

Erweiterung des Instrumentariums der Sozialforschung durch inhaltsanalytische Techniken

Stone, Philip J.; Mochmann, Ekkehard

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Stone, P. J., & Mochmann, E. (1976). Erweiterung des Instrumentariums der Sozialforschung durch inhaltsanalytische Techniken. In M. R. Lepsius (Hrsg.), *Zwischenbilanz der Soziologie: Verhandlungen des 17. Deutschen Soziologentages* (S. 163-174). Stuttgart: Ferdinand Enke. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-187736>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Erweiterung des Instrumentariums der Sozialforschung durch inhaltsanalytische Techniken

Philip J. Stone und Ekkehard Mochmann

In diesem Beitrag werden wir das gestellte Thema unter zwei Aspekten behandeln. Der erste Teil des Beitrages wird eine Übersicht über die Herausforderung der Inhaltsanalyse durch neue sozialwissenschaftliche Forschungsansätze geben. Im zweiten Teil beschreiben wir Inhaltsanalyseinstrumente und Analysemöglichkeiten, die in jüngster Zeit verfügbar geworden sind oder gerade entwickelt werden. Dabei soll auch kurz auf eine Gegenüberstellung der dominierenden Verfahren der automatisierten Inhaltsanalyse eingegangen werden, so wie sie zuletzt im Rahmen des International Workshop on Content Analysis 1974 in Pisa (1) vorgestellt und diskutiert wurden.

I. Die Entwicklung sozialer Indikatoren als Herausforderung an die Inhaltsanalyse

In den letzten Jahren hat das Interesse an der Entwicklung "sozialer Indikatoren" beachtlich zugenommen. Dies ist vergleichbar der Entwicklung ökonomischer Indikatoren, die bereits seit Jahrzehnten gesammelt werden. Viele Forschungsvorschläge auf diesem Gebiet waren jedoch relativ wenig durchdacht, was auch bereits zutreffend von Klages (2) kritisiert wurde. Auch wir wollen uns hier nicht vornehmen, einen neuen Vorschlag vorzustellen. Dennoch bleibt es entmutigend, festzustellen, daß soziale Indikatoren bisher kaum über den Bereich ökonomischer Indikatoren hinaus zu Indexbildungen vorstoßen, die viel besser verdeutlichen könnten, wie sich in verschiedenen gesellschaftlichen Sektoren die Problemstellungen, Wertorientierungen und Weltperspektiven geändert haben.

1. Soziale Indikatoren aus Massenkommunikationsmedien

Massenkommunikation ist heute ein überaus bedeutender Bestandteil des Netzes gesellschaftlicher Integrationsprozesse. Wandlungen der Massenkommunikationsinhalte reflektieren einerseits Veränderungen der Gesellschaft und produzieren andererseits neue Anschauungen. Gerade in den letzten Jahren sind parallel zu den Massenkommunikationsmedien in größerem Umfang auch Publikationen entstanden, die sich an ein bestimmtes Publikum richten. Dies ist zum Teil eine ökonomische Antwort auf den Bedarf an gezielter Werbung auf bestimmten Verbrauchermärkten; in mancher Hinsicht reflektiert dies aber auch die vielfältige Spezialisierung in einer pluralistischen Gesellschaft, in der die Entwirrung verschiedener gesellschaftlicher Komponenten eine der größten Herausforderungen darstellt (3).

Damit werden Archive der Massenkommunikation wie auch Archive für speziell ausgerichtete Kommunikationsmedien eine wertvolle Quelle für die Gewinnung sozialer Indikatoren. Nach den Modellen von *Abelson* und *Bernstein* (4) ist die Person als Schnittpunkt verschiedener Anschauungen und Informationen defi-

niert, die eine Anzahl genereller und spezieller Interessen repräsentiert. Dem sind der Arbeitsbereich, die Freizeit und Erholungspausen, religiöse und seelische Bedürfnisse, politische Werte, Konsumpräferenzen und anderes zuzurechnen. Ein Massenkommunikationsarchiv ist dann besonders wertvoll, wenn gleichzeitig die Querverbindung zu zugehörigen Daten aus Leseranalysen gegeben ist und damit ein Bild gewonnen werden kann, wer welcher Information ausgesetzt ist.

Ein unmittelbares Problem für die Gewinnung sozialer Indikatoren aus der Massenkommunikation ist der immense Materialumfang. Wenn Informationen für verschiedene Sektoren in mehreren Regionen für bestimmte Zeitabschnitte gesammelt werden müssen, wird selbst bei einer sehr beschränkten Stichprobe vermutlich mehr als eine Milliarde Zeichen (characters) Text zu verarbeiten sein. Leider werden optische Lesegeräte, die geeignet wären, auch unterschiedliche, in einem Text gemischt auftauchende Druckzeichen zu lesen, bisher auf dem Markt nicht angeboten. Selbst wenn sie verfügbar wären, bliebe der Beweis anzutreten, daß wir bereits die entsprechenden Auswahlverfahren für ihren sinnvollen Einsatz kennen. Wenn die Massenkommunikation für die Gewinnung sozialer Indikatoren eine Rolle spielen soll, müssen Kriterien für die Auswahl und die Ziehung von Stichproben entwickelt werden.

Analyse von Tageszeitungen

Soziale Indikatoren aus Inhaltsanalysen können sich sowohl auf das völlige Fehlen als auch auf die übermäßige Häufigkeit bestimmter Inhalte stützen. *Johnson, Sears* und *McConahay* (5) analysierten zwei Zeitungen, die in Los Angeles erscheinen, bis zurück zum Jahre 1900, um Maßzahlen für die von ihnen so genannte "schwarze Unsichtbarkeit" (black invisibility) zu gewinnen. Sie weisen nach, daß der prozentuale Anteil an Raum, der der schwarzen Bevölkerung in den Zeitungen gewidmet wird, weit unter deren Anteil an der Einwohnerzahl von Los Angeles liegt und daß diese Rate sich zunehmend verschlechtert. Ein Schwarzer könnte die Los Angeles-Presse lesen und dabei den Eindruck gewinnen, daß es in Los Angeles überhaupt keine Schwarzen gibt. Erwartungsgemäß stieg die Berichterstattung mit den *Watts*-Unruhen an, nach *Watts* fiel die Berichterstattung jedoch wieder weit unter das erwartete Niveau zurück. Anscheinend war *Watts'* Lektion über die Bedeutung der Sichtbarkeit nicht deutlich und nachhaltig genug, obwohl auch *Martin Luther King* und andere darauf hingewiesen haben, daß Sichtbarkeit bzw. das Streben danach ein Hauptgrund dafür war, daß Schwarze ihre Wohnungen selbst in Brand steckten.

Analyse von Schlagzeilen

Deichsel u. Mitarb. (6) wollen die soziale Struktur von Symbolwelten untersuchen. Sie stützen ihre Analysen auf die Auswertung der Schlagzeilen verschiedener deutscher Tageszeitungen. Dabei konnten bereits unterschiedliche, spezifischen Schichten zurechenbare Darstellungen gleicher Sachverhalte festgestellt werden. In Pisa wurde eine internationale Kooperation auf diesem Gebiet vereinbart, so daß Aus-

schnitte der Weltpresse in diese Analysen einbezogen werden können (7). Es liegt nahe, diese Arbeiten auch für die Gewinnung sozialer Indikatoren zu nutzen.

2. Soziale Indikatoren aus Kleingruppendiskussionen

Aber auch außerhalb des Bereiches der Massenkommunikation können soziale Indikatoren gefunden werden. So hat z.B. erst kürzlich in Harvard *Aries* (8) den Inhalt von Kleingruppendiskussionen für drei verschiedene Gruppen – nur Männer, nur Frauen und eine gemischte Gruppe – analysiert. Welche Ergebnisse sind z.B. für gemischte Gruppen zu erwarten? Werden sie eher den reinen Männer- oder aber den Frauengruppen ähneln? *Aries* stellte fest, daß reine Frauengruppen sich intensiver um persönliche Belange kümmern, sich mehr dem Heim, Familienfragen und Beziehungen außerhalb der Gruppe widmen. Männergruppen hingegen diskutierten unpersönlicher, benutzen Geschichten und Metaphern, um eine gewisse Distanz zum Gesprächsgegenstand zu wahren, wobei häufig Konkurrenz motive und aggressive Inhalte anklingen. Gemischte Gruppen konzentrieren sich stärker auf gruppeninterne Prozesse und entwickeln ein stärkeres Gruppenbewußtsein. Der Gesprächsstil ist anspruchsvoller und defensiver.

Eine Reihe dieser Ergebnisse kann als Reflektion der bestehenden Normen gewertet werden, denen zufolge Frauen mehr auf die Vertraulichkeitssphäre und zwischenmenschliche Beziehungen hin sozialisiert werden, Männer hingegen mehr in Richtung Unabhängigkeit, Autonomie und Konkurrenz. Aber sollten sich die Änderungen der Geschlechterrollen in dem Maße, wie sie sich in der heutigen Gesellschaft entfalten, nicht auch in den Kleingruppendiskussionen widerspiegeln? Bilden nicht gerade Indices, die aus Familien- oder Kleingruppendiskussionen gewonnen werden, einen Gegenpol zu der Analyse der Massenmedien? Wieviel spiegelt sich von dem, was extensiv über den Wandel der Geschlechterrollen in den Massenmedien berichtet wird, in den Alltagsgesprächen tatsächlich wider? Indikatoren darüber, wie die Menschen überhaupt interagieren, worüber sie sich austauschen, sollten doch sicherlich mit zum Kern eines jeden Programms für soziale Indikatoren gehören. Wie *Aries* mit ihrer Dissertation demonstriert, liegt hier ein Hauptanwendungsgebiet für die Inhaltsanalyse.

3. Soziale Indikatoren aus der Vercodung offener Fragen

Der Sprache kommt bei der Entwicklung sozialwissenschaftlicher Forschungsinstrumente, insbesondere der Meßinstrumente, insofern eine Sonderstellung zu, als die Auseinandersetzung mit Sprachproblemen in diesem Zusammenhang häufig bewußt vermieden wurde. Vielmehr wurde die Entwicklung der Instrumente so gestaltet, daß nicht sprachliche Äußerungen, sondern codierte Werte das Ergebnis des Meßprozesses bildeten. Ein deutliches Beispiel hierfür ist die Entwicklung geschlossener Fragen mit Antwortenvorgaben. Dem Befragten wird die Aufgabe übertragen, seine Antworten, auch wenn ihm dies unter Umständen nur mit Mühe gelingt, in eine der durch den Forscher vorgegebenen Kategorien einzuordnen. Die

Vor- und Nachteile dieser Technik sind hinlänglich diskutiert. Ein Hauptargument für den Verzicht auf die ursprüngliche Antwort im Klartext war der mit der nachträglichen Codierung verbundene Aufwand, auch wenn dies mehr Freiheitsgrade für die Datenreduktion im Rahmen unterschiedlicher Auswertungsintentionen (z.B. Sekundäranalyse offener Fragen) ließ.

Sudman und *Frisbie* (9) haben unter Verwendung des *General Inquirer* und des *Harvard III Dictionary* die Vercodung der Antworten auf offene Fragen untersucht. Sie kamen dabei insgesamt zu positiven Ergebnissen, da die maschinellen Verfahren mit Ausnahme der Vercodung latenter Dimensionen gleiche Qualität hatten wie die Verschlüsselung durch Coder.

Zur Untersuchung der Forschungsfragen, ob und wie die Parteien in der BRD spontan mit Sachverhalten der Schichtstruktur verknüpft werden, ob Parteipräferenzen ideologisch konditioniert sind oder sachlich begründet werden, hat eine Kölner Forschungsgruppe (*H.D. Klingemann* u. Mitarb., (10) in den letzten Jahren diesen Ansatz wieder aufgegriffen. Antworten auf die offene Frage nach den guten und schlechten Seiten der Parteien wurden mit *Textpack* verschlüsselt, einem im Zentralarchiv speziell für die Verschlüsselung offener Fragen entwickelten Inhaltsanalysesystem, das nicht an eine bestimmte Nationalsprache gebunden ist. Die ersten Vergleiche der automatisierten Verschlüsselung mit konventioneller Verschlüsselung haben die positiven Erfahrungen von *Sudman* und *Frisbie* bestätigt. Trotz der Entwicklungskosten für das Programmpaket und die auf diese Forschungsfragen zugeschnittenen Wörterbücher zeichnet sich hier ein vielversprechender Einsatz dieses Vorgehens ab, zumal diese Frage nicht nur wiederholt in deutschen Wahlstudien, sondern auch in international vergleichenden Studien gestellt wurde. Damit dürfte ein Prototyp für gleichartige Forschungsansätze gegeben sein, die andernfalls wirtschaftlich nicht durchzusetzen wären und bei denen auch eine Umvercodung des Materials aus einer früheren Studie, z.B. in einer Zeitreihe nach neuen Gesichtspunkten, nicht mit so geringem Aufwand möglich wäre.

4. Soziale Indikatoren aus der Analyse historischer Dokumente

Schließlich können soziale Indikatoren aus Inhaltsanalysen auch historische Perspektiven eröffnen, die mit Umfragetechniken nicht zugänglich sind. Ein Beispiel hierfür ist *Namenwirths* Analyse amerikanischer Parteiprogramme in den letzten 150 Jahren (11). *Namenwirth* benutzt die theoretischen Konzepte von *Lasswell*, um den Wandel der Werte zu analysieren. Er fand eine stetige Progression in der Reihenfolge, in der die Betonung der Werte in einem Bereich die Betonungen in einem anderen ablöste, eine Entwicklung, die sich am besten durch eine Sinuskurve beschreiben läßt.

II. Neuere Entwicklungen auf dem Gebiet der Inhaltsanalyse

In der Vergangenheit hat sich der Wörterbuchsatz, so wie er im *General Inquirer* mit unterschiedlichsten Kategoriensystemen erprobt wurde, in der maschinellen Inhaltsanalyse durchgesetzt. Zunehmend gewinnen aber auch Verfahren der empi-

rischen Klassifikation an Bedeutung. Neben der Weiterentwicklung dieser Ansätze ist der maschinellen Umsetzung der Auswertungsprogramme mehr Aufmerksamkeit geschenkt worden. So wurde z.B. in einigen neu programmierten Systemen darauf geachtet, daß die Ausgabeformate der Verschlüsselungsergebnisse direkt für die bestehenden Datenanalysesysteme verwendet werden können. An der Universität Bielefeld wurden Programme für die Bedeutungsfeldanalyse und die Evaluative Assertion Analysis entwickelt (12).

1. Vergleich von Wörterbuchverfahren und empirischen Klassifikationsverfahren

Der Wörterbuchansatz geht davon aus, daß Wörter (types) in theoretisch relevanten Kategorien zusammengefaßt werden können, welche die theoretischen Interessen des Forschers repräsentieren. Probleme, die bisher aus der Homonymie von Wörtern resultieren, sind im *General Inquirer* inzwischen weitestgehend für die englische Sprache durch die Disambiguierungsroutinen, welche anhand des Kontextes die jeweilige Wortbedeutung bestimmen, gelöst (13). Die Kategorien können eine generelle Theorie, etwa die von *Parsons* oder *Lasswell*, repräsentieren, oder sie können mehr an einer Theorie mittlerer Reichweite im Sinne von *Merton* ausgerichtet sein, so wie *McClellands* "need achievement" oder *Ogilvies* Studien des Icarus Motivs.

Die Messung mit Hilfe eines Wörterbuchsystems kann sich entweder direkt auf die Häufigkeitsauswertung der Kategorienbesetzungen konzentrieren, oder aus dem gemeinsamen Auftreten bestimmter Kategorien kann auf Themen geschlossen werden, die noch nach syntaktischen Kriterien differenzierbar sind. So könnte ein Forscher z.B. daran interessiert sein, in welchem Ausmaß das Feindseligkeitsmotiv erwähnt wird, ein anderer könnte versuchen herauszufinden, wie oft dieses Motiv in Zusammenhang mit der Innenpolitik gebracht wird, und ein dritter schließlich möchte unterscheiden, ob Politiker als Initiatoren oder Ziele der Feindseligkeit dargestellt werden.

Einige Forscher, die mit Kategoriensystemen arbeiten, haben die Anschauung vertreten, daß es einen Basissatz von Kategorien gebe, der in verschiedenen Theorien immer wieder auftauche. Solche Kategorien wären z.B. bestimmten Kategorien im *Roget's* Thesaurus vergleichbar, wenngleich sie auch weniger scharf abgegrenzt sind. *Laffal* (14) z.B. behauptet, daß es einen fundamentalen Satz von Inhaltskategorien gebe, der von verschiedenen Forschern entdeckt würde, ganz gleich, welche theoretische Perspektive sie verfolgen. Demgegenüber gehen wir davon aus, daß Spezialwörterbücher Kategorien aus generellen Wörterbüchern übernehmen können und daß es bestimmte Konzepte gibt, die sowohl eine Adaption an verschiedene soziolinguistische Gemeinschaften als auch an bestimmte theoretische Interessen repräsentieren.

Nicht alle Inhaltsanalysekategorien haben den gleichen manifesten Gehalt wie z.B. die Standardkategorien im *Roget's* Thesaurus; häufig werden subtilere Maße entwickelt, um Schlüsse über den Stil und Motive der Autoren eines Textes zu ziehen. Zur Verdeutlichung wollen wir ein Beispiel einer Analyse deutscher Fachliteratur zitieren. *Ertel* (15) hat eine Skala für dogmatischen Stil entwickelt, mit der er die Schriften mehrerer deutscher Sozialwissenschaftler untersuchte. Die Skala wird auf der dogmatischen Seite von *Holzkamp* und auf der undogmatischen Seite von *Topitsch* begrenzt. *Marcuse* wird durch diese Analyse als weniger dogmatisch als

Habermas charakterisiert. Dabei fällt auf, daß die dogmatische Seite der Skala von den "dialektischen Autoren", die undogmatische von den "erfahrungswissenschaftlichen Autoren" belegt wird. Sicherlich kann die Gültigkeit der Meßergebnisse von *Ertef* bezweifelt werden, und einige Forscher haben dies auch bereits getan; die stärkste und konstruktivste Kritik wäre sicher ein anderes Dogmatismusmaß mit größerem Gültigkeitsanspruch.

Ein anderes interessantes Projekt wäre die Untersuchung der Genauigkeit der Fachterminologie. Insofern als das Vokabular der meisten vollentwickelten Wissenschaften durch hohe Präzision gekennzeichnet ist, sei es zur Bezeichnung botanischer Objekte oder physikalischer Prozesse, könnten wir höhere Erwartungen an solche soziologischen Theorien knüpfen, die für ihre Terminologie ein höheres Maß an Präzision aufweisen. Dies gilt speziell für die Theorien, die mehr zu den exakten Wissenschaften als zur Philosophie tendieren. Wie aber ist die Genauigkeit von Konzepten zu beurteilen? Ein Kriterium könnte die semantische Übersetzbarkeit sein. Ist ein Begriff, der aufgrund seiner spezifischen Bedeutung so in komplexe Assoziationen eingebettet ist, die ihrerseits an bestimmte kulturelle Sozialisierungsmuster geknüpft sind, für die Wissenschaft überhaupt noch nützlich? Damit soll natürlich nicht impliziert sein, daß präzise Begriffe nicht für die Beschreibung kultureller Eigenheiten verwendet werden können; Anthropologen beschreiben regelmäßig Verwandtschaftssysteme, die bisher vollkommen fremd waren. Eine Inhaltsanalyse der Präzision von Begriffen wird zu weitaus wertvolleren Ergebnissen führen, als nur zum Erfolg einer nützlichen Übung, und würde zu unserem besseren Verständnis der Sozialwissenschaften beitragen.

Vergangene Studien haben die Wichtigkeit der Identifizierung des soziolinguistischen Geltungsbereiches nachgewiesen. Je besser dieser umschrieben ist, desto mehr kann man auf die Gültigkeit der Schlüsse, die aus der Studie gezogen wurden, vertrauen. — So können z.B. Antworten auf die offene Frage in einem Persönlichkeitswahrnehmungstest ("Ich sehe . . . als . . .") oder der beschränktere Text von Zeitungsschlagzeilen leichter verarbeitet werden als umfangreiche Ausführungen mit unterschiedlicherer Syntax und wechselnden, aber oft nicht direkt zu identifizierenden Kontextverweisen. Mithin sollte ein analytisches System einig Wissen über den Sprachbereich, den es analysiert, haben, speziell im Hinblick auf spezifische Ausdrücke. Mitunter grenzt dies an "Weltverständnis". So wird das Entstehen von Problemen in Vietnam zu einem Beispiel für die Problemgenerierung in Südostasien, oder die Analyse von Kleingruppendiskussionen von Kollegstudenten spiegelt die Regionalsprache und die Bezugspunkte dieser Universität wider.

In einer internationalen, als "Inquirer Projekt" bekannt gewordenen Zusammenarbeit haben die Mitarbeiter aus Australien erst vor kurzem ein neues generelles englisches Inhaltsanalysewörterbuch mit 8700 Wortbedeutungen vorgelegt (16). Dabei wurde auf vergangene Erfahrungen sowie auf neu entwickelte Kontextroutinen zur Disambiguierung verschiedener Wortbedeutungen zurückgegriffen (s.o.). Auch hier wurde besondere Sorgfalt auf die Erstellung und Validierung von Kategorien auf relativ niedriger Inferenzstufe verwendet. Wörterbuchkategorien entsprechen den Variablen in Gleichungen. Daher sollten sie sauber definiert sein. Ist diese Bedingung erfüllt, dann ist auch mehr Spielraum für die Interpretation bei der Analyse, die über die Ebene der Kategorien auf die Ebene der Themen hinausgeht, geschaffen.

Der empirische Ansatz steht den a priori Kategoriensystemen skeptisch gegenüber. Einwände sind nicht nur, daß es offensichtlich problematisch ist, den Sprachgebrauch bestimmter soziolinguistischer Gemeinschaften zu antizipieren, sondern einige Forscher glauben auch, daß mit strikt empirischen Verfahren eine weitaus unparteisichere Beschreibung des Textes gegeben werden kann. Außerdem besteht bei einigen Forschern ein Interesse, empirische Verfahren als komplementäres Instrument zur Bewertung theoretischer Konzeptbildungen und Ableitung von Kategorien anzuwenden.

Empirische Verfahren der Inhaltsanalyse sind ein Gegenstück zu empirischen Beschreibungen, die in anderen sozialwissenschaftlichen Disziplinen, z.B. Persönlichkeitseinschätzungen, verwendet wurden. Statt von einer bestimmten Persönlichkeitstheorie auszugehen, greifen einige Schulen auf eine umfangreiche Fragenbatterie zurück und benutzen dann Faktorenanalysen oder vergleichbare Verfahren, um Persönlichkeitstypen zu identifizieren. Bei Persönlichkeitstests ist die Einheit die Person, bei Inhaltsanalysen ist die Einheit viel schwerer zu bestimmen.

Es können im Extremfall verschiedene Dokumente oder verschiedene Teile eines Dokumentes die Analyseeinheit sein, je nachdem, was der Forscher beabsichtigt. Wenn es ein ganzes Dokument, wie z.B. ein Buch oder ein Tagebuch ist, dann können die Schnittpunkte für die Einteilung in bestimmte Sektionen einen ausschlaggebenden Einfluß auf die Ergebnisse haben. Wenn verschiedene Abschnitte miteinander verglichen werden, wird das Dokument offensichtlich nur durch solche Facetten charakterisiert, die sich von Sektion zu Sektion ändern; der gemeinsame Bestand von Kernthemen, der sich möglicherweise durch das ganze Dokument zieht, wird nicht bei der Charakterisierung berücksichtigt, da die empirischen Verfahren sich ganz auf die Erklärung der Varianz stützen.

Eine ganze Reihe verschiedener empirischer Verfahren, die für die Anwendung im Bereich der Inhaltsanalyse in Betracht kommen, wurden in Pisa von *Coxon* (17) diskutiert. Dazu gehören vor allem neben der Faktorenanalyse Clusterverfahren, wie sie z.B. mit *Wisharts* Clustan Programmen zur Verfügung stehen, und multidimensionale Skalierungsverfahren (18). Einige Forscher sind nach mehreren Tests jedoch zu dem Ergebnis gekommen, daß es ebenso viele unterschiedliche Lösungen wie Verfahren gibt. Dennoch kann *Iker* berichten, daß er eine beachtliche Übereinstimmung zwischen seinen Faktorenanalysen und seinen Clusteranalysen zur Charakterisierung verschiedener Dokumente erzielt hat (19). Ebenso wie in anderen Bereichen sind auch hier die Diskussionen über den Nutzen und Realitätsgehalt dieser Verfahren noch nicht abgeschlossen.

Ungeachtet dessen ist festzustellen, daß empirische Verfahren der Inhaltsanalyse wesentlich von der Entwicklung der Techniken profitiert haben, die keine Normalverteilung voraussetzen. Worthäufigkeiten sind in der Regel nicht normal verteilt, sondern ähneln für die meisten Wörter eher einer Poisson-Verteilung mit einem bestimmten Muster: In vielen Dokumenten wird ein bestimmtes Wort überhaupt nicht auftauchen, wenn es aber einmal auftritt, ist auch ein zweites Erscheinen im betreffenden Text (token) sehr wahrscheinlich. Andererseits werden aber sowohl aus kommunikativen wie auch stilistischen Bedürfnissen Sättigungsbeschränkungen wirksam, deren Umfang vom jeweiligen Konzept abhängt. Die Sprache von Schizophrenen kann tatsächlich durch die Type-Token-Ratio als Abweichung von der Sprachnorm (Sättigungsnorm) charakterisiert werden (20). Bei dieser Art von Sprachver-

teilungsmustern kann es nur vorteilhaft sein, wenn man einen Ausweg aus der Abhängigkeit von der Faktorenanalyse findet.

Jede empirische Analyse steht vor dem Problem, wie viele Variablen einbezogen werden sollen. Einige Verfahren werden unbrauchbar, wenn die Zahl der Variablen die Zahl der Fälle übersteigt. Für die Inhaltsanalyse wird die Zahl der Dokumente bzw. der Abschnitte wesentlich geringer sein als die Zahl der Wörter. Daher und aus Gründen der maschinellen Verarbeitung ist eine gewisse Auswahl sinnvoll. Diese Auswahl kann sich leicht an Wörtern, die im gegebenen Zusammenhang besonders aktuell sind, aber auch an Stilwörtern etc. orientieren. Somit wird die Auswahl der Wörter, die in die Analyse eingehen, sicherlich zu einem entscheidenden Faktor für die zu erwartenden Ergebnisse. *Iker* hat einige interessante empirische Kriterien für die Selektion der Wörter vorgeschlagen (21). Er korreliert alle Wörter miteinander und berechnet dann die Summe der 5. Potenz für die Korrelationen eines jeden Wortes. Die Wörter mit der höchsten Summe werden ausgewählt. Durch die Verwendung der Potenzierung wählt er solche Begriffe aus, die eine Menge hoher Korrelationen aufweisen, statt solcher Begriffe mit hohen Interkorrelationen. Diese "Select"-Prozedur ist Teil des umfassenden Textverarbeitungssystems "WORDS", das unter Führung von *Harway* und *Iker* für die Analyse psychoanalytischer Aufzeichnungen entwickelt wurde, aber auch universell für jeden anderen englischen Text eingesetzt werden kann. Erste Anwendungen für deutsche Texte sind auf dem Inhaltsanalyse-Workshop in Pisa ebenfalls erfolgreich verlaufen. Die Entwicklung von WORDS war von der Absicht bestimmt, bewußt jeglichen Einfluß des Forschers und damit jede a priori Kategorisierung für das sprachliche Material zu vermeiden. Vielmehr sollten die Themen und inhaltlichen Kategorien mit Hilfe statistischer Verfahren aus den vorliegenden Texten selbst bestimmt werden (22). Die kleinste Informationseinheit ist das einzelne Wort. Die Rationale hierfür ist die Grundannahme, daß jedes Wort selbst und darüber hinaus die Verknüpfung mit den anderen Wörtern hinreichend Bedeutung enthält, um den Inhalt des Textes ausreichend wiederzugeben. Das Verfahren läßt sich durch folgende grundlegende Schritte charakterisieren:

1. Das für die Analyse einzugebende Dokument wird in "Segmente" zerlegt. Dies können sein: eine Seite, ein Absatz, eine bestimmte Anzahl von Wörtern, Zeitabschnitte aus Interviews usw.
2. Alle Funktionswörter, wie z.B. Artikel, Konjunktionen werden entfernt.
3. Die verbleibenden Wörter werden auf ihre Grundform zurückgeführt (lemmatisiert).
4. Die Häufigkeit, mit der jedes verschiedene Wort in jedem Textsegment auftritt, wird festgestellt.
5. Ein Teil dieser Wörter wird für die Analyse (wie oben beschrieben) selektiert.
6. Für diese Wörter (types) wird die Interkorrelationsmatrix berechnet.
7. Diese Interkorrelationsmatrix geht dann in eine Faktorenanalyse (Hauptkomponentenverfahren und anschließend Rotation nach dem Varimax-Kriterium) ein.

Daraus wird dann bestimmt, ob es gemeinsame Faktoren gibt, welche die berechneten Zusammenhänge in sinnvoller Weise erklären. In der Extraktion der Faktoren sieht *Iker* (23) die Erkennung des Inhaltes; die Analyse der Beziehungen zwischen den Wörtern mit gemeinsamem Faktor sowie der Beziehungen zwischen den Faktoren und deren Änderung über Zeit, Raum und sich wandelnde Situationen versteht er als die inhaltsanalytische Funktion des Systems.

Eine abschließende Bewertung dieser Verfahren im Vergleich zu Wörterbuchverfahren konnte in Pisa nicht gegeben werden und soll auch hier nicht versucht werden. Einige grundlegende Vor- und Nachteile beider Verfahren dürften deutlich geworden sein. Festzuhalten bleibt, daß weitere multivariate Verfahren für die empirische Ermittlung inhaltlicher Zusammenhänge diskutiert werden und auch z.B. Cluster-Techniken für den Bereich des Information Retrieval in *Saltons* experimentellem SMART-System erfolgreich erprobt wurden (24).

2. Neuere Inhaltsanalyssysteme und Projekte

In Pisa wurden außer dem GENERAL INQUIRER und WORDS weitere in den letzten Jahren entwickelte Inhaltsanalyssysteme vorgestellt. Diese sind weitestgehend an den Wörterbuchverfahren orientiert und leisten daneben grundlegende Sortieroperationen und Worthäufigkeitsauszählungen. Sie sollen hier nicht im einzelnen diskutiert werden, da sie wohl einige technologische Fortschritte, nicht aber grundlegend neue methodische Verbesserungen aufweisen. Interessant sind sie hier speziell unter dem Gesichtspunkt der Sprachunabhängigkeit und der Verfügbarkeit auf verschiedenen Computersystemen: TEXTPACK (25), speziell für die Vercodung von offenen Fragen entwickelt, aber auch schon erfolgreich für die Analyse von Heiratsanzeigen (26) und Forschungsprojektdokumentation (27) eingesetzt, ist verwendbar auf IBM 360/370 und Siemens 4004, EVA (28), für die Analyse von Schlagzeilen entwickelt und inzwischen auch für psychoanalytische Texte getestet (29), läuft auf Telefunken TR 440, die PL/1 Programme von Spence (30), die sich auf Grund ihrer Überschaubarkeit recht gut als Lehrprogramme eignen, sind für die IBM 360/370 programmiert, eine spezielle interaktive Version wurde für Lehrzwecke in Pisa entwickelt (Stone). TEXT (31), dessen Entwicklung noch nicht ganz abgeschlossen ist, wird auf einer CDC 3300 in Oslo benutzt, QUESTER (32), auf Grund seines Ansatzes für die Kontextanalyse besonders interessant, ist auf CDC Cyber 74 und CDC 6000 implementiert und schließlich FORCOD (33), für die Analyse historischer Quellentexte konzipiert, ist zur Zeit in Frankreich verfügbar für SDS 9600 (XDS 9600) und wird umgestellt auf PL/1 für IBM und CII Anlagen. Trotz der Schwierigkeiten, die sich im Einzelfall bei der Implementierung eines Programmpaketes auf einer bestimmten Maschine ergeben können, dürfte angesichts dieses Angebots von Software ein wesentliches Hemmnis für den Schritt zur automatisierten Inhaltsanalyse, nämlich mangelnder Zugriff zu einem entsprechenden System, entfallen. Erste Überlegungen zur Standardisierung der Formate und die Integrierung unterschiedlicher Operationen wurden in Pisa angestellt (34).

Parallel zur Software-Entwicklung sind speziell in der BRD mit den Projekten Fragenvercodung (*Klingemann* u. Mitarb.), Schlagzeilenanalyse (*Deichsel* u. Mitarb.),

Analyse von Schulaufsätzen (*Mohler*), Analyse von Heiratsanzeigen (*Kops*), Analyse psychiatrischer Interviews (*Kächele, Schaumburg*), Analyse der Sprache der Umfragenforschung (*Scheuch, Stone* u. Mitarb.) sowie des Einsatzes inhaltsanalytischer Verfahren für die Rückgewinnung von Fragen für die Sekundäranalyse (*Mochmann* u. Mitarb.), auch die Arbeiten an deutschen Inhaltsanalysewörterbüchern vorangetrieben. So liegen bereits vor: das Hamburger Kommunikationssoziologische Wörterbuch (*Deichsel*), Wörterlisten zur Analyse von Belief Systems (*Klingemann*), das Wörterbuch politischer Einstellungen für die Archivrückgewinnung (*Mochmann, Scheuch*) und dessen englisches Äquivalent (*Scheuch, Stone, Aylmer, Friend*). Als Ergebnis der Diskussionen in Pisa wurde geplant, gemeinsam die Arbeiten an einem generellen deutschen Inhaltsanalysewörterbuch aufzunehmen (35). Ein erster Schritt soll die Integrierung der in den verschiedenen Texten enthaltenen Wortbestände sein. Der intellektuell anspruchsvollste Schritt wird die Erarbeitung eines gemeinsam akzeptierten Kategorienschemas und die Zuordnung der aus der integrierten Textbasis gewonnenen Wörter sein. In Zwischenschritten werden auf der linguistischen Ebene morphologische und syntaktische Probleme der deutschen Sprache für die sozialwissenschaftliche Analyse befriedigend gelöst werden müssen. Dabei werden voraussichtlich Lösungen, wie sie im *General Inquirer* für die englische Sprache realisiert wurden (*Stone*), nicht ohne weiteres übernommen werden können. Alternative Möglichkeiten wurden in Köln und Pisa getestet und diskutiert (36). Besonderes Gewicht bei dieser Entwicklung wird bei der Wörterbuchkonstruktion auf die Analyse und empirische Fundierung des tatsächlichen Sprachgebrauchs gelegt werden.

Mit dieser Übersicht über den heutigen Stand der Inhaltsanalyse in der Sozialforschung sollte deutlich werden, daß eine Reihe neuer Fragestellungen und Forschungsansätze in den letzten Jahren hervorgebracht wurden, deren Lösung mit konventionellen Verfahren der Inhaltsanalyse nur zögernd und dann auch häufig mit eingeschränkteren Auswertungsplänen angegangen wurde. Bemühungen um die systematische Erfassung sozialwissenschaftlich relevanter Dokumente werden unter anderem auch im Zuge der Planung und des Aufbaus von Informationssystemen forciert. Durch die zunehmende Verwendung lochstreifen- oder magnetbandgesteuerter Setzverfahren sowie durch die Weiterentwicklung der Belegleser werden immer mehr Texte bereits maschinenlesbar vorliegen. Daher ist zu erwarten, daß sich der Anteil der inhaltsanalytischen Arbeiten an der Sozialforschung wesentlich erhöhen wird.

Anmerkungen

- 1 Der internationale "Workshop on Content Analysis in the Social Sciences" wurde unter der Schirmherrschaft des Standing Committee on Social Science Data des International Social Science Council vom 5.-13.9.1974 am Centro Nazionale Universitario di Calcolo Elettronico in Pisa abgehalten. Siehe dazu: *P.J.Stone, Report on the ISSC Content Analysis Workshop*, In: *Social Science Information*, 1, 1975.
- 2 *H. Klages*: Assessment of an attempt at a system of social indicators. In: *Policy Sciences* 4, 1973, S. 249-261.
J.M. Firestone: The development of social indicators from content analysis of social documents. In: *Policy Sciences* 3, 1972, S. 249-263.
- 3 *K. Mannheim*: *Ideologie und Utopie*, Bonn 1929.
- 4 *R.B. Abelson, A. Bernstein*: A computer simulation model of community referendum controversies. In: *Public Opinion Quarterly* 27, 1963, S. 93-122.

- 5 *P.B. Johnson, D. Sears, J. McConahay*: Black invisibility, the press and the Los Angeles riot. In: *American Journal of Sociology*, 76, 1971, S. 698-721.
- 6 *A. Deichsel*: Ein Beobachtungsinstrument für Symbolwelten aus Sprache. Referat für den Methodenausschuß der Deutschen Gesellschaft für Soziologie 1973 (vervielfältigtes Manuskript).
- 7 *A. Deichsel*: Newspaper headlines: A multinational content analysis project on textual indicators from mass media. In: *Social Science Information* 1, 1975.
- 8 *E.J. Aries*: Interaction patterns and themes of male, female and mixed groups. Dissertation, Harvard Universität 1973.
- 9 *B. Frisbie, S. Sudman*: The use of computers in coding free responses. In: *Public Opinion Quarterly* 2, 1968, S. 216-232.
- 10 *H.D. Klingemann*: Content Analysis of Open-ended Materials from Survey Research as a Device to Inquire into the Political Beliefs of Mass Publics. Vortrag für die Association for the Study of Advanced Industrial Societies, Bellagio 1974 (vervielfältigtes Manuskript).
- 11 *J.Z. Namenwirth, H.D. Lasswell*: The changing language of American values: A computer study of selected party platforms. Beverly Hills 1970.
J.Z. Namenwirth: Wheels of time and the interdependence of value change in America. In: *Journal of interdisciplinary history* 3 (4) 1973, S. 649-683.
- 12 *A. Weymann*: Bedeutungsfeldanalyse. Versuch eines neuen Verfahrens der Inhaltsanalyse am Beispiel Didaktik der Erwachsenenbildung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 4, 1973, S. 761-776.
- 13 *E. Kelly, P.J. Stone*: Computer identification of English word senses. Amsterdam 1975
- 14 *J. Laffall*: Pathological and normal language. New York 1965.
- 15 *S. Ertel*: Erkenntnis und Dogmatismus. In: *Psychologische Rundschau* 4, 1972, S. 241-269.
- 16 *D.C. Dunphy and Associates*, Harvard IV – 4 Dictionary, General Inquirer Project. University of New South Wales, Computerdruck 1974.
A.P.M. Coxon, A.D. Chalmers: Inquirer Project, Content Analysis as a Social Science resource. In: *Contributed papers: ISSC/CSSD Workshop on content analysis in the Social Sciences*, Pisa 1974, S. 49-68.
- 17 Dieser Vortrag bezog auch die Leistungsfähigkeit existierender Programme zur Multidimensionalen Skalierung mit ein. Eine schriftliche Zusammenfassung mit abschließender Bewertung liegt jedoch nicht vor.
- 18 *D. Wishart*: Clustan Ia User Manual, St. Andrews 1969, Computerdruck.
- 19 *H.P. Iker*, Words System Manual, Rochester, N.Y. 1973, Computerdruck.
- 20 Die Type-Token Ratio (TTR) wird von einigen Programmen standardmäßig ausgewiesen. Sie errechnet sich aus dem Verhältnis der Wortformen (types) zu der Gesamtzahl der Nennungen dieser Wörter (token) im Text.
- 21 *H.P. Iker*: SELECT: A computer program to identify associationally rich words for content analysis, in: *Computers and the Humanities*, Vol. 8, 1974, S. 313-319.
- 22 *H.P. Iker, N.J. Harway*: A Computer Systems Approach toward the recognition and analysis of Content. In: *G. Gerbner und Mitarb. (Hrsg.): The Analysis of Communication Content*, New York 1969, S. 381-405.
- 23 *H.P. Iker*: SELECT: A computer program to identify associationally rich words for content analysis, in: *Computers and the Humanities*, Vol 8, 1974, S. 313-319.
- 24 *G. Salton (Hrsg.)*: The SMART retrieval system. Experiments in automatic document processing, Englewood Cliffs, New Jersey 1971.
- 25 *J. Höhe, H.D. Klingemann, K. Radermacher, D. Fuchs*: TEXTPACK, Benutzer Handbuch Version 2, Köln 1973.
- 26 *M. Kops*: The Identification of Person-Conceptions by the Content Analysis of Marital Advertisements, vervielfältigtes Manuskript für den ISSC/CSSD. Workshop on Content Analysis in the Social Sciences, Pisa 1974.
- 27 *P. Ohly*: Analyse und Vergleich unterschiedlicher Ordnungsschemata für sozialwissenschaftliche Forschungsprojekte. Diplomarbeit am Soziologischen Seminar der Universität zu Köln/Zentralarchiv für empirische Sozialforschung (laufend).
- 28 *R. Tiemann*: Algorithmisierte Inhaltsanalyse, Hamburg 1973 (Dissertation).
- 29 *H. Kächele, C. Schaumburg, H. Thomä*: Changes of speech content during a psycho-

- analytic process. In: Contributed papers: ISSC/CSSD Workshop on content analysis in the Social Sciences, Pisa 1974, S. 99-110.
- 30 *D. Spence*: PL/1 Programs for Content Analysis. In: Contributed papers: ISSC/CSSD Workshop on content analysis in the Social Sciences, Pisa 1974, S. 201-232.
- 31 *J. Fonnes*: TEXT: A general programm package for text processing, In: Contributed papers: ISSC/CSSD Workshop on Content Analysis in the Social Sciences, Pisa 1974, S. 77-83.
- 32 *C. Cleveland, D. McTavern, E. Pirro*: Quester: Contextual analysis methodology, Manuskript für den ISSC/CSSD Workshop on Content Analysis in the Social Sciences, Pisa 1974.
- 33 *M. Couturier, R. Abhassera*: FORCOD C: Manual de référence, Paris, 1972 (vervielfältigtes Manuskript).
- 34 *P.J. Stone*: Standards for computer aided content analysis: The Pisa conventions and recommendations. In: Social Science Information 1, 1975, S. 127-137.
- 35 *E. Mochmann*: A cooperative project for constructing a German content analysis dictionary. In: Social Science Information 1, 1975, S. 117-121.
- 36 *P.J. Stone*: Standards for computer aided content analysis: The Pisa conventions and recommendations. In: Social Science Information 1, 1975, S. 127-137; *E. Mochmann*: Content analysis techniques retrieved from approaches to information retrieval. Vervielfältigtes Manuskript für ISSC/CSSD Workshop on Content Analysis in the Social Sciences, Pisa 1974.