つまり著者の特徴というより、話題からの必然性を強く反映していることになるかもしれない。もっとも、そのような課題を選択する著者の好みということがあるかもしれないが、単語の出現頻度だけから判別を実行しようとするいまの場合には、よい判別子にはならなからうと予想される。

モステラーとワレスの場合には、ハミルトンの48編と、マディソンの50編(『連邦主義者』以外のものも含む)から、詰め語(filler word)と呼ばれる

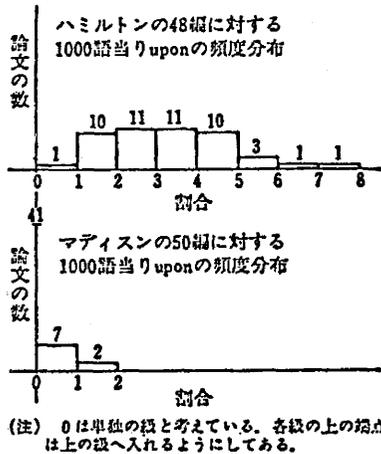


図10.3 ハミルトンとマディソンの比較

前置詞、冠詞、接続詞と、相対的に内容に無関係に思われる副詞、抽象名詞について、1000語当りの(特定の)単語の出現頻度を調べ、最終的に、upon, also, an, by, of, on, there, this, to, although, both, enough, while, whilst, always, though, commonly, apt, according, considerable(ly), consequently, direction, innovation(s), language, vigor(ous), kind, matter(s), particularly, probability, work(s)の30語が選定された。

また単独の判別子として、このなかで最もよいものは upon であって、その割合は、1000語当り、ハミルトンは3、マディソンは1/6である。

(2) 単語の出現頻度分布

文章を1語ずつひろってゆき、あらかじめ決めておいた特定の単語が現われるかどうかをみてゆく。このとき、単語の配列がランダムであって、その単語の出現する確率が、一定の値であるとしたら、 n 語ひろって n_1 語がその単語である確率は、 n_1 が0, 1, 2, ..., n をとる変数であると考え、2項分布になる。変数や2項分布については、近頃は高校の数学の教科書の中でも扱われている。またその n_1 の期待値(平均値)が np であることも知られる。

しかし、単語の配列がランダム、という仮定は、あまり現実味がないと思う

かもしれない。そこで、あらかじめ定めた比較的長い語数について、その内で現われる特定の単語の数としたらどうだろう。つまり文章を特定の長さに、単語を単位に区切ってゆき、各区切りごとに現われる特定の単語の数を数えるのである。2項分布について、 $np=\lambda$ を一定にして、 n をしないで大きくしてゆくと、2項分布は、

$$P(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x=0, 1, 2, \dots \quad (1)$$

という確率分布に近づいてゆく。ここに、 x はあらかじめ定めておいた長さの語数のうちに現われる特定の単語の数であり、 λ はその期待値である。

(1)式の $P(x|\lambda)$ はポワソン分布といわれ、一般にある事柄が起こるか起こらないかを考える場合の、稀現象に関する確率分布と考えられている。図表には、ハミルトンの48編に現われる any の200語当りの出現頻度分布が示してある。視覚的にみて、あてはまりはよい。

さらに、単語の使用の割合を、ガンマ分布に従って変化させると考えると、負の2項分布が得られる。いままでここで例示に使ってきた問題では、an,

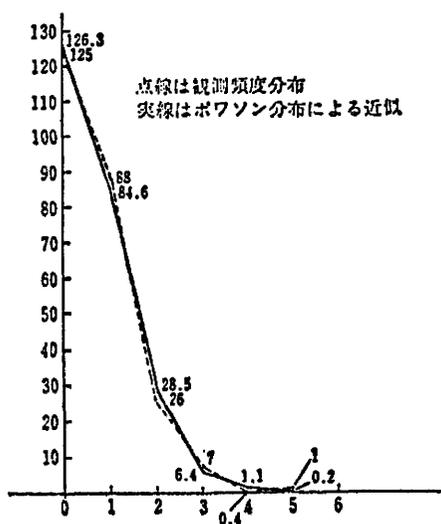


図10.4 ハミルトンの48編中の any の200語当り出現頻度分布

any, upon 等がポワソン分布で, may, his 等が負の2項分布で, かなりあてはまりのよいことが確かめられている.

(3) 勝ち目による推定

2つの仮説, 仮説1と仮説2を考える. いままで述べてきた問題では, それは著者がハミルトンかマディソンかであるが, 仮説 i ($i=1, 2$) が正しいという確率を $w_i = P(\text{仮説 } i)$ としよう. また仮説 i が正しいとき, x を観測する確率を $p_i(x) = P(x | \text{仮説 } i)$ で示そう. このとき, x を観測して仮説 i が正しい確率は,

$$\begin{aligned} w_i(x) &= P(\text{仮説 } i | x) & (2) \\ &= \frac{w_i p_i(x)}{w_1 p_1(x) + w_2 p_2(x)} \quad i=1, 2 \end{aligned}$$

である. これはベイズの定理として知られ, w_i を仮説 i の事前(先験)確率, $w_i(x)$ を事後確率という.

このとき, x の観測に基づく, 仮説2に対する仮説1の勝ち目(odds)は,

$$\begin{aligned} \text{勝ち目}(1, 2 | x) &= \frac{w_1(x)}{w_2(x)} = \left(\frac{w_1}{w_2} \right) \left(\frac{p_1(x)}{p_2(x)} \right) & (3) \\ &= (\text{最初の勝ち目}) \times (\text{尤度比}) \\ &= (\text{最終勝ち目}) \end{aligned}$$

で定義される.

ここで, あらかじめ定めた一定個数の単語を長さの単位に, 文章を区切っていった場合を考え, それを単位に測った論文の長さを L , 単位当りの単語の期待出現頻度を, 仮説 i が正しいときに μ_i ($i=1, 2$) とし, $p_i(x)$ をポアソン分布とすると, 長さ L の論文の1単語に対する尤度比は,

$$\frac{p_1(x)}{p_2(x)} = \left(\frac{\mu_1}{\mu_2} \right)^x e^{-L(\mu_1 - \mu_2)}$$

であって, この対数をとると, すなわち対数尤度比は,

$$x \log(\mu_1 / \mu_2) - L(\mu_1 - \mu_2) \quad (4)$$

である. そこで m 個の独立に生起する単語に対する, 長さ L の論文の対数尤度比は,

$$\sum_{k=1}^m [x_k \log(\mu_{k1}/\mu_{k2}) - L(\mu_{k1} - \mu_{k2})] \quad (5)$$

である。($\sum_{k=1}^m$ は (4) のような m 個の項の和である。) ここに μ_{k1}, μ_{k2} は、単語 k に対する μ_1, μ_2 である。また m 個の単語に対して、

最終対数勝ち目 = 最初の勝ち目 \times (5) の対数尤度比

であるから、判別に適切な単語が、数多く発見できれば、最初の勝ち目にほとんど影響されない最終勝ち目を得ることができる。つまり、判別に主要な役割を果たすのは、(5) の (対数) 尤度比であることになる。

モステラーとワレスの解析では、 μ_k は論説ごとに若干変動するものとみて、そのいくぶん控え目な推定を考える。彼らは論争中の論説すべてが、マディソンの著作とすべき証拠が圧倒的であって、ただ1つだけいくぶん証拠が弱いといえるものがあるが、それでも、80:1 の勝ち目でマディソンを支持するという結果になると報告している。

(4) 判別関数を使う方法

ここで当面している問題について、 (x_1, \dots, x_m) は、選出された m 個の単語の出現頻度であるが、これを一般の問題について考えて、 (x_1, \dots, x_m) が m 種類の種類特性 (項目, 因子) についての測定結果を表わすための変数であるとするとき、その1次関数 $z = a_1 x_1 + \dots + a_m x_m$ を使って、判別を実行しようとするのがよくある。このとき、判別の目的に合致するように、 a_1, \dots, a_m をいかに決めるかが問題になるが、次のような手続きで定められるものを、(1次)判別関数という。それは、

$$\frac{(\text{仮説1のもとでの } z \text{ の期待値} - \text{仮説2のもとでの } z \text{ の期待値})^2}{z \text{ の (級内) 分散}} \quad (6)$$

を最大にするように a_1, \dots, a_m を定めるのである。つまり2つの仮説のもとでの z の期待値が、なるべく離れるように、しかしそれによって、両方の分布の重なりが薄くならなければ無意味であるから、分散による規準化を行なうのである。そのようなわけで、この手続きは、1次関数が最も単純で扱いやすいということを含めて、直観的に実用上合理的にみえる。

さて一方、この問題に対する最適規準は、尤度比に基づくものであることが、一般的に知られている。そこで、 (x_1, \dots, x_m) の仮説 i に対する尤度を $p_i(x_1, \dots, x_m)$ と記すと、それはあらかじめ定められた定数 c に対して、

$$\frac{p_1(x_1, \dots, x_m)}{p_2(x_1, \dots, x_m)} > c \quad (7)$$

となるような (x_1, \dots, x_m) を仮説 1 へ、またそうでない (x_1, \dots, x_m) を仮説 2 へ判別することになる。ただし c は、事前確率 w_1, w_2 に依存する。

さてこのとき、 $p_i(x_1, \dots, x_m)$ が分散・共分散行列(対角成分が x_1, \dots, x_m の分散で、非対角成分がその共分散であるような m 行 m 列の行列)が等しい2つの正規分布であれば、尤度比をもとにして、(1次)判別関数が得られる。

ここでは m 種類の単語に、独立な生起を仮定したから、扱いは比較的単純になった。もちろん形式的に(6)をデータにあてはめて、観測値に対して判別関数をつくることはできる。当然のことながら、このときに独立の仮定は不都合である。しかし、ポアソン分布では、期待値と分散は等しい。これは分散が等しいような単語は、判別に役立たないことを意味するから、判別関数のために、分散・共分散行列が等しいとは仮定できない。

このようなわけで、ここでは独立の仮定に無理がなさそうにみえるので、勝ち目を使う方法に優位が認められよう。

ところで読者は、この章のどこから書き手が変わったのか、おわかりになるであろうか。

4. “三大秘法抄”の著者推定

4.1 『三大秘法抄』の真偽問題の由来と経緯

伊藤 瑞叡

1. 三大秘法抄の教学上の意義と真偽論の由来

日蓮の宗教思想のうち、その教義と実践との根本となる三つの普遍的で秘奥なる真理を、三大秘法という。日蓮が五綱の教判で比較検討の帰結として、また自覚的には法華經の壽量品の文底に秘されたいみじき法として証得したものであり、根本的には一大秘法である。三大秘法は一大秘法である南無妙法蓮華經より開出せられたものであり、本門の本尊・本門の戒壇・本門の題目の三つである。本門の本尊はすべての存在にその主体性として限りなく内在していて、しかも超越的である久遠の本佛（といってもその本質はやはり南無妙法蓮華經）であって、十界曼荼羅として感性的に文字をもって具象化されうる。本門の題目はすべての存在に自覚そのものの眞の活動性として自然に付与されうる具体的普遍である南無妙法蓮華經である。本門の戒壇は本門の本尊に帰依し、本門の題目たる南無妙法蓮華經を唱えるとき、自ら行者の身に無作の具足戒を具えるから、この妙行を修するところが受戒の壇場であることという。

ところで三大秘法は日蓮が佐渡に流されてから後に唱えた説であるが、『三大秘法抄』以外では、三秘の中で本門の戒壇は名目のみで、その説明は皆無である。しかし『三大秘法抄』だけには、本門の戒壇について詳しい説明があり、具体的に戒壇建立の時期とか場所とか、あるいはその意義とかが説示されているため、戒壇法門の帰結は事壇であり、宗門の目標は事壇建立であることが知られる。この書を真作とする派では事壇が建設されるまでは、行者の己心が分の戒壇であり、具体的に建立される事壇が満の戒壇であるとする。一方、偽作とする派では行者が唱題するところ以外に事壇はないというふうに事壇を觀念的に解釈する、いわゆる理壇の立場をとるのである。

このようなわけで、『三大秘法抄』をめぐる、古来より真偽論が争われてきて今日に至っている。

一大秘法は逆縁のものに与えられ、三大秘法は順縁のものが修するものとされている。そして三秘の中で戒壇法門は順縁中の順縁に与えられた密意説として深秘の法門であつたらしく、思われる。したがってその真義は解明し難く、いまだに解明されていないのである。かくして『三大秘法抄』の真偽問題はその解明にとって最重要の意義をもつということができる。

2. 真偽論の概略

『三大秘法抄』は三大秘法稟承事の略称であり通称でもある。弘安4年（または5年説も可）の作とされ、『昭和定本日蓮聖人遺文』では1862頁以下に収められている。日蓮の真筆は現存せず、古来より写本として伝承されてきたためか、成立に関しても内容の上からも、その真偽をめぐる論争がかまびすしく、今日に至っても解決されたとはいえない。

その真偽問題の由来、明治時代以前の真偽論などについて、これを学術的に初めて詳細に論明したのは、国柱会の山川智応博士であり、その「三大秘法抄の真偽問題」という論文である。

すなわち論文中、その4に真偽の異論と題して、古来の真偽の異論を年代順に考察した結果、真撰論者として日蓮宗常門の久遠成院日親（日蓮滅後161年）、日蓮宗向門の行学院日朝（198年以前）、顕本宗什門の妙国寺日悦（209年）、日蓮宗朗門の妙法寺日澄（211年）、本門宗興門の妙本寺日要（233年）、法華宗陣門の本有院日相（463年）、本門法華宗隆門の本興寺日肇（556年）、不受不施講門の安国院日講（415年）、日蓮宗の一妙院日導（508年寂）、日蓮宗の智朗院日賢（535年寂）、日蓮宗の優陀那院日輝（578年寂）を、偽作論者として日蓮宗の日昭門流（180年代）、本門法華宗の本興寺日隆（183年寂）、顕本法華宗の合掌日受（479年）、寂光寺日鑑（585年）を列挙し、「真書としての依用者が、各派各門流に亘りて、大抵、才一流の權威者であり、且つ堂々たる学者であるのに、偽書としての断定者は、日隆師を疑問とせば甚だ狭い範囲で、而も論断も材料も低劣であり、貧弱である」と要約している。

この中、山川博士は日隆には？を附して真偽未決論者と見ているようでもある。

これによると、偽作論者は明治以前では、古い時代の玉詛の昭門流と顕本法華宗とに限られている。

ところでこの山川論文は、偽作論を執った合掌日受の伝統を承ける顕本法華宗の田辺善知師が大正14年に著わした「三大秘法抄を議す」という偽作論に対して、真作論の立場よりした反論である。これとは別に身延山祖山学院教授の塩田義遜博士もまた「三大秘法抄の研究」と題して偽作論を主張したが、山川博士は「三大秘法抄の真偽問題補遺」を論作して反論をなした。

なお田辺論文には日蓮宗の清水龍山師が大正15年に「三大秘法抄真偽弁略」（日宗新報42巻1号）に記載し反論しているし、また塩田論文に対しては日蓮正宗65世日淳師が昭和7年に「塩田義遜師の三大秘法抄の研究を評し併せて偽作説を破す」（昭和7年4月『大日蓮誌』所収）の一文をもって反論している。さらに日蓮宗の北尾日大師も真作論を支持している。

以上が明治以降の真偽論の経緯の概略である。

日蓮宗本山本満寺貫首梅本正雄師は、田辺・山川・塩田の三師の論文を読んで、山川博士の真作論に賛同し、一括して残して置くことを発意し、それらを合録して『三大秘法抄資料』を、昭和44年に公刊した。したがって問題の論争も論争の経緯も、それによって系統的に知られる。

後に鈴木一成博士は『日蓮聖人御遺文』オク巻（昭和8年）の三大秘法抄の解題（319～356頁）において、これらの諸論を総括し、ことに山川論文の研究成果を踏まえ活用しながら、概評を付し新たな資料と考察を加えて私見を提示している。これは学術的な進展を示し、かつ客観的で公平な所論である。とかく偽作論志向の研究者は、単純に真作論者と偽作論者とにわか二分法をとり、真偽未決論者の慶林日隆、小川泰堂をも偽作論者の中に入れて、偽作論者に日昭門流（日実の『宗旨名目』による）、日隆（八品派の祖）、合掌日受（顕本法華宗）、永昌日鑑（同上）、小川泰堂（『高祖遺文録』巻頭の凡例）、田辺善知（『日蓮聖人の本尊論』所収）、塩田義遜（大崎学報77号）

を数え、真作論者に久遠成日親(中山門流)、本成日実(同上)、行学日朝(身延久遠寺11世)、丹明日澄(日朗門流)、一音日暁(日蓮宗)、禅智日好(同上)、一妙日導(同上)、優陀那日輝(同上)、山川智応(『日蓮聖人研究』オノ巻所収)を数えて、偽作論者数を増加せしめようとして公正を欠く場合がありがちであるが、鈴木博士はすでに真作論者・偽作論者・真偽未決論者に三分して学的に厳密を期しているからである。

要するに真偽問題の由来と経緯は、鈴木一成博士の所論において、概略的に列叙し尽されている。加えて最後に博士の私見が付加されてもいるが、それは学的研究の現段階では適正で妥当な見解であるように思われる。

また創価学会編の『佛教哲学大辞典』893~923頁にも、真偽論が詳細に列叙されており、富士門流についての資料についての新たな記述もあり、その中の真作論には一考に値し留意しなければならない所見もあるが、主観に流れて公正を欠く論述なしとはしない矣が残念ではある。

ここではすでに言及し閑説し尽されたこれらの資料により、真偽をめぐる論争の経緯を、真作論者、偽作論者、真偽未決論者に大別三分して整理し、再構成して明示しなければならないと思う。しかし遺憾ではあるが、時間と原稿枚数の限定から、真偽の論案や内容についての詳細は省略をし、他日あらためて論文として纏めることにして、以下、概要骨子のみを摘示しておこう。

3. 上古の古写本

本抄には上古の古写本として三本が現存する。次の如くである。

- (1) 富士門流の日時(聖滅116年)
- (2) 日蓮宗常門(親師法縁々祖)の久遠成院日親(161年)
- (3) 日蓮宗向門の行学院日朝(194年以前)

なおこの中、(1)日時本は昭和29年11月に、日蓮正宗の総本山大石寺の文庫より発見されたものである。これに対しては疑念をいだいている研究者もいるが、正式な反論は出ていない。日時本の末尾では年号が「弘安5年卯月8日」となっているという。しかして富士年表委員会は考証の結果、日時本の信憑に足る理由を確認するに至ったという。真偽論に関係がつかないので、年表委員会の『大日蓮読』(昭和39年11月号)での発表を、転載するべきであるが、長文なので省略する。

4. 真偽に言及する諸文献

本抄を真作として取り扱い、その名目や本文を引用している記録ないし文献は、明治以前のもので、日蓮滅後69年、三位日順の作『摧邪立正抄』を始めとして、江戸末期まで30編に及ぶ(省略)。

偽作とする文献は次の如くである。

- (1) 『当家宗旨名目』上古の日昭門流(聖滅180年代)
- (2) 『本迹自鏡篇』顕本法華宗の合掌日受(497年)
- (3) 『妙宗綱要』日蓮宗の玉沢覚林院日智(573年前)
- (4) 『金剛龜羊弁』顕本法華宗の永昌院日鑑(585年)

本抄を真偽未決とする文献は次の如くである。

- (1)『本門弘経抄』八品派祖日隆(聖滅183年前)
- (2)『高祖遺文録』小川泰堂(584年)

5. 偽作論者(とその所見).

偽作論者を列挙すると、次の如くである。

- (1)上古の日昭門流(玉訳門流)。日蓮宗中山門流の本成日実の『当宗宗旨名目』上巻(西紀2461年)による。
- (2)顕本法華宗の合掌日受(1776年滅)
- (3)顕本法華宗の永昌院日鑑(1869年滅)
- (4)日蓮宗玉訳門流の覚林院日智(1854年滅)
- (5)田辺善知(昭和8年滅)

6. 真偽未決論者(とその所見)

真偽未決論者を列挙すると、次の如くである。

- (1)本門法華宗(八品派)の祖本興寺日隆(1385~1464)
- (2)小川泰堂(1876年滅)

7. 真作論者(とその所見)

主要な真作論者を列挙すると、次の如くである。

- (1)三位阿闍梨日順(1294~1365)
- (2)富士門流の日時(1365~1406)
- (3)中山門流の久遠成院日親(1488年滅)
- (4)中山門流の本成房日実(1461年滅)
- (5)富士門流の日眼(1486年滅)
- (6)身延中興祖の行学院日朝(1422~1500)
- (7)富士門流の左京日教(1428~1489)
- (8)顕本法華宗の慶陽院日悦(1490年)
- (9)六条門流の円明院日澄(1510年滅)
- (10)富士門流の妙本寺日我(1508~1586)
- (11)不受不施講門派の安固院日講(1698年滅)
- (12)一致派の一音院日暁(1710年滅)
- (13)富士門流の日寛(1665~1726)
- (14)禅智院日好(1734年滅)
- (15)法華宗の本有院日相(1756年滅)
- (16)一妙院日導(寛正元年滅)
- (17)本門法華宗の大覚院日肇(1837年)
- (18)優陀那院日輝(1859年滅)
- (19)本門宗の三妙院施命日生(1867年滅)
- (20)富士門流の日露(1861年)

- (21) 日蓮宗の清水龍山 (大正15年)
- (22) 国柱会の山川智応 (大正15年)
- (23) 日蓮正宗大石寺65世日淳 (昭和7年)
- (24) 日蓮宗の北尾日大 (昭和8年)
- (25) 創価学会の統一見解 (『佛教哲学大辞典』所収)
- (26) 立正大学日蓮教学研究所長文学博士宮崎英修 (1972年)
- (27) 法学博士里見岸雄 (昭和47年)
- (28) 国柱会の田中香浦 (昭和51年)

8. ロ決相承説 (=ロ決真意説)

注目すべき新たな特異な見解として、日蓮宗の鈴木一成博士の説がある。博士は「それは本抄の法門は聖人がら老僧や富木・大田等の教団の重立に、深秘の法門としてロ決相承されたもので、その内の一つである太田代の手記が転々伝写されて、真作と伝えられるに至ったものであろうか」と主張する。これは真作論・偽作論・真偽未決論などの既成の三範疇をもって規定されるべき性格のものではない。主旨は日蓮の真意であり、筆録は子弟の相承ないし伝承であるとするこの新説は、ロ決真意説とでもいうべきものである。

かかるロ決真意説ともいうべき見解を認めた場合、筆録伝承伝写の文章の中に、日蓮本人の文章上の特徴がどの程度まで反映しているか、本人の文章と看做しうる可能性の限界との関連はどうか、新しい問題が提起されなければならない。すなわち日隆の「能々之を明らむべし」の言をうけて、ここに文体を科学的に究明することが要請せられることになるのである。

4.2 統計解析

4.2.1 解析に用いた文献について

伊藤 瑞 睿

文体の特徴を系年にわたって総合的に把握するには、規準となる真作を選択する必要がある。

『三大秘法抄』は日蓮の晩年の1281年の作とされている。その1281年に至る20年間の著作の中で、文体が三秘抄と同様に和漢混淆文、あるいはそのように看做しうるものであり、加えて文章量が三秘抄と同程度のものであり、しかも真作として承認されるところのものを、20年間にわたって特定年次に集中することなく、各年次に平均化して総数大体20編前後を選択する必要があるという。

この条件に合致する真作と看做しうる遺文20編は無理であり、ともあれ17編ばかりが可能である。かくして初めに選択した17編は、その中真筆現存のものは7編、曾存のものは2編、現存しないものは8編である。すなわち現存、曾存のものは過半数であり、現存しないものも、古来より、偽作論はこれを聞かないし、偽作することによって特別の意義や効果を期待した特定学派との関係とて考えられないものである。

勿論、現存しないものはこれを用いずとする信念も信念としては可能であろう。しかしそれでは余りに固定的かつ感情的でありすぎるので、今はこの考え方をゆきすぎとして、用いない。

I. 使用文献リスト

・真作・ (和漢混淆文)

1. 二乗作仏の事	(昭和定本番号 19)	1260年
2. 唱法華題目抄	(" 23)	1260年
3. 顕謗法抄	(" 31)	1262年
4. 持妙法華問答抄	(" 33)	1263年
5. 藻玉品得意抄	(" 41)	1265年
6. 善無畏抄	(" 46)	1266年
7. 真言天台勝劣の事	(" 77)	1270年
8. 十章抄	(" 81)	1271年
9. 如説修行抄	(" 124)	1273年
10. 呵責謗法減罪抄	(" 137)	1273年
11. 木絵三像開眼の事	(" 138)	1273年
12. 法蓮抄	(" 175)	1275年
13. 本尊問答抄	(" 307)	1278年
14. 諸経と法華経と難易の事	(" 367)	1280年
15. 諫曉八階抄	(" 395)	1280年
16. 富城入道殿御返事	(" 413)	1281年
17. 曾谷二郎入道殿御報	(" 408)	1281年

..真偽未決.. (和漢混淆文)

0 三大秘法抄 1281年

..偽作.. (和漢混淆文)

21 十八円満抄 1280年

22 法華本門宗要抄 1282年

23 臨終の一心三觀 不明

24 十王讚歎抄 1254年

25 真言見聞 1272年

26 觀心本尊得菴抄 1275年

27 當體蓮華抄 1280年

28 法華大綱抄 1266年

真作の中、真蹟 (=真筆) の現存するものは、次の7書である。

5. 安房保田妙本寺, 外2ヶ所

6. 京都妙顕寺, 外2ヶ所

8. 中山法華経寺

12. 身延山 (曾存), 京都本 寺断片

14. 中山法華経寺

15. 身延山 (曾存), 大石寺断片

16. 中山法華寺

真蹟の曾存のものは、次の2書である。

3. 身延山

11. 身延山

真蹟の現存しないものは、次の7書である。

1

2

4

9. 古写本 日尊筆 茨城富久成寺藏

10

13. 古写本 日興筆 北山本門寺藏

17. 古写本 日興筆 北山本門寺藏

なおこの中、4は日持の代筆という説があり検討を要する。どのような特徴を示すか、興味ぶかい。

II 追加文献リスト

..真作..

1 如来滅後五五百歳觀心本尊鈔 (昭和定本番号 118) 1273年

真蹟現存, 中山法華経寺

2 顕佛未来記 (真蹟曾存, 身延山) (, 125) 1273年

3	法華取要鈔 (昭定番号 145)	1273 年	(真蹟現存, 法華經寺)
4	曾谷入道殿許御書 (" 170)	1275 年	(" 現存, 法華經寺)
5	太田入道殿御返事 (" 197)	1275 年	(" 現存, 豊後鶴崎 法心寺, 外 8ヶ所)
6	始聞佛乘義 (" 277)	1278 年	(" 現存, 中山法華經寺)
7	妙一女御返事 (" 375)	1280 年	(真蹟現存せず)

この中、妙一女御返事のみ和漢混淆文で、他は漢文である。通例により和漢混淆文として読み下した場合の特徴が把握されるべきであるので、読み下した文を使用する。偽作も同様である。

..偽作..

1	問答鈔 (昭定才二輯統編番号 8)	1254 年	和文
2	一念三千法門 (" 14)	1258 年	和漢
3	早勝問答 (" 23)	1271 年	"
4	放光授職灌頂下 (" 28)	1274 年	漢文
5	成佛法華肝心口伝鈔 (" 32)	1275 年	"
6	本寺参詣鈔 (" 46)	1282 年	和文
7	読誦法華用心鈔 (" 48)	1282 年	和漢
8	万法一如鈔 (" 51)	?	和文

..真偽未決..

1	聖愚問答鈔 (昭定才二輯統編番号 43)	1265 年	和漢
2	生死一大事血脈鈔 (" 95)	1272 年	"
3	諸法実相鈔 (" 123)	1273 年	和文
4	日女御前御返事 (" 256)	1277 年	"

III 偽作をなした富士門流の代表的な学匠の著作の一部

- 1 原殿書 (日興真作)
- 2 五重円記 (日興の著述とされる——しかし偽作らしい)
- 3 神天上勘文 (同上)
- 4 本因妙鈔 (同上)
- 5 五人所破鈔 (日澄が日興より相伝したものを日澄の資日順が起草, しかし宮崎英修博士によると日代の著であるという)
- 6 用心鈔 (日順真作)
- 7 心底鈔 (同上)
- 8 従開山伝日順法門 (同上)
- 9 摧邪立正鈔 (日順の著作とされる——偽作らしい)
- 10 富士一跡門徒存知事 (日興の著述とされる——偽作らしい)
- 11 三時弘経次才

4.2.2 解析に用いる情報

文章のいかなる面に著者の特徴が現われるかを知ることが著者推定問題解決の鍵となる。しかしながら一般にはその分野の専門家においても、ある人物の文章の特徴を適格に把握するのは困難であると考えられる。本研究においても日蓮の文章の特徴に関する事前の知識は皆無であった。したがって、過去の研究を参考にして、次に述べるような情報に関して、日蓮の文章の特徴を把握することを試みた。

1. 文の構造に関する情報

文長、単語長、語い量（ユールのK統計量 *etc.*）、各品詞の使用頻度、品詞の接続関係、異なり語数の割合など主として文の構造に関する特徴を把握するための約1000種の統計量（詳細については4.2.5.1を参照）

2. 単語に関する情報

日蓮が好んで用いる単語があるかなど、日蓮の使用単語の特徴を把握するために解析に用いた26の文献（真作17+偽作8+未決1）中に表われる約7200種の異なった言葉の各文献での使用頻度（詳細については4.2.5.2を参照）

3. 論理の展開の仕方に関する情報

仏教の文献には一般的に以下に示すような宗、因、喩、合、結という独得の論理の展開法があるといわれている。

宗 --- 主張 因 --- 主張の理由 喩 --- 引用
合 --- 喩についての適合 結 --- 結論

日蓮の文章においてこのような一般的な論理の展開法がとられているのか、あるいは、彼独自の論理の展開法があるのかを調べる。

（詳細については4.2.5.3を参照）

4. 経年変化のある情報

文章における書き手の特徴は、一生涯不変のものと、年令とともに変化するものがあると考えられる。1, 2, 3の中で年とともに変化すると考えられるものに関しては別に解析する。（詳細については4.2.5.4を参照）

これらの情報を得るために、文章を単語ごとに区切り、各単語を以下の16の品詞のいずれかへ分類した。（数字は計算機での処理の際の品詞コード）

1. 普通名詞 2. 固有名詞 3. 形式名詞 4. 代名詞 5. 数詞
6. 動詞 7. 形容動詞 8. 助詞 9. 接頭語 10. 接尾語

- 11 形容詞 12 感動詞 13 助動詞 14 連体詞 15 副詞
16 接統詞

統計解析の為に使用した統計数理研究所の計算機 (HITAC M180) には漢字の入力出装置がないため、文章はすべて、下記の例のようにローマ字式に読み下し(計算機処理の関係上、ローマ字式に読み下す際はヘボン式を採用せず附録の表1の表記法に従った。ただし“ん”だけは同上の理由から N* として表記した)。

(例) 夫れ法華經第七神力品に云く-----

SORE HOKEKIYAU DAISITI ZIN*RIKIHON* NI IHAKU

なお、解析に用いたテキストは、平樂寺書店版 昭和修 日蓮聖人遺文全集 で、漢字の読み方、句読点等はこのテキストに従った。(ただしこのテキストにおいて明らかにミスと考えられるもの、例えば、同一の漢字の読み方が異なっている場合、あるいは又カッコが閉じていない場合 etc. については、春日、古瀬、伊藤氏等の意見をもとに修正した。)

4.2.3 語の分割と品詞分類

春日 正三
右瀬 順一

1. はじめに

語の分割・認定には『古語辞典』(角川新版・久松潜一・佐藤謙三編)を一応利用した。一応というのは、日蓮聖人の遺文がその用語的特殊性から、文章語すなわち文語的要素と口語—鎌倉時代語—的要素と、それに坂東法師語的要素の位相的性格を秘めていることと、聖人の私信、いわゆる消息文と比較してより公式的文章上から、漢語的仏教語の要素が強く、その為に単純にある種の古語辞典だけでは基準となり得ないという国語史的特殊性を持つと思うからである。そしてまた実際の単語の分割および認定に当っては、『字源』(北辰館・簡野道明著)と『仏教語大辞典』(東京書籍・中村元)を用語検索も含めてあわせ用い、一つの単語を認定したことによるものである。しかし多くの場合、一単語としての認定は、テキスト本文に即ししかも一義を示すと思われる語は、辞書の登録有無にかかわらず一語とし、できうる限り、その情報を豊富ならしめるために細分化した。そしてこの方針は、全作品を通し一貫させた積りである。が、ただ鎌倉時代以前の用法として用例が認められる熟語的合成語についてはその限りでなく、慣用的語句として一語と認定した。ただし、同形の語の識別の都合上、語形変化のある語、すなわち用言ならびに活用連語は、計算機での処理の都合もあって、文法上その基本形(終止形)は同一であると見なされる語であるにもかかわらず異語形として処理されている。従って異なり語数計算上では列々の語として算出される。のでその基本形による語の再構成後算出しなければならない。

漢字・熟語の読みは、テキストの振り仮名表記の記載に従った。テキストには多くの語について振り仮名を付してあるが、違う資料としての御書、ときによっては同一の御書でありながら同一の語に振り仮名を付してある語と施していない語句がある。その場合には振り仮名を施してある語句の読みに従った。しかし「人壽」に「にんじゅ」と「ひとじゅ」の両様の振り仮名が付されている。いささか微妙な問題は存するけれども、これは「にんじゅ」に統一した読み方にした。このような同一熟語、同一漢字に対する振り仮名の異同や「 」, ()が閉じていない、または誤植等の単純な、明らかにミスと思われる語句や振り仮名については、修正を施して資料とした。また「知んぬ」は「知りぬ」に訂正して単語の分割もし、その品詞も断定した。また、計算機での処理上から、記号は訓令式ローマ字綴、ときにはヘボン式表記も併用の音声記号であるので、同一の品詞でしかも同音異義・同音異字の語、例えば「木」「気」「息」「生き」はすべて同一の音節の語、同一品詞の語として集積されるので、これの選別を必要とする。句読点はすべてテキストの記載に従った。

2. 品詞分類基準ならびに品詞コード番号

文法論上の単位として認定された単語を、文法上の性質によって分類したものが品詞である。文法上の性質というのは、ことばのきまりを説明する説明体系のう

之から、単語の形態（形式）に職能の結びついたことば一つ一つのそれぞれが持っている言語使用上の特質を指すのであって、それに意義を加えて分類体系化したものが、品詞の分類であり品詞分類表である。従って品詞分類の基準は、巨視的には、文法論上、語の単位として認定された単語の、意義と形態と職能を基準にして分類される。巨視的立場で分類した日本語の規範的品詞は、以下に示す10品詞である。

- | | | | | | |
|---|--------|---------|--------|---------|-------|
| 詞 | 1. 名詞 | 2. 動詞 | 3. 形容詞 | 4. 形容動詞 | 5. 副詞 |
| | 6. 連体詞 | 7. 接続詞 | 8. 感動詞 | | |
| 辞 | 9. 助詞 | 10. 助動詞 | | | |

以上の品詞についての概念は、常識的に理解されていることなのでその説明をここでは省略することにする。しかし本研究ではその資料を豊富ならしめる為と、より細かな分析のために、上記10品詞以上の品詞を設定したので以下そのことについて説明することにする。

巨視的立場の文法上の性質からは、共通の文法的性格を備えているということの為に、その品詞は名詞として一括されて、固有名詞・普通名詞・代名詞・数詞・形式名詞という品詞名は通常認めていない。しかし、名詞に所属する単語は、他の品詞に属する単語と比較して、実に数多く、しかも文法的内容も複雑、多種多様である。日常の言語生活の中で使用されているこれら名詞に一括されている単語は、その用途に応じて、それぞれに異なる文法的機能や職能、または語義を持って、しかも群れをなして使用されている。ただ雑然と混在しているのではない。たとえば、一つ一つのものが他と区別する為に使われている固有名詞と、同類に属するすべてのものを一切含めて表現するのに可能な普通名詞は、その語の特性において基本的・根本的に違うものであることが認められる。同様に、代名詞も、数詞も、形式名詞も、それぞれその所属する単語の特性を持ってしていると認められる。ということは、巨視的な品詞分類上からは、一つの名詞に属する単語の中により細かな基準を設定することによって、これら名詞を更に細かく分類することが可能であることが分かる。そしてときには、いささか文法的機能というより語義的な面も考慮に入れて、名詞の内部をいくつかの種類に下位分類する。

たとえば、単語の示す対象を客体化しながら一般概念化することの出来る語群と、対象を客体化する文法的機能については前者と変わりがないが、一般概念化するのではなくて、個別概念化する語群とに分類することができる。前者が普通名詞といわれる品詞で、後者が固有名詞といわれる品詞である。この二つの品詞に対する判断は、ときによって、または語によって、いずれにも決しかねるときがある。極端な言い方をすれば、主体者の言語主観に左右されるとも言える。本研究資料の品詞断定でもこの二つの品詞に属する語に対しては、文章の中でその表現内容の概念が、一般概念なのか、個別概念なのか、いささか判然としない面もあって曖昧さも残っている。しかし、単純に、テキストに表現されている語の構成と文脈と語義からして、少なくともその内容が個体表現であり個別概念を示す語であると認識された場合には固有名詞とし、一般概念であると認識された語の場合には、普通名詞とした。

例えば、「法華經」「神力品」「普賢」「天台大師」「疏記」「梵天」「帝釈」「大日經」「法蓮」「禪宗」「念仏宗」「真言宗」等の人名、地名、作品名等属性觀念ならびに個別概念要素の強い語を固有名詞とした。一方、「閻浮提」「御室」「本文」「別經」「圓經」「阿難」「無間地獄」「無間」「真言」「圓」「穢土」「応身」「權實」「權教」「禪法」「相好」「教主」「極樂」「由旬」「本尊」「法華」等一般的事物の一般概念化の職能を持つ語を普通名詞とした。また、対象を客体化する文法的職能は既述の固有名詞、普通名詞と変わらないが、対象を個別概念化し、しかも直接的に指示するという、名詞に属する語群でもっとも体系的なまとまりを示すのが、代名詞である。例えば、「われ」「きみ」「これ」「それ」「あれ」「どれ」「たれ」「かれ」「いずれ」「いすこ」「なに」等々であって、本研究資料には、主として人代名詞と指示代名詞が殆んどで、反射代名詞は、まれである。なお、文語文において「代名詞+助詞(格)」とされる、「この」「その」「あの」「どの」のような語は、これらの語の持つ文法的職能から連体詞とした。以上の連体詞のほか、本研究資料で使用されている連体詞の語彙は、「かの」「いわゆる」「ある」「いかなる」「いぬる」「去る」「餘の」である。

以上のほかに、名詞の下位分類は、数詞と形式名詞がある。数詞には、「一」「三」「五」「七」「九」等の基数詞と、「一つ」「三人」「五回」「七日」「三世」「五字」等の助数詞、「二つめ」「第四番」「第六番」「八番目」「三等」「十四品」等々の序数詞すべてを含めて数詞とした。

形式名詞とは、語形(語音)上普通名詞的でありながら、その内容において対象を客体化することのできない、文法的職能のうすらいだ語群のことである。たとえば、「まま」「ほど」「ころ」「もの」「ところ」「ゆゑ」「ため」「はず」「あひだ」「やう」「かたち」「ごと」等であるが、ときによつては「間」「物」が普通名詞になり、「形」「姿」「様」が形式名詞になった。しかしその場合には、漢字書きの語と仮名書きの語の問題の残った語もあったが、できうる限り全研究資料を通して一貫させた。

以上のように、巨視的立場で分類すると名詞として一括される品詞も、その名詞内部の若干の小異を基準にして分類すると、既述のような下位分類も可能である。そしてこの名詞の下位分類は、名詞の所属語彙それぞれが独自の語彙体系を構成しているものでもあるので、その基準の取り方いかんによつては、名詞所属語群の上下左右各類に分類し、その命名も可能である。

今回の本調査研究では、各語ならびに各品詞の使用割合や品詞間の接続関係等々を数量化し、個々の単語の使用割合に関する情報の蒐集に主目的があったので、巨視的立場からだけの語認定、品詞分類ではなく、微視的立場からの特に名詞にあつてはその語義性も考慮して分類、命名を行なつた。

また、固有名詞に意を注いだのは、筆者が作品の中で固有名詞を使用する場合その固有名詞の表わす属性觀念が、他の個体とどのように、意識の中で区別しているか、または区別しようとしたかの意図を探ぐるのに有効な、しかも重要な品詞であると思つているからである。我々現代人もその言語生活の話しことばにおいては、その使用する語句の、アクセント、アロミネンス等々の強弱によつて、話し手の特殊な感情(喜・怒・哀・楽・敬・卑)要素と、聞き手に対する待遇意識としての敬・卑要素の表現がなされるのと同様、文字言語における固有名詞に付加する「がの」

待遇意識が、各御書にどう表われているかを、調査分析したかったからでもある。上記待遇意識有無と関連するのが、接頭・接尾語である。従ってこれにも一品詞としての場を与えて数量化の対象とした。例えば、「御」「大」「諸」「小」「等」「あひ」「打ち」などである。

巨視的立場で分類した他の品詞については、規範文法をはじめ様々な文法書に説明されていることだし、常識的なことであるとも考えて、あえて紙面を費すこともなかろうと思う。ただ既に述べたことでもあるが、用言に属する語は、その語形が實際上（資料としての御書の中で）どう表現されていようと、省略、脱字とも考え、また一つの語の形態上の変化と見なして、その語群はすべて同一の品詞として数量化した。巨視的、微視的両方の立場から基準を求めて単語を分類した品詞コードは以下の通りとなる。（コード数値には、特に意味はない。語の分析・整理に便利ならしめるために付されたまでである。）

- | | | | | |
|---------|---------|---------|--------|--------|
| 01 普通名詞 | 02 固有名詞 | 03 形式名詞 | 04 代名詞 | 05 数詞 |
| 06 動詞 | 07 形容動詞 | 08 助詞 | 09 接頭語 | 10 接尾語 |
| 11 形容詞 | 12 感動詞 | 13 助動詞 | 14 連体詞 | 15 副詞 |
| 16 接続詞 | | | | |

3. その他

単語認定ならびに品詞の断定にいささか問題点があったけれども、あえて統一的に処理した様々な語句の諸相について述べておく。

1) 「真言宗」は固有名詞としたが、「真言」は普通名詞とした。その理由は、即ちそれぞれの品詞分類の項で、その判断基準について述べてはあるが、「真言宗」は個別概念と見なし、「真言」は一般概念と見たからである。

2) 「三蔵」には、①小乗教を意味する場合、②三蔵に精通した僧に対する尊称としての、いわゆる三蔵法師を意味する場合、③経・律・論の三蔵を意味する場合、④三乗のことを意味する場合がある。そこで、①②の意味に使用している場合には固有名詞とし、③④の意味を持つと判断された場合には普通名詞とした。「如来・菩薩・阿修羅・大龍」等々も一般概念を示していると判断した。

3) 「一念三千・一向・一如・一義・一切経・一代聖教・三無差別」等々も普通名詞として数詞にはとらなかつたが「三千大千世界」の場合には、「三千」が数詞、「大」が接頭語、「千世界」を普通名詞として処理しておいた。

4) 頻度数の高い語、ならびに統一的に処理した品詞を列挙することにして本報告書の結びに代えることにする。順序は、語頭音を五十音図に従って記入する。

「あて・あつて・ありて」接続詞、「ありのまま・ありのままに」「あに」「あまつさへ」副詞、「あざやかに」「安穩なら」「強ちに」「哀れなり」形容動詞、「あらあら」「或は」副詞、「哀れ・あはれ」感動詞、「非ず・あらず」助動詞、「怪しめ/ず/して」動詞+助動詞+助詞、「相/似/たる」接頭語+動詞+助動詞、「況や」「争でか」「いかが」「如何」「いかんぞ」「いずくんぞ」「幾ばく」副詞、「雖/ども・いへ/ども」動詞+助動詞、「何/様に」副詞+形容動詞、「得がたき」形容詞、「於て」「及び」「還つて」「重ねて」接続詞、「各々」「恐らく」「おゝよそ」「をすをす」「かつて」「かくの」「必ずしも」

「殊に」副詞, 「大いなる」形容動詞, 「大いに」副詞, 「おろそかに」副詞, 「おろかに」副詞, 「肝要たる」形容動詞, 「加様に」形容動詞, 「同じ」形容詞, 「どうじ」動詞, 「構え/て」動詞+助詞, 「如く」形容詞, 「如き」形容詞, 「如くん」助動詞, 「とば」助動詞, 「とが」助動詞, 「をや」助動詞, 「をぞ」助動詞, 「かは」助動詞, 「かな」助動詞, 「かし」助動詞, 「先に」副詞, 「更に」副詞, 「しばしば」副詞, 「しばらく」副詞, 「しかしながら」副詞, 「しかるに」副詞, 「しかも」副詞, 「したがって」副詞, 「次第に」副詞, 「夫れ」代名詞, 「そもそも」副詞, 「総じて」副詞, 「それに」接続詞, 「盛んなる」形容動詞, 「左様に」形容動詞, 「速かに」副詞, 「ぜんぜんに」副詞, 「形容動詞」, 「しかる/に」動詞+助詞, 「充滿せ/る」動詞+助動詞, 「加之」(「しか/のみ/なら/ず」)副詞+助詞+動詞+助動詞, 「立どころに」副詞, 「たしかに」副詞, 「互に」副詞, 「多分に」副詞, 「直ちに」副詞, 「たとへば」副詞, 「たとい(へ)」副詞, 「たまたま」副詞, 「ただ」副詞, 「度々」副詞, 「つがさに」副詞, 「常に」副詞, 「ついに」副詞, 「つくづく」と「つらつら」と「とかく」副詞, 「時々」副詞, 「ただし」副詞, 「次に」副詞, 「ついでに」副詞, 「ついでに」副詞, 「同じて」接続詞, 「忽ちに」副詞, 「大なる」形容動詞, 「只今」副詞, 「度」普通名詞, 「為/に」名詞+助詞, 「誑らかさ/れ/て」動詞+助動詞+助動詞, 「なと」副詞, 「何と」副詞, 「何か」副詞, 「何ぞ」副詞, 「何とか」副詞, 「願くは」副詞, 「何しに」副詞, 「にわか」副詞, 「ないし」副詞, 「並びに」接続詞, 「なら/せ/給ひ/けれ」動詞+助動詞+補助動詞+助動詞, 「早々」副詞, 「ひとえに」副詞, 「初めて・始めの」接続詞, 「遙に」副詞, 「分明に」副詞, 「不審に」副詞, 「誠に」副詞, 「又」副詞, 「まさに」副詞, 「まして」副詞, 「全く」副詞, 「自ら」副詞, 「専ら」副詞, 「若し」副詞, 「もともと」副詞, 「勿論」副詞, 「むげに」副詞, 「以/て」接続詞, 「ままに」副詞, 「希に」副詞, 「まちまちに」副詞, 「みだりに」副詞, 「無益なり」形容動詞, 「滅せ/る」動詞+助動詞, 「みたれ/かわしく」動詞+形容詞, 「やがて」副詞, 「やや」副詞, 「能く」副詞, 「餘に」副詞, 「よも」副詞, 「ようやく」副詞, 「ゆめゆめ」副詞, 「よし」副詞, 「故に」副詞, 「依/て」副詞, 「例せば」接続詞, 「様」形式名詞, 「様に」副詞, 「様なれ」副詞, 「ゆるやかに」副詞, 「餘年なり」形容動詞, 「譲ら/せ」動詞+助動詞.

4.2.4 統計解析

4.2.4.1 文の構造に関する情報

前述のように日蓮の文章における特徴に関して事前の知識がない為に、過去の研究を参考にし、特徴がでる可能性のある以下に記すような約1000の統計量を計算した。これらの中には互いに相関のある統計量が多く存在する。(表4:1)

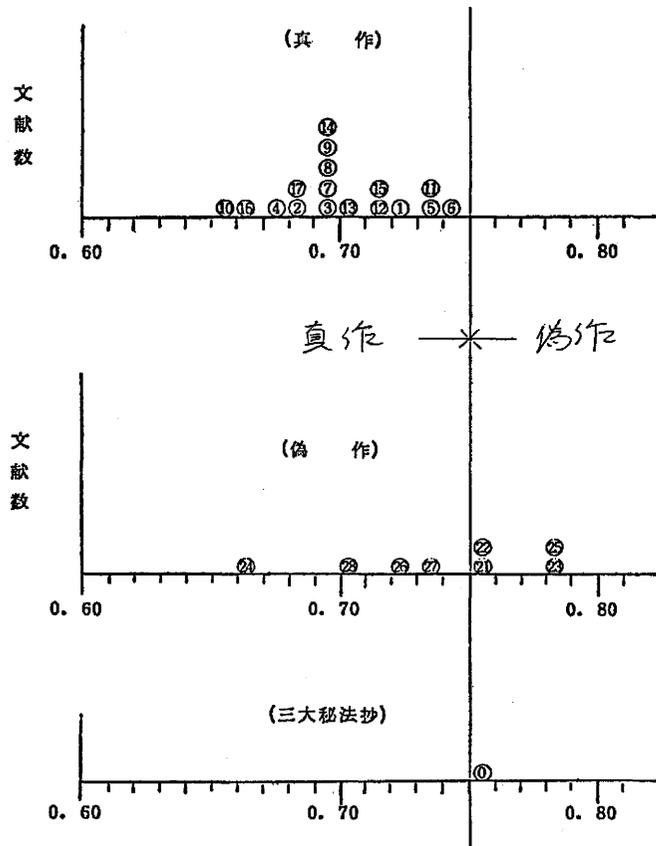
これらの情報を利用して、A判別分析(2群判別, 多群判別) B主成分分析の結果と利用した判別分析 C主成分分析の結果を利用したクラスター分析によって解析を行った。

A 判別分析 (discriminant analysis) による解析

1) 2群判別

解析に用いた文献を真作と偽作の2グループに大別し、“三大秘法抄”がどちらのグループに属するかを推測する。

前記の約1000の変数の中には1つの変数で真作グループと偽作グループを完全に分離できるものは存在しない。両グループの分離に最も有効な変数は、“助詞の前に名詞がくる割合”(コードNo891)であるが(図1)仮に、この変数の0.75を分離点としても、偽作の4編が真作と誤判断されることになる。(891) 助詞の前に名詞がくる割合 (図1)



コード	統制番号 or 変数名	コード	統制番号 or 変数名	コード	統制番号 or 変数名
1	文献番号	31	1~5語から成る文の割合	61	異なり語数の割合 (550語)
2	著作年	2	6~10語 "	2	" (600)
3	総単語数	3	11~15語 "	3	昔名 → 昔名への接続の割合
4	本文中の単語数	4	16~20語 "	4	固名 → "
5	(本文中の単語数)/(総単語数)	5	21~25語 "	5	形名 → "
6	引用文中の単語数	6	26~30語 "	6	代名 → "
7	(引用文中の単語数)/(総単語数)	7	31~35語 "	7	数詞 → "
8	異なり語数	8	36~40語 "	8	動詞 → "
9	(異なり語数)/(総単語数)	9	41~45語 "	9	形動詞 → "
10	文の数	40	46~50語 "	70	助詞 → "
1	平均の文長(単位:語)	1	1~10文字から成る文の割合	1	接頭語 → "
2	" のS.D.	2	11~20文字 "	2	接尾語 → "
3	平均の文長(単位:字)	3	21~30文字 "	3	形容詞 → "
4	" のS.D.	4	31~40文字 "	4	副詞 → "
5	漢語の割合	5	41~50文字 "	5	助動詞 → "
6	総文字数(カナ読み1=2)	6	51~60文字 "	6	連体詞 → "
7	本文中の文字数	7	61~70文字 "	7	副詞 → "
8	引用文中の文字数	8	71~80文字 "	8	接続詞 → "
9	(本文中の文字数)/(総文字数)	9	81~90文字 "	9	昔名 → 固名への接続の割合
10	(引用文中の文字数)/(総文字数)	50	91~100文字 "	80	固名 → "
1	1字より成る単語の出現割合	1	異なり語数の割合 (50語)	1	形名 → "
2	2	2	" (100)	2	代名 → "
3	3	3	" (150)	3	数詞 → "
4	4	4	" (200)	4	動詞 → "
5	5	5	" (250)	5	形動詞 → "
6	6	6	" (300)	6	助詞 → "
7	7	7	" (350)	7	接頭語 → "
8	8	8	" (400)	8	接尾語 → "
9	9	9	" (450)	9	形格詞 → "
30	10	60	" (500)	90	成動詞 → "

のS.D. ... 標準偏差

の"2"は終る単語の割合であって、必ずしも漢語とは限らない。

コードNo	統訂量 or 変数名	コードNo	統訂量 or 変数名	コードNo	統訂量 or 変数名
91	助動詞 → 固名への接続の割合	121	形容詞 → 代名への接続の割合	151	接続語 → 動詞への接続の割合
2	連体詞 → "	2	感動詞 → "	2	接尾語 → "
3	副詞 → "	3	助動詞 → "	3	形容詞 → "
4	接続詞 → "	4	連体詞 → "	4	感動詞 → "
5	普名 → 形名への接続の割合	5	副詞 → "	5	助動詞 → "
6	固名 → "	6	接続詞 → "	6	連体詞 → "
7	形名 → "	7	普名 → 数詞への接続の割合	7	副詞 → "
8	代名 → "	8	固名 → "	8	接続詞 → "
9	数詞 → "	9	形名 → "	9	普名 → 形容動詞への接続割合
100	動詞 → "	130	代名 → "	160	固名 → "
1	形動詞 → "	1	数詞 → "	1	形名 → "
2	助詞 → "	2	動詞 → "	2	代名 → "
3	接続語 → "	3	形動詞 → "	3	数詞 → "
4	接尾語 → "	4	助詞 → "	4	動詞 → "
5	形容詞 → "	5	接続語 → "	5	形動詞 → "
6	感動詞 → "	6	接尾語 → "	6	助詞 → "
7	助動詞 → "	7	形容詞 → "	7	接続語 → "
8	連体詞 → "	8	感動詞 → "	8	接尾語 → "
9	副詞 → "	9	助動詞 → "	9	形容詞 → "
110	接続詞 → "	140	連体詞 → "	170	感動詞 → "
1	普名 → 代名への接続の割合	1	副詞 → "	1	助動詞 → "
2	固名 → "	2	接続詞 → "	2	連体詞 → "
3	形名 → "	3	普名 → 動詞への接続の割合	3	副詞 → "
4	代名 → "	4	固名 → "	4	接続詞 → "
5	数詞 → "	5	形名 → "	5	普名 → 助詞への接続の割合
6	動詞 → "	6	代名 → "	6	固名 → "
7	形動詞 → "	7	数詞 → "	7	形名 → "
8	助詞 → "	8	動詞 → "	8	代名 → "
9	接続語 → "	9	形動詞 → "	9	数詞 → "
120	接尾語 → "	150	助詞 → "	180	動詞 → "

下NO	統訂量 or 変数名	上NO	統訂量 or 変数名	下NO	統訂量 or 変数名
181	形動詞 → 助詞の接続の割合	211	数詞 → 接尾語の接続の割合	241	形名 → 感動詞の接続の割合
2	助詞 → "	2	動詞 → "	2	代名 → "
3	接頭語 → "	3	形容詞 → "	3	数詞 → "
4	接尾語 → "	4	助詞 → "	4	動詞 → "
5	形容詞 → "	5	接頭語 → "	5	形動詞 → "
6	感動詞 → "	6	接尾語 → "	6	助詞 → "
7	助動詞 → "	7	形容詞 → "	7	接頭語 → "
8	連体詞 → "	8	感動詞 → "	8	接尾語 → "
9	副詞 → "	9	助動詞 → "	9	形容詞 → "
190	接続詞 → "	220	連体詞 → "	250	感動詞 → "
1	普名 → 接頭語の接続の割合	1	副詞 → "	1	助動詞 → "
2	同名 → "	2	接続詞 → "	2	連体詞 → "
3	形名 → "	3	普名 → 形容詞の接続の割合	3	副詞 → "
4	代名 → "	4	同名 → "	4	接続詞 → "
5	数詞 → "	5	形名 → "	5	普名 → 助動詞の接続の割合
6	動詞 → "	6	代名 → "	6	同名 → "
7	形動詞 → "	7	数詞 → "	7	形名 → "
8	助詞 → "	8	動詞 → "	8	代名 → "
9	接頭語 → "	9	形動詞 → "	9	数詞 → "
200	接尾語 → "	230	助詞 → "	260	動詞 → "
1	形容詞 → "	1	接頭語 → "	1	形動詞 → "
2	感動詞 → "	2	接尾語 → "	2	助詞 → "
3	助動詞 → "	3	形容詞 → "	3	接頭語 → "
4	連体詞 → "	4	感動詞 → "	4	接尾語 → "
5	副詞 → "	5	助動詞 → "	5	形容詞 → "
6	接続詞 → "	6	連体詞 → "	6	感動詞 → "
7	普名 → 接尾語の接続の割合	7	副詞 → "	7	助動詞 → "
8	同名 → "	8	接続詞 → "	8	連体詞 → "
9	形名 → "	9	普名 → 感動詞の接続の割合	9	副詞 → "
210	代名 → "	240	同名 → "	270	接続詞 → "

統計表 10-1

ナNo	統計号 or 変数名	ナNo	統計号 or 変数名	ナNo	統計号 or 変数名
27	1 普名 → 連体詞の接続の割合	30	1 副詞 → 副詞への接続の割合	33	1 助動詞の後に普名が来る割合
2	2 固名 → "	2	2 接続詞 → "	2	2 連体詞 → "
3	3 形名 → "	3	3 普名 → 接続詞の接続の割合	3	3 副詞 → "
4	4 代名 → "	4	4 固名 → "	4	4 接続詞 → "
5	5 数詞 → "	5	5 形名 → "	5	5 普名の後に固名が来る割合
6	6 動詞 → "	6	6 代名 → "	6	6 固名 → "
7	7 形動詞 → "	7	7 数詞 → "	7	7 形名 → "
8	8 助詞 → "	8	8 動詞 → "	8	8 代名 → "
9	9 接頭語 → "	9	9 形動詞 → "	9	9 数詞 → "
28	10 接尾語 → "	31	10 助詞 → "	34	10 動詞 → "
1	1 形容詞 → "	1	1 接頭語 → "	1	1 形動詞 → "
2	2 感動詞 → "	2	2 接尾語 → "	2	2 助詞 → "
3	3 助動詞 → "	3	3 形容詞 → "	3	3 接頭語 → "
4	4 連体詞 → "	4	4 感動詞 → "	4	4 接尾語 → "
5	5 副詞 → "	5	5 助動詞 → "	5	5 形容詞 → "
6	6 接続詞 → "	6	6 連体詞 → "	6	6 感動詞 → "
7	7 普名 → 副詞への接続の割合	7	7 副詞 → "	7	7 助動詞 → "
8	8 固名 → "	8	8 接続詞 → "	8	8 連体詞 → "
9	9 形名 → "	9	9 普名の後に普名が来る割合 ¹⁾	9	9 副詞 → "
29	10 代名 → "	32	10 固名 → "	35	10 接続詞 → "
1	1 数詞 → "	1	1 形名 → "	1	1 普名の後に形名が来る割合
2	2 動詞 → "	2	2 代名 → "	2	2 固名 → "
3	3 形動詞 → "	3	3 数詞 → "	3	3 形名 → "
4	4 助詞 → "	4	4 動詞 → "	4	4 代名 → "
5	5 接頭語 → "	5	5 形動詞 → "	5	5 数詞 → "
6	6 接尾語 → "	6	6 助詞 → "	6	6 動詞 → "
7	7 形容詞 → "	7	7 接頭語 → "	7	7 形動詞 → "
8	8 感動詞 → "	8	8 接尾語 → "	8	8 助詞 → "
9	9 助動詞 → "	9	9 形容詞 → "	9	9 接頭語 → "
30	10 連体詞 → "	33	10 感動詞 → "	36	10 接尾語 → "

1) (普名 → 普名の接続の割合) / (普名 → 全ての品詞への接続の割合)

トNo	統制番号 or 変数名	トNo	統制番号 or 変数名	トNo	統制番号 or 変数名
361	形容詞の後に形名が来ると結合	391	接頭語の後に数詞が来ると結合	421	形動詞の後に形動詞が来ると結合
2	感動詞	2	接尾語	2	助詞
3	助動詞	3	形容詞	3	接頭語
4	連体詞	4	感動詞	4	接尾語
5	副詞	5	助動詞	5	形容詞
6	接続詞	6	連体詞	6	感動詞
7	普名の後に代名が来ると結合	7	副詞	7	助動詞
8	同名	8	接続詞	8	連体詞
9	形名	9	普名の後に動詞が来ると結合	9	副詞
370	代名	400	国名	430	接続詞
1	数詞	1	形名	1	普名の後に助詞が来ると結合
2	動詞	2	代名	2	国名
3	形動詞	3	数詞	3	形名
4	助詞	4	動詞	4	代名
5	接頭語	5	形動詞	5	数詞
6	接尾語	6	助詞	6	動詞
7	形容詞	7	接頭語	7	形動詞
8	感動詞	8	接尾語	8	助詞
9	助動詞	9	形容詞	9	接頭語
380	連体詞	410	感動詞	440	接尾語
1	副詞	1	助動詞	1	形容詞
2	接続詞	2	連体詞	2	感動詞
3	普名の後に数詞が来ると結合	3	副詞	3	助動詞
4	国名	4	接続詞	4	連体詞
5	形名	5	普名の後に形動詞が来ると結合	5	副詞
6	代名	6	国名	6	接続詞
7	数詞	7	形名	7	普名の後に接頭語が来ると結合
8	動詞	8	代名	8	国名
9	形動詞	9	数詞	9	形名
390	助詞	420	動詞	450	代名

1-F No	統计量の变数名	2-F No	統计量の变数名	3-F No	統计量の变数名
45	数詞の後に接頭語のまと割合	48	形名の後に形容詞のまと割合	51	普名の後に助動詞のまと割合
2	動詞	2	代名	2	固名
3	形動詞	3	数詞	3	形名
4	助詞	4	動詞	4	代名
5	接頭語	5	形動詞	5	数詞
6	接尾語	6	助詞	6	動詞
7	形容詞	7	接頭語	7	形動詞
8	感動詞	8	接尾語	8	助詞
9	助動詞	9	形容詞	9	接頭語
46	連体詞	49	感動詞	52	接尾語
1	副詞	1	助動詞	1	形容詞
2	接統詞	2	連体詞	2	感動詞
3	普名の後に接尾語のまと割合	3	副詞	3	助動詞
4	固名	4	接統詞	4	連体詞
5	形名	5	普名の後に感動詞のまと割合	5	副詞
6	代名	6	固名	6	接統詞
7	数詞	7	形名	7	普名の後に連体詞のまと割合
8	動詞	8	代名	8	固名
9	形動詞	9	数詞	9	形名
47	助詞	50	動詞	53	代名
1	接頭語	1	形動詞	1	数詞
2	接尾語	2	助詞	2	動詞
3	形容詞	3	接頭語	3	形動詞
4	感動詞	4	接尾語	4	助詞
5	助動詞	5	形容詞	5	接頭語
6	連体詞	6	感動詞	6	接尾語
7	副詞	7	助動詞	7	形容詞
8	接統詞	8	連体詞	8	感動詞
9	普名の後に形容詞のまと割合	9	副詞	9	助動詞
48	固名	51	接統詞	54	連体詞

2-F-40	統計量の変数名	2-F-40	統計量の変数名	2-F-40	統計量の変数名
54	副詞の後に連体詞が来り割合	57	助動詞の後に接続詞が来り割合	60	国名の前に形容詞が来り割合
2	接続詞	2	連体詞	2	感動詞
3	普名の後に副詞が来り割合	3	副詞	3	助動詞
4	国名	4	接続詞	4	連体詞
5	形名	5	普名の前に普名が来り割合	5	副詞
6	代名	6	国名	6	接続詞
7	数詞	7	形名	7	形名の前に普名が来り割合
8	動詞	8	代名	8	国名
9	形動詞	9	改詞	9	形名
550	助詞	580	動詞	610	代名
1	接頭語	1	形動詞	1	改詞
2	接尾語	2	助詞	2	動詞
3	形容詞	3	接頭語	3	形動詞
4	感動詞	4	接尾語	4	助詞
5	助動詞	5	形容詞	5	接頭語
6	連体詞	6	感動詞	6	接尾語
7	副詞	7	助動詞	7	形容詞
8	接続詞	8	連体詞	8	感動詞
9	普名の後に接続詞が来り割合	9	副詞	9	助動詞
560	国名	590	接続詞	620	連体詞
1	形名	1	国名の前に普名が来り割合	1	副詞
2	代名	2	国名	2	接続詞
3	数詞	3	形名	3	代名の前に普名が来り割合
4	動詞	4	代名	4	国名
5	形動詞	5	数詞	5	形名
6	助詞	6	動詞	6	代名
7	接頭語	7	形動詞	7	改詞
8	接尾語	8	助詞	8	動詞
9	形容詞	9	接頭語	9	形動詞
570	感動詞	600	接尾語	630	助詞

① (格名→普名の接続詞格)/(全20の品詞→普名の接続詞格)

コード	統計量の変数名	コード	統計量の変数名	コード	統計量の変数名
63	代名の前に接頭語が来る割合	66	動詞の前に形動詞が来る割合	69	助詞の前に数詞が来る割合
2	接尾語	2	助詞	2	動詞
3	形容詞	3	接頭語	3	形動詞
4	感動詞	4	接尾語	4	助詞
5	助動詞	5	形容詞	5	接頭語
6	連体詞	6	感動詞	6	接尾語
7	副詞	7	助動詞	7	形容詞
8	接続詞	8	連体詞	8	感動詞
640	数詞の前に普名が来る割合	670	副詞	69	助動詞
1	同名	1	接換詞	9	連体詞
2	形名	2	形動詞の前に普名が来る割合	700	副詞
3	代名	3	同名	1	接続詞
4	数詞	4	形名	2	接頭語の前に普名が来る割合
5	動詞	5	代名	3	同名
6	形容詞	6	数詞	4	形名
7	助詞	7	動詞	5	代名
8	接頭語	8	形動詞	6	数詞
9	接尾語	9	助詞	7	動詞
650	形容詞	680	接換語	8	形動詞
1	感動詞	1	接尾語	9	助詞
2	助動詞	2	形容詞	1	接頭語
3	連体詞	3	感動詞	2	接尾語
4	副詞	4	助動詞	3	形容詞
5	接続詞	5	連体詞	4	感動詞
660	動詞の前に普名が来る割合	690	副詞	5	助動詞
1	同名	6	接換詞	6	連体詞
2	形名	7	助詞の前に普名が来る割合	7	副詞
3	代名	8	同名	8	接続詞
4	数詞	9	形名	9	接尾語の前に普名が来る割合
660	動詞	690	代名	220	同名

1-F20	統計量の变数名	2-F20	統計量の变数名	3-F20	統計量の变数名
72	接尾語・前1: 形名 加和割合	75	感動詞・前1: 普名 加和割合	78	助動詞・前1: 副詞 加和割合
2	" 代名 "	2	" 固名 "	2	接頭語 "
3	" 数詞 "	3	" 形名 "	3	連体詞・前1: 普名 加和割合
4	" 動詞 "	4	" 代名 "	4	" 固名 "
5	" 形動詞 "	5	" 数詞 "	5	" 形名 "
6	" 助詞 "	6	" 動詞 "	6	" 代名 "
7	" 接頭語 "	7	" 形動詞 "	7	" 数詞 "
8	" 接尾語 "	8	" 助詞 "	8	" 動詞 "
9	" 形容詞 "	9	" 接頭語 "	9	" 形動詞 "
730	" 感動詞 "	760	" 接尾語 "	790	" 助詞 "
1	" 助動詞 "	1	" 形容詞 "	1	" 接頭語 "
2	" 連体詞 "	2	" 感動詞 "	2	" 接尾語 "
3	" 副詞 "	3	" 助動詞 "	3	" 形容詞 "
4	" 接頭語 "	4	" 連体詞 "	4	" 感動詞 "
5	形容詞・前1: 普名 加和割合	5	" 副詞 "	5	" 助動詞 "
6	" 固名 "	6	" 接頭語 "	6	" 連体詞 "
7	" 形名 "	7	助動詞・前1: 普名 加和割合	7	" 副詞 "
8	" 代名 "	8	" 固名 "	8	" 接頭語 "
9	" 数詞 "	9	" 形名 "	9	副詞・前1: 普名 加和割合
740	" 動詞 "	770	" 代名 "	800	" 固名 "
1	" 形動詞 "	1	" 数詞 "	1	" 形名 "
2	" 助詞 "	2	" 動詞 "	2	" 代名 "
3	" 接頭語 "	3	" 形動詞 "	3	" 数詞 "
4	" 接尾語 "	4	" 助詞 "	4	" 動詞 "
5	" 形容詞 "	5	" 接頭語 "	5	" 形動詞 "
6	" 感動詞 "	6	" 接尾語 "	6	" 助詞 "
7	" 助動詞 "	7	" 形容詞 "	7	" 接頭語 "
8	" 連体詞 "	8	" 感動詞 "	8	" 接尾語 "
9	" 副詞 "	9	" 助動詞 "	9	" 形容詞 "
750	" 接頭語 "	780	" 連体詞 "	810	" 感動詞 "

統計表作成

コード	統計量の変数名	コード	統計量の変数名	コード	統計量の変数名
811	副詞の前に助動詞がある割合	841	副詞→名詞への接続の割合	871	名詞の後に形容詞がある割合
2	“ 連体詞 ”	2	接続詞→ “	2	“ 感動詞 ”
3	“ 副詞 ”	3	名詞→動詞への接続の割合	3	“ 助動詞 ”
4	“ 接続詞 ”	4	“ →形動詞 ”	4	“ 連体詞 ”
5	接続詞前に普通名がある割合	5	“ →助詞 ”	5	“ 副詞 ”
6	“ 固有名 ”	6	“ →接頭語 ”	6	“ 接続詞 ”
7	“ 形名 ”	7	“ →接尾語 ”	7	名詞の前に名詞がある割合
8	“ 代名 ”	8	“ →形容詞 ”	8	“ 動詞 ”
9	“ 数詞 ”	9	“ →感動詞 ”	9	“ 形動詞 ”
820	“ 動詞 ”	850	“ →助動詞 ”	880	“ 助詞 ”
1	“ 形動詞 ”	1	“ →連体詞 ”	1	“ 接頭語 ”
2	“ 助詞 ”	2	“ →副詞 ”	2	“ 接尾語 ”
3	“ 接頭語 ”	3	“ →接続詞 ”	3	“ 形容詞 ”
4	“ 接尾語 ”	4	名詞の後に名詞がある割合	4	“ 感動詞 ”
5	“ 形容詞 ”	5	動詞 ”	5	“ 助動詞 ”
6	“ 感動詞 ”	6	形動詞 ”	6	“ 連体詞 ”
7	“ 助動詞 ”	7	助詞 ”	7	“ 副詞 ”
8	“ 連体詞 ”	8	接頭語 ”	8	“ 接続詞 ”
9	“ 副詞 ”	9	接尾語 ”	9	動詞の前に名詞がある割合
830	“ 接続詞 ”	860	形容詞 ”	890	形動詞 ”
1	名詞→名詞への接続の割合	1	感動詞 ”	1	助詞 ”
2	動詞→ “	2	助動詞 ”	2	接頭語 ”
3	形動詞→ “	3	連体詞 ”	3	接尾語 ”
4	助詞→ “	4	副詞 ”	4	形容詞 ”
5	接頭語→ “	5	接続詞 ”	5	感動詞 ”
6	接尾語→ “	6	名詞の後に動詞がある割合	6	助動詞 ”
7	形容詞→ “	7	“ 形動詞 ”	7	連体詞 ”
8	感動詞→ “	8	“ 助詞 ”	8	副詞 ”
9	助動詞→ “	9	“ 接頭語 ”	9	接続詞 ”
840	連体詞→ “	870	“ 接尾語 ”	900	

1) コード No 831 ~ コード No 899 は「名詞」は普通名詞, 固有名詞, 形式名詞, 代名詞, 数詞 を含めた
 の

コード No 831 と コード No 854 の違いは?

コードNo	統計量 or 変数名	コードNo	統計量 or 変数名	コードNo	統計量 or 変数名
90	普通名詞の使用割合	93	接頭語の使用数/動詞の使用数	96	副詞の使用数/接頭語の使用数
2	固有名詞	2	接尾語の使用数/	2	接続詞の使用数/
3	形式名詞	3	形容詞の使用数/	3	形容詞の使用数/接尾語の使用数
4	代名詞	4	感動詞の使用数/	4	感動詞の使用数/
5	数詞	5	助動詞の使用数/	5	助動詞の使用数/
6	動詞	6	連体詞の使用数/	6	連体詞の使用数/
7	形動詞	7	副詞の使用数/	7	副詞の使用数/
8	助詞	8	接続詞の使用数/	8	接続詞の使用数/
9	接頭語	9	助詞の使用数/形動詞の使用数	9	感動詞の使用数/形容詞の使用数
910	接尾語	940	接頭語の使用数/	970	助動詞の使用数/
1	形容詞	1	接尾語の使用数/	1	連体詞の使用数/
2	感動詞	2	形容詞の使用数/	2	副詞の使用数/
3	助動詞	3	感動詞の使用数/	3	接続詞の使用数/
4	連体詞	4	助動詞の使用数/	4	助動詞の使用数/感動詞の使用数
5	副詞	5	連体詞の使用数/	5	連体詞の使用数/
6	接続詞	6	副詞の使用数/	6	副詞の使用数/
7	名詞 ¹⁾	7	接続詞の使用数/	7	接続詞の使用数/
8	動詞の使用数/名詞の使用数	8	接頭語の使用数/助詞の使用数	8	連体詞の使用数/助動詞の使用数
9	形動詞の使用数/	9	接尾語の使用数/	9	副詞の使用数/
920	助詞の使用数/	950	形容詞の使用数/	980	接続詞の使用数/
1	接頭語の使用数/	1	感動詞の使用数/	1	副詞の使用数/連体詞の使用数
2	接尾語の使用数/	2	助動詞の使用数/	2	接続詞の使用数/
3	形容詞の使用数/	3	連体詞の使用数/	3	接続詞の使用数/副詞の使用数
4	感動詞の使用数/	4	副詞の使用数/	4	出現度数1の単語の割合
5	助動詞の使用数/	5	接続詞の使用数/	5	" 2 "
6	連体詞の使用数/	6	接尾語の使用数/接頭語の使用数	6	" 3 "
7	副詞の使用数/	7	形容詞の使用数/	7	" 4 "
8	接続詞の使用数/	8	感動詞の使用数/	8	" 5 "
9	形容詞の使用数/動詞の使用数	9	助動詞の使用数/	9	" 6 "
930	助詞の使用数/	960	連体詞の使用数/	990	" 7 "

1) 普通名詞 + 固有名詞 + 形式名詞 + 代名詞 + 数詞

統計学 第 1 巻

コード	統計量の変数名	コード	統計量の変数名	コード	統計量の変数名
19	出現度数8の単語の割合	1661	ユールのK統計量(全品詞)	1691	
2	" 9 "	2	" (名詞)	2	
3	" 10 "	3	" (普通名詞)	3	
4		4	" (固有名詞)	4	
5		5	" (形式名詞)	5	
6		6	" (代名詞)	6	
7		7	" (数詞)	7	(動詞, 助動詞, 助詞, 助格詞, 助詞, 助格詞, 助詞, 助格詞)
8		8	" (動詞)	8	動詞, 助動詞, 助詞, 助格詞, 名詞, 以外の品詞の使用割合
9		9	" (形容動詞)	9	動詞, 助動詞, 助詞, 助格詞, 名詞, 以外の品詞の使用割合
0		1670	" (助詞)	1700	動詞, 助動詞, 助詞, 助格詞, 名詞, 以外の品詞の使用割合 (単位)
1		1	" (接頭語)	1	文の長さの系列相関係数(通れ)
2		2	" (接尾語)	2	" (" 2)
3		3	" (形容詞)	3	" (" 3)
4		4	" (感動詞)	4	" (" 4)
5		5	" (助動詞)	5	" (" 5)
6		6	" (連体詞)	6	" (" 6)
7		7	" (副詞)	7	" (" 7)
8		8	" (接統詞)	8	" (" 8)
9		9	ユールのDIVERSITY(全品詞)	9	" (" 9)
0		1680	シニエルの "	1710	" (" 10)
1		1	"	1	(文章単位) 文の長さの系列相関係数(通れ)
2		2		2	" (2)
3		3		3	" (3)
4		4		4	" (4)
5		5		5	" (5)
6		6		6	" (6)
7		7		7	" (7)
8		8		8	" (8)
9		9		9	" (9)
0		1690		1720	" (10)

統計学入門

このように単独の変数での判別は不可能であり、そこで判別に有効な複数個の変数の情報を総合して判断する必要がある。判別に有効な変数の選択は次の2段階で行った。

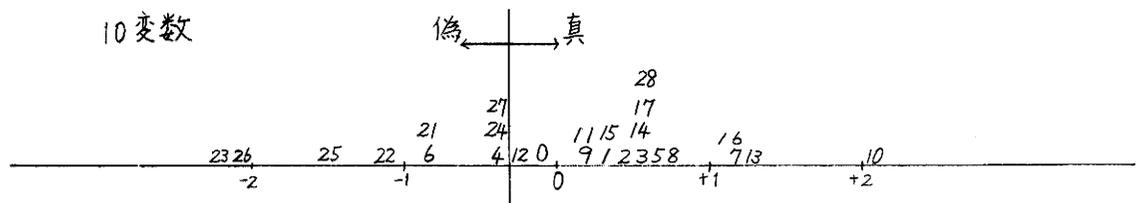
オ1段階：前記の約1000の変数の中から、真作グループに関し、変動係数の小さい変数を N 個選択。（以後この N 個の変数を候補変数と呼ぶ）

オ2段階： N 個の候補変数の中から、Wilksのラムダ統計量を用いて、判別効率の高い変数を選択。（“SPSS統計パッケージ”Ⅱ（東洋経済）P30を参照）

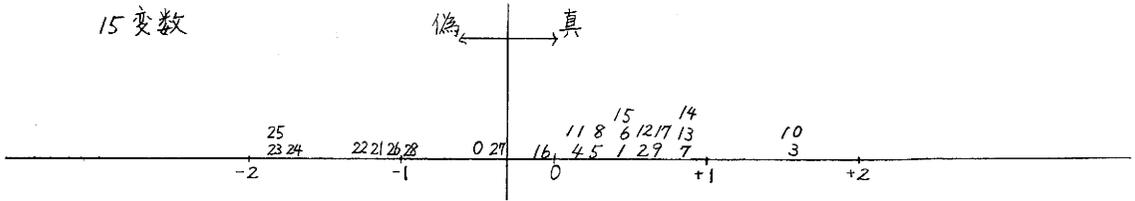
分析は $N_1=10, N_2=15, N_3=20, N_4=25, N_5=30, N_6=35, N_7=40, N_8=45$ で計8回行った。

候補変数の数	候補変数コードNo	判別に用いられた変数の数とコードNo	“三秘”イタナへの判別結果 正解率		相関比
$N_1=10$	1700, 891, 935, 516, 952 913, 15, 855, 692, 687	3変数 891, 855, 687	真	88%	4, 6 → 偽作 } 28 → 真作 } 誤判別
$N_2=15$	N_1 の10個 + 1697, 918, 917, 865, 1670	9変数 891, 935, 516, 913, 15, 692, 687, 1697, 1670	偽	100%	
$N_3=20$	N_2 の15個 + 906, 901, 845, 406, 1699	10変数 891, 935, 516, 952 1697, 865, 1670, 906, 901, 406	偽	100%	
$N_4=25$	N_3 の20個 + 930, 22, 880, 1679, 1661	10変数 891, 935, 516, 692 1697, 865, 1670, 906, 901, 880	真	100%	
$N_5=30$	N_4 の25個 + 5, 150, 772, 582, 180	7変数 516, 901, 22, 880, 150, 582, 180	真	96%	12 → 偽作 と誤判別
$N_6=35$	N_5 の30個 + 70, 834, 326, 175, 857	2変数 455, 175	真	88%	4, 6 → 偽作 } 28 → 真作 } 誤判別
$N_7=40$	N_6 の35個 + 57, 53, 431, 13, 868	2変数 855, 175	真	88%	4, 6 → 偽作 } 28 → 真作 } 誤判別
$N_8=45$	N_7 の40個 + 710, 863, 920, 877, 19	20変数 891, 935, 15, 855 687, 1697, 918, 865, 906 880, 1679, 1661, 582, 834 326, 857, 57, 53, 877, 19	真	100%	

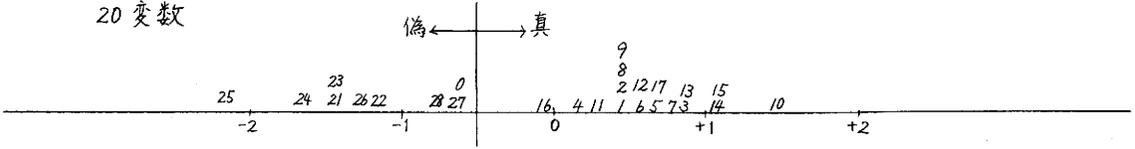
各 N_i ($i=1\sim 8$)における判別関数値の分布は次のヒウリである。



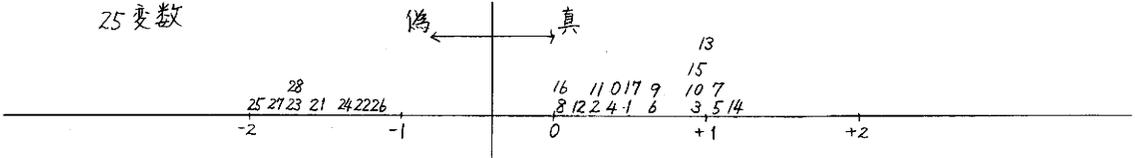
15 变数



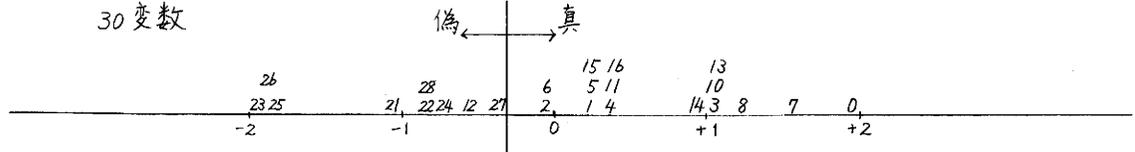
20 变数



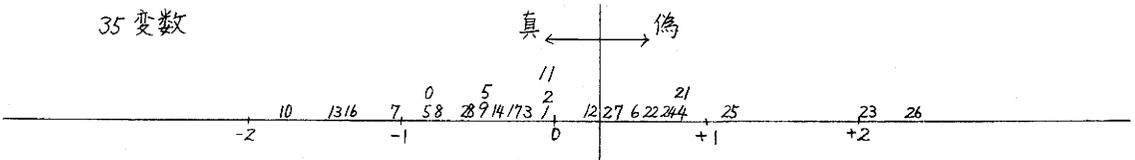
25 变数



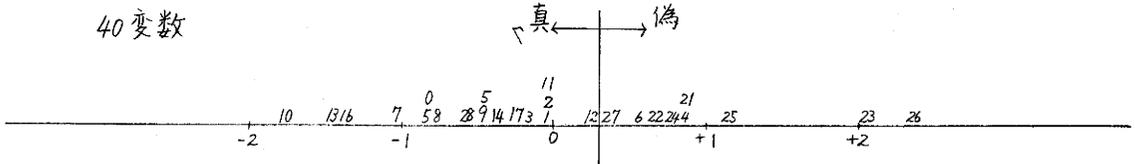
30 变数



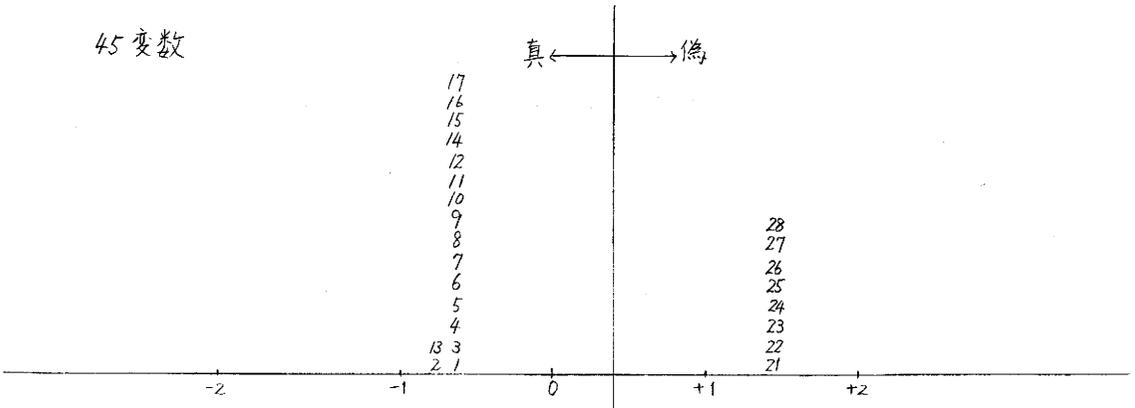
35 变数



40 变数



45 变数



2) 多群判別

2群判別では偽作8編をまとめて1グループとしたが、偽作8編の著者はすべて異なっている可能性が高く、したがってここでは偽作は1編ずつを1グループとみなし、“三大秘法抄”が真作グループをも含めた9つのグループのいずれに属するかを推測する。変数選択は次のように2段階で行った。

オ1段階：真作17編での標準偏差 S_1 と偽作8編の標準偏差 S_2 の比 S_2/S_1 の大きい変数(候補変数) N 個を前記の約1000の変数の中から選択。

オ2段階：2群の判別での変数選択のオ2段階と同じ。

判別分析は、 $N_1=10, N_2=15, N_3=20, N_4=25, N_5=30, N_6=35, N_7=40, N_8=45$ で計8回行ったが、 $N_7=40$ の場合のみ計算不能(原因不明)であった。

文の構造に関する情報 判別・多群

候補変数の数	候補変数コードNo	判別に用いられた変数の数とコードNo	“三秘”判別結果	似たりけの正解率	
$N_1=10$	61, 443, 855, 22, 985 1670, 19, 406, 918, 1699	8変数 61, 443, 855, 22 985, 19, 406, 918	偽作 (No.25) と判別	64.7%	真作の中で 2, 7, 12 13, 14, 16 が偽作 27 と誤判別
$N_2=15$	N_1 の10変数+ 917, 930, 906, 1697, 70	12変数 61, 443, 855, 22, 70 985, 19, 406, 1699, 917, 906, 1697	偽作 (No.27) と判別	76.5%	真作の中で 7, 8, 12, 16 が偽作 27と誤判別
$N_3=20$	N_2 の15変数+ 891, 1700, 25, 857, 935	14変数 61, 443, 855, 22, 19 985, 406, 918, 917, 930, 857 1697, 25, 935	偽作 (No.28) と判別	100%	
$N_4=25$	N_3 の20変数+ 952, 913, 15, 1661, 326	16変数 61, 443, 22, 985, 19 1670, 406, 918, 1699, 1697 930, 891, 1700, 857, 935, 952	真	100%	
$N_5=30$	N_4 の25変数+ 692, 834, 858, 150, 877	15変数 443, 855, 985, 19 406, 918, 1697, 25, 913, 15 326, 692, 858, 150, 877	偽作 (No.28) と判別	100%	
$N_6=35$	N_5 の30変数+ 516, 23, 175, 180, 920	15変数 61, 443, 855, 1670 19, 1699, 70, 25, 913, 150 877, 516, 175, 180, 920	真	100%	
$N_8=45$	N_5 の30変数+ 1679, 1680, 845, 901, 687 582, 772, 432, 1698, 51	16変数 61, 443, 855, 985, 19 1670, 930, 906, 891, 25, 952 150, 23, 687, 432, 51	真	100%	

B. 主成分分析の結果を利用した判別分析

判別分析ではあらかじめ選択しておいた N 個(最大45)の候補変数の中から判別に有効な幾つの変数をワルファスのラムダ統計量を基準に選択して判別を行った。しかしながらこの場合には最も多くても24の変数の情報しか利用できない。(25(三大秘法抄を除く文献数)-1=24)。そこで、主成分分析法によって多くの変数の情報を少数の基本的因子に集約し、その因子情報にもとずいて、判別分析を行った。

1) 2群判別

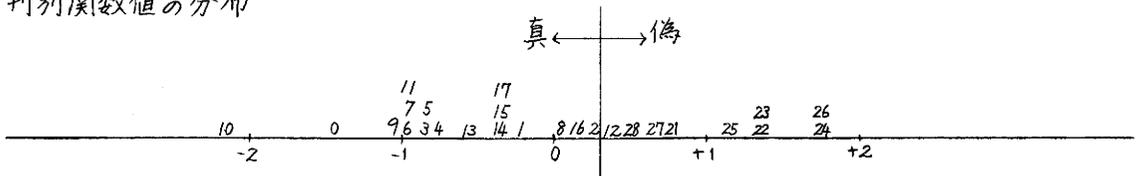
まず真作17編に関し変動係数の小さな以下に記す69個の変数を用いて主成分分析を行いオ1~オ16因子得点を求め次にこの情報を利用して、偽作8編と三大秘法抄に関してオ1~オ16因子得点を計算した。こうして求めた26の文献に関するオ1~オ16因子を変数と考え、判別を行った。変数選択にはWilksのラムダ統計量を用いた。判別結果は次のとおりである。

判別に用いられた変数(因子)はオ7, オ13, オ15, オ16因子であり、インターネットチェックでの正解率は96% (真作のNo.12のみが偽作と誤判別された。), で三大秘法抄は真作のグループに分類された。

主成分分析に用いた変数(コードNo.)

13, 15, 19, 21, 22, 23, 24, 25, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 70, 150, 175, 180, 326, 327, 332, 406, 431, 432, 443, 516, 582, 662, 687, 692, 710, 772, 834, 845, 855, 857, 858, 863, 865, 868, 877, 880, 891, 893, 901, 906, 908, 913, 917, 918, 920, 930, 935, 952, 984, 985, 1661, 1670, 1679, 1680, 1697, 1698, 1699, 1700

判別関数値の分布



2) 多群判別

真作17編での標準偏差 S_1 と偽作8編の標準偏差 S_2 の比 S_2/S_1 の値の大きい以下に記す39個の変数を用いて、真作17編のデータで主成分分析を行いオ1~オ16因子の因子得点を求め、次にこの情報を利用して、偽作8編と三大秘法抄に関してオ1~オ16因子の因子得点を計算した。こうして求めた26の文献に関するオ1~オ16因子を新たに変数と考え、判別を行った。

変数選択はWilksのラムダ統計量を用いて行った。結果は次のとおりである。主成分分析に用いた変数(コードNo.)

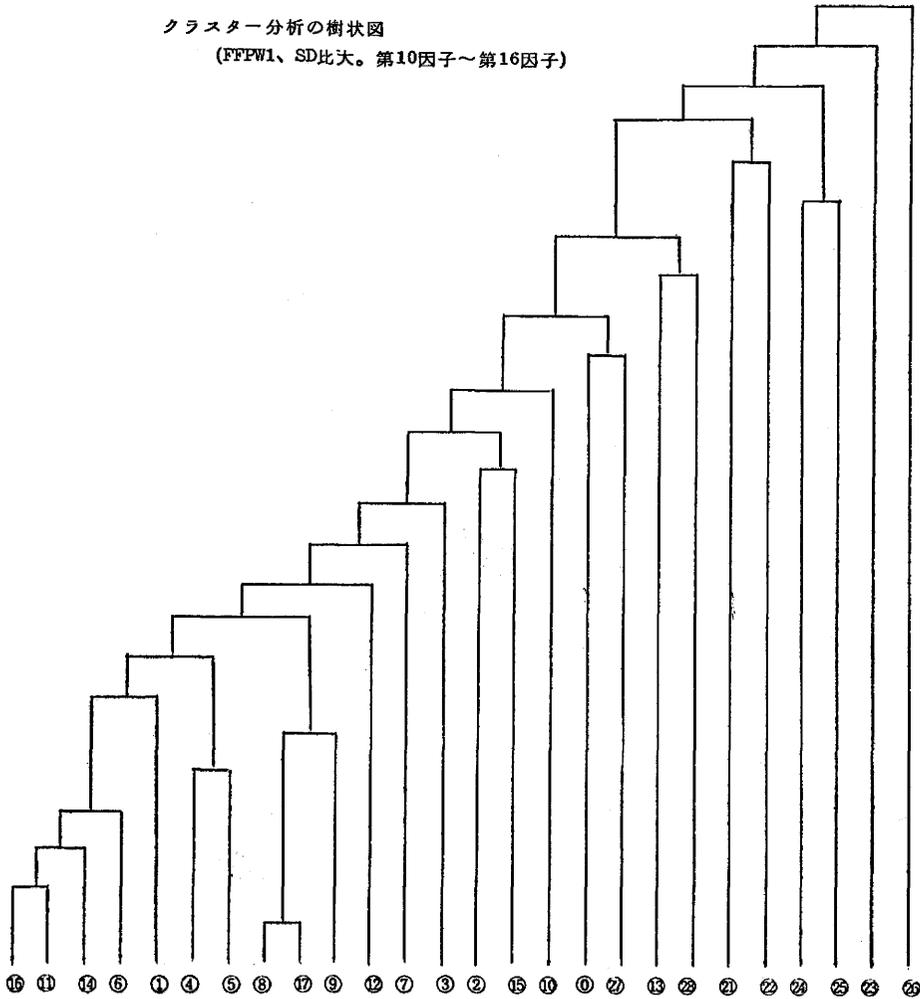
4, 5, 7, 8, 10, 21, 22, 23, 24, 25, 26, 29, 32, 33, 36, 37, 39, 42, 43, 44, 48, 50, 52, 53, 55, 56, 57, 58, 59, 60, 61, 63, 96, 97, 98, 99, 106, 108, 109

判別に用いられた変数(因子)はオ11, オ12, オ13, オ15因子で三大秘法抄は真作のグループに分類された。

なお真作の中No.12とNo.16の2編が偽作27のグループへNo.2, No.13, No.15の3編が偽作28のグループへ誤分類された。

- C. 主成分分析の結果を利用したクラスター分析
 前述の因子分析で得た各文献のオ1~オ16因子の中オ10~オ16までの7因子を用いてクラスター分析を行った。

クラスター分析の樹状図
 (FFPW1、SD比次。第10因子~第16因子)



4.2.4.2 単語に関する情報

26の文献には計90322語の単語が出てくる(引用文を除いた本文だけでは79397語)。各文献の総単語数は表1の通りである。

(表1)

文献No	総語数	本文中の語数	引用文中の語数
1	2601	2169 (83.4%)	432 (16.6%)
2	8078	7346 (90.9%)	732 (9.1%)
3	8508	7500 (88.2%)	1008 (11.8%)
4	3687	3054 (82.8%)	633 (17.2%)
5	2421	2261 (93.4%)	160 (6.6%)
6	2052	1917 (93.4%)	135 (6.6%)
7	2285	2061 (90.2%)	224 (9.8%)
8	1382	1283 (92.8%)	99 (7.2%)
9	2104	1993 (94.7%)	111 (5.3%)
10	3450	3188 (92.4%)	262 (7.6%)
11	1164	1021 (87.7%)	143 (12.3%)
12	7362	6826 (92.7%)	536 (7.3%)
13	4272	3960 (92.7%)	312 (7.3%)
14	787	596 (75.7%)	191 (24.3%)
15	6396	4969 (77.7%)	1427 (22.3%)
16	1023	996 (97.4%)	27 (2.6%)
17	1892	1672 (88.4%)	220 (11.6%)
0	1319	1115 (84.5%)	204 (15.5%)
21	2670	2258 (84.6%)	412 (15.4%)
22	7104	5664 (79.7%)	1440 (20.3%)
23	774	576 (74.4%)	198 (25.6%)
24	8551	8493 (99.3%)	58 (0.7%)
25	3876	2527 (65.2%)	1349 (34.8%)
26	615	526 (85.5%)	89 (14.5%)
27	2671	2524 (94.5%)	147 (5.5%)
28	3278	2902 (88.5%)	376 (11.5%)

また79397語は6607種の異なる単語*から成っているがこれらと品詞別に見たのが表2である。

品詞	引用文も含めた 場合の異なり 語数	本文中の 異なり語数	品詞	引用文も含めた 場合の異なり 語数	本文中の 異なり語数
普通名詞	2821	2552	接頭語	14	14
固有名詞	1037	970	接尾語	7	7
形式名詞	15	14	形容詞	210	199
代名詞	40	38	感動詞	3	3
数詞	645	614	助動詞	90	89
動詞	1870	1690	連体詞	18	18
形容動詞	140	132	副詞	171	158
助詞	54	52	接続詞	59	57

(表 2)

(注) 異なり語数の意味について

文法上の問題はあろうが計算機での処理上、活用のある単語は活用語尾が異なるとすべて異なった単語として認識され処理されている。例えば、「好み」(動)と「菓(このみ)」(名)は異なった言葉として処理される。また同一品詞で同音異語の言葉(例えば普通名詞(01)の“木”“気”)は同じ言葉として処理されている。

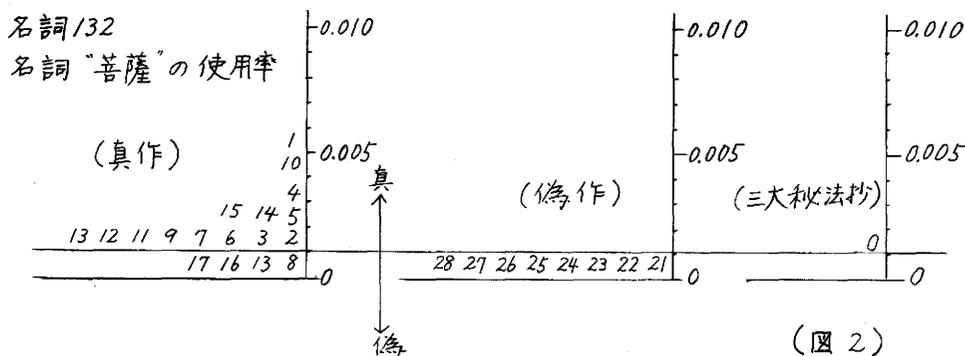
単語に関する情報の解析には次の4つの方法を用いた

- A. 判別分析(2群判別, 多群判別)
- B. 主成分分析の結果を利用した判別分析(2群判別, 多群判別)
- C. 主成分分析の結果を利用したフラスター分析
- D. 延べ語数と異なり語数の比の分析

A. 判別分析(discriminant analysis)による解析

1) 2群判別

6607種の単語の中にその使用率によって真作グループと偽作グループを完全に分離できるものは存在しない。両グループの分離に最も有効なものは普通名詞“菩薩”という語の使用率(図2)であるが、仮にこの変数の使用率0.001を分離点としても、真作の4編が偽作と誤判断されることになる。



このように単独の変数での判別は不可能であり、したがって、判別に有効な複数個の変数を次の2段階

第1段階：6607種の単語の中から真作グループに関し、使用率の変動係数の小さい変数を N 個選択（以後この N 個の変数を候補変数と呼ぶ）、

第2段階： N 個の候補変数の中から Wilks のラムダ統計量を用いて判別効率の高い変数を選択、

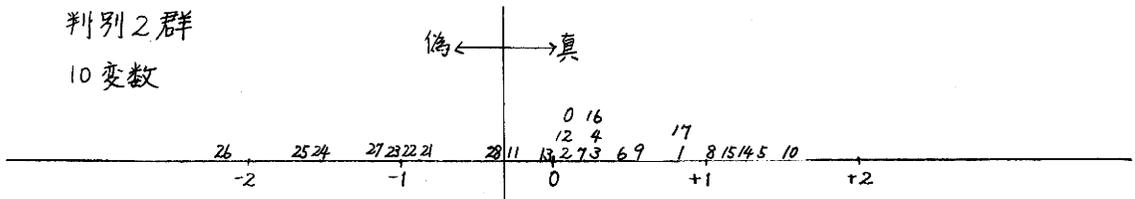
を行い、こうして選ばれた複数個の変数の情報を総合して判断する。解析は $N_1=10$, $N_2=15$, $N_3=20$, $N_4=25$, $N_5=30$, $N_6=35$, $N_7=40$ の計7回行った。

候補変数の数	候補変数(内は品詞コード)	判別に用いられた変数の数と変数	"三秘"判別結果	内部正解率	相関比
$N_1=10$	(10) TŌU (1) BOSATU (16) NARABINI (8) NITE (7) IKANI (1) MŌTŌ (1) NŌTI (5) ITI (2) HOKEKIYAU (13) SE	6変数 (10) TŌU (1) BOSATU (8) NITE (7) IKANI (1) NŌTI (5) ITI	真	100%	
$N_2=15$	$N_1 +$ (8) DŌMŌ (13) TARU (4) KŌRERA (13) GŌTŌKU (1) BUTUPŌHU	6変数 (1) BOSATU (2) HOKEKIYAU (7) IKANI (8) DŌMŌ (5) ITI (13) GŌTŌKU	真	96%	28 → 真作と誤判別
$N_3=20$	$N_2 +$ (8) NŌ (15) IMADA (3) TŌKŌRŌ (13) GŌTŌSI (9) DAI	6変数 (1) BOSATU (2) HOKEKIYAU (7) IKANI (8) DŌMŌ (5) ITI (13) GŌTŌKU	真	96%	28 → 真作と誤判別
$N_4=25$	$N_3 +$ (6) TAMAH I (8) WŌBA (8) MŌ (8) KA (13) TARI	9変数 (5) ITI (8) DŌMŌ (10) TŌU (13) GŌTŌKU (1) BOSATU (6) TAMAH I (8) NITE (8) MŌ (7) IKANI (8) MŌ	真	100%	
$N_5=30$	$N_4 +$ (6) ŌMŌHU (8) BA (1) HŌHU (13) RE (1) KŌN*Z IYAU	22変数 (9) DAI (15) IMADA (10) TŌU (13) GŌTŌSI (1) BOSATU (6) TAMAH I (7) IKANI (8) MŌ (1) MŌTŌ (13) TARI (5) ITI (8) WŌBA (8) DŌMŌ (8) WŌBA (4) KŌRERA (6) ŌMŌHU (13) TARU (8) BA (13) GŌTŌKU (1) HŌHU (8) NŌ (13) RE (3) TŌKŌRŌ	真	100%	
$N_6=35$	$N_5 +$ (4) KANŌ (16) YUWEN I (1) NAKA (6) NARA (1) HITŌ	21変数 (8) NŌ (10) TŌU (3) TŌKŌRŌ (13) GŌTŌSI (1) BOSATU (8) MŌ (16) NARABINI (8) WŌBA (7) IKANI (6) ŌMŌHU (1) MŌTŌ (1) HŌHU (1) NŌTI (13) RE (5) ITI (4) KANŌ (2) HOKEKIYAU (1) NAKA (4) KŌRERA (1) HITŌ (13) GŌTŌKU	偽	100%	

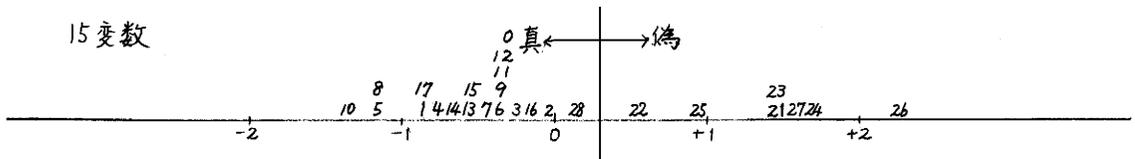
候補変数の数	候補変数()内は品詞コード	判別に用いられた変数の数と変数	"三秘"判別結果	イタル正解率	相関比
$N_7=40$	$N_6 +$ (1) M \bar{O} N \bar{O} (2) K \bar{O} T \bar{O} (1) Y \bar{O} (2) K \bar{O} S \bar{O} (1) I HAKU	22変数 (10) T \bar{O} U (1) B \bar{O} SATU (5) I TI (13) SE (4) K \bar{O} RERA (13) G \bar{O} T \bar{O} KU (3) T \bar{O} K \bar{O} R \bar{O} (9) DA I (15) I MADA (6) TAMAH I (8) W \bar{O} BA (8) KA (6) \bar{O} M \bar{O} HU (8) BA (1) H \bar{O} HU (12) RE (4) KAN \bar{O} (1) NAKA (6) NARA (1) HI T \bar{O} (2) K \bar{O} S \bar{O} (1) I HAKU	真	100%	

判別2群

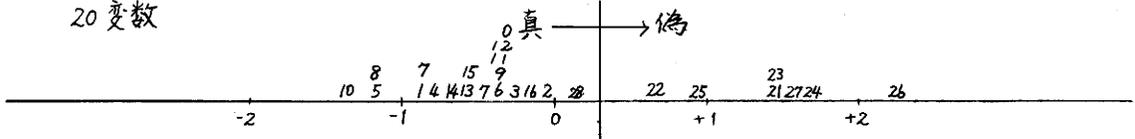
10変数



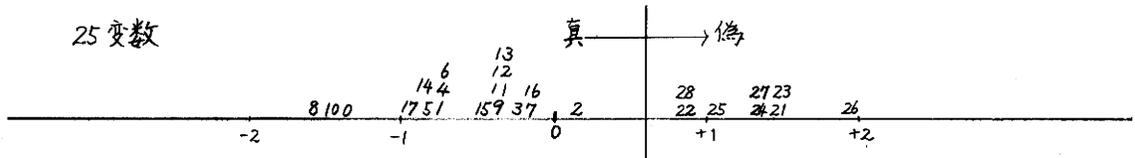
15変数



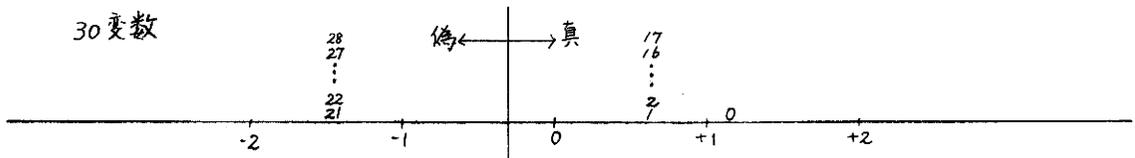
20変数



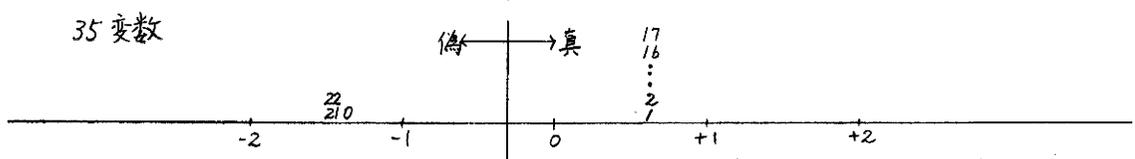
25変数



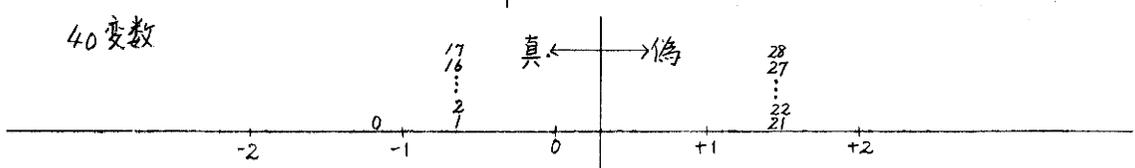
30変数



35変数



40変数

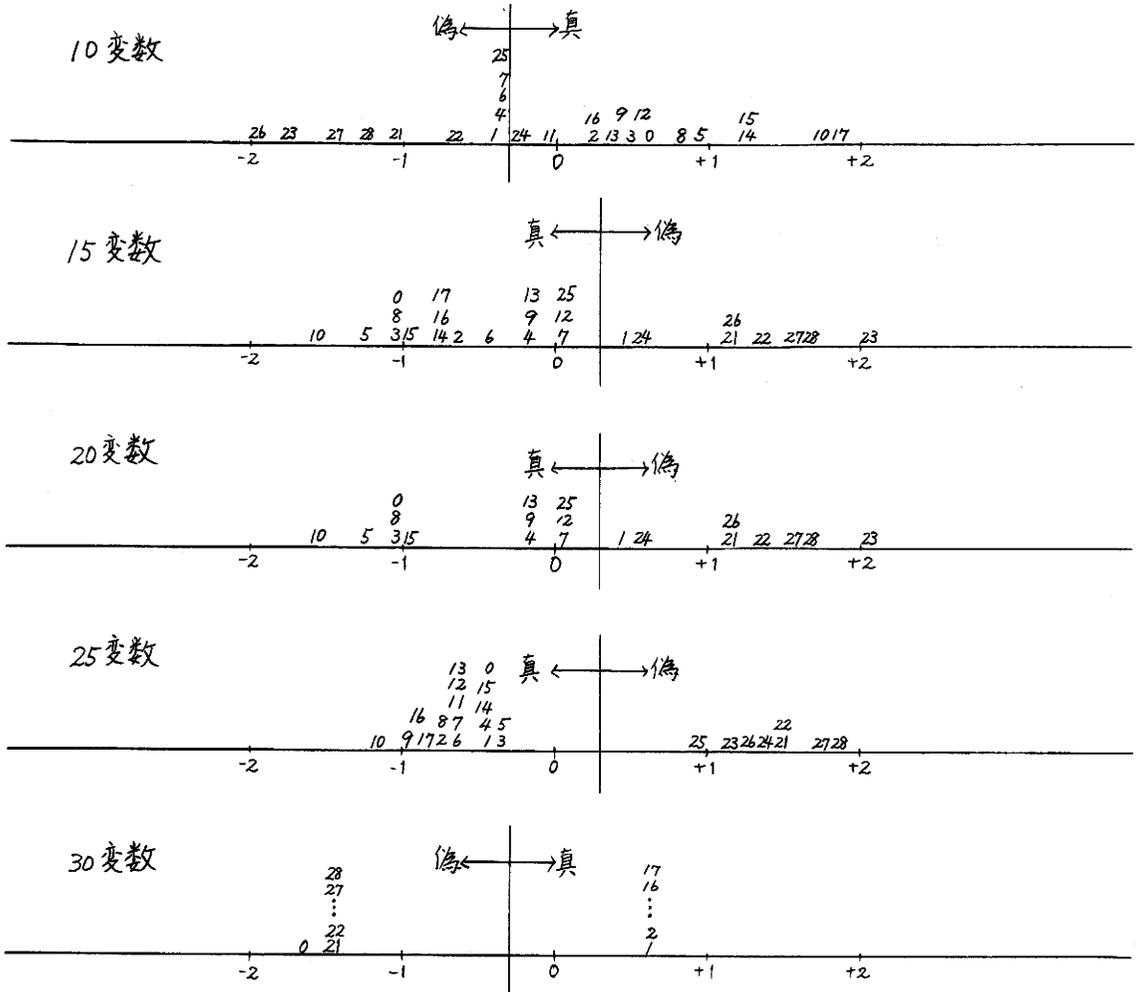


次に名詞、及び活用のある品詞を除いた単語だけを用いて今と同じ方法で2群判別を行ってみた。解析は $N_1 = 10$, $N_2 = 15$, $N_3 = 20$, $N_4 = 25$, $N_5 = 30$ の計5回行った。

単語に関する情報 (FF) 判別2群

候補変数の数	候補変数()内は品詞コード	判別に用いられた変数の数とコード N_0	"三秘"判別結果	インターナルチェック正解率	相関比	
$N_1 = 10$	(10) TÔU (16) VARABINI (8) NITE (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA	4変数 (10) TÔU (8) NITE (8) NÔ (8) MÔ	真	80%		真作の内、1, 4, 6, 7が偽作と、偽作の内、2, 4が真作と誤判別
$N_2 = 15$	(10) TÔU (16) VARABINI (8) NITE (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA (8) BA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (14) SÔNÔ	4変数 (10) TÔU (8) BA (8) KÔSÔ (14) KAKUNÔ	真	92%		真作の内、1が偽作と、偽作の内、25が真作と誤判別
$N_3 = 20$	(10) TÔU (16) VARABINI (8) NITE (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA (8) BA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (14) SÔNÔ (11) NASI (15) MATA (8) MADE (16) ARUHIHA (9) ÔN*	4変数 (10) TÔU (8) BA (8) KÔSÔ (14) KAKUNÔ	真	92%		真作の内、1が偽作と、偽作の内、25が真作と誤判別
$N_4 = 25$	(10) TÔU (16) VARABINI (8) NITE (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA (8) BA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (14) SÔNÔ (11) NASI (15) MATA (8) MADE (16) ARUHIHA (9) ÔN* (15) TADA (8) NI (16) MÔTUTE (9) SIYÔ (11) NAKI	12変数 (10) TÔU (16) VARABINI (8) WÔBA (8) KA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (8) MADE (15) TADA (8) NI (16) MÔTUTE (9) SIYÔ	真	100%		
$N_5 = 30$	(10) TÔU (16) VARABINI (8) NITE (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA (8) BA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (14) SÔNÔ (11) NASI (15) MATA (8) MADE (16) ARUHIHA (9) ÔN* (15) TADA (8) NI (16) MÔTUTE (9) SIYÔ (11) NAKI (8) HA (8) TÔMÔ (8) TE (8) GA	23変数 (10) TÔU (8) DÔMÔ (8) NÔ (9) DAI (15) IMADA (8) MÔ (8) WÔBA (8) KA (8) BA (16) YUWENI (8) KÔSÔ (14) KAKUNÔ (11) NASI (16) ARUHIHA (9) ÔN* (15) TADA (8) NI (16) MÔTUTE (9) SIYÔ (11) NAKI (8) TÔMÔ (8) TE (8) GA	偽	100%		

単語に関する情報 (FF) P_1 判別2群



2) 多群判別

三木秘法抄が真作グループと偽作8グループ（偽作ひとつひとつを、1グループとみなす）のいずれに帰属するかを推測する。変数選択は文の構造に関する情報の場合と同じく2段階である。判別分析はまず全品詞の単語の情報を利用して $N=10, 15, 20, 25, 30$ の計5回行った。

単語に関する情報 (FF) 判別・多群

候補変数の数	候補変数()内は品詞コード	判別に用いられた変数の数とコードNo	"三秘"判別結果	内部的な正解率	
$N_1 = 10$	(1) BUTU (13) BEKARA (1) HŌHU (13) TARU (2) TAME (16) YUWEMI (5) ITI (1) SIYUZIYAU (8) NITE (1) MI	10変数 (13) MI (1) BUTU (13) TARU (1) HŌHU (16) YUWEMI (2) TAME (1) SIYUZIYAU (5) ITI (1) MI (8) NITE (1) MI	偽作 No. 28の グループA 分類	94.1%	真作の中 16が偽作24の グループと誤判別
$N_2 = 15$	(1) BUTU (1) SIYUZIYAU (1) HŌHU (1) MI (1) TAME (8) YA (5) ITI (1) ŌMŌHI (8) NITE (1) KIYAUMŌN* (13) MI (13) BEKARA (13) TARU (8) NO (16) YUWEMI	12変数 (16) YUWEMI (1) BUTU (1) SIYUZIYAU (1) HŌHU (1) MI (1) TAME (8) YA (5) ITI (1) KIYAUMŌN* (8) NITE (1) BEKARA (13) TARU (13) BEKARA	偽作 No. 28の グループA 分類	100%	
$N_3 = 20$	(1) BUTU (8) YA (1) HŌHU (4) ŌMŌHI (2) TAME (1) KIYAUMŌN* (5) ITI (13) BEKARA (8) NITE (8) NO (13) MI (3) TŌKŌRŌ (13) TARU (1) IHAKU (16) YUWEMI (15) TADA (1) SIYUZIYAU (6) ARU (1) MI (13) BEKI	16変数 (1) SIYUZIYAU (1) BUTU (8) YA (1) HŌHU (1) KIYAUMŌN* (1) TAME (8) NO (5) ITI (3) TŌKŌRŌ (8) NITE (1) IHAKU (13) MI (15) TADA (13) TARU (13) BEKI (16) YUWEMI	真	100%	
$N_4 = 25$	(1) BUTU (13) BEKARA (1) HŌHU (8) NO (1) TAME (3) TŌKŌRŌ (5) ITI (1) IHAKU (8) NITE (15) TADA (13) MI (6) ARU (13) TARU (13) BEKI (16) YUWEMI (13) ŌTŌKU (1) SIYUZIYAU (14) SŌNO (1) MI (1) YUWE (8) YA (1) NAKA (6) ŌMŌHI (13) NARI (1) KIYAUMŌN*	16変数 (8) YA (1) BUTU (6) ŌMŌHI (1) HŌHU (6) ŌMŌHI (1) TAME (8) NO (5) ITI (3) TŌKŌRŌ (8) NITE (1) IHAKU (13) TARU (6) ARU (16) YUWEMI (13) ŌTŌKU (1) SIYUZIYAU (13) NARI	真	100%	
$N_5 = 30$	(1) BUTU (8) NO (1) HŌHU (3) TŌKŌRŌ (1) TAME (1) IHAKU (5) ITI (15) TADA (8) NITE (6) ARU (13) MI (13) BEKI (13) TARU (13) ŌTŌKU (16) YUWEMI (14) SŌNO (1) SIYUZIYAU (1) YUWE (1) MI (1) NAKA (8) YA (1) NAKA (6) ŌMŌHI (13) NARI (1) KIYAUMŌN* (6) TAMAHU (13) BEKARA (8) TE (15) MATA (8) MŌ (8) MI	16変数 (6) ŌMŌHI (1) BUTU (3) TŌKŌRŌ (1) HŌHU (1) IHAKU (1) TAME (13) BEKI (5) ITI (6) TAMAHU (8) NITE (15) MATA (13) TARU (8) MI (1) SIYUZIYAU (8) MŌ (8) YA	真	100%	

次に2群判別の場合と同じく、名詞及び活用のある品詞を除いた単語だけを用いて、多群判別を行った。判別は $N_1=10$, $N_2=15$, $N_3=20$ の3回行った。

単語に関する情報 (FF) 判別多群

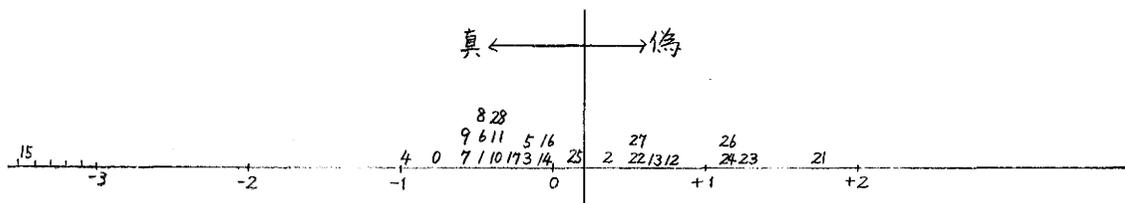
候補変数の数	候補変数()内は品詞コード	判別に用いられた変数の数とコード" N_0	"三秘"判別結果	インターナルチェック正解率	相関比	
$N_1=10$	(8) NITE (16) YUWENI (8) YA (8) NŌ (15) TADA (8) WŌBA (14) SŌNŌ (8) TE (8) NI (15) MATA	6変数 (15) TADA (8) NITE (16) YUWENI (8) YA (14) SŌNŌ (15) MATA	偽作 No. 27の グループ 分類	64.7%		真作の内 2. 4. 6. 16 が 偽作 27 真作の内 7. 11 が 偽作 22 のグループと誤判別
$N_2=15$	(8) NITE (16) YUWENI (8) YA (8) NŌ (15) TADA (8) WŌBA (14) SŌNŌ (8) TE (8) NI (15) MATA (8) MŌ (8) MŌ (8) BA (15) IMADA (16) MŌTUTE (14) KAKUNŌ	9変数 (8) NI (8) NITE (16) YUWENI (8) YA (15) TADA (14) SŌNŌ (8) NI (8) BA (16) MŌTUTE (14) KAKUNŌ	偽作 No. 27の グループ 分類	47.1%		真作の内 2 が 偽作 27 真作の内 3. 7 が 偽作 22 真作の内 5. 8. 9. 11. 13. 17 が 偽作 23 のグループと誤判別
$N_3=20$	(8) NITE (16) YUWENI (8) YA (8) NŌ (15) TADA (8) WŌBA (14) SŌNŌ (8) TE (8) NI (15) MATA (8) MŌ (8) BA (15) IMADA (16) MŌTUTE (14) KAKUNŌ (8) KŌSŌ (8) TŌ (8) MADE (9) SIYŌ (8) GA	12変数 (15) IMADA (16) YUWENI (8) YA (8) NŌ (14) SŌNŌ (8) BA (8) NI (8) TŌ (8) MADE (9) SIYŌ (8) GA	偽作 No. 23の グループ 分類	76.5%		真作の内 3. 7. 9. 15 が 偽作 27 のグループと誤判別

B. 主成分分析の結果を利用した判別分析

分析方法は文の構造に関する情報の場合の主成分分析の結果を利用した判別分析と同じである。

1) 2群判別

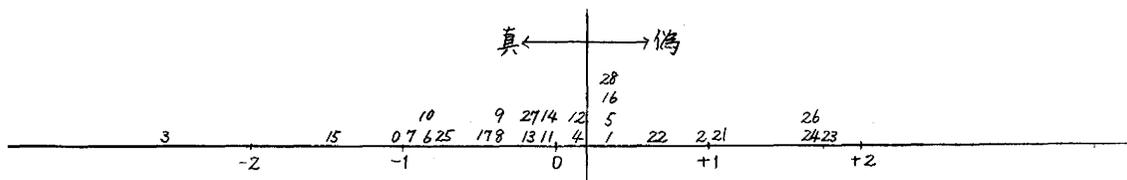
- 主成分分析に用いた変数は次ページのリストにある101変数である。判別に用いた変数(因子)はオ14, オ16因子でインターナルチェックの正解率は80% (真作の中で No. 2, No. 12, No. 13 が偽作と誤判別され、また偽作の中で No. 25, No. 28 が真作と誤判別された。)で三秘法抄は真作のグループに分類された。



1	1の132	BOSATU	36	6の813	NARA	71	13の10	BEKI
2	142	BUTU	37	826	NARU	72	12	BESI
3	158	BUTUPOHU	38	942	OMOHI	73	14	GOTOKU
4	496	HITO	39	945	OMOHU	74	16	GOTOSI
5	518	HOHU	40	1115	SIKARU	75	29	N*
6	647	IHAKU	41	1136	SIRA	76	35	NARA
7	952	KIYAU	42	1307	TAMAHI	77	37	NARI
8	958	KIYAUMON*	43	1308	TAMAHU	78	38	NARU
9	986	KOKORO	44	7の44	IKANI	79	40	NI
10	1019	KON*ZIYAU	45	8の4	BA	80	42	NU
11	1195	MI	46	11	DOMO	81	48	RARE
12	1206	MINA	47	12	GA	82	51	RE
13	1247	MONO	48	13	HA	83	54	RI
14	1251	MOTO	49	18	KA	84	55	RU
15	1310	NAKA	50	23	KOSO	85	63	SE
16	1396	NOTI	51	24	MADE	86	66	SI
17	1872	SIYUZIYAU	52	25	MO	87	78	TARI
18	1971	TAME	53	30	NI	88	79	TARU
19	2175	UHE	54	32	NITE	89	89	ZARU
20	2185	UN*NUN*	55	34	NO	90	91	ZU
21	2199	UTI	56	38	SITE	91	14の9	KAKUNO
22	2288	YO	57	40	TE	92	12	KONO
23	2330	YUWE	58	41	TO	93	18	SONO
24	2の248	HOKEKIYAU	59	43	TOMO	94	15の30	IMADA
25	3の10	KOTO	60	45	WO	95	66	MATA
26	14	TOKORO	61	46	WOBA	96	113	SUDENI
27	4の10	KANO	62	50	YA	97	121	TADA
28	19	KORE	63	52	YORI	98	16の6	ARUHIHA
29	20	KORERA	64	9の5	DAI	99	23	MOTUTE
30	5の144	ITI	65	11	ON*	100	26	NARABINI
31	6の60	ARI	66	15	SIYO	101	60	YUWENI
32	61	ARU	67	10の10	TOU			
33	380	ITARU	68	11の120	NAKI			
34	738	MI	69	124	NASI			
35	740	MIE	70	13の6	BEKARA			

(表3)

ロ次に前記(表3)の101変数の中から名詞及び活用のある品詞の単語を除いた単語だけを用いて、同様に主成分分析を行って2群判別を行った。判別に用いた変数(因子)はオ13, オ14因子で、インターナルチェックの正解率は76%(真作の中でNo.1, No.2, No.5, No.6が偽作と誤判別され、また偽作の中でNo.25, No.27が真作と誤判別された)で、三大秘法抄は真作のグループに分類された。



2) 多群判別

判別方法は文の構造に関する情報の場合の主成分分析の結果を利用した判別分析と同じである。

1) 主成分分析に用いた変数は前記の101変数である。判別に用いた変数(因子)はオ14, オ15, オ16因子である。

判別結果は

真作の中で	偽作の	真作の中で	偽作の
No. 2	→ No. 22	No. 10	→ No. 25
No. 3	→ No. 25	No. 12	→ No. 24
No. 4	→ No. 28	No. 13	→ No. 27
No. 5	→ No. 28	No. 15	→ No. 28

のグループに誤判別され三大秘法抄は偽作No.25のグループに分類された。

ロ) 前記の101変数の中から名詞及び活用のある品詞(単語)を除いた変数を用いて主成分分析を行った。判別に用いた変数(因子)はオ12, オ14, オ15, オ16因子である。

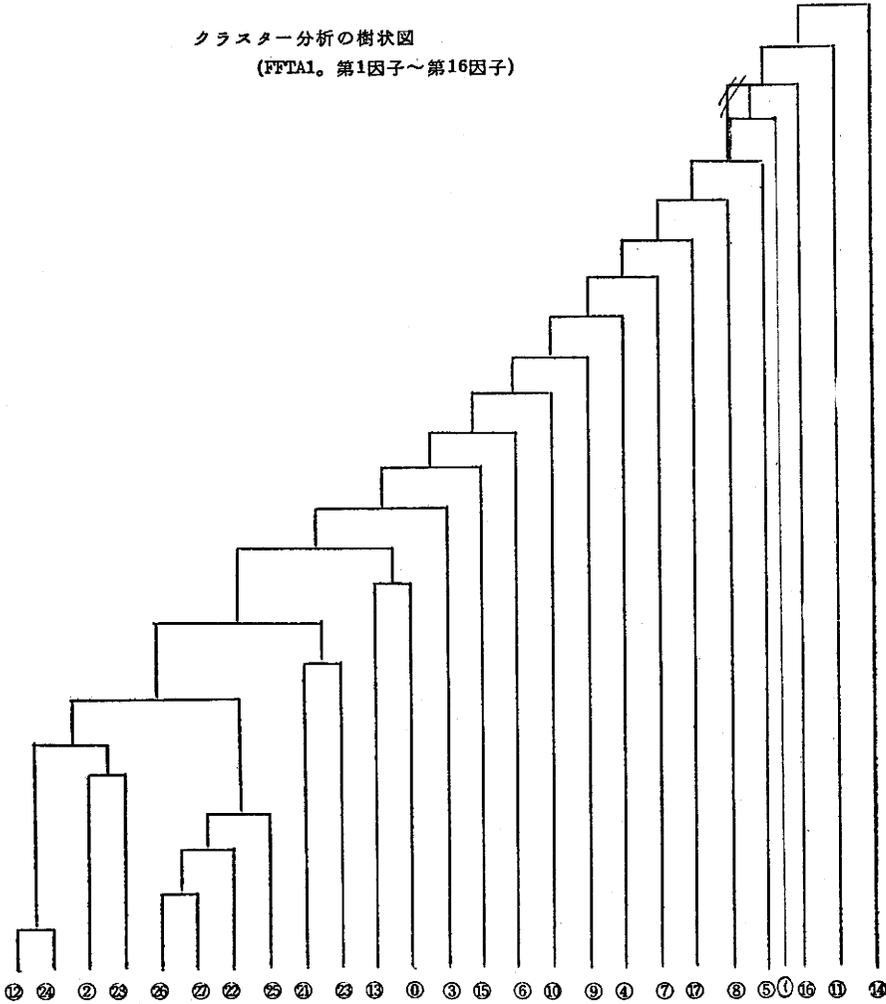
判別結果は

真作の中で	偽作の	真作の中で	偽作の
No. 1	No. 22	No. 9	No. 22
No. 2	No. 21	No. 11	No. 22
No. 4	No. 22	No. 12	No. 28
No. 5	No. 22	No. 15	No. 23
No. 6	No. 22	No. 16	No. 22
No. 7	No. 23		

のグループに誤判別され三大秘法抄は偽作No.23のグループに分類された。

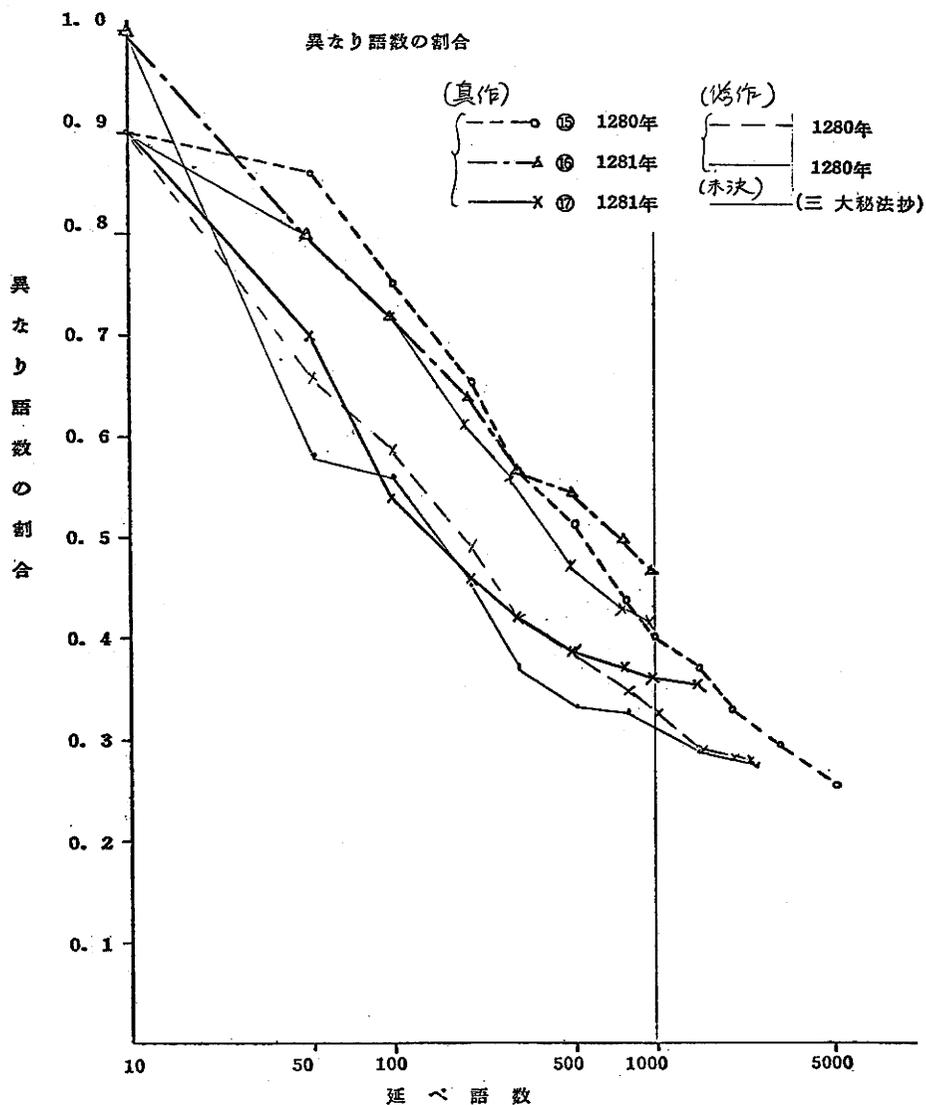
- C. 主成分分析の結果を利用したフラスター分析
 Bの1)で得た各文献のオ1~オ16主成分の情報を用いて、フラスター分析を行った。

クラスター分析の樹状図
 (FFTA1。第1因子~第16因子)



D. 異なり語数と延べ語数の割合

三大秘法抄での(異なり語数) / (延べ語数)の変化を三大秘法抄が書かれた頃の真作3編(No. 15, 16, 17)と偽作2編(No. 21, 27)と比較した(図, 横軸は対数目盛)。三大秘法抄は引用文を除くと1115語の単語から成っているが、延べ語数1000での異なり語数の割合を見ると真作3編はいずれも偽作よりも高い割合を示し、また三大秘法抄はNo. 16の真作とほぼ同じ値となっている。



4.2.4.3. 経年変化のある情報

文章における書き手の特徴はその人の (イ) 一生を通して不変のもの と (ロ) 年令とともに変化 (経年変化) するものが考えられる。(3.1節参照)
 ここではロについての解析を行うがその目的は経年変化が認められる統計量について1281年に書かれたとされる三大秘法抄か、日蓮の1281年頃の文章の特徴を有しているか否かを調べることにある。経年変化の有無の判断は著作年との相関係数 r に基づいている。(幾分任意ではあるが) r が0.605以上のものについては経年変化があると考える。($r = 0.605$ は帰無仮説 H_0 : 経年変化はない ($\rho = 0$) の 0.005% 棄却値である)

文の構造に関する情報に関して著作年との相関係数 r が0.605以上である変数は次の通りである。

コードNo.	統計量 or 変数名	r 値
21	1字から成る単語の使用割合	-0.716
24	4字から "	0.626
83	数詞→固有名詞への接続の割合	0.609
208	固有名詞→接尾語への接続の割合	0.608
431	普通名詞の後に助詞の来る割合	-0.661
518	助詞の後に助動詞が来る割合	0.612
779	助動詞の前に助動詞が来る割合	-0.618
847	名詞→接尾語への接続の割合	0.638
868	名詞の後に助詞が来る割合	-0.684
898	名詞の前に副詞が来る割合	0.649
908	助詞の使用割合	-0.694
910	接尾語の "	-0.609
920	助詞の使用数 / 名詞の使用数	-0.705
949	接尾語の使用数 / 助詞の使用数	-0.636
1671	ユールのK統計量(接頭語)	-0.605
1676	" (連体詞)	-0.656

また単語に関する情報に関して著作年との相関係数 r が0.605以上である変数は次の通りである。

品詞コード	単語名	r 値	品詞コード	単語名	r 値	品詞コード	単語名	r 値
1. 普通名詞	文 (もん)	-0.734	2. 固有名詞	叡山(えいざん)	0.608	6. 動詞	説か(たか)	-0.608
1. "	中 (なか)	-0.651	2. "	日本国	0.723	8. 助詞	が	0.627
1. "	二乗(にじょう)	-0.607	5. 数詞	四十餘年	-0.678	8. "	の	-0.667

4.2.4.4 論理の展開の仕方に関する情報

各文章を以下の29のいずれかに分類しその接続関係を調べることによって日蓮の文章において彼独自の宗・因・喩・合・結という論理の展開法のパターンがあるか否かを調べる。

- | | |
|--|--|
| <p>(宗) 1 主張 (問答一段の中心的主張
* 問いに應じない場合もある)</p> <p>2 主張の解説</p> <p>3 主張に付属する事柄解説の導入</p> <p>4 " の主張</p> <p>5 " の理由</p> <p>6 " の結論</p> <p>7 主張に関する選択的な標示</p> <p>8 標示中の選択されない主張</p> <p>9 " の選択される主張</p> | <p>(合) 18 「喩」についての適合</p> <p>(結) 19 結論</p> <p>(問) 20 前段の「宗」等に関する設問</p> <p>21 " に付属する事柄に関する設問</p> <p>22 " の領解・承認</p> <p>23 " に関する経証を求める設問</p> <p>24 " に対する反対主張</p> <p>25 反対主張の否定</p> <p>26 " の理由</p> <p>27 " の結論</p> <p>(その他)</p> <p>28 論理的構造上、上記いずれにも属さない導入</p> <p>29 主張・結論等に基づく推定・演繹</p> |
|--|--|
- (因) 10 (主張)の理由
- 11 理由の解説
- (喩) 12 譬喩 (直喩)
- 13 " (隱喩)
- 14 経・論文の引用
- 15 注釈等の著作文の引用
- 16 外典文の引用
- 17 歴史的事実等の引用

4.3 今後の研究計画について

本研究は昭和53年～55年の3年間にまたがって行なわれたが、この3年間の研究活動では当初の目的を完全に果したということとはできない。今後の研究計画としては主として次の二つを考えている。

1. 解析に用いる文献の追加

これまでに述べた解析結果はデータチェックが終了している26の文献に関してであったが、統計解析をより正確にすべく、以下に記す真作・偽作及び真偽未決の文献を追加する予定である。(この内の数編については現在作業中。)

追加文献名

真作		偽作		真偽未決	
如来滅後五五百歳観心本尊鈔	1273年	問答鈔	1254年	聖愚問答鈔	1265年
		一念三千法門	1258年	生死一大事血脉鈔	
頭佛未來記	1273年	早勝問答	1271年		1272年
法華取要鈔	1274年	成佛法華肝心口伝鈔		諸法実相鈔	1273年
曾谷入道殿許御書	1275年		1275年	日女御前御返事	1277年
太田入道殿御返事	1275年	本寺参詣鈔	1282年		
始聞佛乘義	1278年	読誦法華用心鈔			
妙一女御返事	1280年	万法一如鈔	?		
		放光授職灌頂下	1274年		

また、これまでの解析結果をみると三大秘法抄の文章には、日蓮の文章と同じような特徴が見られるが、しかしながらこれだけでは三大秘法抄が日蓮の手によるものと結論するのは早急すぎる。なぜなら、日蓮の高弟の中に日蓮と同じような特徴をもつ文章を書きえた人物が存在したかもしれないからである。日蓮の弟子は多いが、しかしながら三大秘法抄はその内容がきわめて高度なものであり、したがって、もしこれが弟子の手によるものであれば、それを書きえた人物の数は限られ一応次の三人の人物がその著者としての可能性が高いと考えられる。

日順 (年), 日算 (年), 日興 (年)

そこでこの三人の著作

用心抄 (年) 日順,	富士一跡門徒存知事 (年) 日興
本門心底抄 (年)	原殿御返事 (年)
五人所破抄 (年)	

と三大秘法抄の比較を行う。

ロ解析方法の検討及び開発

これまでの解析は、イ判別分析、ロクラスター分析、ハ主成分分析、ニ回帰分析を主な武器として用いてきた。今後は追加文献の情報をも加え、これらの方法でもう一度解析を行うと同時にこれらの解析に用いた変数の妥当性等について再検討する予定である。またこれ以外に次のような解析法の使用も予定している。

- 安本の方法 1.
- 安本の方法 2.

5. 計算プログラムと情報の入出力

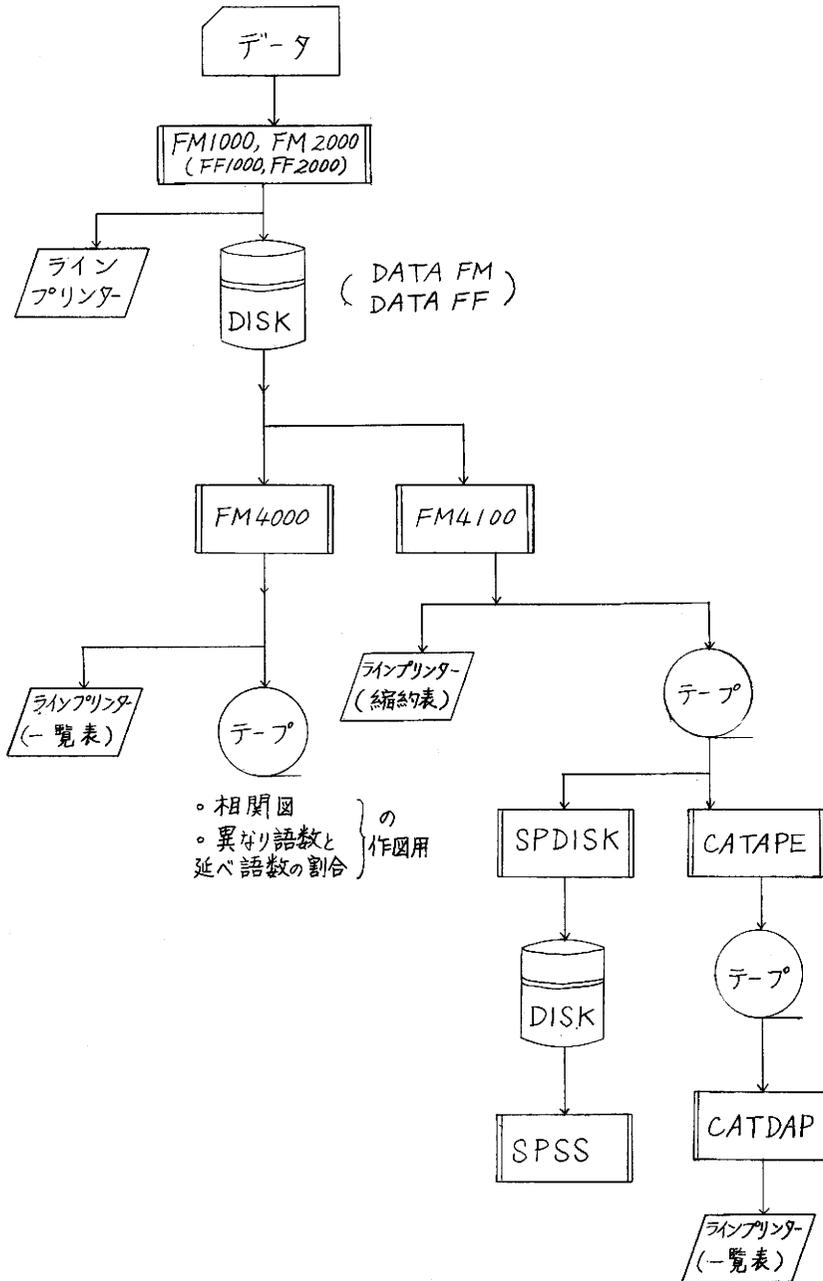
現在統計解析には既存のパッケージ(主としてSPSS)を用いているが、本研究を進めるにあたり、データの集計、作表、作図、あるいは変換の為に開発した幾つのサブプログラムがある。以下にこれらのサブプログラムの機能及び使用法の概略を示す。なおプログラムは全てFORTRAN言語でプログラミングされている。

5.1 データ処理用サブプログラムとその主な機能

サブプログラム名	主な機能
FM1000 (FF1000), FM2000 (FF2000) FM1000, FM2000は引用文も含めて 解析用 FF1000, FF2000は地の 文(引用文と除去)のみの解析用	各文献の文の構造に関する種々の 統計量及び単語に関する種々の 統計量の計算
FM4000	<ul style="list-style-type: none"> 全文献の文の構造に関する統計量の 一覧表の作成。 著作年との相関の高い統計量に関して 相関図の作図 異なり語数と延べ語数の関係を作図
FM4100	全文献の文の構造に関する統計量の 縮約表の作成
FM5000	単語別使用頻度一覧表作成 (分母は全単語の使用頻度:略記 P1)
FM6000	単語別使用頻度一覧表作成 (分母はその単語が帰属する品詞の 使用頻度:略記 P2)
FM7000	単語別使用頻度縮約一覧表作成 の準備
FM8000	単語別使用頻度縮約一覧表作成
AW SORT	全単語をアルファベット順に並べ変える
SPDISK	SPSS用データをディスクへ出力
CATAPE	CATDAP(変数選択プロ)用 データをテープへ出力

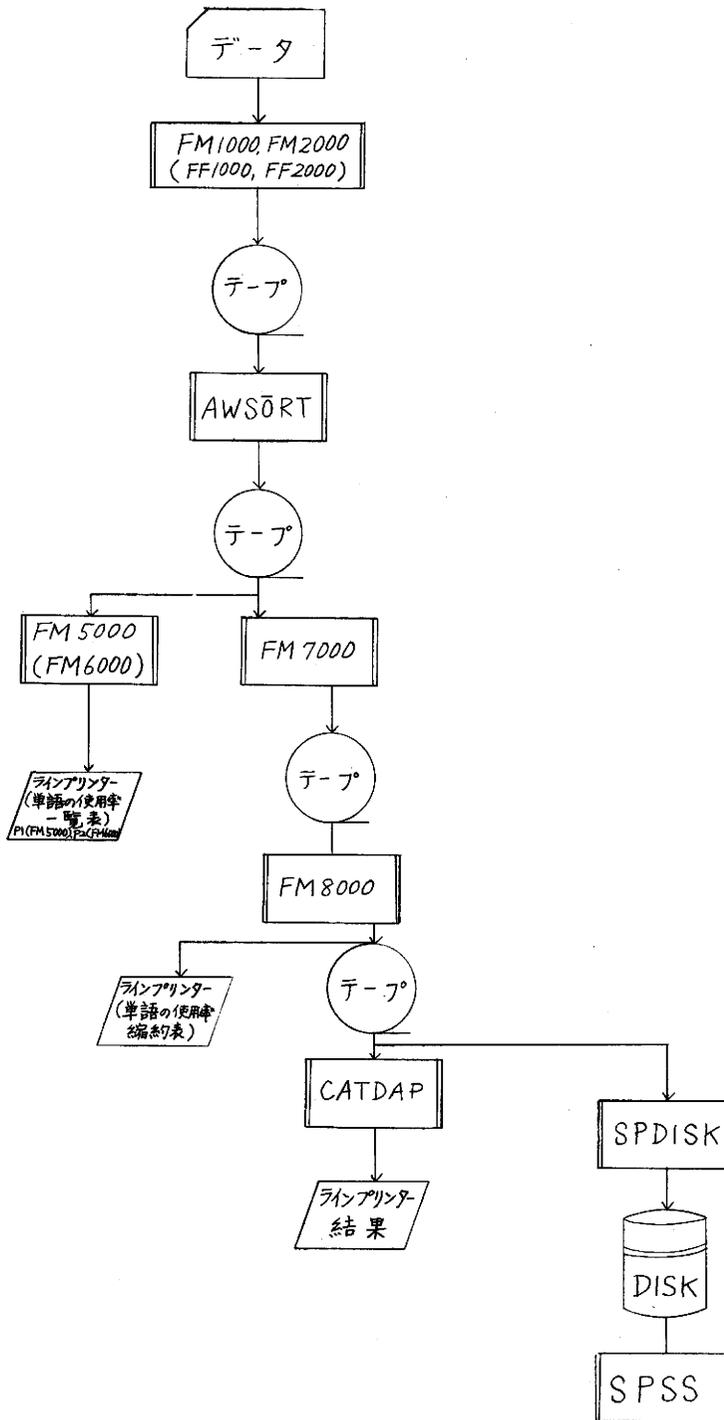
データの流れ

・文の構造に関するデータ



データの流れ

・単語に関するデータ



P. FORT FM1000, FM2000 (FF1000, FF2000)

```

//
//GO.SYSIN DD *
//GO.FT10F001 DD DSN=DATAFF, DISP=MOD
//GO.FT09F001 DD DSN=&WORK5, DISP=OLD
// UNIT=DISK, DCB=(RECFM=FB, LRECL=44, BLKSIZE=4400)
//GO.FT01F001 DD DSN=&WORK2, SPACE=(TRK, (30, 5)), DISP=(NEW, PASS),
//GO.FT04F001 DD DSN=&WORK4, DISP=OLD
//GO.FT02F001 DD DSN=&WORK1, DISP=OLD
MSTARTC FM2000 (FF2000)
MPROCESS NOSOURCE
//FLG EXEC FLG, SLIB=PROGRAM
//*
//SORT FIELDS=(15, 32, CH, A), SIZE=E1000, DYNALOC=(8589-11), 10)
// UNIT=DISK, DCB=(RECFM=FB, LRECL=46, BLKSIZE=460)
//SORTOUT DD DSN=&WORK1, SPACE=(TRK, (30, 5)), DISP=(NEW, PASS),
//SORTIN DD DSN=&WORK, DISP=OLD
//SORT EXEC SORT
//GO.SYSIN DD DSN=FMDATA(S14), LABEL=(, , IN), DISP=SHR;
// UNIT=DISK
//GO.FT09F001 DD DSN=&WORK5, SPACE=(CYL, (1, 1)), DISP=(NEW, PASS),
//GO.FT03F001 DD DUMMY
// UNIT=DISK
//GO.FT04F001 DD DSN=&WORK4, SPACE=(TRK, (30, 5)), DISP=(NEW, PASS),
// UNIT=DISK, DCB=(RECFM=FB, LRECL=46, BLKSIZE=460)
//GO.FT01F001 DD DSN=&WORK, SPACE=(TRK, (30, 5)), DISP=(NEW, PASS),
MSTARTC FM1000 (FF1000)
MPROCESS NOSOURCE
//FLG EXEC FLG, SLIB=PROGRAM
//*JOBPARM S=1000
//MURAKAMI JOB CLASS=D

```

b b b b b b

0	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
---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

誤 正

P28 ↓ 16 玉誤 → 玉沢

P29 ↑ 2 玉誤覚林院 → 玉沢覚林院

P30 ↓ 6 玉誤門流 → 玉沢門流

P33 ↓ 16 京都本寺断片 → 京都本園寺断片

P74 ↓ 16 日算(年)は削除

日順(1294年~1354(又は1365)年)

日興(1246年~1332年)

↓ 18 用心抄(建武3(1336)年)日順

本門心底抄(貞和5(1349)年)日順

五人所破抄(文和3(1354)年)日順

富士一跡門徒存知事(徳治1(1306)年)日興相伝

原殿御返事(正応1(1288)年(?))日興

Research Report

General Series No. 53

STATISTICAL INFERENCE
IN
AN AUTHORSHIP PROBLEM

by

H. Hudimoto, M. Murakami,

Z. Itô, S. Kasuga, J. Kose

TÔKEI-SÛRI KENKYÛZYO

Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku,

Tôkyô 106, Japan