Hitotsubashi Journal of Arts and Sciences 42 (2001), pp.19-29.

DAS EIGENTÜMLICHE IM STIL VON KLEISTS *ERZÄHLUNGEN* EIN STILSTATISTISCHER VERSUCH

HIROSHI ARAI

Zunächst möchte ich kurz zusammenfassen, was ich durch stilstatistische Untersuchungen der japanischen, deutschen und englischen Prosa feststellen konnte.[1]

(1) Die Satzlänge hat die Tendenz, im allgemeinen lognormaler Verteilung zu folgen, d.h. die Längen der Sätze sind in den meisten Fällen $\Lambda(\mu, \sigma^2)$-verteilt. Diese Tendenz wird meistens noch klarer, wenn man sie als gleitende Durchschnitte beobachtet.

(2) Die Frequenz-Verteilung der Wörter (Erscheinungsform) kann man benutzen, um die Homogenität verschiedener Texte (Stichproben) zu prüfen oder u.U. das Autorschaftsproblem zu lösen. Dabei können die sog. „noncontextual" Wörter oder „function words",[2] vor allem Partikeln wie Konjunktionen, Präpositionen, Adverbien

---

[1] *Stilstatistische Untersuchungen des Goethe-Schiller-Briefwechsels (1794-98) und des "Lalebuchs" (1597) mit der Methode der multivariaten Analysen*, Arai,H., [Bericht der KAKEN 1998-2000, japanisch]. Da dieser Bericht schwer zu erhalten ist, möchte ich hier das Resümee zitieren : This research has two purposes: (1) to provide a comparison of some statistical, in particular multivariate analysis methods applied to philological objects and estimate their efficiency; (2) using most effective techniques, to solve the authorship question of the ***Lalebuch***, published anonymously in 1597.

The representative philological objects for methodological experiments is the collected letters of Goethe and Schiller corresponded in the years 1794-98, in which the greatest German classical authors exchanged frankly and friendly their critical opinions on the same themes and topics concerning literature and thought. The same genre, almost the same form and theme, two quite different personalities of the same ability-rank etc, are best conditions for methodological examination of style-statistical approach.

After an investigation of sentence-length finding that the sentence-length follows the log-normal distribution, I have selected 20 non-contextual *function words* for multivariate analysis : the count of these words are the variables used for discriminant analysis, principal component analysis and cluster analysis of G-S correspondence (divided into 25 Goethe- and 26 Schiller-samples). PCA with component-score plots shows on the one side G-groups and on the other side S-groups; one can distinguish one group from the other more effectivly with the help of the result from discriminant analysis.

The conclusions about the authorship problem of the ***Lalebuch*** have rather some similarity to Honneger's hypothesis that J.Fischart wrote this satirical masterpiece of *Volksbuch* (folktales), but the reasons for the conclusions are quite different.

He regards the one year later appeared and more proliferous clone book ***Schildbürger*** for more authoritative than ***Lalebuch*** and assumes further the existence of an unfound original print.

For lack of evidence and persuasive power these theses are not to support.

The statistical analysis showed above all that the first and the second part of the tale are not homogeneous, so I have decided to devide the text-data of ***Lalebuch*** into two samples (L1,L2) and to compare with samples of Fischart's works, namely ***Eulenspiegel Reimenweiß, Geschichtklitterung*** etc. After the comparative analysis and statistical tests of the sentence-length distribution, of the binomial predicates in subordinate clauses, and of the first symbol of words, the component-score plot of the PCA (25 variables) showed most evidently that L1 is surrounded by Fischart's groupe samples, while L2 stands quite apart from others. On the basis of the mentioned data, Fischart is extremely likely to have written the first part of the disputed ***Lalebuch***, but the second part is supposedly left by him as unfinished story which some one posthumously revised and published.

[2] *Inference in an Authorship Problem*, Mosteller,F/Wallace,D.L., in *Journal of the American Statistical Association*, June 1963, p.280; dazu *Applied Baysian and Classical Inference — The Case of "The Federalist Papers"*, Mosteller/Wallace, ²1984.

u.s.w. größere Rolle spielen als die grammatisch und textinhaltlich wichtigeren Wörter wie Substantive und Verben. Von den multivariaten Analysen sind besonders die Hauptkomponenten-Analyse und Diskriminanzanalyse erfolgversprechend. Die Testverfahren des X^2-Homogenitätstest bezüglich der Wortlänge-Häufigkeit der Stichproben[3] ist dagegen nicht immer effektiv. Sie scheint vor allem gegen den Genre-Unterschied sehr empfindlich zu sein.

In diesem Aufsatz handelt es sich um die Analyse der acht Novellen oder „Erzählungen", (wie sie der Autor selbst genannt) von Heinrich Kleist, der 1811 im Alter von 34 Jahren am Wannsee Doppelselbstmord beging. Fast zufälligerweise habe ich die Satzlänge-Verteilung von ***Michael Kohlhaas*** observiert, nachdem ich im vorigen Jahr die merkwürdige Abweichung von der Lognormalität im Stil der letzten Lebensjahre von AKUTAGAWA Ryunosuke, der sich auch, nebenbei bemerkt, „in der Mitte des Lebenslaufs" das Leben nahm, bemerkt und im Vergleich zum Stil von DAZAI Osamu, dem wohl bekanntesten Doppelselbstmörder in der neueren japanischen Literaturgeschichte, betrachtet hatte. ***Kohlhaas*** mit seinem unbestechlichen Rechtsgefühl ist mir schon seit meiner Studentenzeit ein fesselndes Werk gewesen, und was die zufällig angestellte statistische Messung zeigte, war zu interessant, um es beiseite zu legen. Neben den Kleist-Erzählungen sind also über 20 Novellen und Geschichten von fünf repräsentativen Autoren kontrastiv beobachtet und statistisch analysiert worden.[4] Das Ergebnis soll im folgenden mit Hilfe der Graphik visuell dargestellt werden.

---

[3] Vgl. *Mark Twain and the Quintus Curtius Snodgrass Letters : A statistical Test of Authorship*, Brinegar,C.S., in *Journal of the American Statistical Association*, March 1963.

[4] (1) *Die Leiden des jungen Werther* (Goethe; 2.Fassung)
(2) Die Geschichte der Eltern Mignions (Goethe; aus *Wilhelm Meisters Lehrjahre*)
(3) aus *Die Wahlverwandtschaften* (Goethe)
(4) aus *Die Wahlverwandtschaften* (Goethe)
(5) *Sanct-Rochus-Fest zu Bingen* (Goethe)
(6) *Die Belagerung von Mainz* (Goethe)
(7) *Die Novelle* (Goethe)
(8) *Eine großmütige Handlung aus der neuesten Geschichte* (Schiller)
(9) *Spiel des Schicksals* (Schiller)
(10) *Der Verbrecher aus verlorener Ehre* (Schiller)
(11) *Merkwürdiges Beispiel einer weiblichen Rache* (Schiller)
(12) *Der Geisterseher* (Schiller; 1.Teil)
(13) *Das Bettelweib von Locarno* (Kleist)
(14) *Die heilige Cäcilie oder die Gewalt der Musik* (Kleist)
(15) *Der Findling* (Kleist)
(16) *Das Erdbeben in Chili* (Kleist)
(17) *Der Zweikampf* (Kleist)
(18) *Die Verlobung in St.Domingo* (Kleist)
(19) *Die Marquise von O.* (Kleist)
(20) *Michael Kohlhaas* (Kleist)
(21) *Der Sandmann* (Hoffmann)
(22) *Das steinerne Herz* (Hoffmann)
(23) *Das Fräulein von Scuderi* (Hoffmann)
(24) *Eine Nacht im Jägerhaus* (Hebbel)
(25) *Anna* (Hebbel)
(26) *Kinder* (Hebbel)
(27) *Der Vater* (Kafka)
(28) *Das Urteil* (Kafka)
(29) *Die Verwandlung* (Kafka)

T01

	Satzlänge		Observiert (reelle Zahl)				Logarithmen			
author	title	n	mean	St.D.	KS-Test	Lilliefo.	mean	StD.	KS-Test	Lilliefo.
G1	Werther	705	19.89	15.31	0	0	2.68	0.84	0.005	0
G2	MigElt	129	28.08	18.49	0.012	0	3.15	0.61	0.920	0.20*
G3	Wahlv1	832	22.98	15.50	0	0	2.93	0.66	0.079	0.001
G4	Wahlv2	871	21.11	13.85	0	0	2.85	0.63	0.044	0
G5	StRoc	370	22.70	14.34	0	0	2.95	0.58	0.124	0.002
G6	Belage	414	27.06	15.57	0.023	0	3.11	0.66	0.002	0
G7	Novelle	239	29.79	18.79	0.005	0	3.17	0.75	0.004	0
S1	Grossm	70	14.39	10.44	0	0	2.43	0.72	0.438	0.059
S2	Spiel	131	29.90	14.74	0.178	0.005	3.28	0.51	0.466	0.073
S3	Verbre	450	16.77	11.44	0	0	2.58	0.73	0.001	0
S4	WeibRa	765	15.68	11.57	0	0	2.47	0.80	0.003	0
S5	Geister	1132	16.77	12.54	0	0	2.54	0.80	0.001	0
Kl1	Locar	20	48.00	18.13	0.888	0.20*	3.79	0.42	0.844	0.20*
Kl2	Caecilie	71	60.10	25.13	0.612	0.20*	3.97	0.58	0.089	0.001
Kl3	Findling	136	43.80	20.82	0.924	0.20*	3.62	0.65	0.006	0
Kl4	Chili	164	32.74	17.59	0.142	0.003	3.33	0.61	0.074	0
Kl5	2Kamp	219	52.37	32.02	0.221	0.009	3.72	0.78	0.03	0
Kl6	Doming	376	35.19	23.29	0.003	0	3.33	0.75	0.006	0
Kl7	Marquis	598	24.78	18.99	0	0	2.96	0.73	0.050	0
Kl8	Kohlhs	725	47.18	31.42	0	0	3.58	0.83	0	0
Ho1	Sandm	527	23.18	14.91	0.001	0	2.91	0.73	0.002	0
Ho2	Herz	339	25.41	16.38	0.001	0	3.03	0.67	0.336	0.032
Ho3	Scuderi	1039	22.53	13.61	0	0	2.92	0.67	0	0
He1	Nacht	197	17.82	12.79	0.001	0	2.65	0.70	0.354	0.035
He2	Anna	86	26.03	19.53	0.088	0.001	2.99	0.82	0.791	0.20*
He3	Kinder	623	23.03	17.77	0	0	2.88	0.74	0.145	0.003
Ka1	Vater	81	29.37	22.16	0.101	0.001	3.15	0.50	0.330	0.355
Ka2	Urteil	222	18.01	16.16	0	0	2.65	0.70	0.526	0.200
Ka3	Verwan	689	27.79	20.28	0	0	3.07	0.75	0.003	0

1.1 Definition von „Wort" und „Satz"

Ein „Wort" ist — nach W.Fuchs — definiert als jede Buchstabengruppe des Textes, die durch einen Zwischenraum (*space*) von der nächsten getrennt ist. Eine getrennte Vorsilbe z.B. wird hier als ein „Wort" betrachtet, ein zusammengefügter zu-Infinitiv dagegen wird auch als ein „Wort" betrachtet, wenn es sich um Satzlänge handelt.

Ein „Satz" wird nach der Anzahl der Wörter gemessen, und die Satzlänge wird zuerst durch diese Anzahl (ganze Zahl), dann logarithmiert angegeben. Die Satzgrenze wird prinzipiell durch den Punkt, das Fragezeichen und Ausrufezeichen bestimmt, solange sich der Sinn/Inhalt damit vollendet. Vor allem bei der direkten Rede aber ist die Satzgrenze nicht eindeutig klar, so daß das mit AWK-script entwickelte Meßverfahren manchmal durch die „manuelle Methode" kontrolliert werden muß. Der erste Satz der direkten Rede wird dabei prinzipiell mit dem Satz vor dem Kolon zusammen als eine Periode gerechnet. Hier zwei Beispiele :

1) Babekan stand auf und sagte, mit einem Ausdruck von Unruhe, indem sie den Zettel in den Wandschrank legte : „Herr, wir müssen Euch bitten, Euch sogleich in Euer Schlafzimmer

zurückzuverfügen." [Satzlänge 29]

2) „Es wird alles besorgt werden", fiel ihm die Alte ein, während, durch ihr Klopfen gerufen, der Bastardknabe, den wir schon kennen, hereinkam; und damit befahl sie Toni, die, dem Fremden den Rücken zukehrend, vor den Spiegel getreten war, einen Korb mit Lebensmitteln, der in dem Winkel stand, aufzunehmen ; und Mutter, Tochter, der Fremde und der Knabe begaben sich in das Schlafzimmer hinauf. [62]

1.2 Verteilung der Satzlänge

Die Tabelle 01 (im folgenden T01) zeigt die Statistiken der Satzlänge von 29 Werken : die 3. Spalte (im folgenden C3) zeigt die Zahl der Sätze jedes Werks (Beobachtungs-Menge ; Größe der Stichprobe, wenn man ein Werk als Stichprobe aus der Grundgesamtheit oder Population betrachtet), die Spalten 4-7 betreffen die beobachteten absoluten Daten in reellen Zahlen, die Spalten 8-11 betreffen die logarithmierten Daten. Die Spalten 4 (C4) und 8 (C8) zeigen das arithmetische Mittel, während die Spalten 5 (C5) und 9 (C9) die Standardabweichung. KS-Test (C6 & C10) bedeutet asymptotisches Signifikanzniveau als Ergebnis des Kolmogorov-Smirnov-Tests der Normalverteilung, Lilliefo. (C7 & C11) ebenfalls das des strengeren Kolmogorov-Smirnov-Lilliefors-Tests (beide durch SPSS gerechnet). Falls die angegebene Zahl in C6, C7, C10, C11 null ist, soll die Null-Hypothese der Normalverteilung abgelehnt werden.

Unter allen Titeln gibt es nur zwei, deren C6, C7, C10, C11 alle null angeben : nämlich *Kohlhaas* von Kleist und *Scuderi* von Hoffmann. Dieses Stück aber, wenn man das sog. moving-average (gleitender Durchschnitt, Intervall 20) errechnet, folgt ganz der

T02 (observed)

author	Goethe1	Goethe2	Schiller	Kleist	Hoffmann	Hebbel	Kafka
n	3598	2893	2548	2121	1905	906	992
mean	23.05	23.82	17.06	39.512	23.221	22.235	25.729
SD	15.61	15.59	12.536	28.365	14.532	17.169	19.648
KS	0	0	0	0	0	0	0
KS/L	0	0	0	0	0	0	0
			(observed -> moving average)				
mean	23.03	23.76	17.022	39.538	23.214	22.267	25.599
SD	5.78	5.67	5.974	17.090	4.893	5.805	8.950
KS	0.014	0.101	0	0	0.166	0	0
KS/L	0	0.002	0	0	0.006	0	0
			(observed -> logarithmiert)				
author	Goethe1	Goethe2	Schiller	Kleist	Hoffmann	Hebbel	Kafka
n	3598	2893	2548	2121	1905	906	992
mean(log)	2.91	2.970	2.555	3.391	2.936	2.839	2.980
SD(log)	0.7065	0.66	0.804	0.822	0.689	0.743	0.759
KS(log)	0	0	0	0	0	0.076	0.005
KS/L(log)	0	0	0	0	0	0.001	0
			(observed -> logarithmiert -> moving average)				
mean(log)	2.91	2.97	2.552	3.39	2.936	2.839	2.976
SD(log)	0.2736	0.24	0.380	0.49	0.229	0.258	0.375
KS(log)	0.009	0.365	0.233	0	0.036	0.312	0.027
KS/L(log)	0	0.049	0.015	0	0	0.029	0

Normalverteilung, während bei jenem auch alle null angeben. Unter Kleists Stücken findet man noch drei Ausnahmen, nähmlich *Findling, Erdbeben in Chili* und *Zweikampf* ; bei ihnen sind die Angaben von C6 und C7 größer als die von C10 und C11, d.h. die Verteilung der beobachteten Daten in reellen Zahlen (eigentlich ganze Zahlen) paßt sich besser der Normalverteilung an als die in die Logarithmen transformierten Daten, was selten geschieht.

Die Tabelle 02 (T02) zeigt die entsprechenden Daten nach dem Autor (wegen Raummangels senkrecht): wir setzen die beobachteten Meßreihen der Satzlänge des einzelnen Werks je nach dem Autor zusammen und errechnen erneut das Mittel usw.

Goeth1 behandelt die Gesamtheit einschließlich *Werther*, Goeth2 ohne *Werther*, denn der Stil des *Werther* ist allzu deutlich verschieden von den anderen, also statistisch als „outlier" (Ausreißer) zu betrachten.

Der Größe nach gereiht ist die durchschnittliche Satzlänge: Kleist, Kafka, Goethe2, Hoffmann, Goethe1, Hebbel und Schiller. Verglichen mit den anderen bildet Schiller wesentlich kürzere Sätze, was ein unvermutetes Ergebnis ist. Aber es ist auch deutlich, daß Kleist wieder bei jedem KS- und KS/L-Test null zeigt. Die Satzlänge Kleists folgt also weder der Normalverteilung noch der Log-Normalverteilung, und zwar auch bei „smoothed data".[5] Dies visuell zu machen ist wohl das Histogramm am geeignetsten. Die Abbildung 01 zeigt kontrastiv je 3 Histogramme von Goethe (links) und Kleist.

Besonders charakteristisch und einzigartig ist das zweipolige Histogramm von Kleists moving-average (GD). Weder Schiller noch Kafka im ganzen zeigen etwas Ähnliches, geschweige denn Hoffmann und Hebbel. Im Großen und Ganzen bilden sie alle eine glockenförmige symmetrische Verteilung wie Goethe, obwohl wir unter den einzelnen Werken drei Ausnahmen finden, deren Verteilungen der gleitenden Durchschnitte auch zweipolige Histogramme bilden. Es sind Kafkas *Verwandlung*, Schillers *Verbrecher aus verlorener Ehre* und Kleists *Kohlhaas*. Die zwei letzten haben hinsichtlich des Themas und der Erzählungsstruktur etwas Gemeinsames, aber daß *Kohlhaas* als Erzählung bei weitem dynamischer, spannender und komplizierter als *Verbrecher* ist, können wir auch aus der Abb. 02 absehen. Abb.02 zeigt neben den Histogrammen der beiden die Schwankung und Wandlung der aufeinanderfolgenden Satzlänge-GD ; wir behandeln dabei die Datenreihe sozusagen als „time series".[6]

1.3 t-Test und Levene-Test

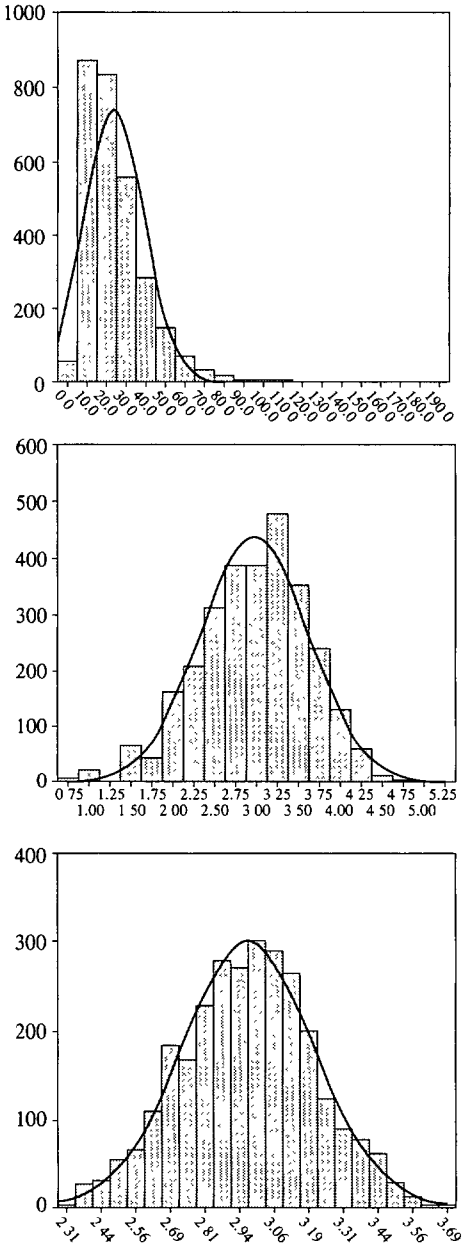
Im Fall Kleist können wir, wie oben gesehen, die Log-Normalität der Verteilung der Satzlänge nicht einfach voraussetzen. So wollen wir hier kurz das Ergebnis des robusten Levene-Tests berichten. Mit diesem Test kann man die Homogenität der Varianz der verschiedenen Werken oder Autoren prüfen. Unter Kleists Erzählungen ist wieder *Kohlhaas* etwas eigentümlich : er ist nur mit *Zweikampf* homogen, d.h. die Nullhypothese der Homogenität der Varianz wird nur bei dieser Kombination angenommen (*accepted*), sonst werden die Nullhypothesen alle abgelehnt (*rejected*) (Signifikanzniveau 0.001). Zum

[5] Die Satzreihe kann man in gewissem Sinne als "time series" betrachten. Moving average gehört zu den "smoothing"-Methoden und ist m.E. ziemlich effektiv, um die Tendenz der Satzreihe zu ermitteln.

[6] Bei der Methode der gleitenden Durchschnitte, die auf die Zeitreihe (time-series) angewendet wird, ist es wichtig, ein mit dem zugrunde liegenden Zyklus übereinstimmdes Zeitintervall zu finden. Hier ist das Intervall zunächst fixiert, und es wird versucht, die Entwicklungstendenz anschaulich zu machen.

ABB.01

Goethe (observed; logarithmiert; log-moving-average)



Kleist (observed; logarithmiert; log-moving-average)

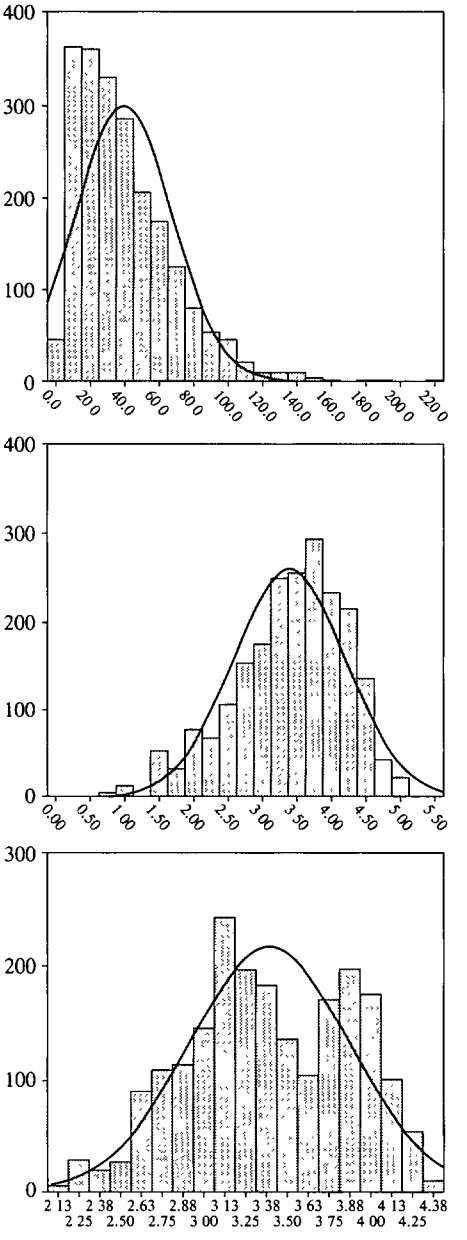
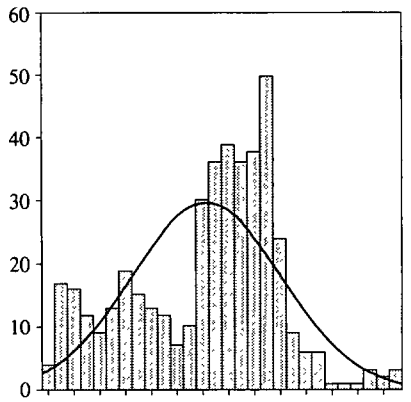
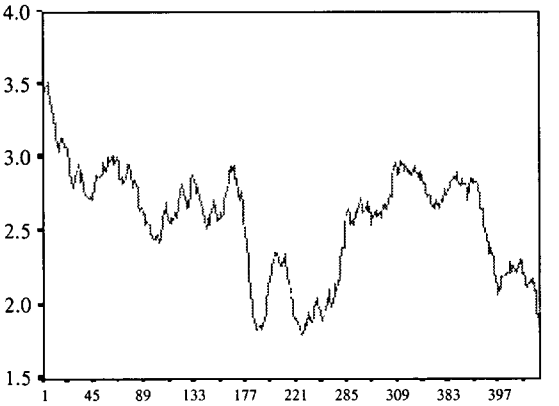


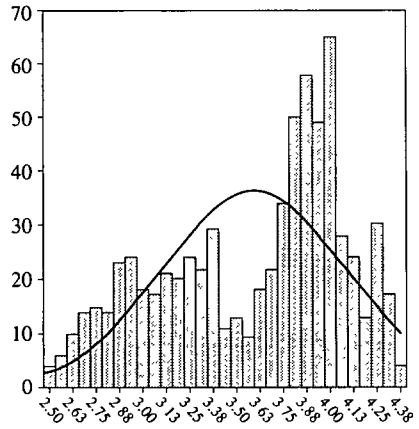
ABB.02

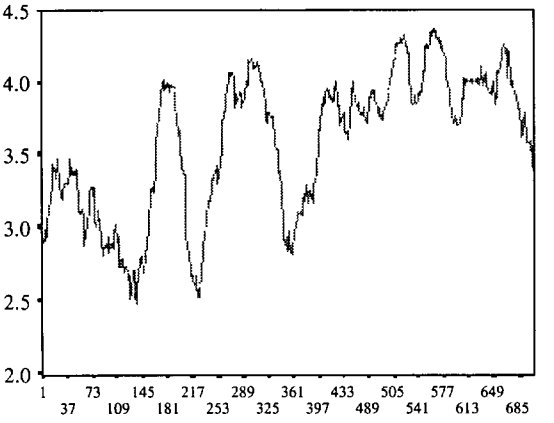
Schiller: *Der Verbrecher aus verlorener Ehre*





Kleist: *Michael Kohlhaas*





T03

Autor	Goethe	Schiller	Kleist	Hoffm.	Hebbel	Kafka
Goethe	****	*17.97*	*−22.7*	*3.86*	*2.75*	*−2.52*
Schiller	−6.38	****	*−35.4*	*−16.88*	*−9.55*	*−14.33*
Kleist	−7.23	−1.28	****	*42.30*	*−83.52*	*13.46*
Hoffmann	1.06	6.57	7.38	****	*3.28*	*−1.53*
Hebbel	−1.79	2.56	3.44	−2.39	****	*−4.07*
Kafka	−3.32	1.30	2.27	−3.83	−3.83	****

Vergleich : bei Goethe ist *Werther* ganz eigenartig: die Nullhypothese wird konsequent abgelehnt, bei Schiller ist die Nullhypothese der Homogenität der Varianz von *Verbrecher* und *Weibliche Rache* — übrigens freie Übersetzung aus Diderots Werken — abgelehnt.

Während der Levene-Test zur Prüfung der Varianz zu benutzen ist, wird mit dem t-Test die statistische Gleichheit des Durchschnitts geprüft, solange man ungefähr die Normalität der Verteilung voraussetzen kann.

Die Tabelle T03 zeigt im oberen Dreieck das Ergebnis des t-Tests zwischen den Autoren und im unteren Dreieck das des Levene-Tests. Der t-Test Kafka/Hoffmann ergibt die Nullhypothese als nicht abgelehnt, während die Nullhypothese des Levene-Tests im Fall Goethe/Hoffmann, Goethe/Hebbel, Schiller/Kleist und Schiller/Kafka auch nicht abgelehnt ist. Wir dürfen natürlich daraus nicht sofort folgern, daß z.B. die Durchschnitte von Kafka und Hoffmann gleich seien etc, aber eine Art Tendenz können wir daraus wohl ablesen. Gäbe es irgendeine Autoren-Kombination, bei der die Nullhypothese der beiden Teste zugleich nicht abgelehnt würde, müßte man entweder annehmen, daß sie zu derselben Grundgesamtheit (*population*) gehören, oder die Gültigkeit der Methode bezweifeln, wenn es sich wenigstens um die Verteilung der Satzlänge handelt. Hier gibt es keinen solchen bedenklichen Fall.

2. *Hauptkomponenten-Analyse*

Als Variablen sind folgende 15 „nonkontextuale" Wörter gewählt : *aber, als, da, da* (Konj. im folgenden *dA*), *daß, denn, doch, man, nicht, so, und, wenn, wie, zu* (Prä.), *zu* (Inf. im folgenden */zu/*). Die Tabelle 04 zeigt die relativen Frequenzen dieser Wörter beim Autor, die Tabelle 05 im Werk. Die absolute Frequenz ist vorher in die relative Frequenz per 1000 Wörter umgerechnet. In Kleists „Erzählungen" z.B. erscheint die Konjunktion *und* insgesamt 2980 mal unter 87 928, also umgerechnet 33.89 per 1000. Diese Summe 87 928 aber bedeutet nicht direkt das gesamte Vorkommen (i.e. die Gesamtmenge, die im Text erscheint), sondern die Anzahl der Wörter ausschließlich der in Eigennamen und Zahlwörtern. Die Kriterien der Auswahl der obengenannten Wörter sind 1) Häufigkeit, 2) Unabhängigkeit von dem Textinhalt (*nonkontextual*) und 3) relative Gemeinsamkeit innerhalb der Stücke Kleists.

Anhand dieser Daten kann man die Prozedur der Hauptkomponenten-Analyse so zusammfassen:[7]

T04

Autor	aber	als	da	dA	daß	denn	doch	man	nicht	so	und	wenn	wie	zu(Pr	zu(In
Goethe	4.28	6.29	2.84	1.11	6.74	7.54	2.56	7.51	11.41	8.80	35.53	4.01	5.58	4.01	15.10
Schiller	5.84	6.18	1.47	0.41	7.86	7.81	1.82	4.18	10.50	7.40	27.21	3.03	5.17	3.03	14.85
Kleist	1.79	6.00	1.25	1.75	9.76	10.79	1.80	2.80	6.65	4.54	33.89	2.06	3.94	2.06	19.40
Hoffmann	5.12	6.93	2.35	0.33	9.35	5.12	2.61	1.93	8.53	6.58	29.63	1.55	7.50	1.55	9.26
Kafka	9.29	6.42	2.87	0.52	7.89	4.31	3.35	3.43	14.71	5.38	30.62	3.59	5.50	3.59	11.92

[7] Hierzu Manly,B.F.J., Multivariate Statistical Methods, London 1986.

1) die Variablen standardisieren (z-transformieren), so daß jede einzelne Variable den Mittelwert 0 und die Varianz 1 hat.
2) die Varianz-Kovarianz-Matrix, d.h. Korrelationskoeffizient-Matrix, errechnen ;
3) die Eigenwerte und die denen entsprechenden Eigenvektoren suchen ;
4) durch das sog. Eigenwert-Kriterium die wichtigsten Faktoren bestimmen ; in diesem Fall trifft die kumulative Beitragsrate der drei Eigenwerte über 80%, so daß diese drei Faktoren allein betrachtet und die übrigen ignoriert werden können.
5) die z-transformierte Variablen-Matrix sowie die Korrelationskoeffizient-Matrix mit den Eigenvektoren multiplizieren ;
6) das Ergebnis in Punktdiagrammen graphisch darstellen.

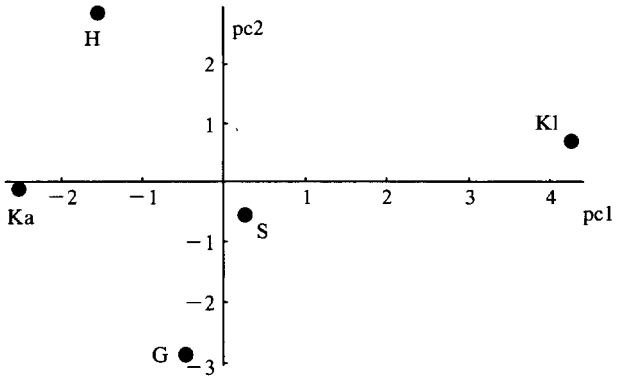
Die Abb.03 zeigt einen Teil der Ergebnisse der Hauptkomponenten-Analyse von fünf Autoren im Punktdiagramm (die erste und zweite Hauptkomponente). Man muß nicht hervorheben, daß Kleist deutlich den andern fernsteht.

Nun sind die Frequenzen der Wörter *daß, zu* (Präp. sowie Infinitiv) und *da* (Adverb sowie Konj.) bei Kleist am stärksten, während die der Wörter *aber, doch, denn, nicht, so, wenn*

T05

Werke	aber	als	da	dA	daß	denn	doch	man	nicht	so	und	wenn	wie	zu(Pr	zu(In
Werther	1.65	5.15	3.08	2.36	7.01	1.86	2.72	2.72	13.95	10.38	37.50	6.44	8.23	2.93	6.01
MigEltem	1.28	6.59	1.06	2.13	9.35	1.70	1.28	8.93	11.27	6.16	35.71	3.40	5.53	8.72	17.86
Wahlv-1	4.13	6.14	1.96	0.95	7.84	3.28	3.71	7.47	13.51	10.86	32.31	4.82	5.14	9.64	16.26
Wahlv-2	4.28	7.52	1.26	0.16	6.97	2.75	2.14	7.41	12.03	8.79	32.73	3.68	5.33	9.61	21.36
StRochus	6.30	6.54	1.21	1.45	5.69	4.00	2.42	10.66	7.15	8.84	37.07	2.42	4.00	7.51	9.81
Belagerg	5.20	5.29	1.09	0.73	5.75	2.83	1.46	11.86	7.39	5.66	39.51	2.19	3.74	7.94	16.24
Novelle	7.64	6.79	2.26	0.99	3.68	3.25	2.97	5.94	10.47	6.79	39.33	2.97	6.93	4.39	16.41
Verbrech	5.32	5.32	0.67	0.00	5.32	2.26	1.33	3.59	6.25	4.39	31.40	2.79	5.72	4.39	13.84
Weiblich	7.99	4.62	1.09	0.17	8.41	1.93	2.86	4.45	13.87	8.15	24.46	4.12	5.97	6.47	14.21
Geister	4.50	6.44	1.21	0.42	8.39	1.02	1.39	3.85	9.27	6.95	24.20	2.60	4.59	8.07	14.37
LocaCaeci	2.13	5.61	2.32	3.68	8.51	1.16	0.19	4.06	3.09	3.29	32.69	1.74	3.87	5.03	16.83
Findling	1.91	7.80	4.34	2.77	5.72	1.56	2.77	3.12	5.72	4.68	35.72	0.17	4.51	5.38	18.55
EB-Chili	0.59	10.16	2.54	1.95	6.64	0.59	3.52	3.71	7.23	5.28	36.73	2.15	4.88	4.30	10.94
2Kampf	2.43	6.11	2.25	1.17	5.93	0.81	0.90	1.98	5.39	3.59	33.79	0.54	4.76	5.21	16.00
Domingo	1.57	4.79	1.88	1.49	8.55	1.10	1.10	2.12	6.43	3.37	36.09	2.51	3.69	4.55	15.53
Marquise	1.44	6.17	3.77	0.21	13.50	1.58	3.36	2.88	9.11	5.48	31.44	3.36	5.07	25.83	36.99
Kohlhs	1.92	5.48	1.59	2.22	11.23	1.38	1.50	2.90	6.71	4.94	33.59	2.19	3.02	11.29	16.17
Sandmann	6.79	7.63	2.29	0.25	10.10	3.90	2.54	2.04	8.06	8.23	35.04	2.80	8.74	3.05	8.14
Steinherz	6.96	5.52	2.28	0.60	6.84	2.76	2.16	2.40	8.17	5.76	30.62	0.72	7.56	2.76	5.64
Scuderi	3.56	7.09	2.01	0.27	9.89	1.38	2.81	1.69	8.91	6.02	26.42	1.20	6.82	7.09	11.18
Vat-Urteil	10.07	4.47	1.44	0.16	10.07	1.44	4.79	3.20	17.42	7.03	23.17	4.63	5.91	4.95	6.87
Verwandl	9.03	7.07	2.50	0.64	7.17	3.35	2.87	3.51	13.81	4.83	33.10	3.24	5.37	4.09	13.60
AnNacht	4.57	5.27	3.51	0.53	3.86	1.93	2.99	1.93	10.18	4.92	34.77	2.28	7.37	5.97	7.90
Kinder	5.76	8.18	2.78	0.07	7.04	1.57	2.78	0.57	10.82	4.70	34.94	3.34	5.55	6.40	9.18
BasNacht	7.16	5.77	2.68	0.60	7.85	1.89	1.09	1.29	5.86	5.57	47.71	1.89	6.06	6.06	12.62

ABB.03



und *als* am geringsten unter den fünf Autoren. Bei Goethe finden sich die Wörter *man, so, und, wenn* am häufigsten, bei Kafka die Modalpartikeln *aber, doch, denn* und *nicht*, bei Hoffmann *als* und *wie*. Die weiteste Ferne, die zwischen Kleist und Kafka, scheint den Gegensatz der „staccatoartigen" Stil- und Satzbauelemente und der modalen Expressionselemente zu spiegeln.

Zum Schluß versuchen wir noch, bei den einzelnen Werken die Hauptkomponenten-Analyse durchzuführen. Die Grunddaten sind, wie erwähnt, in Tabelle 05 gezeigt. Nach der Prozedur ergeben sich die Hauptkomponenten in folgenden Formeln :

$$Z_1 = 0.347X_1 - 0.079X_2 - 0.127X_3 - 0.287X_4 - 0.029X_5 + 0.281X_6 + 0.325X_7 + 0.075X_8 + 0.422X_9 + 0.368X_{10} - 0.154X_{11} + 0.338X_{12} + 0.292X_{13} - 0.085X_{14} - 0.198X_{15}$$

$$Z_2 = -0.176X_1 - 0.077X_2 - 0.037X_3 - 0.074X_4 + 0.460X_5 - 0.135X_6 + 0.136X_7 + 0.084X_8 + 0.173X_9 + 0.104X_{10} - 0.255X_{11} + 0.223X_{12} - 0.204X_{13} + 0.543X_{14} + 0.456X_{15}$$

ABB.04

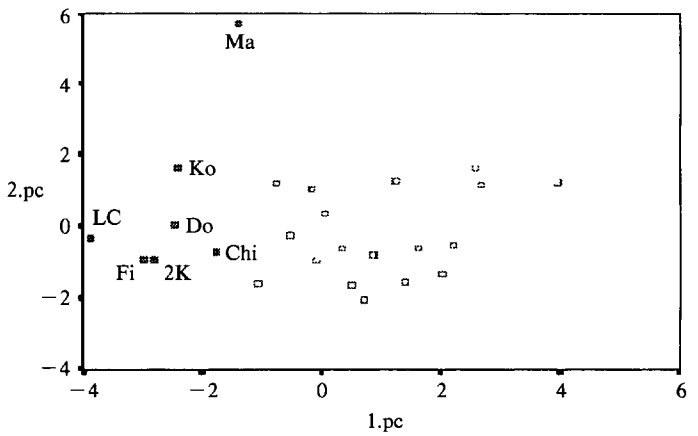
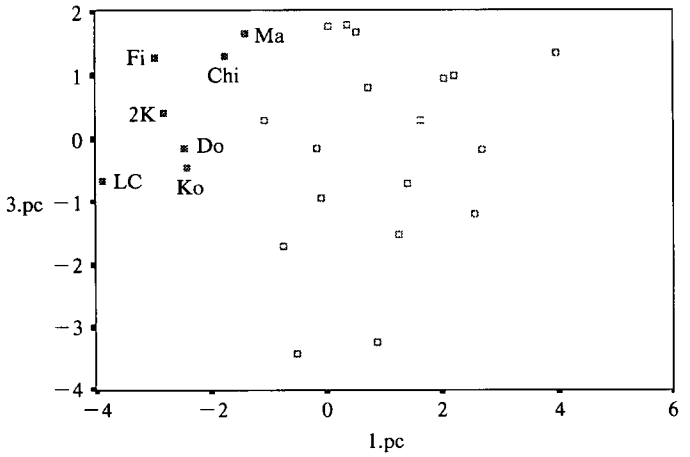


ABB.05



$Z_3 = -0.038X_1 + 0.140X_2 + 0.463X_3 - 0.098X_4 + 0.167X_5 - 0.312X_6 + 0.281X_7 - 0.613X_8 + 0.092X_9 - 0.162X_{10} - 0.160X_{11} - 0.049X_{12} + 0.318X_{13} - 0.017X_{14} - 0.102X_{15}$

Die Diagramme der Abb.04 und Abb.05 zeigen wieder deutlich, wie die Erzählungen Kleists doch zusammen eine Gruppe bilden und sich nicht mit den andern mischen, obwohl *Marquise O* in Abb.04 etwas entfernt steht.

Schon beim ersten Lesen der Kleistschen Erzählungen nehmen wir etwas Eigentümlich-Charakteristisches, sei es ästhetisch, sei es sprachlich, wahr. Dies ist hier stilstatistisch analysiert und untermauert worden.[8]

HITOTSUBASHI UNIVERSITY

[8] Meinem Kollegen und Freund Rainer Habermeier danke ich für seine wertvollen Verbesserungsvorschläge.