

文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (二)

一 分析の対象

ほかならぬ『ゲーテ・シラー往復書簡集』を計量的分析の対象に選んだ理由は、「文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (一)」(以下「特性 (一)」と略す)において既に述べたが、要は往復書簡の性質上ジャンル(書簡文)およびテーマ(話題)が共通し、試掘的作業において両者の文体が予想外に似通った特徴を示したため、多変量解析など種々の分析手法の有効性を試す恰好の素材であること、また、偉大な二つの精神の交流記録という、ひとりゲルマニステイクのみならず、精神文化史一般に関心をもつ者にとって興味尽きない内実がそこにはある、という二点に尽きよう。

新 井 皓 士

一七四九年八月二八日生まれのヨハン・ヴォルフガング・ゲーテと、一七五九年十一月十日生まれのフリードリヒ・シラーの文通は、一七九四年六月十三日に始まり、一八〇五年五月にシラーが早世するまで間断なく続いている。情報と思想の交換手段としての手紙のやりとりが大きな比重をもっていた十八世紀・十九世紀のことであり、ゲーテから発信された書簡だけでも膨大な量のもので残っているにしても、この書簡集は特別なものである。なによりもそれは受信した手紙類を焼却することもあったゲーテの⁽²⁾、往復書簡集の形で公刊された書信往来は、音楽家ツェルターとのものと、このシラーとのもの、だけであることが示している。しかも一七九九年に始ま

り両者没年(一八三二)に終焉するゲートル・ツェルター書簡集⁽³⁾では、最初期を除けば、友人知己の間で普通用いられる親称(Du)をむしろ年長のゲートルが促す形で使ったのに対し、偉大な同志であり交友であると互いに認めあうゲートルとシラーの間では、一貫して敬称(Sie)が用いられ親称で呼び掛け合うことが遂になかった。その意味についてはさまざまな解釈も可能であろうが、年齢、身分、親疎の度合などで単純に割り切ることだけは不可能である。

しかし我々はこのことは事実の指摘にとどめ、またこの珠玉の書簡集の全貌や人文的ないし定性分析の特徴にふれることはここでは避けて、もっぱら計量分析の素材としてこれに接することにしよう。そのような素材としていま手元に用意されているのは、一七九四年から一七九七年まで四年間について、ゲートルとシラーそれぞれに分け、機械可読形式にしたテキスト・データである。底本としたのは、いわゆるアルテミス版ゲートル全集に依拠したフィッシャー文庫本であり、一枚一枚はがした頁をOCRで読み取りテキスト・データに変換したものを、手作業で誤読を改め分析用に整形・修正したものである。

コンパクト・ディスク版各種テキストが急速に普及しつつある情勢にあつては、数年前にはなお多少とも先端的であつたこのようなテキスト・データ作成作業も、もはや些か迂遠であり非効率なものになりつつある。しかし、文頭以外でも名詞や敬称を大文字で書き始めるなど、やや特殊な書記法をとる近・現代のドイツ語文献を計量分析の対象とする限り、正常なテキストといえども分析の手法や目的に応じてその都度、整形ないし修正の必要があることには変りない。

二 文章長の分析

二・一 文章長の分布は、対数正規分布をなす
始めに図一、図二をご覧いただきたい。図一はゲートル書簡、図二はシラー書簡の、文章長分布をヒストグラムで表わしたものであり、上から順次一七九四年、九五年、九六年、九七年となっている。左側のヒストグラムはいずれも尖度が高く鋭峰をなし、歪度からみれば左傾しているのに対し、右側のヒストグラムはほぼ釣鐘型の正規分布を示していると言ってよいであろう。左右は同じデータに基づくものだが、左側は頻度の実数データを使っ

(3) 文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (二)

ているのに対し、右側はその自然対数をとっている。文章長の分布が対数正規分布に近いことはつとに指摘されているようだが、ゲーテ・シラー書簡集の場合もくっきりとそれが形にあらわれたのである。書簡集の特徴として結びの常套句、たとえば「*バギげんよう*」(Loben Sie mich)などがほぼ毎回現れるので、実数データによる左図の左端部(数語からなる文の頻度)にそれが如実に反映している。このような「外れ値」はテューキーの提唱に従って5%トリムなどを行えば多少すっきりするが、ここでは図表としてはあえて示さなかった。年次毎の主な記述統計量は表一から表四にまとめてある。

文通開始当初二年間のみを扱う「特性(一)」をまとめた時点では、半年足らずと期間も短い第一年目は両者の間に量的のみならず文章長平均などに当然差異がみられるのに対し、両者の親密度が急速に増し往来も頻繁になる第二年目は少なくとも文章長平均などはほとんど同じ値となる意外さに驚くとともに、第三年目以降も同様の傾向がむしろ強まるかと予測したが、こうして四年間を通算してみるとむしろ第二年目が例外である可能性が浮かんできくる。出自、経歴、文学的な資質や世代が相当

に異なる両者が接近する直接のきっかけは、シラーが企画した高踏的文学誌『ホーレン』にあり、この時期の書簡にはこれをめぐる話題も多いので、できれば『ホーレン』が廃刊となる一七九八年まで通算五年間を通してみたかったが、残念ながら本稿執筆時点ではデータをそろえることができなかった。平均の変化など時系列的観察には手持ちデータはその意味でまだ不十分なわけだが、後で述べる判別分析などでは現在得られているデータによる帰結(判別式)の妥当性を検証するため未知の一七九八年資料を利用できる希望もある。

二・二 カイ二乗検定、 t 検定、分散分析

さて、文章長の頻度分布は、たとえば作者判別のデータとしてどの程度有効であろうか。対数正規分布をすることが確認できたので、ここでノンパラメトリックなカイ二乗検定(一様性検定)と正規分布を前提とする平均値の t 検定を比べてみよう。試みに一七九六年度のデータ(対数)をゲーテ、シラーともそれぞれ六個のファイルに分割し、長さにより九つのクラス(階級)に分けて頻度を調べてみる。これを合算した年度全体のデータに

ついで、ゲートとシラーの文章長頻度分布は一様性をもつ(換言すれば、ゲート・シラーという分類と長さによる分類は互いに独立である)という帰無仮説を立て、カイ二乗検定統計量を計算すると自由度六で十一・三三五となるから、仮説は棄却されない。一方、それぞれの年度平均値についてt検定を行うと、自由度一五一一で二・三〇五となり、五%水準で有意である。またゲート・シラー計十二個の小ファイル(群)について分散分析を行ってみると、分散比(F値)は十・三四(自由度は十一と一五〇〇)となり、この値は一%水準で有意に大きいから、十二群の平均は変化があるということができる。

次に、ゲートとシラーそれぞれ六個のファイルについて別々に分散分析を行ってみると、ゲート方は分散比が四・九六一、シラー方は十七・四一五となるゆえ、ともに一%水準で有意に大きい。表五より明らかのように、両者とも同じ時期、すなわち七月および年末の書簡文の長さの平均が高いのだが、七月は特にゲートの小説『ヴェルヘルム・マイスターの修行時代』、年末はシラーが手掛けている戯曲『ヴァレンシュタイン』や、両者合作

の寸鉄詩『クセニエン』や『マイスター』がしきりに話題にのぼり、やや論評的内容の書簡文が多い。翌一七九七年では八月のゲートの書簡文が異常に長いこと、その背景にゲートの私生活における転機があることを、分散分析応用の一例として『計量言語学の方法と実践』⁽⁵⁾で指摘したが、ここでは文学史上のいわゆるワイマール古典期に関する精神的転機が書簡文に反映しているとみることができようか。文章長の分散分析は、インターネット経由で入手した報道文(『ジューマン・ニューズ』、『ドイチェ・ヴェレ』)についても試みてきたが、これらは平均が極めて安定している。その一方、たまたまテキスト・データの用意があったゲートの『マインツ包囲滞陣記』や『ビンゲンの聖ロフス祭』を分析してみると、その一部に平均の異常に大きい(文が長い)部分が見られ、分散比はいずれも有意となる。前者の場合は陣中生活が終わって帰路に故郷フランクフルトに立ち寄る記事の部分で、九七年書簡で異常値を示した八月も生母の暮らすフランクフルト滞在期であることを思い起こさせる。また後者の場合は旅行記の体裁をとる文脈の中に、善男善女大衆の信ずる聖者伝を意識的に物語として挿入した部

(5) 文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (二)

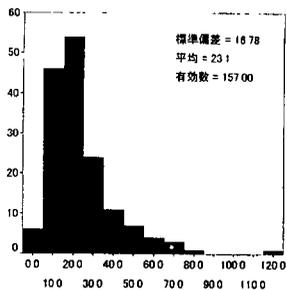


図 1

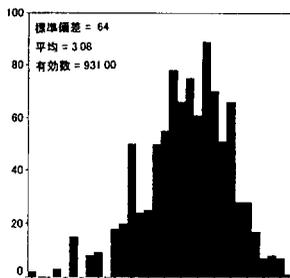
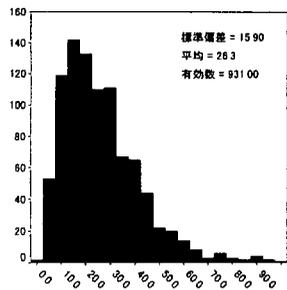
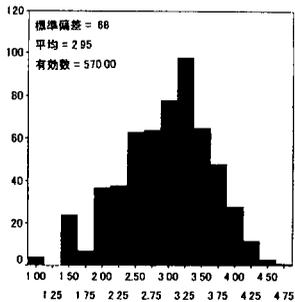
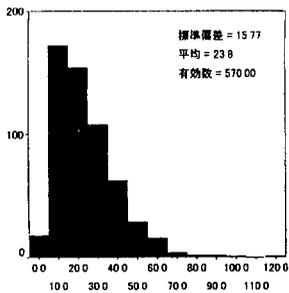
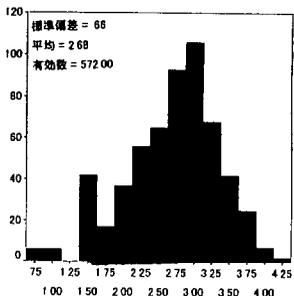
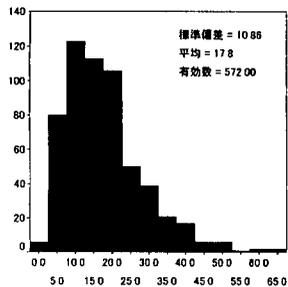
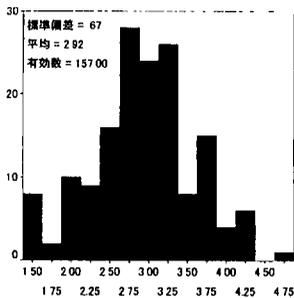
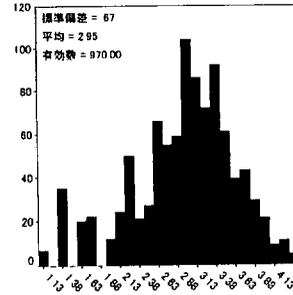
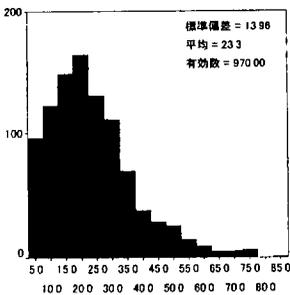
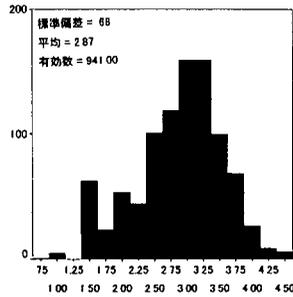
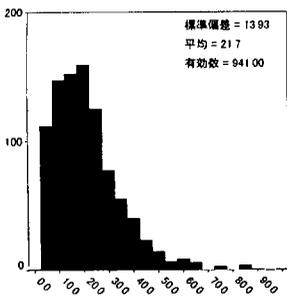
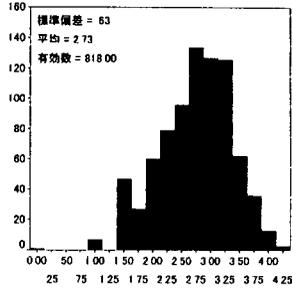
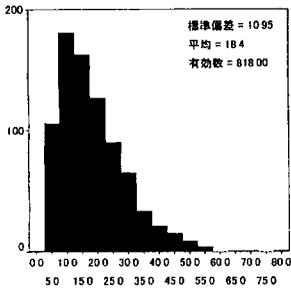
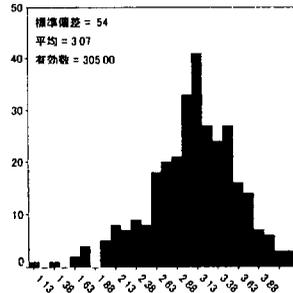
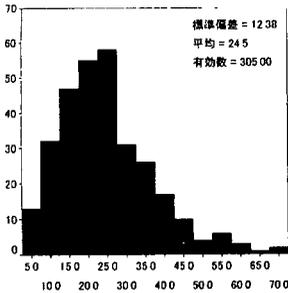


図2



(7) 文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (二)

表1 ゲーテ書簡の文章長

	n	mean	SD	Range	*	対数		
						mean	SD	Range
1794	157	23.08	16.78	4-120	*	2.919	0.672	1.386-4.787
1795	572	17.76	10.86	2-67	*	2.680	0.663	0.693-4.205
1796	570	23.77	15.77	3-67	*	2.953	0.683	1.099-4.787
1797	931	26.28	15.90	2-98	*	3.081	0.644	0.693-4.585

表2 シラー書簡の文章長

	n	mean	SD	Range	*	対数		
						mean	SD	Range
1794	305	24.54	12.38	3-68	*	3.068	0.539	1.099-4.220
1795	818	18.38	10.95	3-78	*	2.731	0.627	1.099-4.357
1796	941	21.73	13.93	2-101	*	2.870	0.679	0.693-4.615
1797	970	23.33	13.96	3-83	*	2.954	0.681	1.099-4.419

表3 ゲーテ書簡の文章長 (トリムド平均)

	n	mean	SD	Range	*	対数		
						mean	SD	Range
1794	142	21.57	11.70	6-60	*	2.934	0.528	1.792-4.094
1795	516	16.95	8.30	4-40	*	2.699	0.534	1.386-3.689
1796	514	22.45	11.66	6-54	*	2.968	0.550	1.792-3.989
1797	840	25.01	11.83	7-56	*	3.100	0.505	1.946-4.025

表4 シラー書簡の文章長 (トリムド平均)

	n	mean	SD	Range	*	対数		
						mean	SD	Range
1794	275	23.72	9.35	8-49	*	3.084	0.417	2.079-3.892
1795	738	17.55	8.329	5-40	*	2.745	0.506	1.609-3.689
1796	848	20.54	10.18	5-47	*	2.886	0.548	1.609-3.850
1797	875	22.31	10.63	5-51	*	2.977	0.533	1.609-3.932

分が、他とかけ離れた平均値を示しているのである。その一方、短編物語『ノヴェレ』では分散比は二・〇四九で五％水準で有意ではなく、平均に変化があるとはいえない。この場合は物語の流れが一定していて、意識のないし無意識的に文体に長さの変化があらわれる必然性がないと解釈されよう。いずれにせよ口述筆記が多いゲーテは文の長さが比較的安定しているような印象がこれまで漠然とあったが、先入観を改める必要と変化の理由を考察する必要があるというのである。

(一) 文章長の分布は、対数正規も暫定的に次のような結論を下すことができる。即ち、

分布をなす。

(二) 文章長の分析は、相当量の資料を必要とするが、作者判別には必ずしも有効とはいえない。但し、カイ二乗一様性検定より、平均値による検定の方がやや有効である。

(三) 同一作者の文体の変化は、文章長を、変量とする分散分析を手掛かりに検出することが可能である。

三 単語次元の分析

我々は次に単語の出現頻度に基づく分析を試みよう。

但し、ここにいう「単語」とは見出し語(タイプ、レンマとも)ではなく、水谷⁶⁾のいう単位語(トークン)に類し、かつドイツ語特有の動詞の分離前綴りや「連糸(ストリング)化した分離動詞のZu不定詞なども一語とみなすものである。換言すれば、前後の空白や句読点などで区切られた一連の文字群を単語とみなすことにする。

このような前提にたった上で、単語の長さによる頻度分布を調べ、分割表に基づくカイ二乗検定によって、異なるテキストの一様性を判定する方法は、すでに何度か試みたところであり、データ量がある程度以上あればそ

れなりに有効であることも明らかになっている。しかし、今回はこの単語長によるカイ二乗一様性検定は省略し、不変化詞の頻度を多変量として、比較的語数の少ないテキストの判別を可能にする方法の確立を試みる。というのは、単語長による一様性検定では一万語程度のデータがそろえばかなり確実に判定できることが経験的に明らかになっているが、現実に作者不明テキストの帰属性を判定したい場合、そのテキストは一万語ほどの長さをもつことは期待できないことが多いからである。このような問題ではベイズ定理を適用したモステラーとウォレスによる『連邦主義者』の大規模な分析⁷⁾がすでに画期的評価を得ているが、我々はなお「古典的」多変量解析手法を用いて、なるべく少ない資源で効率的な判別ができる方法を模索してみることにする。

三・一 不変化詞二十語の選択

ゲートに関してはずでに不変化詞二五語を選び、主成分分析によって『若きウェルテルの悩み』、『親和力』、『ニコライの『ウェルテルの喜び』、『修士セバルドゥス』の距離関係を図示してみたが、その場合用いた二五語は

『ファウスト』コンコーダンスも参照しつつ頻度を優先して選択したので、当然のことながら前置詞も含むものであった。しかし前置詞は最も頻度の高い“zu”が上に述べたように分離動詞と一単位語化したり、前綴りと同形であったりするので、厳密に区別し数え上げるのはかなり面倒である。そこで今回は前置詞は一切外すことにし、『書簡集』の中でのなるべくゲーテとシラーの使用度に違いのあるものに着目しながら、改めて何度か選択を繰り返した結果、表六にある二十語を選び基礎データとすることにした。表中、“man”は不定代名詞であり本来の不変化詞とはいえないが、特にゲーテにおいて顕著な使用例がみられるので、あえて取り上げている。頻度を表す数値は、固有名詞(地名、書名などを含む)や数字、記号を除いた総語数をもとに、千語当りの頻度に換算したものである。

三・二 判別分析

この不変化詞の表を眺めながら、始めに試みたのは線形判別式による判別分析である。この手法がテキストの判別にかなり有効である可能性は、戯曲『ヘンリー六

世』三部作(シェイクスピアないしマローへの帰属問題)の分析⁽⁹⁾において示したが、その場合、テキストの量も大きく、また判別式の変数も十個と多かった。計算の手間を厭うわけではないが、変数の数はなるべく少なくし、少数標本でも判別できるようにしたい、というのが今回の目的である。試行錯誤を繰り返した末、変量として“aber auch man und”四語の頻度をとり、判別式

$$y = -1.215X_1 - 0.13X_2 + 1.687X_3 + 0.66X_4 - 22.50$$

を得た。この判別式を作るのにあたって利用したデータは、一七九四年、一七九五年、一七九六年のデータで、九五年のゲーテ・テキストは四分割、シラー・テキストは六分割、九六年のテキストは共に六分割して、上記四語の千語当りの頻度をそれぞれ算出し、平均値のベクトルと標本共分散行列の逆行列を用いて、判別関数を導出したものである。計算にはこれまで同様、AWKと「マセマテイカ」を使った。

一七九七年のデータを判別式作成に利用しなかったのは、この判別式の有効性、つまり誤判別率を検証するためのデータとして残しておく為である。一七九七年は文章長の分散分析を通じてゲーテの八月テキストに特異性

表5 ゲーテ・シラー書籍の文章長頻度分布(1796年)

a) ゲーテ							
文章長クラス	1-5月	6月	7月	8月	10-11	11-12月	
log1-log4	5	1	2	9	2	0	
log5-log7	4	5	3	15	8	1	
log8-log12	17	18	11	21	24	11	
log13-log20	27	21	9	32	25	14	
log21-log33	28	27	15	27	27	36	
log34-log54	14	13	21	23	14	15	
log55-log90	3	2	5	6	2	5	
log91-log148	0	0	0	1	0	2	
n(sum)	98	87	66	134	102	84	
** mean ***	2.910	2.913	3.146	2.824	2.838	3.202	
*** SD ****	0.662	0.648	0.738	0.756	0.653	0.571	
** SoS ****	42.896	36.562	35.976	76.540	43.501	27.416	
b) シラー							
文章長クラス	1-5月	6月	7月	8月	10-11	11-12月	
log1-log4	11	10	12	5	5	2	
log5-log7	12	10	11	4	18	12	
log8-log12	28	33	26	12	30	18	
log13-log20	36	39	71	28	40	35	
log21-log33	26	37	106	29	35	40	
log34-log54	8	8	66	12	12	25	
log55-log90	0	1	18	1	2	4	
log91-log148	0	0	2	0	0	0	
n(sum)	123	138	312	91	142	136	
** mean ***	2.600	2.650	3.100	2.871	2.700	2.969	
*** SD ****	0.661	0.633	0.841	0.667	0.661	0.749	
** SoS ****	57.783	54.999	142.598	33.876	60.675	57.835	

がみられたので、不変化詞による判別分析でも同様の結果がでるか、興味があった。また、テキスト・データ整備の過程で、所属不明のテキストが旧ファイルに残っていることに気付いたので、これにもこの判別式を適用してみることにした。

判別の結果は、表七に示されている。ゲーテのテキストならプラス値、シラーのテキストならマイナス値をとるのが正しい判別で、判別式を作る際に利用したテキスト群はすべて正しく判別されている。しかしこの判別関数の機能をコントロールするため残した一七九七年テキストについては、ゲーテ方、シラー方も一つづつ誤判別が出ている。ゲーテの方は多少可能性を予期した八月テキストで

はなく、年初期のテキストであり、“man”と“und”の頻度が低いため、この結果を招いたのであろう。一方シラーの場合は六月から八月にかけてのテキストで、こちらはシラーとしては“und”の頻度が並外れて高い。変量を少なくすると、このように特異値に影響される可能性が増すが、一定の目安をつける意味では四個位の变量で十分ではないかと思われる。なお、帰属不明ファイルは、判別式の数値が示すように、ゲーテのものであった。

三・三 主成分分析

今回、最後に試みたのは、不変化詞二十の頻度を変量とみなした主成分分析である。まず年次毎のデータ(表六)を用いて分析した結果、第一主成分を横軸に、第二主成分を縦軸にとり、ゲーテ、シラーそれぞれ四個づつの標本の主成分得点により、散布図(布置図)を描いたものが、図三で、その下の表には固有値、および固有ベクトルを表示した。固有ベクトルと各標本のデータ行列の積が「年次毎の主成分得点」欄で、布置図はこの中の第一、第二主成分得点を用いている。第一、第二主成分の累積寄与率をみると、五八・六四であるから、やや物

足りない感があるが、それでも図ではゲーテとシラーがはっきり二つのグループに分かれている。

この結果に勇気づけられ、問題の一七九七年度について、ゲーテ七分制ファイル、シラー六分制ファイルを標本とし、同じ不変化詞二十の頻度に基づく主成分分析を行い、第一主成分得点と第二主成分得点により描いた布置図が図四である。この場合は図三の場合ほどゲーテの標本グループとシラーの標本グループが明確に分離せず、G一とS三などはかなり接近した形になっている。G一とS一は一七九七年の一月から四月にかけての、またS三は六月から八月にかけての書簡文テキストだが、前節の判別分析でもG一とS三は、微妙な差ではあるが、誤判別であった。その一方、文章長の解析では特異値を示したG三は図四では最右翼の一角に位置し、まぎれもなくゲーテ的ではあるが、他の標本とはかけ離れた恰好である。そこで少し伝記的事実を調べてみると、一七九七年春には、二月二十日から三月三十一日まで、約四十日間にわたって、ゲーテがシラーの居るイェーナに滞在し、しばしば対面するとともに執筆中の『ヘルマンとドロテア』について手紙の中で頻繁に言及している。また七月

表6

	G1794	G1795	G1796	G1797	S1794	S1795	S1796	S1797
	3504	9834	13213	23828	7329	14492	19720	22092
aber	2.85	2.85	2.57	2.98	5.59	5.11	4.94	4.71
als	4.85	3.86	3.33	5.41	5.18	4.28	4.94	4.16
auch	3.71	6.30	7.42	6.88	4.23	7.80	8.58	8.33
bald	3.14	3.25	2.04	1.76	0.96	1.52	0.50	1.36
da	3.42	3.05	3.48	3.23	3.00	3.04	2.98	2.63
dass	9.70	9.15	9.08	9.48	10.37	10.01	10.39	10.14
denn	1.71	2.54	3.25	3.06	3.00	1.93	1.61	1.90
doch	3.71	4.78	4.77	4.03	3.00	3.38	3.94	3.03
hier	1.43	2.85	2.27	2.27	3.14	3.17	3.78	2.58
man	1.43	5.90	6.74	5.96	2.73	2.76	3.03	3.49
nicht	11.99	10.37	9.69	8.35	10.10	12.01	10.80	11.81
noch	5.99	6.41	5.98	6.88	4.91	7.80	8.73	4.66
nun	2.00	2.64	1.82	2.85	1.50	1.10	1.82	1.40
nur	4.85	3.86	5.52	5.29	3.27	3.38	4.39	3.85
recht	5.71	4.58	6.21	4.62	1.09	4.00	3.58	4.57
schon	2.57	2.64	2.27	2.06	2.32	2.21	3.78	2.72
so	11.13	8.95	11.43	11.41	9.96	11.45	12.06	10.91
und	36.53	35.79	34.74	37.31	29.34	25.46	30.98	33.09
wenn	4.28	5.08	5.52	4.62	3.55	5.56	4.04	3.80
wohl	5.99	6.10	5.22	4.53	1.39	3.17	4.04	3.94

十一日から十八日にかけてはシラーがワイマールのゲーテ邸に滞留し、ゲーテはその後もなく、内縁の妻クリスティアーネと息子アウグストの為に遺言書の書き換えをし、故郷フランクフルトにおいて母と妻を引合せている。実生活上の距離の接近が必ずしも書簡の文体に直接影響するものではないことは、同じ様に約四十日間ゲーテがイエーナに滞在した一七九六年の主成分分析布置図(図五)が示している。しかし九六年はいずれの作か容易に判断できぬほど合作ないし共作の実をあげた『クセニエン』の年であるのに、主成

(13) 文体統計論的にみた『ゲーテ・シラー往復書簡集』の特性 (二)

表7 判別関数[$y = -1.215X_1 - 0.13X_2 + 1.687X_3 + 0.66X_4 - 22.5$]による判別結果

初期標本		判別すべき標本		
標本 (ファイル)	判別得点	標本 (ファイル)	判別得点	正誤
Goethe 1794	0.077	Goethe 97-1	-1.687	誤
Goethe 95-1	4.517	Goethe 97-2	6.782	正
Goethe 95-2	3.861	Goethe 97-3	6.997	正
Goethe 95-3	11.547	Goethe 97-4	11.717	正
Goethe 95-4	6.891	Goethe 97-5	12.520	正
Goethe 96-1	3.799	Goethe 97-6	5.375	正
Goethe 96-2	9.023	Goethe 97-7	6.634	正
Goethe 96-3	6.845			
Goethe 96-4	10.562	Schiller 97-1	-3.448	正
Goethe 96-5	4.813	Schiller 97-2	-12.686	正
Goethe 96-6	9.923	Schiller 97-3	0.649	誤
		Schiller 97-4	-5.196	正
Schiller 1794	-5.872	Schiller 97-5	-2.420	正
Schiller 95-1	-5.763	Schiller 97-6	-0.548	正
Schiller 95-2	-11.785			
Schiller 95-3	-8.040	Unbekannt	10.143	
Schiller 95-4	-11.846			
Schiller 95-5	-4.504			
Schiller 95-6	-9.618			
Schiller 96-1	-8.545	[X ₁ aber の千語当り 頻度]		
Schiller 96-2	-1.645	[X ₂ auch "]		
Schiller 96-3	-3.437	[X ₃ man "]		
Schiller 96-4	-0.173	[X ₄ und "]		
Schiller 96-5	-2.786			
Schiller 96-6	-7.977			

分分析の結果ははっきり両者の差異を示すのに対し、「バラードの年」といわれる九七年が接近(G一)と離反(G三)をとともに示しているのは何か意味がありそうに思える。人生の重要案件の介在が書簡の文体に影響を与えていると即断することはできないが、なお今後の課題として検討に値しよう。

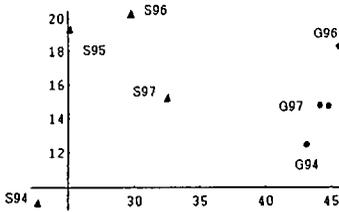
以上の試みから、我々は単語次元の解析に関し暫定的に次のようにいうことができよう。

(一) 頻度がある程度高い不変化詞群は、多変量解析のデータ(変量)として、文体特徴をとらえるのに有効である。

(二) 判別関数分析によれば、比較的少ない変量により、かなりの程度正しく作者の判別をすることができる。

(三) 変量が二十ほどあれば、主成分分析により、相当程度正確に作者識別も

図3



上図は、不変化詞20語の頻度分布に基づく主成分分析の結果、第一・第二主成分得点の散布図(布置図)を描いたもの。横軸は第一主成分、縦軸は第二主成分。なお、固有値、固有ベクトルは次表(左)に示す。

変量 I Vector II-V III-V IV-V

aber	-0.985	-0.004	0.140	-0.086
als	-0.394	-0.349	0.070	0.765
auch	-0.135	0.789	0.354	-0.144
bald	0.748	-0.227	-0.483	-0.234
da	0.709	-0.021	-0.068	0.195
dass	0.459	-0.463	0.699	-0.136
denn	-0.963	-0.005	-0.735	0.252
doch	0.832	0.331	0.226	-0.058
hier	-0.653	0.368	0.525	-0.121
man	0.693	0.100	0.640	-0.195
nicht	-0.456	0.231	-0.794	-0.328
noch	-0.059	0.758	0.167	0.285
nun	0.713	-0.159	0.223	0.421
nur	0.761	0.216	-0.044	0.475
recht	0.770	0.411	-0.384	-0.045
schon	-0.358	0.545	-0.153	0.365
so	-0.136	0.611	-0.034	0.466
und	0.797	-0.209	-0.133	0.394
wenn	0.565	0.441	0.131	-0.508
wohl	0.818	0.281	-0.407	-0.061

* 年次毎の主成分得点

	第1	第2	第3	第4
G94	43.19	12.41	-8.10	19.95
G95	44.83	14.76	-1.77	16.71
G96	45.56	18.15	-0.50	17.53
G97	44.18	14.81	1.19	21.79
S94	22.78	9.02	1.24	16.87
S95	25.09	19.30	-0.17	13.28
S96	29.88	20.19	1.47	18.84
S97	32.63	15.14	-1.88	16.31

固有値	8.606	3.122	2.70	2.205
寄与率	43.03	15.61	13.50	11.03
累積 ¹⁾	43.03	58.64	72.14	83.17

図4

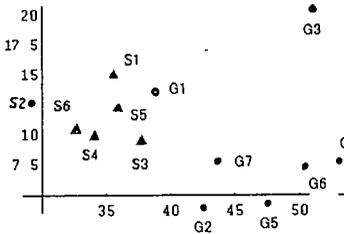
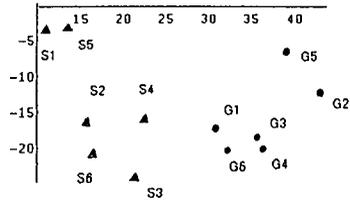


図5



可能であり、その結果は視覚的にとらえることができる。

筆者が計量言語学とともに関心を寄せる近世ドイツの言語文化資料には作者不明の文献が少なくないが、その一つに十七世紀末に刊行された『ラーレブーフ』という諷刺物語がある。近年その作者を、シュトラースブルク出身の多芸多才な法学博士ヨハン・フィッシャルトに擬する説が、書誌学方面から浮上し有力視されているが、いまだ決着をみていない。フィッシャルトは法曹人としての職務の傍ら、ラブレエの『ガルガンチュア』を倍ほどの長さに翻案するような奇妙な男である。その存在に對する興味と統計学的アプローチへの興味が、筆者を『ラーレブーフ』作者問題にも向かわせるのだが、そのためにはまず方法の有効性と限界を確かめなければならぬ。『ゲーテ・シラー往復書簡集』は腕試しにはまさにうってつけの相手である。また、よしやこの実験が徒勞に終わらうとも、フランス革命と欧州動乱の迫りくる時代に生きた精神界の巨人の葛藤に一端でも触れることができるのは、学徒冥利に尽きるというものである。ゲーテに做っていえば、願わくば、もっと時間を。

* 本稿は、内容の貧しさにも拘わらず、学の道において先輩であり多年同僚の榮をかたじけなくした諏訪功教授の御退官を記念しつつ、文部省科学研究所補助金を受けた研究成果の公表義務を果たし、かつは一研究者として「どこい、おいらは生きている」事を証しするという、欲のほつた多目的に従うものであることを付言する。

(1) 『ゲーテ年鑑』第三十八巻(一九九六年)、一五五頁〜一七四頁。なお、底本としたのは、*Goethe-Schiller Briefwechsel*, Frankfurt/Hamburg 1950.

(2) たとえば、一七九七年七月九日の日記に、「手紙類焼却。用紙が鉄線格子近くで焼けるとき炎の美しい緑の色」とも(『Tagebuecher 1775-1809, div. Gesamtausgabe 43, p. 173』)。この時は、故郷フランクフルトへの出立を前にして、一七七二年から九二年にかけて受け取った約八百通の手紙を、二日と九日の日曜日に焼いたらしく(『*Goes Lebens von Tag zu Tag*, Bd. 3, Hg. R. Steiger, Zurich/Muenchen 1984, p. 603 f.』)。一種の覚悟の念が感じられる。

(3) *Briefwechsel zwischen Goethe und Zelter in den Jahren 1799 bis 1832*, Hg. H.G. Ottenberg/E. Zehm, Muenchen/Wien 1991)

(4) 文章長は、本稿第三節にいう「単語」の数で測っている。

- (5) 『一橋論叢』第一二一卷第四号(一九九九年)、八一頁
～九七頁。
- (6) 『数理言語学』(水谷静夫、第二刷、東京一九八七、二
六頁)
- (7) *Applied Bayesian and Classical Inference—The
Case of "The Federalist Papers", 2nd Edition, New
York/Berlin/Heidelberg/Tokyo 1983. #27, 1983*
- (8) いては、『一橋論叢』第一一五卷第三号(一九九六年)、一
四頁～二九頁。
- (9) 前掲『一橋論叢』第一二一卷第四号(一九九九年)。
『一橋論叢』第一一九卷第三号(一九九八年)、一頁～
一九頁。
- (一橋大学大学院言語社会研究科教授)