

多変量解析の文献学的応用： 『カルストハンス』の事例

新井 皓 士

1. 『カルストハンス』の作者問題

1.1 『カルストハンス(Karsthans)』とは、1520年と1521年の交りにかけてシュトラースブルクで刊行され、その後当市はもちろんバーゼル、アウクスブルクなどでも版を重ねた、宗教改革初期のパムフレット文書(Flugschriften)の代表作である。ウルリヒ・フッテンの対話体啓蒙闘争文書の流れを汲み、数名の登場人物の口を通じて、ローマ・カトリック教に対峙するマルティン・ルターを擁護し、当時最大の論敵とみなされたトマス・ムルナーを揶揄する、人文主義的色彩の濃い文書である。対話体をとるところから、ラテン語の直接的影響を受けやすい文語的文体を避け、極力口語体に近づこうとする方向性がみられ、初期近代ドイツ語発展史上も興味ある資料といえよう。カルストは農具の一種で、二又または三又の鋤²⁾とってよいが、穀さおとともに、南ドイツでは農民の象徴とみなされていたらしい。ハンスはむろんナントカ太郎に類する、やや侮蔑的な一般固有名詞のような使用方法がある。つまりカルストハンスとは、愚昧朴訥の農民の謂となるが、ここではそれが逆転の発想から使われ、会話の中心人物カルストハンスは、いわば健全な理性、良識の代表者であり、篇の前半では書物に無縁であるような印象を与えるが、後半に至ると相当の識字力と学殖を有していなければありえない発言をしている。匿名作者は人文主義的教育を受けた知識人であり、この文書の流布も、象徴的作用は別として、直接には市部を中心とし、村落部でも上層の識字農民に限られた、と推測される所以である。

会話はカルストハンスを中心に、ケルンの大学に在籍する息子と、メルクリウスと称する客人風コメンテーター、これにムルナー、および、これと入れ替わりに登場するルター

1) このことは矢継ぎ早に発表されたムルナー文書のうち、20年12月24日付けで公刊された『ドイツ国民の貴族諸侯に寄せて』は『カルストハンス』で軽く言及されているのに対し、21年2月7日付けの『ルター博士が誤れる原因に動かされて教会法を焼き捨てた次第』については言及がないこと、21年1月13日に『カルストハンス』発禁の要望書がムルナーからシュトラースブルク市当局に提出されていることから、推測される。

2) 初版のタイトル頁の版画では、柄の部分と刃の部分がほぼ直角に交差する二又のカルストを右肩に担いだカルストハンスが、メルクリウス、ムルナー、シュトゥデンスと並び立つ姿が描かれている。これに対して『フッテン全集』(Vol. 4, p. 615)やS. ミュンスターの『コスモグラフィア』(Bd. 2, p. 717)には、柄全体がゆるやかに撓んだ先に三又の刃が直結するカルストを杖のようにして孤立する農民像が掲載されている。好みにもよろうが、後者の版画は一層印象深い。

の間で交わされる。歴史上のムルナー³⁾は、フランシスコ会修道士としてカトリック信仰を擁護するが、宗教改革以前はむしろ教会制度もふくめて当時の社会にひそむ問題を諷刺する作品群や高等教育関連書で知られ、神学と法学の博士号をもち、かつ桂冠詩人でもある。しかし、この会話篇では雄猫まがいの風体で登場し旧弊墨守の傲慢無知な一派代表とされ、対照的にルターは謙虚な有徳の師として尊崇をうけ、「ドイツ語で」真の教えをひろめるよう要望されている。息子(Studens)ははじめこそ大学仕込みの知識をひけらかそうとするが、次第に愚直なはずの父の良識と知識に太刀打ちできなくなる格好である。メルクリウスなる人物はもっぱらラテン語しか口にせず、やや斜に構えた格好でちらりちらりと警句めいた注釈をはさみ、大学生がときどきカルストハンスとの仲介役となる。メルクリウスはルキアノスの作品に多くみられ、またエラスムスの『俚諺集成』(Adagia Selecta)とも関係するとみられること、息子が『無メイ氏書簡集』(Epistolae Obscurorum Virorum)を連想させる「ケルン」で学んでいることなどからも、作者が人文主義的知識人であることが推測されるのである。また、このように紹介すると、いかにもスマートで口当たりのよい近代作品を思わせるが、実際は近代ドイツ語生成発展期特有の熱気と短絡に富む、いわば胃もたれする作品であることはいうまでもない。

1.2 匿名作者の正体についてはざっと次のような名があがっている。『新カルストハンス』の作者であり宗教改革者としても知られるマルティン・ブツァー(Martin Butzer)、シュトラースブルクの人文主義史家ニコラウス・ゲルベル(Nikolaus Gerbel)、シュレットシュタットの高等学校長ヨハン・サピドゥス(Johann Sapidus)、ザンクトガレンの博学な宗教改革派人文主義者ヨアヒム・ヴァット(Joachim v. Watt; Vadianus; 1484-1551)、そしてシュトラースブルクの宗教改革派説教師マティアス・ツェル(Matthias Zell)、などがそれぞれである。このうちヴァット(ヴァディアヌス)について、筆者は15年以上の昔に、『カルストハンス』の作者は通称ヴァディアン(…)である、と断定してもほぼ間違えないと考える。」と書いている。それは筆者が、文学ないし歴史的作品に対して「philologisch-statistischな接近」を試みた最初のステップであり、未知の海に漕ぎ出す探索者のような、やや危うい自負と高揚感をともなうテーゼであった。パーソナル・コンピュータが漸く個人研究室でも使用可能になり、手作業で試行錯誤のテキスト入力をし、PCフォートランで簡単なプログラムを書いて統計処理を試みた時代のことである。そのとき応用した統計手法は、シェークスピア・ペーコン説を否定した「特定の著作者のテキストは少なくとも大規模な次元で見る限り無意識裏に使用されている単語の長短の比率が一定であることを示す」とするメンデンホール説に則り、ブラインガーが『マーク・トゥエーンといわゆるQCS文書』で範例を示したカイ二乗分布に基く統計的仮説検定であった。しかし「ほぼ」と留保

3) 以下、ムルナー、フッテンその他、『カルストハンス』の周辺文書については、不備なものだが筆者に『近世ドイツ言語文化史論——「祖国」と「母語」が意識されゆくころ』(1994)、また、文体統計論的試みについては『多変量解析による「ゲーテ・シラー書簡集」と「ラーブーフ」の文体統計論的研究』(2001)などがある。

を置きはしたが、その後さまざまな手法でさまざまなテキストを実験的に分析する可能性が情報機器の進化とともに広がるにつれ、単一の方法で判定する危険性の認識や表記法のばらつきもふくめた時代的制約やジャンルの変動の影響に関する知見の積み重ねなどが、かつての暫定的テーゼをあらためて見直す必要を促すのである。

今回ヴァエディアン作者説を再検討するために、対照データとしてあらためて用意したのは、ルター、ムルナー、ミュンツァーである。すなわち、問題の『カルストハンス』テキスト・データ、ヴァエディアン作であることがほぼ間違えない『豺狼歌』、『新旧信仰論』、およびブツァー作であることがほぼ確定している『新カルストハンス』のテキスト・データに加えて、ルターの『ドイツ貴族に寄せる』の一部、ムルナーの『ドイツ貴族諸侯に』の一部、ミュンツァーの『やむをえざる弁護論』の一部、さらに現代人のために標準化されたルター・テキストの『カルストハンス駁論』、『農民の12か条に関する平和勧告』、『農民暴徒鎮圧論』、『農民厳罰勧告回状』、『ハンスヴォルスト』を比較対照のため、分析用に整備⁴⁾した。このうちルターが相対的に多いのは、CDによるテキストを別途分析用に加工してあったので転用できる便宜と、キーボード入力せざるをえなかった原文テキストとこ

4) テキスト資料リストの順序と番号は後段の表や図と対応している。

01-04 Judas Nazarei (Pseudonym): Vom alten und neuen Gott, Glauben und Lehre (1521), Halle a.S., 1896 (Hg: Eduard Kück, Flugschriften aus der Reformationszeit 12)

05-06 Judas Nazarei : Das Wolffgesang (in: Flugschriften Sammlung G. Freytag, 3940, Microfilm)

07-10 anonym: Karsthans (Hg: Eduard Böcking, Ulrichs von Hutten Schriften, 4.Bd. p. 615-647, Aalen 1963, Neudruck der 1859-1861)

11-12 anonym: Gesprächsbiechlin Neüw Karsthans (Böcking, p. 649-681)

13 Murner, Thomas: An den Großmechtigsten vnd durchlüchtigsten adel tütscher nation (. . .), (in : Deutsche Flugschriften zur Reformation 1520-1525, hg.v. K. Simon, Stuttgart 1980, p. 110-126)

14 Müntzer, Thomas : Hochverursachte Schutzrede vnd antwort wider das Gaistlose Sanfft lebende fleysch zu Wittenberg (. . .), (in: Flugschriften des Bauernkrieges, hg.v. K. Kaczerowsky, Hamburg 1970, p. 145-167)

15 Luther, Martin: An den Christlichen Adel deutscher Nation (. . .), (Martin Luther Studien Ausgabe, hg.v. H-U. Delius, Bd. 2, p. 97-108)

16-22 Hg. Aland, Kurt: Gesammelte Werke (Digitale Bibliothek 63), davon

16 An den christlichen **Adel** deutscher Nation (1520)

17 Eine treue Vermahnung an alle Christen, sich zu hüten vor Aufruhr und Empörung (1522)(als **ggKHans**)

18 Ermahnung zum Frieden auf die zwölf **Artikel** der Bauernschaft in Schwaben (1525)

19 Wider die räuberischen und mörderischen **Rotten** der Bauern (1525)

20 Ein Sendbrief von dem **harten** Büchlein wider die Bauern (1525)

21 Warnung an seine lieben **Deutschen** (1531)

22 Wider **Hans Worst** (1541)

Nr. 16の略題を「ggKHans」としたのは、まぎらわさを避けるためであり、この書の中に次のような文言があり、ルターはすでに『カルストハンス』を読んだ上で暴動を未然に警告しているとみられるからである。(.) mit Flegeln und Kolben dreinzuschlagen, wie der Karsthans drohet.⁴⁾

の標準化テキストの一部を比較対照し⁵⁾、今後の研究の参考にするためでもある。分析用テキストは、この時代の書記法の慣例に反して、名詞は大文字で始め、固有名詞や地名は大文字で全書し、文中の典拠表示(たとえば、Matt. 6.7のような)や略語表記などは一定の処理を施すなどして、同一の基準ではかれるよう前処理をし、文章の長さや語の頻度などを測定した。いわゆる、zu 不定詞は大概の場合、切れ目なく表記されているが、さほど多くない分離動詞ではそのまま1語扱いに、それ以外ではzuと不定詞の2語扱いとしている。要するに、スペースや句読点などで区切られた一連の文字綴りを語とみなす、という原則を、これまでの試行錯誤の結果、採用している。

1.3 『カルストハンス』が前半と後半でやや趣を変えていることはすでに指摘したが、文の長さの分布にそれは明らかである。図1は、文の長さを語数で測定し、測定値(正の整数)を対数変換(自然対数)した値を、擬似的な時系列データとみなしてグラフ化したものである。少し見にくいだが、上下動の激しい全データのグラフの上に、15項移動平均グラフを重ねて表示させている。移動平均は細かなデータではとらえにくい傾向をみるために行う平滑化の代表的手法である。

図の上でも260番のあたりにひとつの切れ目が認められるが、内容的にもこの直前に「ムルナー」及び入れ替わりに登場した「ルター」も退場し、以後は1520年末に相次いで出版されたムルナーの対ルター論争書が批判的話題の焦点になる。つまり前半は会話基調が強く、後半は(書評的)論議中心となるが、終始一貫して槍玉にあげられているのは、ムル

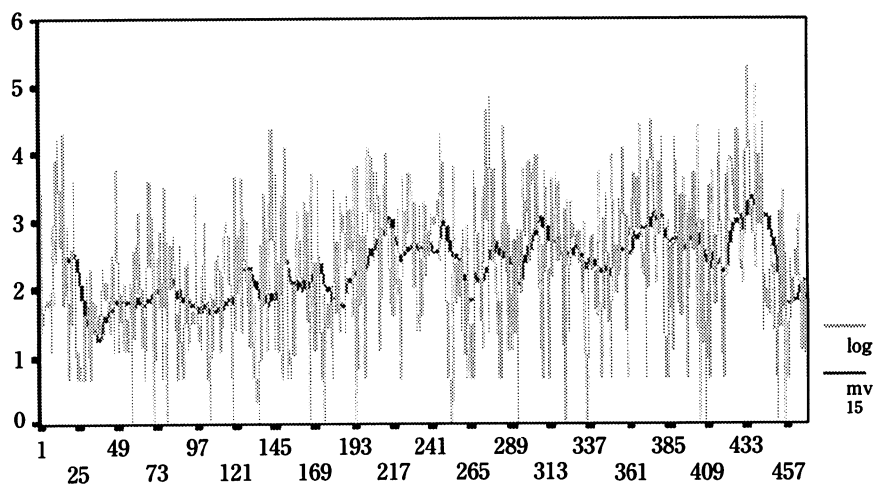


図 1

5) 注4のNo. 15, 16がこれに相当する。同一箇所についてみると、原文版にくらべアーラント版は読みやすくする為の句読点などの影響で文の数は増えているが、文長分布の検定では有意差はみられなかった。

ナーであり、実際ムルナーは農民的愚物を風刺するタイプとしての「カルストハンス」生みの親ともいえるから逆に揶揄されても仕方ないともいえようし、後には『ルター派大阿房』において「カルストハンス」に反撃している。おそらくムルナー自身は『カルストハンス』の作者についてある程度想像がついていたのだろう。しかし後世にあげられた作者候補の中ではヴァディアン説が比較的有力と思われるので、さしあたりヴァディアンの宗教改革関連文書と前後二分した『カルストハンス』との比較を試みることにする。図2は文長の分布を視覚的にとらえる為の、箱ひげ図(box and whisker plot)またはボックス・プロットと呼ばれるもので、第1四分位から第3四分位までがボックスで表され、四分位偏差の一倍半までが線(ひげ)で表されている。その上下の点は、はずれ値(outlier)と呼ばれるもので、たとえば1番左の上部にあるものは、全体の序文にあたり、際立って長く、実測値では244語を数える。1, 2列目(Kh 1, 2)はラテン語文も含む原文の分布であり、3, 4列目のdeの付されたものは、ラテン語文や完全引用文を除いたドイツ語文のみの分布、5列より8列までは『新旧信仰論』、9, 10列は『豺狼歌』である。

ボックスの中央からややずれた位置にメディアン(中央値, 中位数)を示す線があるのは、分布が歪んでいることを意味している。この図2からは、『カルストハンス』のグループと残りのグループはやや異なる印象をうけるのだが、このことはヴァディアン匿名作者説には不利である。というのは、『新旧信仰論』と『豺狼歌』は「ユダス・ナザレイ(Judas Nazerei)」という同一の偽名で公表されたが、これが一見似た名のレオ・ユートではなく、

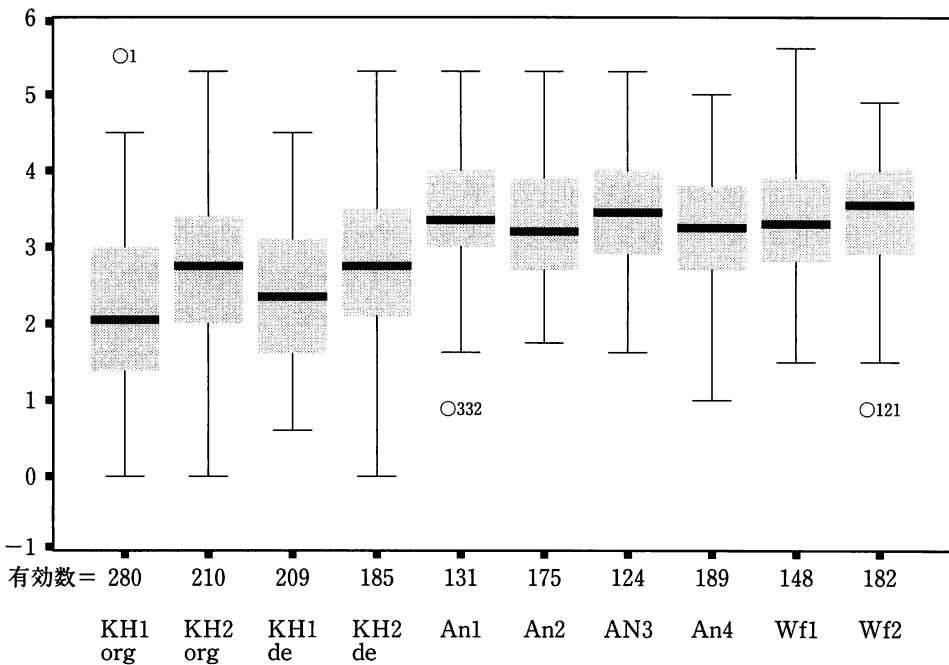


図 2

ヴァディアン⁶⁾の隠れた名であることはE. キュックによって文献学的に論証されており、その正しさは図2や次節以下の分析によって裏打ちされるからである。

視覚的印象ばかりでなく、平均に変動はない、とする仮説が認められるか、いわゆる分散分析(一元配置)を次におこなってみよう。表1は『カルストハンス』原文と『新旧信仰論』、『豺狼歌』計8群の分散分析、表2は『カルストハンス』ドイツ語文と『新旧信仰論』、『豺狼歌』計8群の分散分析の結果である。いずれも、平均に変わりはない、という仮説は棄却されている。

表 1

変動要因	変 動	自由度	分 散	観測された分散比	P-値	F境界値
グループ間	341.8631	7	48.83759	63.36177138	2.83E-79	2.016151
グループ内	1072.146	1391	0.770774			
合 計	1414.009	1398				

表 2

変動要因	変 動	自由度	分 散	観測された分散比	P-値	F境界値
群 間	185.8929	7	26.55613	39.85084	5.09E-51	2.016531
群 内	876.3006	1315	0.666388			
合 計	1062.194	1322				

一般に文長(対数変換値でも)の分散分析では、上下5%位をトリムしない限り、仮説受容(あるいはP値が0.05以上)となることは比較的少ない。その意味ではこれらの結果はさほど驚くべきことではないといえよう。しかし『カルストハンス』の前後篇の間ではt-検定の結果、平均が等しいとする仮説が棄却されること、その一方『新旧信仰論』、『豺狼歌』の6群の間では分散分析の結果、0.1%水準とはいえ、一応仮説が受容されることは無視することはできない。すなわち『カルストハンス』の前半と後半は文長においても異なるとみてよいこと、『新旧信仰論』と『豺狼歌』の文長は全体として変動が少なく、ほぼ同一とみてよいことが、これらの分析から結論付けられる。なお、これらの文長分布が対数正規分布に従うことは、コルモゴロフ・スミルノフ検定から明らかになっている。

1.4 語の次元では、語長分布がひとつの目安となることもあるが、正書法等が確定していない時代を対象にする場合は、必ずしも決定力をもちえないことは始めに触れたとおりである。いわゆる標準化を厳密な方針と手順に従って完遂しない限り、個人差(植字工の癖も加わる)や地方差、ジャンル差や微妙な時代差などがバイアスとなって、文体分析の結

6) 注4の01-04にあげた書のEinleitung IV。„Judas Nazarei“は„J. Vadian arzet“のアナグラムとする。

果を不安定なものにする。たとえば『カルストハンス』から目に付く一例をとってみれば、(1) sy, (2) syg, (3) sig, (4) syg など頻出するが、(1) は人称代名詞 sie 及び sein 動詞接続法 sei に、(2) は同じく sei に、(3) (4) は名詞 Sieg に相当する、といった具合で、16 世紀 20 年代前半と期間を限っても解読能力をもつスタッフ養成をふくむ相当大掛かりなプロジェクトの形で集中的かつ綿密にテキスト資料整備をする必要があり、そのような条件は現下の大学や研究環境では望むべくもない。

文体や語彙頻度がテキストの主題と関連することはいうまでもないことである。特に名詞においてそれは顕著であるが、一般に意味のある語、すなわち名詞、動詞、形容語彙などが、テキストそのものから切り離せない関係をもつものに対し、相対的に「ノンコンテクスチュアル」とみなされるのは、接続詞や前置詞、副詞の一部など、いわゆる不変化詞 (Partikel) とドイツ文法上よばれる語のグループである。日本語文法では自立語に対する付属語ないし辞がこれにほぼ相当するといえよう。日本語の助動詞や助詞が歴史的には複雑な機能をもつと同様、ドイツ語の場合も前置詞などは近世に微妙な変化が認められ意味論の視点からの検討と慎重な扱いが必要であるが、接続詞類などは形態的に判別しやすく「ノンコンテクスチュアル」な語として比較的安定しており、出現頻度も高いため、統計的分析対象として好適である。すなわちテキストの作者が比較的無意識に使用する可能性が意味ある語彙類より高いので、異なる作者間の差を反映しやすいのである。この点に着目して、多変量解析、特に判別分析や主成分分析、また質的データへのその応用としての数量化手法などを、作者問題などに活用し、文献学や文体論に多少とも貢献することが、目下の筆者の願いであり、本稿の試みである。

判別分析がもっとも有効となるケースは、二群ないし複数群の、作者がすでに判明しているテキスト群があり、ある不明文書がそのいずれに属するか、決定したいときである。筆者はかつてシェークスピア初期の作品に属するとされるが、論議のたえない『ヘンリー 6 世』3 部作について、作者候補をマーローとシェークスピアに限定して判別を試み、少なくとも第 3 部はマーロー作の可能性が高い、という結果をえた。また全体がハミルトン、マジソン、ジェイの 3 人の分担執筆であることは明らかであるが、その全 85 篇のうちマジソンによるかハミルトンによるか問題とされていた、アメリカ合州国成立期の政治文書『連邦主義者』中の 10 篇ほどについて、主成分分析と判別分析を組み合わせた比較的簡便な方法により、モスラーとウォレスの大規模で総合的な統計的分析と同じ結論に達している。この場合も、作者候補をあらかじめ限定できる利点があり、そのようなケースにおいて判別分析 (及び数量化 II 類) が分析手法として有効に働きうる。主成分分析 (及び数量化 III 類) はそれに対して、候補がより漠然と想定される場合に、特定の統計的モデルを前提とせず応用することができる利点があり、付置図のかたちで視覚的に結果を示すことが可能である。付置図は最近では個体の位置関係ばかりでなく、変量のベクトル表示もあわせておこなうバイプロットの形で示す例が増えているようだ。

変量としてノンコンテクスチュアルな機能語を利用するとして、次に問題になるのはどの語を選ぶか、変数の数はいくつにするかである。統計的分析をおこなうのであるから、

出現頻度が低いものはさしあたり考慮対象からはずし、候補として残る多出機能語 30 位の中から 5 乃至 10 程度の語を選び、変量 (変数) として利用するのであるが、その場合、頻度情報は元のテキストの大きさの影響を極端に受けないように千語当たりの相対出現頻度とする。「千語」ではなく「百語」当たりに換算しても良いようなものだが、それでは頻度の桁が小さくなって扱いにくいので、語の頻度情報は伝統的に千語当たりに換算している。一般にもっとも頻度の高い語は並列接続詞 und (and) だが、千語当たり換算でこれが 30 から 50 位の頻度となり、頻度上位語でも 2 桁になるものはそれほど多くないからである。また英語テキストなどでは定冠詞や不定冠詞もある程度無意識的に使われる機能語とみなされるが、ドイツ語では名詞との関わりを無視できないので、冠詞類はとりあえず対象外とする。日本語のように冠詞をほとんど使わない言語もあるので、比較の上でもその方が実際の感がある。もちろん総語数では重要なファクターであるが。

表 3 は今回の分析にあたり最終的に採用した 7 語のテキスト別の相対頻度である。

この段階まで絞る過程で最終的に準拠した基準は、『新旧信仰論』と『豺狼歌』計 6 個の分割テキスト (これをファイルと略称し、これら 6 ファイルをヴァディアン群とする) の平均と標準偏差から変動係数を算出し、この変動係数が 30 以下、というものである。すなわ

表 3 千語当たりの頻度

	sum	aber	dass	nicht	so	und	wenn	wie
Ang123	4853	3.71	9.89	9.07	11.95	34.00	5.56	3.09
Ang456	5450	2.57	11.56	9.36	9.54	40.55	4.22	4.22
Ang789	4940	4.66	10.73	13.97	19.64	25.71	2.83	4.86
Ang1022	5949	3.53	7.73	14.29	19.33	34.46	5.55	4.87
Wolf123	4953	4.44	14.74	14.74	21.60	38.36	3.43	3.84
Wolf456	6018	2.82	12.63	12.63	17.12	37.55	3.82	4.32
KHans1	3445	2.32	14.51	19.45	13.35	34.25	4.93	5.22
KHans2	4378	4.57	16.22	17.36	16.90	26.72	6.85	6.62
KHans1d	3216	2.49	14.61	20.83	13.68	33.27	5.29	5.91
KHans2d	4216	4.74	16.13	17.55	14.94	26.09	6.64	6.40
NewKH12	6997	7.00	13.86	21.29	5.15	52.02	2.57	8.43
NewKH34	7017	7.98	14.54	18.10	5.27	46.32	2.28	8.27
Murner	4406	8.62	25.42	15.89	10.89	34.95	0.91	4.54
Muentzer	2943	7.48	15.97	20.73	8.83	25.82	1.36	16.31
Adel01	4274	3.04	16.14	19.65	12.40	45.39	2.81	7.02
Adel-m	4568	3.06	14.45	18.17	12.48	45.53	4.16	7.88
ggKHans	4693	2.98	15.98	21.10	12.78	52.84	4.90	8.74
Artikel	9171	4.03	17.12	19.74	10.69	58.23	5.78	9.16
Rotten	2044	3.42	16.14	16.14	10.76	66.54	3.42	8.32
harte	7986	6.39	18.16	20.66	12.65	54.60	4.38	10.27
Deutsch	6052	3.97	15.04	19.17	14.87	59.15	3.14	7.93
HWorst	9553	5.13	13.40	16.43	9.63	53.70	3.87	9.63

ち『カルストハンス』の作者として一応最も有力視される、ヴァディアン群のテキストにおいて、相対的に最も安定している7語(aber, dass, nicht, so, und, wenn, wie)を主成分分析の変量としている。但し, dass については dz, daß, dasz など, nicht については nit, nitt, nüt など, und については vnd, vnnd, undt などの形があり, wenn についてはこの形より wan, wann のような変形がむしろ一般的, また so は副詞, 定関係代名詞, 従属接続詞(wenn)の機能を兼ね, 時には also や als の意味でも用いられているが, ここではあえて機能分類は行っていない。総数(sum)は, 出現語の中から固有名詞, 地名, 書名, 非ドイツ語文の語, 数字などを除いたものである。

このデータに基づいて分散共分散行列を求め, その固有値 $\lambda_1, \lambda_2, \lambda_3$ 等と, これらの固有値に対応する固有ベクトル $a_1, a_2, (\dots), a_p$ を求めて, 第1主成分, 第2主成分等(の計算式)を決定し, 個々の値を散布図(付置図)にあらわす, という手順が, 基本的な主成分分析の手法である。図3はその結果をバイプロット図で表したものだが, 数字の1から6までがヴァディアン群, 7から10までが『カルストハンス』, 11と12は『新カルストハンス』(ブツター), 13がムルナー, 14はミュンツァー, 15から22までがルター文書で, そのうち15と16は『ドイツ貴族に与える』の原典テキストと現代の読者が読みやすいようにある程度「標準化」したものである。この図3が示すところは, 同一作者のテキストはほぼ一箇所に固まっていること, 帰属が問題となっている『カルストハンス』の前後2分割ファイルは, 無削除版の7, 8とドイツ語文のみの9, 10がそれぞれ対応するファイルに近く, かつ全体として, ヴァディアン群, ルター群, ブツター群, あるいはムルナー, ミュンツァー文書からも距離があるということである。

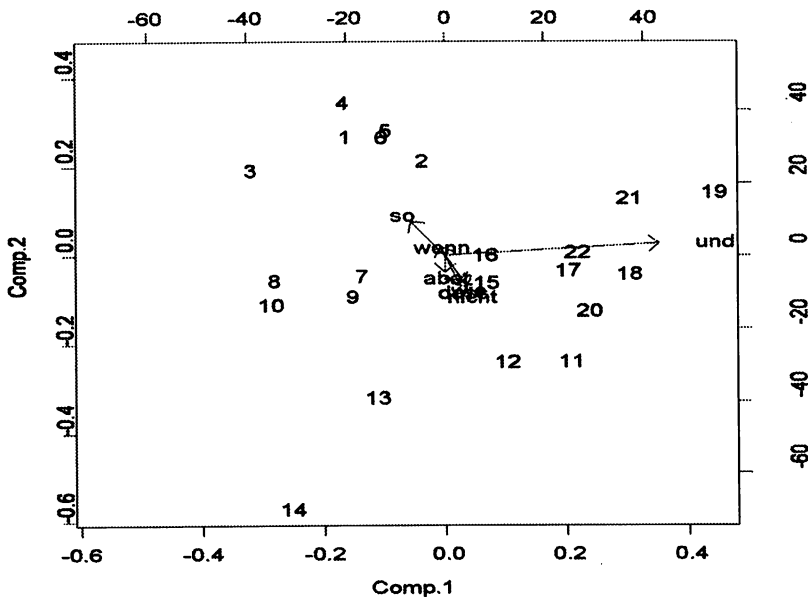


図 3

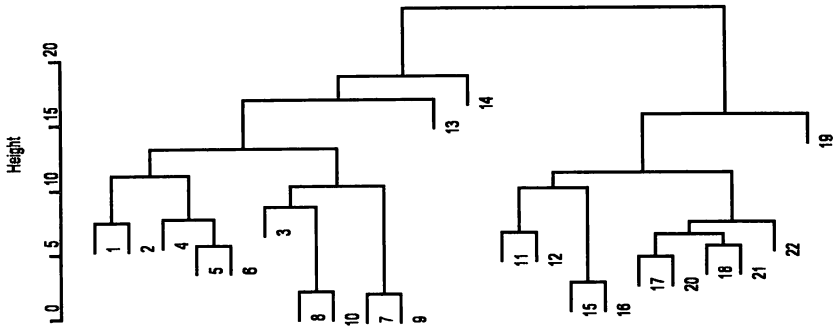


図 4

この結果は筆者にとっては以前のテーゼに固執する限り、いささか期待に反するものである。しかし文長分布の分析結果とはむしろ同じ傾向を示すといつてよい。横軸方向の第1主成分と縦軸方向の第2主成分の累積寄与率は計86.8%に達しているから、全分散の9割弱がこの二つの主成分で説明されたといえるが、中でも第1主成分は75%に近く、so以外はプラス値であるから、図では右方向のベクトルがそれを表している。なお同じデータにより、ユークリッド距離による凝集的・群平均法によるクラスター分析を行い樹形図で表すと、図4のようになる。ヴァディアン群の中では『新旧信仰論』の3が他とやや離れ、『カルストハンス』群に近い関係を示している。

1.5 数量化III類は、名義尺度によって測られる質的データをダミー変数に変換し量的操作の対象とするため考案された「林の数量化理論」のひとつで、ダミー変数(0, 1)を使った主成分分析の変種ともいえるものである。数値データでも次元の異なるものを同じ次元で扱いたい場合に、一定の基準を立て、その基準を満たすかどうかとして見れば名義尺度で測るのと同じことになるので、たとえば語の頻度情報、スペル特徴、文章の特徴などを総合的にみる方法として応用が可能である。但し元来が数値データであるものをわざわざダミー変数化する部分もふくむので、その意味では情報の劣化をまねく一面もあるが、ここではすでに利用した7語の頻度情報に加えて、スペルや使用度に特徴のある *sondern* 及び *wenn* の対応語と、名詞使用率、文長の正規分布性をとりあげ、ノンコンテクスチュアルな語の頻度情報だけを利用した主成分分析とは違った観点から『カルストハンス』と他のテキストの関係を分析してみる。

まず8機能語(*aber, dass, nicht, so, sondern, und, wenn, wie*)についてはその頻度について、中央値より小なる場合は0を、そうでない場合は1を与える。また *sunder*[n] というスペルが *sonder*[n] より多い場合は0、そうでない場合は1、仮定・条件の従属接続詞が現代とほぼ同様に *wenn* であらわされる場合は0、もっぱら *wan*[n] である場合は1、名詞率が中央値より小である場合は0、逆は1、そして自然対数による文長分布が5%水準で正規分布に従う場合は1、そうでない場合は0というダミー変数を与えることにする。その結果

表 4

	aber	dass	nicht	so	sond	und	wenn	wie	so/su-	wa/we-	名詞	文長
Ang1	0	0	0	0	0	0	1	0	1	1	1	0
Ang2	0	0	0	0	0	1	1	0	1	1	1	1
Ang3	1	0	0	1	0	0	0	0	1	1	1	1
Ang4	0	0	0	1	0	0	1	0	1	1	1	1
Wf1	1	1	0	1	1	1	0	0	1	1	1	1
Wf2	0	0	0	1	0	1	1	0	1	1	1	1
Kh1	0	0	1	1	0	0	1	1	0	1	0	0
Kh2	1	1	0	1	1	0	1	1	0	1	1	0
Kh1d	0	1	1	1	0	0	1	1	0	1	0	0
Kh2d	1	1	1	1	1	0	1	1	0	1	1	0
NKh1	1	0	1	0	1	1	0	1	0	1	0	0
Nkh2	1	0	1	0	1	1	0	1	0	1	0	0
Mur	1	1	0	0	1	1	0	0	0	1	0	1
Muent	1	1	1	0	0	0	0	1	1	1	0	1
Adel01	0	1	1	0	1	1	0	1	1	0	0	1

えられたダミー変数行列が表 4 で、この分析ではある程度現代ドイツ語風に標準化されたルター・テキスト群 (7 編) は文長分布や wenn の扱いが他とは異なるので、分析対象から外している。

このデータに基づき所定の手続きに従って数量化Ⅲ類の計算を行うと、変量に関する「カテゴリー数量」とテキスト群に関する「個体数量」を得るが、図 5 はそのうちの個体数量のみを散布図の形で視覚化したものである。これによれば、ヴァディアン群はすべて x 軸

Ang1	-1.796	-0.638
Ang2	-1.299	0.707
Ang3	-1.698	0.514
Ang4	-1.607	-0.36
Wolf1	-0.225	1.019
Wolf2	-1.265	0.227
KH1	0.349	-2.298
KH2	0.193	-0.647
KHd1	0.515	-1.863
KHd2	0.460	-0.831
NKH1	1.514	0.365
NKH2	1.514	0.365
Murner	0.681	1.709
Muentzer	0.513	0.191
Adel01	0.793	0.855

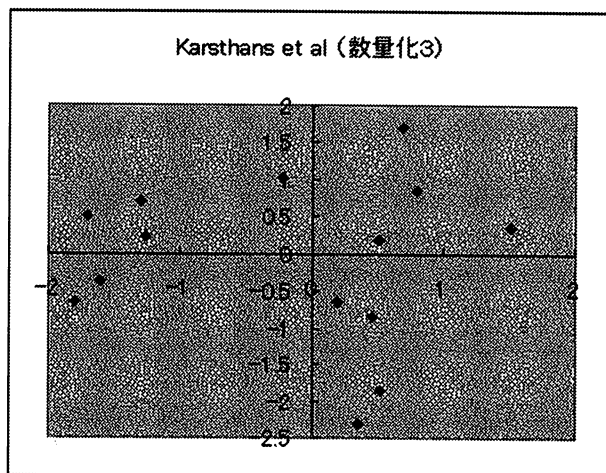


図 5

の負領域にあり、ブツァーの『新カルストハンス』2ファイルはまったく同一値をとってx軸でみて最も右側の位置において重なりあい、ムルナー、ルター、ミュンツァーが第1象限に散在するのに対し、『カルストハンス』群は1, 2がそれぞれ対応する相手と近接しながら第4象限に集まっている。すなわち数量化Ⅲ類による分析においても、ヴァディアン群と『カルストハンス』群は同一の集団に属するとはいえない結果となった。

以上、文長分布と語の出現頻度に関する分散分析、主成分分析、数量化Ⅲ類による多変量解析の結果をふまえて筆者は、かつての暫定的テーゼを撤回し、『カルストハンス』文書はヨアヒム・ヴァット、通称ヴァディアンの匿名作品と直ちに断定することはできない、という結論に至った。この結論の歯切れの悪さは、ひとつには『カルストハンス』が対話編の形をとるのに対し、『豺狼歌』と『新旧信仰論』は論議書の体裁をとるフルークシュリフトであり、ジャンル差の影響を測りきれないからである。とはいえ後者2点の作者は共にヴァディアンである、とするE・キュッケのテーゼは本稿によっても裏打ちされたものとみることができよう。テキストそのものがなかなか入手しがたいのだが、今後もヴァディアンはじめ匿名作者の候補に名があがっているもののみならず、その周辺にも探索的研究対象を広げつつ、方法論的検討を重ねて、非母語話者としての立場からドイツ語ドイツ文学研究における貢献の道をさぐっていきたいと考える。

Zur Anwendung statistischer Methoden auf philologische Probleme—am Beispiel von *Karsthans*

ARAI Hiroshi

Die Verfasserschaft der um die Jahreswende 1520/21 in Straßburg anonym erschienenen Flugschrift *Karsthans*, einer der populärsten unter ca. 2000 Stücken, ist bis heute nicht überzeugend ermittelt worden, obwohl Joachim von Watt (Vadianus, 1484–1551) als verborgener Verfasser öfter genannt wird. In diesem kleinen Aufsatz habe ich versucht, die Vadian-Hypothese, die ich selbst vor etwa fünfzehn Jahren aufgrund einer einfachen Analyse der Wortlänge-Verteilung unterstützte, von einem anderen Standpunkt aus und mit verschiedenen Methoden statistisch zu überprüfen.

Die Satzlänge, die im allgemeinen die Tendenz hat, der lognormalen Verteilung zu folgen, wird mit drei Methoden analysiert und visualisiert, nämlich als Boxplot, Serien-Bild der gleitenden Durchschnitte und Tafel der ANOVA (Abb. 1,2 und Tab. 1,2). Daß *Karsthans* de facto aus zwei Teilen besteht, kann ein sorgfältiger Leser ohne weiteres erkennen, und die Analyse der Sätze belegt noch dazu, daß die Vadian-Hypothese nicht zwingend ist, während die beiden Vadian-Werke (*Das Wolffgesang* sowie *Vom alten und neuen Gott, Glauben und Lehre*) offensichtlich derselben Population

angehören müssen.

Die Häufigkeit der „nonkontextualen“ Wörter als Daten zugrunde legend, habe ich als nächstes Verfahren die Hauptkomponenten-Analyse (PCA) und dann, einige andre stilistische Elemente in die Betrachtung einbeziehend, die sog. Hayashi-Quantifikationstheorie III benutzt. Als Kontrollgruppe habe ich Samples aus den Werken von Butzer, Murner, Müntzer und Luther vorbereitet (s. Anmerkung 4). Die Tab. 3 zeigt die Daten (relative Häufigkeit pro 1000 Wörter) für die PCA, die Abb. 3 (Biplot) und 4 (Kluster-Baum) zeigen die Ergebnisse : die *Karsthans*-Gruppe (7-10) steht in der Abb. 3 unabhängig von der Vadian-Gruppe (1-6), in der Abb. 4 jedoch relativ näher zur Vadian-Gruppe als andere. Das Ergebnis der Quantifikationstheorie III, deren Daten als Dummy-Variablen in der Tab. 4 gezeigt werden, ist insgesamt ähnlich : in der Abb. 5 stehen alle Samples von *Karsthans* im 4. Quadranten, während sich Samples der Vadian-Texte im 2. und 3. und die anderen im 1. Quadranten sammeln.

Dieser Beitrag präsentiert nur Zwischenresultate meiner laufenden stilstatistischen Untersuchungen und erhebt keinen Anspruch auf Vollständigkeit. Wie aber die diesmal erhaltenen Ergebnisse belegen, kann man *Karsthans* nicht einfach dem Autor von *Wolffgesang* etc. zuschreiben.