

Methodische Überlegungen zu Inhaltsanalysen von Mailinglisten

Von Manuela Pötschke (Bremen)

1. Problemstellung

Die Ausbreitung Neuer Medien geht mit einer gesteigerten Textproduktion einher, die immer mehr Menschen als Autoren und/oder Leser einbezieht. War es vor der Zeit des Internet für Einzelpersonen eher schwierig, einen Text mit angemessenen Kosten veröffentlichen und verbreiten zu können, sind die Möglichkeit durch die Vernetzung heute sehr viel größer. Die entstehenden Texte sind öffentlich zugänglich und von ihnen sind auch Rückschlüsse auf allgemein interessierende gesellschaftliche Fragen zu ziehen.

Ungeachtet der Abwägung, wann diese freie Zugänglichkeit verschiedener Ideen und Gedanken in allen denkbaren Qualitäten als Vor- oder Nachteil¹ zu bewerten ist, eröffnet sie dem sozialwissenschaftlich orientierten Forscher einen unvergleichlich preiswerten Zugang zu einer Fülle ganz unterschiedlicher Quellen. Abgesehen von der reinen Materialmenge liegt ein weiterer Vorteil von internetbasierten Texten möglicherweise darin, dass sie nicht nur die Anwendung der bekannten Verfahren empirischer Sozialforschung zu ihrer Interpretation erlauben, sondern dass darüber hinaus mit ihrer Hilfe auch methodische Probleme behoben werden können.

Dieser Beitrag behandelt Möglichkeiten, Mailinglisten angemessen inhaltsanalytisch auszuwerten und versteht sich in erster Linie als Beitrag zur Methodenentwicklung. Im Zentrum steht dabei die Frage, welchen Besonderheiten die Inhaltsanalyse von Mailinglisten unterliegt.

Die Inhaltsanalyse von Mailinglisten ordnet sich ein in das allgemeinere Feld der sogenannten Online-Forschung. Im Allgemeinen wird die Erforschung *des* Internets von der Erforschung *mit Hilfe* des Internets unterschieden.

Wird das Internet zur Datengewinnung eingesetzt, können beispielsweise Onlinebefragungen von Textanalysen auf Internetseiten oder in Mailinglisten unterschieden werden. In der Diskussion um die Methodik von Online-Erhebungen stehen zwar in aller Regel Befragungen im Mittelpunkt, in diesem Beitrag soll aber gezeigt werden, dass auch netzbasierte Inhaltsanalysen spezifische Erklärungen – im Vergleich zu anderen empirischen Verfahren - bieten können und dabei speziellen Bedingungen – im Vergleich zur „klassischen“ Inhaltsanalyse - unterworfen sind.

Die wachsende Hinwendung der Sozialforschung zu internetbasierten Untersuchungen² resultiert zum Teil aus den Vorteilen dieser Vorgehensweise gegenüber anderen Erhebungsverfahren der empiri-

¹ Grundsätzlich ist der Zugang zum Internet und die Beteiligungsmöglichkeiten an dessen Gestaltung sicherlich positiv zu bewerten. Nachteile entstehen z.B. durch fehlende Kontrolle der Textqualität, weitreichende Möglichkeiten der Plagiarierung und fehlende rechtliche Sanktionsmöglichkeiten. Der letzte Gedanke bezieht sich unter anderem auf Internetauftritte die nach deutschem Recht Straftatbestände wie Volksverhetzung erfüllen, deren Betreiber durch die Verbreitung der Inhalte über ausländische Provider aber nicht belangt werden können.

² Insgesamt sind die Anteile an netzbasierter Forschung im Bereich der Sozialwissenschaften und Marktforschung sicher noch gering. Es ist aber der Bereich, für den große Entwicklungspotentiale prognostiziert werden (vgl. Korff 2000; Dialego AG o.J.; Zimmermann/ Gadeib/ Lürken 2001).

schen Sozialforschung. Für die Inhaltsanalyse von Mailinglisten ist dabei der Kostenvorteil von besonderer Bedeutung. Das Textmaterial, auf denen die Analysen basieren, steht nämlich bereits in digitalisierter und meist archivierter Form leicht zugänglich zur Verfügung. Aber netzbasierte Untersuchungen weisen auch spezifische Probleme, wie Unsicherheiten über die Befragungssituation, Non-response oder systematische Verzerrungen auf. Das mit Abstand größte Problem besteht generell jedoch in der Schwierigkeit, die Grundgesamtheit zu definieren und eine korrekte Stichprobe zu ziehen.³ Beim Vergleich der internetbasierten Untersuchungsverfahren zeigt sich aber, dass die Analyse von Mailinglisten mit diesem Problem nicht konfrontiert scheint. Durch das Prinzip des Abonnements kann jederzeit ohne besonderen Aufwand festgestellt werden, wer Mitglied der Mailingliste ist und demnach zur Grundgesamtheit gehört. Für diese Grundgesamtheiten sind Zufallsauswahlen oder Vollerhebungen gleichermaßen realisierbar. Werden die Abonnenten einer Liste als Analysegegenstand gewählt, stehen zumeist Fragen nach Gruppenprozessen und Kommunikationsstrukturen im Mittelpunkt.

In Abhängigkeit von der Forschungsfrage können aber auch die eingehenden Mails selber als Grundgesamtheit definiert werden. Die bereits angesprochene Archivierung der Mitteilungen vereinfacht dann nicht nur den Zugang zum Analysematerial, sondern impliziert auch die Bestimmung der Grundgesamtheit. Aus dieser Grundgesamtheit kann nach den üblichen Verfahren eine Stichprobe gezogen werden. Für die hier vorgestellten Analysen wird eine Vollerhebung der in einem bestimmten Zeitraum eingegangenen Mails der Liste zugrunde gelegt.

2. Beiträge aus Mailinglisten: Eine spezifische Datenbasis

Das Internet ist durch eine umfangreiche Protokollierung gekennzeichnet. So sehr es beispielsweise die räumlich und zeitlich unabhängige Kommunikation befördert, so stark sind die Möglichkeiten der Einbeziehung der später in die Kommunikation Tretenenden von der Nachvollziehbarkeit des bisherigen Dialogs abhängig. Das bedeutet, dass eine Kommunikation, die im realen Leben in der Regel mündlich erfolgen würde und nach ihrem Abschluss unwiederbringlich erscheint, in der netzbasierten Variante zeitunabhängig und „für alle Ewigkeit“ zur Verfügung steht. Dadurch kann in neuer Abonnement der Mailingliste „im Original nachlesen“, wie die Diskussion vor seiner Teilnahme ablief. In mündlichen face-to-face Situationen würde die Integration des neu Hinzugekommenen dadurch gelöst, dass eine Kommunikation über die vorangegangenen Kommunikation stattfinden müsste. In Abhängigkeit vom Verständnis dieser Kommunikation und des Gesprächspartners, der die Szene schildert, kann es dabei zu starken Verzerrungen kommen.

In Mailinglisten liegen Informationen in Form von Nachrichten vor. Diese Nachrichten werden in der Regel nicht mit dem Ziel ihrer empirischen Analyse verfasst; sie sind demnach Leserschriften oder Zeitungsbeiträgen vergleichbar. Das heißt, dass die Inhalte nicht mit Blick auf die Operationalisierungsangebote sozialwissenschaftlicher Problemstellungen hin entstehen, sondern in anderen Kontexten. Diese Kontexte erschließen sich zum Teil durch die Themenstellung der Mailingliste selber oder sie werden in den Beiträgen explizit dargestellt. Mails sind in dieser Hinsicht anderen Texten

³ Bisher gibt es als Lösungsvorschlag dazu für die Onlinebefragung die Idee des Befragtenpools. Dieser Pool entspricht in wesentlichen Merkmalen den Internetnutzern und soll diese repräsentativ abbilden. Aus dem Befragtenpool werden für die jeweiligen Studien Stichproben nach den bekannten Verfahren gezogen. Für die Analyse von Internetseiten wird eine Listenauswahl aus Suchmaschinenergebnissen diskutiert. Beide Vorgehensweise sind problembehaftet und stellen im Grunde noch kein letztlich befriedigendes Ergebnis dar.

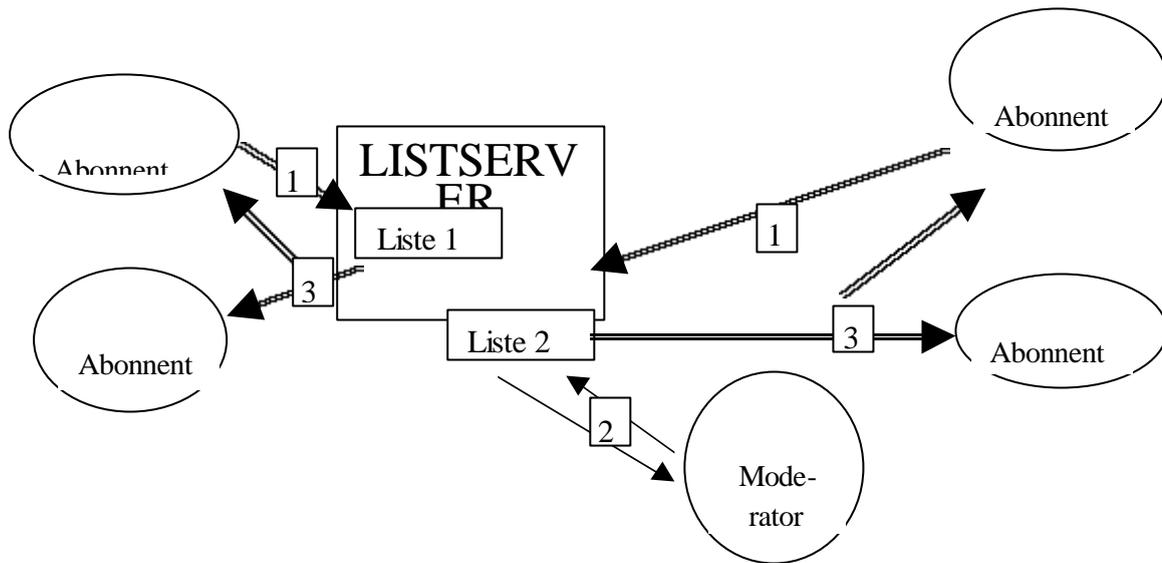
ähnlich; sie unterscheiden sich aber in einem wesentlichen Merkmal: Sie können nämlich gleichzeitig als einseitige Äußerung einer Person und als Bestandteil eines diskursiven Prozesses aufgefasst werden. Das erhöht ihr Interpretationspotential.

Als Analyseeinheit bieten sich für Inhaltsanalysen Wörter, Abschnitte und ganze Texte in Abhängigkeit der konkreten Fragestellung an. Bei der Untersuchung von Mails tritt nun die Besonderheit auf, dass sie – zumindest in fachbezogenen Wissenschaftskontexten - meist sehr kurz sind. Im Vergleich zu anderen wissenschaftlichen Texten handelt es sich bei Mails selbst in Diskussionen meist nicht um eine argumentative Auseinandersetzung, sondern vielmehr um eine Ansammlung von Verweisen, um die knappe Benennung vermeintlicher Fakten und schlicht um Informationen.

2.1 Prinzip von Mailinglisten

Das Prinzip der Mailingliste besteht darin, dass Mitteilungen von *einem* Absender an *viele* andere Teilnehmer gleichzeitig versandt werden. Handelt es sich dabei um eine unmoderierte Liste dann wird die Mitteilung sofort nach ihrem Eingang (1 in der Abbildung) im sogenannten Listserver an alle Abonnenten verteilt (3 in der Abbildung). Es gibt jedoch auch die Variante, in der jede eingehende Mitteilung zuerst an einen Moderator zur Freigabe geschickt wird (2 in der Abbildung). Dieser Moderator entscheidet dann, ob die Mitteilung veröffentlicht wird oder nicht. Auch Mischformen sind möglich: Die eingetragenen Leser einer Liste können ihre Nachrichten frei veröffentlichen, die Mitteilungen anderer Autoren werden von einem Moderator gesichtet.

Abbildung 1: Prinzip Mailingliste



Um in einer Mailingliste lesen zu können, muss sich der Teilnehmer dort anmelden. Dabei werden offene Listen, in die sich jede Person selber ohne Reglementierung einschreiben kann, von geschlossene Listen unterschieden, in die der Moderator ausgewählte Personen einträgt oder zumindest ihre Mitgliedschaft bestätigen muss.

Das bedeutet, dass der Textproduktion, aber auch der Textrezeption in Mailinglisten ein Akt der „Selbstrekrutierung“ vorangehen muss.

2.2 Das Beispiel GIR-L

Das Anwendungsbeispiel hier bezieht sich auf die Mailingliste der Deutschen Gesellschaft für Online-Forschung. Gir-L steht dabei für German Internet Research-List. Sie wurde 1996 gegründet und zum Zeitpunkt der Analyse waren etwa 760 Leser eingeschrieben. Die Liste wird zwar betreut, ist aber unmoderiert.

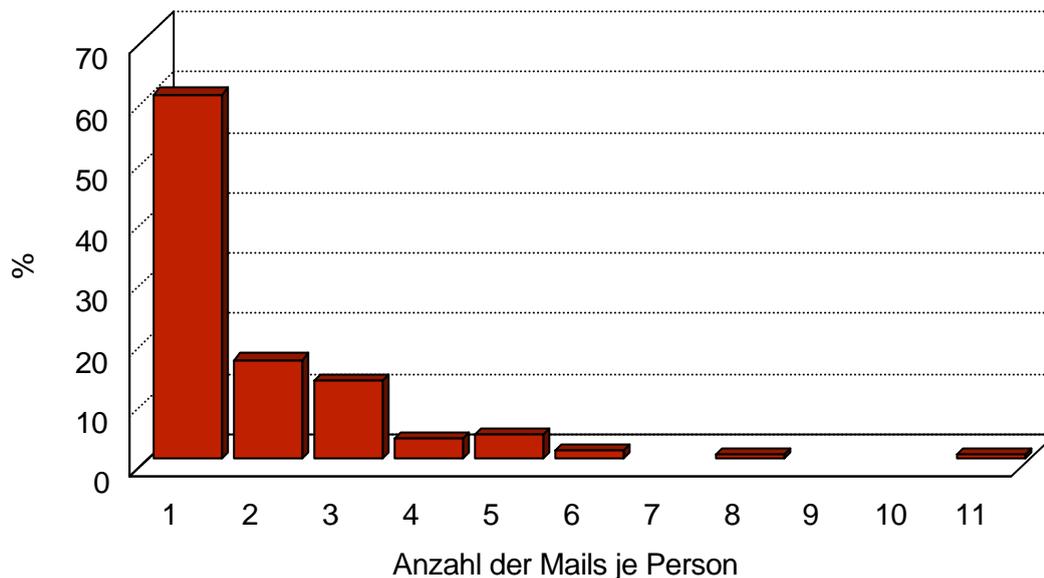
Die Liste ist als ein Diskussionsforum für alle Interessierten an Sozial-, Kommunikations- und Marktforschungsfragen rund um das Internet gedacht. Sie soll helfen, Kontakte zwischen Forschern aus unterschiedlichen Zusammenhängen und Arbeitsfeldern herzustellen und den Austausch über Fragen im Bereich der Online-Forschung zu befördern. Die Themen sind dabei recht vielfältig und reichen von methodischen und ethischen Fragen in Online-Erhebungen über Qualitätsaspekte netzbasierter Befragungen bis zu spezifischen Inhalten wie computervermittelter Kommunikation, Online-Marketing oder E-learning. Damit ist das Spektrum der Diskussion im Vergleich zu anderen fachbezogenen Listen von vorn herein sehr weit gefasst.

Die Mailingliste wurde für eine umfassendere Forschungsfrage nach der Verwendung empirischer Methoden in der Onlineforschung ausgewählt. Für die hier zu beschreibenden Möglichkeiten der Lösung methodischer Probleme in Inhaltsanalysen ist die Auswahl gerade dieser Liste nicht zwingend. Sie wird vielmehr als Materialquelle für die Exploration genutzt.

Die Datenbasis für die nachfolgenden Analysen bilden 274 Mails aus den Monaten Juli, August und September 2001. Es handelt sich für diesen Zeitraum um eine Vollerhebung aller eingegangenen und

archivierten Mails der Liste. Insgesamt haben 146 Personen die 274 Mails verfasst. Das sind im Durchschnitt 1,8 Mails pro Person.

Diagramm 1: Anzahl der Mails pro Personen insgesamt



Die Auswahl der drei Untersuchungsmonate ist der Datenerhebung für die oben angesprochene Frage nach der Methodenverwendung in der Onlineforschung geschuldet, die im Herbst 2001 abgeschlossen war. Die zuletzt verfügbaren Daten bezogen sich auf den Monat September. Da die Datenerhebung nicht weiter fortgeführt wurde, handelt es sich bei diesen drei Monaten um die aktuellsten vorliegenden Daten. Im Zusammenhang mit der explorativen Fragestellung hier ist die Datenbasis unproblematisch, weil keine verallgemeinerbaren Aussagen getroffen werden sollen (und übrigens auch nicht können). Schwierig scheint auf den ersten Blick die Einbeziehung der Einträge aus September 2001, weil für diesen Monat eine deutlich reduzierte Zahl an Mails konstatiert werden muss.⁴ Beim Vergleich der Anzahl eingegangener Einträge über einen Zeitraum von Juli 1999 bis Oktober 2001 zeigt sich nun aber, dass – ohne das dafür ein plausibler Grund auf der Hand zu liegen scheint – in mehreren Monaten eine deutlich geringere Zahl von Einträgen zu verzeichnen ist. Deshalb scheint die Wahl der drei Monate gerechtfertigt, auch wenn sich die durchschnittliche Zahl der Einträge mit 91 Mails vom Durchschnitt aus dem Jahr Oktober 2000 bis September 2001 mit 126 Mails unterscheidet.

Für die Zahl und Struktur der Schreiber (vgl. Tabelle 1) oder der angesprochenen Themen (vgl. Tabelle 2) gibt es keine Hinweise auf systematische Verzerrungen.

Tabelle 1: Anzahl der Mails und Schreiber in den Untersuchungsmonaten

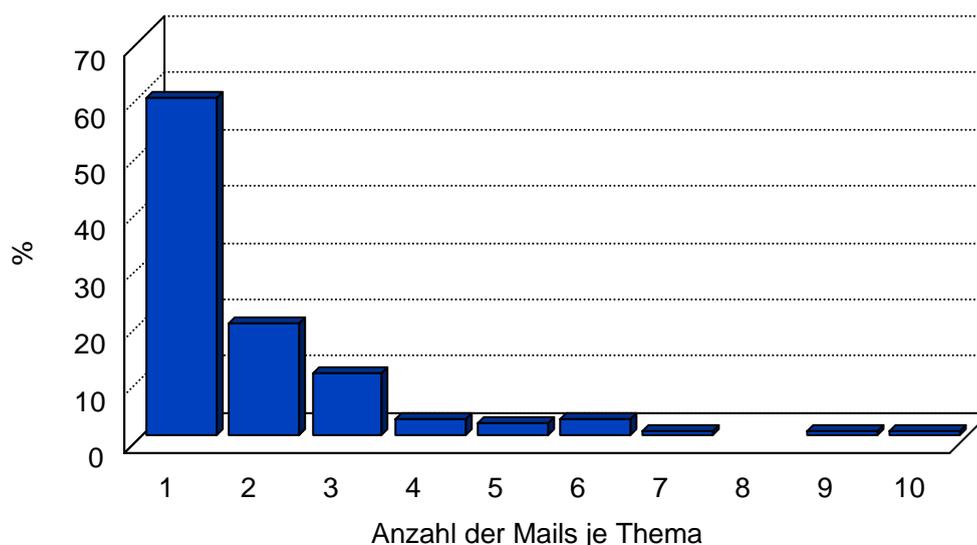
⁴ Möglicherweise haben die Ereignisse des 11. September die geringe Beteiligung an der Mailingliste mit hervorgerufen. Es könnte unterstellt werden, dass für einen gewissen Zeitraum andere Themen als Onlineforschung die Diskussionen unter den Menschen dominierten und dafür andere Foren gewählt wurden als die GIR-L. Die Ereignisse wurden in dieser Liste jedenfalls nicht thematisiert. Der empirischen Evidenz eines Zusammenhangs zwischen den Anschlägen und der aktive Teilnahme in der Liste kann in dem Zusammenhang hier jedoch nicht nachgegangen werden.

		Anzahl der Schreiber		
		Juli	August	September
Anzahl der Mails	1	54 (70,1%)	53 (74,7%)	32 (71,1%)
	2	14 (18,2%)	11 (15,5%)	10 (22,2%)
	3	7 (9,1%)	4 (5,6%)	3 (6,7%)
	4	1 (1,3%)	2 (2,8%)	
	5		1 (1,4%)	
	6	1 (1,3%)		
Gesamt		77 (100%)	71 (100%)	45 (100%)

Obwohl die Gesamtzahl derjenigen, die sich mit einer Mail an der Liste beteiligten in den einzelnen Monaten unterschiedlich ist, bleibt die generelle Struktur erhalten: Die meisten Schreiber haben sich im jeweiligen Monat einmal zu Wort gemeldet; die Anteile betragen hier bis zu drei Viertel der Schreiber. Bis etwa ein Fünftel der Schreiber verfassen zwei Mails in den Monaten. Die Anteile derjenigen, die sich mehr als zweimal zu Wort melden, werden dann sehr klein. In weiteren Analysen muss untersucht werden, ob diese Schreiber spezielle Aufgaben in der Mailingliste erfüllen. Es könnte sich beispielsweise im Mitglieder handeln, die oft auf Fragen antworten, weil sie zu den Experten der Liste gehören.

Auch in der Gesamtzahl angesprochener Themen pro Monat gibt es einige Unterschiede. Insgesamt wurden in den drei Monaten 150 Themen angesprochen. Dabei wird der überwiegende Anteil der Themen nur durch eine Mail repräsentiert. Für etwas mehr als zehn Prozent der Themen kommt eine größere Diskussion mit drei und mehr Mitteilungen zustande.

Diagramm 2: Anzahl der Mails je Thema insgesamt



Im September 2001 war dabei die Zahl mit 42 Themen wesentlich kleiner als im Juli oder im August 2001. Gleichzeitig ist der Anteil der Themen, die nur mit einer Mail angesprochen wurden mit fast 70% am größten. Unterstellt man, dass eine der zentralen Aufgaben von Mailinglisten der Austausch zwischen ihren Lesern ist, dann scheint sich diese Aufgabe in den Verteilungen nach Themen nicht adäquat wiederzuspiegeln. Im Juli 2001 wird nicht einmal die Hälfte der Themen in einer Antwortmitteilung wieder aufgegriffen. Im August und September 2001 reduziert sich dieser Anteil weiter bis auf 31%. Eine mögliche Erklärung dieser auf den ersten Blick etwas seltsam anmutenden fehlenden Dialogik wird später mit der Unterscheidung nach Funktionen einzelner Mitteilungen gegeben.

Tabelle 2: Anzahl der Mails und der diskutierten Themen in den Untersuchungsmonaten

	Anzahl der Themen		
	Juli	August	September
1	26 (53,1%)	36 (61%)	29 (69%)
2	9 (18,4%)	13 (22%)	9 (21,4%)
3	6 (12,2%)	6 (10,2%)	2 (4,8%)
4	2 (4,1%)	1 (1,7%)	2 (4,8%)
5	1 (2%)	2 (3,4%)	
6	2 (4,1%)	1 (1,7%)	
7	1 (2%)		
8			
9	1 (2%)		
10	1 (2%)		
Gesamt	49 (100%)	59 (100%)	42 (100%)

Wie eingangs beschrieben, besteht die Datenbasis aus den archivierten Beiträgen der Mitglieder der Mailingliste. Das bedeutet jedoch, dass nur Aussagen über die Teilnehmer der Liste formuliert werden können, die sich im Untersuchungszeitraum wenigstens einmal zu Wort gemeldet haben. Lurking und andere Phänomene der Gruppendynamik in Mailinglisten können deshalb hier nicht erfasst werden und sind auch nicht Ziel der Analyse.

3. Zur Spezifik der Inhaltsanalyse von Mailinglisten

Die Inhaltsanalyse zeichnet sich durch eine besondere Breite ihrer Anwendungsmöglichkeiten aus: Stehen formale oder formalisierbare Aspekte im Mittelpunkt kann das Textmaterial in Form von Daten für fortgeschrittene statistische Analyseverfahren aufbereitet werden. Gleichzeitig eröffnet sich die Möglichkeit der sinnhaften Interpretation der Absichten und impliziten Erwartungen der Schreiber einer Textnachricht. Wie weiter unten zu beschreiben sein wird, bietet die Inhaltsanalyse in Mailinglisten über diese allgemeine Mächtigkeit des Ansatzes hinaus eine Reihe von verbesserten Möglichkeiten der Datenabsicherung.

Im Kern kann jede Inhaltsanalyse als Methode verstanden werden, die fixierte Kommunikation systematisch, regel- und theoriegeleitet analysiert und Rückschlüsse auf bestimmte Aspekte dieses Prozesses zieht (vgl. Mayring 2000: 12-13).

In einem ersten Schritt müssen die für die Inhaltsanalyse relevanten Analyseeinheiten festgelegt werden. Je nach Forschungsfrage und nach der Struktur des Materials kann es sich dabei um Worte, Wortgruppen, Halbsätze, Sätze, Abschnitte oder ganze Dokumente handeln.

Da die – in Abhängigkeit von der Mailingliste stärker oder weniger stark eingeführte - ungeschriebene Norm gilt, dass Mails themenzentriert geschrieben und bezeichnet werden, kann im Unterschied zu Textanalysen von Zeitschriftenartikeln oder ähnlichen Quellen in der Untersuchung von Mailinglisten die Mail als solche als Analyseeinheit definiert werden.

Neben der Bestimmung der Analyseeinheit ist es in Inhaltsanalysen wichtig, die Kontextbedingungen aufzudecken, unter denen das Material entstand. Hier liegt sicher ein Manko in der Analyse von Mailinglisten. Bis auf die Bedingungen, die durch die Liste selber gesetzt werden, gibt es fast keine Kontextbeschreibung für die einzelnen Mitteilungen. Eine Ausnahme ist dort festzustellen, wo die Absender der Mails selber eine Kontexteinbindung vornehmen; z.B. „Ich schreibe gerade an meiner Diplomarbeit ...“.

Die zu untersuchenden Mitteilungen können nach Diekmann (1997: 486) grundsätzlich nach formalen oder inhaltlichen Aspekten untersucht werden. Zu einer formalen Analyse könnten beispielsweise die Häufigkeiten auftretender Wortarten, die Länge der Beiträge oder die Struktur der Mails gehören. Auf eine solche Untersuchung wird in diesem Kontext hier verzichtet. In der Diskussion um die Möglichkeiten der Behebung spezifischer methodischer Probleme wird jedoch eine inhaltliche Darstellung von Verteilungen integriert sein. Zu diesen methodischen Problemen gehören an zentraler Stelle eine reliable Codierung und Möglichkeiten der Validierung der Inhaltsanalyse. Beiden Themenbereichen ist im folgenden jeweils ein Abschnitt gewidmet.

3.1 Das Problem reliabler Codierung in der Inhaltsanalyse

Das Kategoriensystem bildet das Kernstück jeder Inhaltsanalyse. Diekmann beschreibt die Anforderungen an das Kategoriensystem in Analogie zu den Anforderungen an Kategorien von Variablen in Befragungen. Die Kategorien sollen sich ausschließen, ein Konstrukt erschöpfend abbilden und präzise formuliert sein.

Eine erste Kategorisierung der analysierten Beiträge in der GIR-L könnte sich an den Themen der Mailingliste orientieren (vgl. II.2). Bei genauer Betrachtung zeigt sich nun aber, dass diese Themen die Anforderungen an Kategoriensysteme nicht erfüllen.

So lassen sich die Themen „Methoden in der Online-Forschung“, „Qualitätsaspekte“ und „Online-Panelforschung“ nicht klar abgrenzen. Solche Begrifflichkeiten wie „Panel“ sind darüber hinaus in der Online-Forschung nicht eindeutig definiert. Der Rückgriff auf die Ansammlung möglicher Themen durch die Initiatoren der Liste scheint also für die Analyse nicht angezeigt zu sein. Andererseits fehlt

aber auch eine weitergehende Theorie über mögliche Themen.⁵ Es wird deshalb vorgeschlagen, in Anlehnung an Mayring (2000: 13) folgendermaßen vorzugehen:

1. Da die Mails in der Regel sehr kurz sind, kann der Schritt der Zusammenfassung von Textabschnitten hier vernachlässigt werden.
2. Es können aber Mails gesucht werden, die genau einen Themenbereich der Diskussion beispielhaft darstellen.
3. Danach werden die anderen Mails diesen Beispielen zugeordnet und
4. es wird eine kontrastreiche Struktur erzeugt,
5. die dann noch der Begrifflichkeit bedarf.

Es wurden typische, kontrastreiche Mitteilungen herausgegriffen, denen dann die anderen Mails jeweils inhaltlich passend zugeordnet werden konnten. Eine Gruppe bildete sich beispielsweise durch die Zuordnung zu folgender Mastermail:

Liebe gir-ls,

ich möchte Euch kurz auf ein neues Buch zur "Online-Marktforschung" hinweisen, das in diesen Tagen endlich auf dem Markt erscheinen wird. Es wurde von [...] herausgegeben und ist im Gabler Verlag erschienen.

Wir glauben, einen wirklich schönen Querschnitt an bekannten Autorinnen und Autoren von verschiedenen Firmen und Institutionen gewonnen zu haben, die sich in ihren Beiträgen sowohl von der theoretischen als auch der praktischen Seite dem Thema nähern.

Weitere Infos zu Inhalten und Themen gibt es auf der Webseite des Buches: [...]

Viele Grüße [...]

Diese Mail wurde zur Mastermail für den Bereich der Quellen- und Literaturempfehlung. Hier wird die Spezifik der Mailingliste genutzt, um schnell und recht bequem über die aktuellsten oder besonders hilfreiche Publikationen zu informieren und in geringen Ausmaßen für die eigene Arbeit zu werben.

Eine andere Mastermail lautet:

Hallo liebe gir-ls,

ich bin auf der Suche nach einem Tool das Bildschirmaktivitäten protokolliert (wie z.B. den Screenrorder von MatchWare). Eine Eigene Recherche ergab bislang nur Programme die lediglich die Bildschirm-Bilder aufzeichnet, den Ton jedoch nicht. Kann mir hier vielleicht jemand weiterhelfen? Wäre nett!

Viele Grüße,

[...]

⁵ Diekmann (1997: 489) schlägt weiter eine theoriegeleitete Kategorisierung in Abhängigkeit der Hypothesen vor. Für die Analysen hier wurden aber keine Hypothesen entwickelt; eine theoriegeleitete Strukturierung kann deshalb nicht realisiert werden.

Die Zuordnung zu diesem Master führte zur Kategorie Tools und Software. Im Mittelpunkt stehen technische Fragen und Informationen zu Computerprogrammen im weitesten Umfeld der Online-Forschung.

Auf diesem Wege wird jede Mail einer Mastermail zugeordnet, wobei die Ausdifferenzierung nur durch die Anzahl der Mails begrenzt ist. Mastermails sollen eineindeutig sein und nur genau eine Kategorie abbilden. Die Anzahl der Kategorien wird durch den Forscher in Abhängigkeit der Fragestellung festgelegt. Die zuzuordnenden Mails können eine Doppelfunktion ausfüllen, dann werden sie gleichermaßen zwei Mastermails zugeordnet. Das heißt, dass die Summe aller Zuordnungen nicht der Gesamtzahl der Mails entspricht, sondern darüber liegt.

Über die Notwendigkeit adäquater Regeln für die Codierung hinaus wird in der Literatur das Problem der Reliabilität diskutiert. Dabei geht es im Wesentlichen darum, dass Sachverhalte unabhängig vom Instrument codiert werden und damit die Replizierbarkeit des Kategoriensystems gegeben ist.⁶ Das Problem reliabler Codierungen tritt in jeder Inhaltsanalyse – unabhängig vom Textmaterial – auf. Hier kann also nur den üblichen Empfehlungen gefolgt werden. Diekmann (1997: 492) stellt zwei unterschiedliche Strategien als angemessen vor: a) zwei Codierer arbeiten parallel und die Übereinstimmung der Codierungen gibt ein Maß für ihre Reliabilität (Intercoderreliabilität); b) eine Person codiert das Material zu zwei Zeitpunkten und die Übereinstimmung dieser Codierungen gibt Auskunft über ihre Reliabilität (Intracoderreliabilität). In der Inhaltsanalyse hier wurde durch eine Person zwei mal codiert. Für die Berechnung der Intracoderreliabilität kann in Anlehnung an Diekmann (1997: 493) die folgende Formel Verwendung finden:

$$\text{Intracoderreliabilität} = \frac{Ü}{Kt_1 + Kt_2}$$

Hier steht Ü für die Anzahl der übereinstimmenden Codierungen für beide Zeitpunkte; Kt_1 steht für die Gesamtzahl der Codierungen zum ersten Zeitpunkt und Kt_2 für die Gesamtzahl der Codierungen zum zweiten Zeitpunkt.

Im Ergebnis zeigte sich mit 0,90 – das entspricht 90% der Codierungen - eine große Übereinstimmung zwischen den beiden Codiervorgängen. Das ist jedoch auch der Tatsache zu verdanken, dass die Kategorien theoretisch sehr deutlich abgegrenzt worden waren und die gleichen Mastermails zugrundegelegt wurden.

Nicht übereinstimmende Codierungen wurden nach einer Kontrollsichtung korrigiert. Im Ergebnis wurden neun Kategorien gebildet, die sich wie folgt verteilen (vgl. Tabelle 3).

⁶ Die folgenden Ausführungen sind ausführlicher in Simonson/Pötschke (2002) enthalten.

Tabelle 3: Verteilung der Mails nach inhaltlichen Kategorien

Kategorie	<i>Häufigkeit</i>
Methodische Fragen	51 (18,6%)
Technische Umsetzung	12 (4,4%)
Internetspezifische Fragen	72 (26,3%)
Organisatorisches	35 (12,8%)
Literaturempfehlung	94 (34,3%)
Quellensuche	43 (15,7%)
Software	26 (9,5%)
Fehler	1 (0,4%)
Sonstiges	76 (27,7%)
Gesamt	410*

* Anmerkung: Die Prozentierung bezieht sich auf die Zahl der untersuchten Mails (N=274), nicht auf die Gesamtzahl der Zuordnungen. Letztere ist größer, weil Mehrfachzuordnungen möglich waren.

Im Gegensatz zur reliablen Codierung gibt es für die Frage nach der Validität der erhobenen Daten eine textspezifische Lösungsmöglichkeit in der Inhaltsanalyse von Mailinglisten.

3.2 Das Problem der Validierung

Validität bedeutet die Übereinstimmung des Gemessenen mit dem tatsächlichen Sachverhalt. Das heißt beispielsweise, wenn eine Mail nach den oben erläuterten Regeln mehrmals reliabel zugeordnet wurde, weil im Body der Begriff Literaturrecherche aufschien, sie aber gar nicht als Suche nach Literatur verstanden werden kann, ist die Zuordnung nicht valide.

Das Problem der Validität besteht für jedes Instrument empirischer Sozialforschung. In Mailinglisten ergibt sich aber eine spezifische Möglichkeit der Kontrolle. Durch die Kurzbeschreibung der Mail im Subject nimmt der Schreiber der Mail eine ähnliche Kategorisierung vor wie der Forscher, der die Inhalte im Nachhinein codiert.

Auch wenn die beiden Systeme sich in wesentlichen Merkmalen unterscheiden – das Subject impliziert in der Regel nur eine Funktion oder einen Inhalt, während die Codierung der tatsächlichen Mail mehrere Funktionen und mehrere Inhalte zulässt – kann die Einordnung durch die Subject-Bezeichnung als eigenes Kategoriensystem verstanden werden. Ein Abgleich der Zuordnungen gibt dann einen Hinweis zur Validierung.

Tabelle 4: Funktionale Dimension

	<i>Funktion aus dem Subject</i>	<i>Funktion nach der Mail</i>
Frage	26 (9,5%)	87 (31,8%)
Kommentar	72 (26,3%)	141 (51,5%)
Mitteilung	14 (5,1%)	45 (16,4%)
Keine Funktionszuordnung	162 (59,1%)	5 (1,8%)
Gesamt	274 (100%)	278*

* Anmerkung: Die Prozentierung bezieht sich auf die Zahl der untersuchten Mails (N=274), nicht auf die Gesamtzahl der Zuordnungen. Letztere ist größer, weil Mehrfachzuordnungen möglich waren.

Von den 274 eingegangenen Mails haben 265 (96,7%) eine einzige Funktion, 4 (1,5%) haben zwei Funktionen, für die verbleibenden 5 Mails konnte keine Funktion im Rahmen der Definitionen angegeben werden. Die doppelten Funktionen werden jedoch nicht aus der Codierung im Subject deutlich. Die Kategorisierung der Subjects ergibt sich durch die Frage, ob aus dem Subject auf die Absicht des Absenders geschlossen werden kann. Wenn ein Bezug durch Replay hergestellt wurde, dann wird davon ausgegangen, dass der Absender sich auf eine vorherige Mail bezieht. Ob es sich dann aber tatsächlich um einen Kommentar oder aber um eine Nachfrage handelt, bleibt offen. Eine Frage wird nur durch die direkte Beschreibung z.B. durch „dringend Hilfe gesucht“, „Frage“ oder „wer weiß ...“ als solche erkannt.

Wenn der Absender die funktionale Dimension nicht zur Subject-Bezeichnung heranzieht, sondern einen inhaltlichen Bezug wählte, ist die Funktionen nicht zu eruieren. Deshalb ist der Anteil der Mails ohne Funktionszuordnung hier besonders hoch. Durch die Kreuztabellierung der Kodierungen aus dem Subject und aus dem Mailtext ergeben sich die folgenden Übereinstimmungen und Abweichungen.

Tabelle 5: Übereinstimmung und Wechsel der Funktion in Subject und Mailtext

		<i>Kodierung des Mailtextes</i>			
		<i>Frage</i>	<i>Kommentar</i>	<i>Mitteilung</i>	<i>keine Funktion</i>
Kodierung nach dem Subject	Frage	19 (6,9%)	8 (2,9%)	1 (0,4%)	
	Kommentar	8 (2,9%)	58 (21,2%)	3 (1,1%)	3 (1,1%)
	Mitteilung		1 (0,4%)	13 (4,8)	
	Keine Funktion	60 (21,9%)	74 (27,0%)	28 (10,2%)	2 (1,1%)

Für ein Drittel der eingegangenen Mitteilungen gibt es eine übereinstimmende Kodierung der Funktion der Mail. Wie aus der Tabelle 5 sehr deutlich wird, ist für die geringe Übereinstimmung aber nicht eine fehlerhafte Zuordnung verantwortlich, sondern eher die große Zahl an Mails, die im Subject keine Funktion aufweisen.

Aber auch Fehlkodierungen können in beiden Systemen auftreten. Der Forscher kann eine Mail unkorrekt zuordnen oder die Beschreibung der Mail im Absender durch den Schreiber kann fehlerhaft oder unspezifisch sein.

Die Funktionalität einer Mail in einer Mailingliste ist deshalb nur sehr unzulänglich über das Subject zu validieren.

Generell beinhalten die Mails einer Mailingliste neben der funktionalen eine inhaltliche Dimension. Die Mails verteilen sich nach den inhaltlichen Kategorien (vgl. III.2) in Abhängigkeit vom Kategorienbezug folgendermaßen.

Tabelle 6: Häufigkeit inhaltlicher Kategorien im Subject und im Mailbody

	<i>Häufigkeit im Subject</i>	<i>Häufigkeit in den Mails</i>
Methodische Fragen	62 (22,6%)	51 (18,6%)
Technische Umsetzung	11 (4,0%)	12 (4,4%)
Internetspezifische Fragen	88 (32,1%)	72 (26,3%)
Organisatorisches	49 (17,9%)	35 (12,8%)
Literaturempfehlung	20 (7,3%)	94 (34,3%)
Quellensuche		43 (15,7%)
Software	15 (5,5%)	26 (9,5%)
Fehler		1 (0,4%)
Sonstiges	29 (10,6%)	76 (27,7%)
Gesamt	274 (100%)	410*

* Anmerkung: Die Summe der Prozentangabe ergibt nicht 100%, da die Prozentuierung auf die Gesamtzahl der Mails (274) Bezug nimmt und nicht auf die Gesamtzahl der Nennungen.

Die Gesamtzahl der Kategorienzuordnung unterscheidet sich wieder dadurch, dass die Mail selber mehreren Kategorien zugeordnet werden kann, durch das Subject aber nur ein Bezug hergestellt wird. Aber auch die Anteile der Kategorien verändern sich. Besonders deutlich ist, dass im Subject meist keine Unterscheidung danach getroffen wird, ob Literatur empfohlen oder gesucht wird. Diese Kategorien wurden deshalb für die Subjectkodierung zusammengefasst. Der Anteil an Literaturempfehlungen in den Mails liegt deutlich höher als im Subject; dort wurde möglicherweise das konkretere Thema benannt, woraus die größere Häufigkeit bei themengebundenen Inhalten wie den internetspezifischen Fragen resultieren mag.

In der Gruppe der „Sonstigen“ befinden sich für die Subject-Codierungen die Mails, die ausschließlich eine funktionale Zuschreibung wie „kurze Frage“ wählten.

Die Zahl der in Tabelle 7 dokumentierten Kategorisierungen enthält nicht die Mails, die nach dem Mailtext als Sonstige klassifiziert wurden.⁷

Tabelle 7: Inhaltliche Kategorisierung nach Subject und Mailtext

		<i>Kategorien nach Mailtext</i>					<i>Software</i>
		<i>Methodische Fragen</i>	<i>Technische Umsetzung</i>	<i>Internetspezifische Fragen</i>	<i>Organisatorisches</i>	<i>Literatur</i>	
Kategorien nach Subject	Methodische Fragen	41 (15,0%)	2 (0,7%)	11 (4,0%)		32 (11,7%)	1 (0,4%)
	Technische Umsetzung		2 (0,7%)				7 (2,6%)
	Internetspezifische Fragen	2 (0,7%)	3 (1,1%)	44 (16,1%)		65 (23,7%)	
	Organisatorisches	5 (1,8%)	2 (0,7%)	5 (1,8%)	32 (11,7%)	7 (2,6%)	2 (0,7%)
	Literatur	1 (0,4%)		6 (2,2%)		20 (7,3%)	
	Software			1 (0,4%)	1 (0,4%)	2 (0,7%)	9 (3,3%)
	Sonstiges	2 (0,7%)	3 (1,1%)	5 (1,8%)	3 (1,1%)	11 (4,0%)	8 (2,9%)
	Gesamt	51 (18,6%)	12 (4,4%)	72 (26,3%)	35 (12,8%)	137 (50%)	26 (9,5%)

Anmerkung: Die Prozentuierung ergibt sich als Anteile zu den 274 analysierten Mails.

Im Ergebnis zeigt sich, dass mehr als die Hälfte der Mails (54,1%) nach Subject und Mailtext in die gleiche Kategorie fielen. Die größte Abweichung ergibt sich in der Zuordnung von fast einem Viertel der Mails zum Themenbereich internetspezifische Fragen, wenn die Subjects zugrunde liegen und ihrer Zuordnung zur literaturbezogenen Gruppe wenn der Mailtext herangezogen wird. Diese Nicht-Übereinstimmung ergibt sich wieder aus der zweifachen Dimensionalität von Funktion und inhaltlichem Thema.

⁷ Die Kategorie ‚Sonstige‘ diente im Datensatz für die Mailtexte neben der Sammelfunktion für restliche Inhalte auch der Erläuterung und Spezifizierung von Mailtexten, die ansonsten in anderen Kategorien zugeordnet werden konnten. Letztere würden doppelt in die Übersicht eingehen, obwohl nicht zwei unterschiedliche Kategorien gegeben waren, sondern die Konkretisierung einer Zuordnung vorgenommen wurde.

Es kann für die inhaltliche Zuordnung also von einer etwas besseren Übereinstimmung in den Zuordnungen zu den Kategoriensystemen ausgegangen werden als das für die funktionale Dimension der Fall war. Gleichwohl scheint die Subject-Bezeichnung des Absenders zwar im Zweifel hilfreich für die Interpretation der Kommunikationsabsicht zu sein, als Validierungsinstrument erfüllt es aber die Erwartungen nur sehr eingeschränkt.

4. Schlussbemerkungen: Potentiale einer Inhaltsanalysen von Mailinglisten

Abschließend lässt sich sagen, dass die Inhaltsanalyse von Mailinglisten eine gut zu realisierende Möglichkeit darstellt, Text aus dem Internet zu interpretieren und Schlussfolgerungen für listenbezogenen Sachverhalte zu ziehen.

Im Unterschied zu anderen Verfahren der Online-Forschung sieht sich der Forscher mit einer Inhaltsanalyse von Mailinglisten nicht mit dem Problem mangelnder Grundgesamtheits- oder Stichprobendefinitionen konfrontiert. Das bedeutet, dass die üblichen Standards der induktiven Statistik eingehalten werden können.

Durch die Interpretationsmöglichkeit der Subject-Bezeichnung durch den Absender als eigenes Kategoriensystem sollte theoretisch die Möglichkeit gegeben sein, dass die Validität des konstruierten Systems überprüft werden kann. Die empirische Analyse hat aber gezeigt, dass hier nur sehr eingeschränkte Aussagen möglich sind. In Mailinglisten mit stärkeren Regeln bezüglich der Verwendung von Subject-Bezeichnungen mögen hier aber zusätzliche methodische Potentiale liegen.

Darüber hinaus bestehen die besonderen Potentiale der Inhaltsanalyse von Mailinglisten in der Möglichkeit, Kommunikationsstrukturen nachzuzeichnen. Dies war hier nicht der Gegenstand, aber es scheint auf der Hand zu liegen, dass schon einfache formale Informationen, wie die Zahl und die Länge der Antworten auf Fragen, die Länge von Diskussionsbeiträgen oder eine spezifische Analyse der beteiligten Personen Hinweise auf die Interaktionen geben können.

Literatur:

Dialego AG o.J.: Prognose zur Entwicklung des deutschen Online-Marktforschungs-Marktes. <http://www.dialego.de> vom 17. August 2001.

Diekmann, Andreas. 1997: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.

Korff, Jörg. 2000: Management Summary. Akzeptanzanalyse Online-Forschung. Köln: SKOPOS.

Mayring, Philipp. 2000: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Deutscher Studien Verlag.

Simonson, Julia/ Pötschke, Manuela. 2002: Die Bedeutung empirischer Methoden im Spiegel ausgewählter soziologischer Fachzeitschriften. In: Uwe Engel (Hrsg.): Praxisrelevanz der Methodenausbildung. Bonn: IZ, S. 171-182.

Zimmermann, Matthias/ Gadeib, Andera/ Lürken, Alexander. 2001: Marktforschung Online – Schöne neue Welt? In: Planung & Analyse 2/2001, S. 38-43.

Kontakt zur Autorin:

Manuela Pötschke mpoetschke@sfb186.uni-bremen.de