

Von der manuellen zur elektronischen Datenerhebung: Informationsquellen und Textanalysen

Galliker, Mark

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Galliker, M. (1998). Von der manuellen zur elektronischen Datenerhebung: Informationsquellen und Textanalysen. *ZUMA Nachrichten*, 22(43), 54-72. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208238>

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.

VON DER MANUELLEN ZUR ELEKTRONISCHEN DATENERHEBUNG: INFORMATIONSQUELLEN UND TEXTANALYSEN

MARK GALLIKER

In diesem Beitrag wird abgeklärt, ob bzw. inwieweit die neuen elektronischen Informationsquellen für die empirische Sozialforschung nutzbar sind. Ausgehend von der Auswahl und Bearbeitung traditioneller Datenquellen werden manuell erstellte computerlesbare Dokumente, Setzbänder von Verlagen, CD-ROMs von Printmedien, Online-Datenbanken und das WWW als Informationsquellen behandelt. Die in Suchmasken, Suchmaschinen sowie externen Programmen (wie z.B. TEXTPACK) integrierten Booleschen Operatoren eröffnen auch Perspektiven für Textanalysen. Abschließend wird auf methodologische Probleme bei der Auswertung elektronischer Informationsquellen hingewiesen.

This article considers whether and to what extent new data sources can be used for empirical social research. A brief review of selection and analysis procedures with traditional sources of data is followed by a discussion of electronic corpora such as individual collections of 'machine-readable' documents, typeset volumes of daily newspapers, CD-ROMs of print media, Online databanks, and the WWW. Textual analysis of these can utilise the Boolean operators in search engines, search frames and in software packages such as TEXTPACK. The article concludes with a discussion of methodological problems in the analysis of electronic data sources.

1. Einleitung

In der vorliegenden Arbeit wird auf die neuen elektronischen Informationsquellen und ihre Suchsysteme sowie auf extern verwendbare Such- und Verarbeitungsmöglichkeiten hingewiesen. Es stellt sich die Frage, ob sich elektronische Informationsquellen für die empirische Sozialforschung als nützlich erweisen, oder ob sie vorläufig ausschließlich für andere Zwecke (z.B. für journalistische Recherchen) dienlich sind.

Von linguistischer Seite wird die wissenschaftliche Verwendbarkeit von CD-ROMs eher zurückhaltend beurteilt (Wagner 1997). Allerdings wird der aktuelle technologische Stand der CDs verschiedener (u.a. amerikanischer und deutscher) Printmedien nicht differenziert. Die Möglichkeit, externe Textverarbeitungsprogramme für die Behandlung komplexerer Problemstellungen beizuziehen, wird nicht wahrgenommen. Ferner werden die hinsichtlich von Textanalysen relevanten Booleschen Operatoren nicht berücksichtigt. In der vorliegenden Arbeit werden die neuen elektronischen Untersuchungsmöglichkeiten (u.a. CDs) an den vorelektronischen gemessen und die Probleme aus der Perspektive empirisch-analytischer Wissenschaft betrachtet.

In den Sozialwissenschaften und insbesondere in der hier beleuchteten Sozialpsychologie wird der Analyse computerlesbarer Dokumente immer noch wenig Bedeutung beigemessen. Dabei soll nicht bestritten werden, daß es einige fruchtbare empirische Ansätze gibt, vor allem bezüglich gesellschaftlich relevanter Attributionen sowie Verkörperungen sozialer Urteile und Einstellungen (vgl. u.a. Schütz/Six 1994). Doch werden in einigen Bereichen der Medien- und Kommunikationswissenschaften die Zeichen der Zeit in theoretischer sowie praktischer Hinsicht besser erkannt. In Universitätsinstituten, Fachhochschulen und praxisbezogenen Forschungseinrichtungen werden Kurse für Media-Research und Online-Journalismus angeboten. Bei diesen Kursen werden jedoch selten Techniken zur systematischen Inhaltsanalyse diskutiert. Im Vordergrund steht die Sichtung thematisch vorgegebener Texte hinsichtlich der Produktion neuer Texte. Die sozialwissenschaftlichen Möglichkeiten von systematischen Recherchen von Volltexten werden nicht oder höchstens am Rande thematisiert.

Seit den ersten computergestützten text- und inhaltsanalytischen Verfahren (vgl. u.a. Klingemann 1984) gab es immer wieder Fortschritte in den Auswertungsmöglichkeiten von Textmaterial zu verzeichnen (vgl. Nissan/Schmidt 1995). Qualitativ neue inhaltsanalytische Möglichkeiten eröffneten jedoch erst die an klassischen empirischen Untersuchungen zur Worthäufigkeit und Themenbildung (u.a. Stone 1966; Iker/Harway 1969) orientierten Programme, mit denen Kombinationen verbaler Einheiten nicht nur bezüglich des Gesamttextes, sondern primär bezüglich kleinerer Text- bzw. Zählseinheiten erfaßt werden.

Von den meisten elektronischen Informationsquellen sind Items abrufbar, die auch unabhängig vom Forschungsprozeß existieren und nicht erst durch Abfragemethoden oder Tests konstituiert werden. Nach erfolgtem Zugriff liegen sie zwar ebenfalls in selektiver Form vor, doch das zugrundeliegende Datenmaterial ist in seiner ursprünglichen Gestalt konserviert und weiteren Auswertungen im Kontext gleicher, ähnlicher oder neuer Fragestellungen zugänglich. Damit lassen elektronische Datenquellen die in den letzten Jahren

vielfach erhobene methodenpluralistische Forderung als realisierbar erscheinen, gleiche Materialien mit verschiedenen Auswertungsmethoden zu untersuchen.

2. Informationsquellen

2.1 Manuell ausgewählte und gesichtete Dokumente

Bei Analysen traditioneller Dokumente wird das Untersuchungsmaterial in der Regel wie folgt erstellt: Zunächst wird das Material nach thematischen Gesichtspunkten gesichtet, ausgewählt und sortiert. Anschließend werden Originale kopiert und zur weiteren Bearbeitung vollständig, ausschnittsweise oder nur punktuell (u.a. im Sinne von Antworten auf einen vorgefertigten Abfragebogen) in den Rechner eingetippt. Hierbei besteht die Gefahr von Übertragungsfehlern, selbst wenn das Original nicht einem Tonträger entnommen und in schriftsprachliche Form transkribiert werden muß.

Maßgebliche Dokumentenanalysen und bibliometrische Untersuchungen werden nach wie vor manuell durchgeführt (vgl. u.a. Cronin/Davenport/Martinson 1997). Allerdings werden in einzelnen Untersuchungen stellenweise computerlesbare Datenquellen mitberücksichtigt (u.a. Davenport 1995). Da die klassischen Inhaltsanalysen auch bei kleinem Umfang aufwendig sind und sich das Verhältnis von Aufwand und Ertrag oft als eher ungünstig erweist, erstaunt es kaum, daß solche Untersuchungen relativ selten durchgeführt werden. Ökonomischer scheint es zu sein, die Daten mit Skalen zu konstituieren, die sich auf die subjektive Seite eines Diskurses beziehen. Häufig ist es aber schwer abzuschätzen, ob bzw. inwieweit die in den Skalen aufscheinenden positiven und negativen Diskriminierungen überhaupt gesellschaftlich relevant sind (vgl. u.a. Blank/Heinrich 1998). Im Extremfall haben sie realiter keine Bedeutung (mehr) und werden ausschließlich methodisch (wiederum) nahegelegt, während aktuelle sprachliche Diskriminierungen ignoriert werden. Die von den Medien verbreiteten Meinungen werden zwar zuweilen bei der Konstruktion von Fragebogen mitberücksichtigt, doch werden sie selten in ihrer objektiven oder vergegenständlichten Form ausgewertet, ähnlich wie sich die herkömmliche quantitative Therapieforschung selten primär auf Verhalten und Interaktion (vgl. u.a. Galliker/Herman 1994), sondern vielmehr vorwiegend auf Antworten des Klienten im Postsession outcome konzentriert (vgl. u.a. Grawe/Caspar/Ambühl 1990).

Zu den Ausnahmen im Bereich der medienpsychologischen Forschung gehört eine Studie über die Repräsentation von AIDS in der Presse (Bock et al. 1992). In dieser Studie werden thematisch relevante Beiträge heterogener Printmedien für die Jahrgänge 1982-1990 erhoben. Allerdings können mit manuellem Zugriff in der Regel nur selektive Text-

korpora erstellt werden, die pro Thema und Jahrgang nicht vollständig sind, sondern eine hinsichtlich der Fragestellung relevante Auswahl von Artikeln beinhalten, wobei die Selektionskriterien nicht immer genügend ausgewiesen sind. Die Kriterien divergieren zwischen den verschiedenen Themengruppen sowie zwischen den einzelnen Textkorpora. Trotz einem systematischen Anspruch werden die Textausschnitte oft erst im Verlaufe eines Projektes bereitgestellt und als - empirisch-analytisch durchaus fragwürdige - Belegstellen für theoretische Vorannahmen verwendet (vgl. u.a. Van Dijk 1991).

2.2 Manuell erstellte computerlesbare Dokumente

Zeitungskorpora finden sich unter anderem in Bibliotheken, Archiven und Forschungsinstituten. Oft handelt es sich um historisch relevante Dokumente, die für Längsschnittuntersuchungen mit vergangenheitsbezogenen Fragestellungen verwendbar sind. Darunter befinden sich auch Dokumente, die häufig in langwieriger Kleinarbeit in eine computerlesbare Form gebracht wurden. Als Beispiel eines elektronischen Textkorpus sei hier der *Wendekorpus (WK)* des *Institutes für Deutsche Sprache (IDS)* in Mannheim angeführt (IDS, R 5,6; 68159 Mannheim). Dieser Korpus entstand im Rahmen des Projektes *Gesamtdeutsche Korpusinitiative* in Zusammenarbeit zwischen IDS und dem *Zentralinstitut für Sprachwissenschaft* der Akademie der Wissenschaften in der ehemaligen DDR (Herberg/Stickel 1992). Der Wendekorpus ist eine Dokumentensammlung, bestehend aus 3.387 Artikeln (circa 3.3 Millionen Wortformen), die aus ost- und westdeutschen Printmedien sowie aus Protokollen öffentlicher politischer Debatten in Ost und West zur Zeit der Wende (1989/90) zusammengestellt wurden. Auswahlkriterium der Texte war der inhaltliche Bezug zu den historischen Ereignissen in der DDR und zur Vereinigung beider deutschen Staaten (Herberg 1993).

Der Wendekorpus ist mit dem korpuspezifischen Programm COSMAS lesbar (al-Wadi 1994). Dieses Suchsystem eignet sich besonders für linguistische und sprachpsychologische Detailanalysen. Einzelne Suchwörter lassen sich im Sinne der Booleschen Logik (siehe unten) kombinieren und bezüglich minimaler Texteinheiten orten. Untersucht man beispielsweise sprachliche Diskriminierungen von Migrantengruppen in den im Wendekorpus gespeicherten Aussiedlungs- und Übersiedlungsdebatten des Deutschen Bundestages, können entsprechende Keywords für den Diskursmodus, für die besprochenen Personengruppen sowie für Vergleichsgruppen eingegeben und in ihrem verbalen Kontext quantitativ sowie qualitativ verglichen werden (näheres in Galliker/Wagner 1995).

2.3 Setzbänder von Verlagen

Verschiedene Verlage stellen die Setzbänder ihrer Produktionen wissenschaftlichen Institutionen zur Verfügung. So übermittelte der Mannheimer Morgen (MM) dem IDS für das

Handbuch Korpus (HBK), das auch Texte anderer Printmedien beinhaltet (insgesamt 11 Millionen Wortformen), die Setzbänder der Jahrgänge 1985-88. Die Artikel der Rubriken *Wirtschaft, Behörde-Soziales, Technik, Kultur, Politik, Umwelt, Medizin, Pädagogik* sowie *Psychologie* wurden in eine FIDAS-Datenbank aufgenommen. In dieser Datenbank sind alle bibliographischen und dokumentarischen Angaben zu jedem Artikel verzeichnet (Kolvenbach 1987). Des weiteren erhielt das IDS vom Mannheimer Morgen die Setzbänder der Jahrgänge 1989, 1992 und 1994 (jeweils circa 15 Millionen Wortformen). Für den Zeitraum vom Mai 1995 bis Ende 1996 wurden die kompletten Satzdaten ins IDS überspielt und jeweils zu Monats- und Jahresbeständen kumuliert. Die vorliegenden Daten sind extern abrufbar und zur Weiterverarbeitung auf projekteigene Rechner importierbar (vgl. u.a. Galliker/Herman/Wagner/Weimer 1996).

2.4 Auf CD-ROMs gespeicherte Printmedien

Wo Korpora von Printmedien auf CD-ROM vorliegen, steht das vollständige Material sämtlichen Interessenten - also auch Wissenschaftler/innen verschiedener theoretischer Ausrichtungen oder Disziplinen - zur Verfügung. Heterogene theoretische Ansätze und methodische Zugänge können bezüglich der gleichen Datenbasis verglichen werden. Die Suchmasken erlauben nicht nur Einwortabfragen. Mit den Booleschen Operatoren können auch komplexere Suchaufträge formuliert werden. In methodologischer Hinsicht ist relevant, daß Ergebnisse sozialwissenschaftlicher Untersuchungen erstmals generell zugänglich und überprüfbar sind.

Im deutschsprachigen Raum liegt bis jetzt als einzige Zeitung *Die Tageszeitung (TAZ)* vollständig auf CD-ROM vor. Seit 1993 erscheinen die jeweiligen Jahrgänge der *Frankfurter Allgemeine Zeitung (FAZ)* und der *Neue Zürcher Zeitung (NZZ)* in computerlesbarer Form. Die seit 1994 auf CD-ROMs erhältliche *Süddeutsche Zeitung* bietet in einer Gesamtausgabe einen bis 1933 zurückreichenden Teil an, der für die Einstellungsforschung von besonderem Interesse ist.

Von den italienischen Printmedien sind *Il Soe 24 Ore* (1988-92; 1995-1997) und *La Stampa* (ab 1992) auf CD-ROMs erhältlich. In Frankreich existieren CD-ROMs von *Le Monde* sowie von *Le Monde diplomatique*. Die englischen Zeitungen *Times* sowie *Sunday Times* (ab 1990) sind ebenfalls in dieser Form zugänglich. In den USA erscheinen u.a. *New York Times* (ab 1981), *Washington Post* (ab 1989), *Wall Street Journal US-Edition* (1989), *USA Today* (ab 1990), *Wall Street Journal European* und *Asian Edition* (ab 1991) sowie *Business Dateline* (ab 1995) in CD-Ausgaben.

Mit dem neuen Datenträger CD-ROM hat sich die sozialwissenschaftliche Datenlage sprunghaft verbessert. Forschungsvorhaben werden möglich, die früher in personeller und

zeitlicher Hinsicht zu aufwendig gewesen wären. Mittels CD-Volltexten wird die qualitative und quantitative Präsenz von Einzelpersonen, Personengruppen, Sachverhalten und Problemen in Medien verschiedener politischer Richtung, Auflagenstärke, Reichweite, Länder usw. zugänglich, vergleichbar und analysierbar.

2.5 Online-Datenbanken

Online-Datenbanken wurden bisher selten für Inhaltsanalysen verwendet. Bei einem der ersten Versuche wurden Berichte der Olympiade 1996 von *Die Welt* und *Welt online* verglichen (Ring 1997). Für sozialwissenschaftliche Fragestellungen relevante Datenbanken sind über verschiedene Online-Dienste zugänglich. Die Online-Ausgaben von Zeitungen enthalten jedoch aus naheliegenden Gründen meistens nicht die vollständige Printausgabe, wobei neuerdings die Zeitungsverleger eher nach einem eigenen Profil für die Online-Ausgabe suchen als lediglich versuchen, eine direkte Konkurrenz auszuschließen (Cornelissen 1997). In Zukunft werden lange Artikel kaum mehr auf dem Bildschirm erscheinen. Die eigentlichen interaktiven Möglichkeiten des Internet werden allerdings von den Verlagen noch kaum ins Auge gefaßt.

Relativ ökonomische Recherchen bietet *CompuServe Information Services (CIS)*. Dieser thematisch hierarchisch organisierte Online-Dienst mit 2 Millionen Mitgliedern in 185 Ländern (Juni 1998) benutzt HTML und andere Web Standards für die Präsentation der Informationen. CIS ermöglicht in erster Linie Zugriffe auf Datenbanken amerikanischer Printmedien. Datenbanken europäischer Zeitungen werden nur vereinzelt vermittelt (vor allem englische, französische und russische Zeitungen).

Iquest ist innerhalb von CIS ein Recherchesystem, das den Zugang zu circa 600 Datenquellen ermöglicht (u.a. *AP News*, *Fisheries Abstracts*, *BBC Summary of World*, *Chicago Tribune*, *Detroit Free Press*, *Los Angeles Times*, *Memphis Commercial*, *New York Times*, *Seattle Times*, *Time Magazine*). Das System schirmt die Informationssuchenden von den ungleichen Standards der verschiedenen Datenbanken ab, indem es Suchaufträge nach einheitlichen Kriterien definiert. Nachdem Keywords, Autorennamen, Jahrgänge usw. abgefragt sind, werden die entsprechenden Referenzen präsentiert. Wie bei den CD-ROMs können die mit Maskenzeichen variabel definierbaren Suchbegriffe mit den Booleanischen Operatoren kombiniert werden. Fundstellen lassen sich paketweise in eigene Dateien übertragen.

Kostengünstiger als *Iquest* ist der abends und an Wochenenden zugängliche *Knowledge Index*. Dieser CIS-Recherchedienst erlaubt den Zugriff auf eine Teilmenge von *Iquest* (circa 100 Literatur- und Volltextdatenbanken bestehend aus Fachzeitschriften, Magazinen und Tageszeitungen).

Der Recherchedienst *Newspaper Archive* beinhaltet die Freitextarchive von circa 60 amerikanischen und britischen Tageszeitungen (u.a. *Daily News*, *New York Times*, *USA Today*, *Washington Post*, *Washington Times*). Der Vorteil dieses CIS-Dienstes besteht darin, daß er relativ weit in die Vergangenheit zurückreichende Recherchen und damit ausgehende Längsschnittuntersuchungen ermöglicht. Die meisten Datenquellen beginnen mit den späten 80er Jahren; einzelne bereits 1983 (u.a. *Miami Herald*, *Philadelphia Daily News*, *Philadelphia Inquirer*) oder 1980 (*Boston Globe*).

Mit CompuServe ist ein direkter Zugriff auf laufende Nachrichten möglich. *Executive News Service (ENS)* bündelt die Informationen international relevanter Nachrichtenagenturen (u.a. *UPI*, *AP*, *dpa*, *Reuters*). Diese Informationsquellen können einzeln, gruppenweise oder vollständig genutzt werden. Die Artikel werden mit den einstellbaren Suchkriterien von ENS-Ordnern verglichen, bei Übereinstimmung mit den Kriterien in dieselben kopiert und dort vorübergehend aufbewahrt. Nutzern stehen drei Ordner mit Verfalldaten bis zu einem Jahr zur Verfügung. Diese Ordner werden laufend ergänzt, so daß im Laufe der Zeit zu bestimmten Wissensgebieten spezifische Datenbanken aufgebaut werden.

Die Suche nach Artikeln aus ENS-Nachrichtendiensten läßt sich durch CIS-Dienst *Rosebud* ökonomisch und vielseitig gestalten. Hierbei handelt es sich um ein Hintergrundprogramm, das sich selbst täglich mehrmals mit dem Ausgangsprogramm verbindet, die gewünschten Nachrichtenquellen mit bis zu sieben, mittels Klammerausdrücken komplex kombinierbaren Suchbegriffen abfragt und die gefundenen Artikel auf die heimische Festplatte überträgt, so daß sie mit methodisch weiterführenden Programmen extern bearbeitbar sind (siehe unten).

2.6 Das WWW als Datenquelle

Von den über die Online-Dienste zugänglichen Datenbanken des Internet ist das World Wide Web (WWW) als Datenquelle zu unterscheiden (Galliker/Männel 1998). Zur Recherche von Netzressourcen stehen über 100 verschiedene Suchmaschinen zur Verfügung (siehe <http://www.albany.net/allione/>).

Eine für Inhaltsanalysen geeignete Suchmaschine ist die seit dem 15.2.1995 im Einsatz stehende AltaVista (<http://www.altavista.digital.com/>). Bei einer guten Internet-Anbindung liefert sie binnen weniger Sekunden einschlägige Informationen. Das Programm indexiert momentan (Juni 98) 17 Milliarden Wörter von circa 32 (der insgesamt circa 50 Millionen) Webseiten auf über 275.600 Servern sowie von vier Millionen Artikeln der 15.000 USENET-Newsgroups (insgesamt circa 200 Gigabytes). Pro Webseite sind circa 500 Wörter indexiert. Die Datenbasis wird von automatischen Such-Robots auf

dem neuesten Stand gehalten. Das Indexprogramm Ni2 erstellt Links zu jedem Wort auf jeder von den Robots zurückgebrachten Webseite. Die Robots reagieren auch auf die für das Auffinden der Seiten relevanten unsichtbaren Zusammenfassungen der Meta-Tags im HTML-Code (Werner 1997). Pro Tag suchen und sammeln die Robots Daten von circa drei Millionen Webseiten. Die Aktualität der Nachrichten basiert auf der hohen Hit-Quote (20 Millionen pro Tag).

3. Textanalysen

3.1 Analysen mit Hilfe der Suchmasken von CD-ROMs

Nicht alle Problemstellungen der Einstellungs- und Diskriminierungsforschung erfordern systematische Textanalysen. Oft sind eingehende Analysen von Beispielsätzen (vgl. u.a. Graumann 1995) oder qualitative Analysen einschlägiger Textmaterialien (Jäger 1996) einer bestimmten Fragestellung durchaus angemessen und wissenschaftlich sicherlich vertretbar. Freilich erlauben an Einzelbeispielen gewonnene Erkenntnisse nur Es-gibt-Aussagen, die als solche nicht falsifizierbar sind (Popper 1966). Insbesondere Arbeiten, die mit einem theoretischen Anspruch auftreten, beinhalten in der Regel generelle Aussagen, die ihrerseits systematische Textanalysen erfordern. Fehlen Angaben über Auswahl, Häufigkeit, Repräsentativität sowie Vergleichbarkeit der beigezogenen Textstellen, ist es schwer zu beurteilen, ob es sich um typische oder willkürliche Beispiele handelt bzw. ob die durch die Beispiele nahegelegenen Folgerungen einen theoretischen oder eher spekulativ-ideologischen Charakter haben.

Systematische Textanalysen sind mit den quellenspezifischen Instrumenten durchführbar. Bei den CD-ROMs werden die Suchmasken genutzt. In die Eingabefelder sind in der Regel mehrere Suchbegriffe bestimmbar und mit den *Boolschen Operatoren* verknüpfbar:

Mit dem UND-Operator erhält man Zugang zu allen Dokumenten, die den Suchbegriff vor sowie den Suchbegriff nach diesem Operator beinhalten. Nur wenn beide Ausdrücke wahr sind, ist der Gesamtausdruck wahr. Der UND-Operator ist für die Inhaltsanalyse besonders relevant, erbringt er doch Fundstellen, in denen zwei oder mehrere Themen miteinander kombiniert werden.

Verbindet man zwei oder mehrere Suchbegriffe mit dem ODER-Operator, werden sämtliche Dokumente aufgerufen, die mindestens einen der berücksichtigten Begriffe beinhalten. Wenn nur einer der angeführten Ausdrücke wahr ist, dann ist auch der Gesamtausdruck wahr. Der ODER-Operator ist inhaltsanalytisch insofern von Bedeutung, als er die Bildung von Kategorien erlaubt, die mehrere (z.B. synonyme oder familienähnliche) Wörter enthalten.

Mit dem NICHT-Operator werden bestimmte Wörter aus der Analyse entfernt. Damit lassen sich verschiedene Kategorien voneinander abgrenzen und präzisieren. Wenn bei Verwendung eines Platzhalters (z.B. *bank) Wörter erfaßt werden, die nicht zur intendierten Kategorie (z.B. Kreditinstitut) gezählt werden können (z.B. Sitzbank), sind dieselben mit dem NICHT-Operator ausschließbar.

Bereits mit diesen einfachen logischen Mitteln lassen sich relativ komplexe empirische Untersuchungen durchführen. So wurde anhand von drei Jahrgängen der Frankfurter Allgemeinen Zeitung (FAZ 1993-95) die sozialpsychologisch orientierte Delegitimierungstheorie (vgl. u.a. Bar-Tal 1989) überprüft, indem festgestellt wurde, ob bzw. inwieweit Migranten kriminalisiert werden und ob bzw. inwieweit ein Zusammenhang mit der Thematisierung von Abschiebung besteht. Hierbei wurden folgende Personenkategorien verglichen: *Ausländer**, *Asylbewerber** (*Asylsuchende**, *Asylant**), *Immigrant** (*Einwanderer**, *Zuwanderer**), *Flüchtling**, *Fremde** und als Kontrollgruppen: *Arbeitnehmer**, *Arbeitslose**, *Einwohner** (*Bewohner**), *Deutsche**. Abgeklärt wurde, wie häufig die Verhaltenskategorie *Kriminalität** (*kriminell**, *Straftat**, *Straftäter**, *straffällig**, *Delikt**, *delinquent**, *Verbrechen**, *Verbrecher**, *illegal**, *illegitim**) die Personenkategorie kontextualisiert. Im weiteren wurde das gemeinsame Auftreten der Personenkategorien mit *Abschieb** (u.a. *Ausweisung**, *Rückführung**, *Deportation**) untersucht. Die mit dem UND-Operator feststellbare Kontiguität von Personen- und Verhaltenskategorien wurde pro Artikel, Absatz, Satz, Wortpaar ausgezählt (näheres in Galliker 1996).

Das skizzierte Vorgehen bezieht sich auf verbale Ereignisse (Occurrences), die in mehr oder weniger enger Begleitung mit anderen Ereignissen (Co-occurrences) erscheinen. Unter *Co-occurrence* wird die Kontiguität verschiedener Occurrences (bzw. Wörter) pro Zählereinheit verstanden. Die systematische Analyse, die computergestützt für große Textkorpora durchführbar ist, wird Co-occurrence-Analyse genannt (näheres u.a. in Galliker/Herman/Wagner/Weimer 1996). Neben der hier behandelten *horizontalen* Co-occurrence-Analyse wurde auch die *vertikale* Co-occurrence-Analyse entwickelt, mit der verbale Übereinstimmungen in aufeinanderfolgenden Beiträgen verschiedener Dialogpartner/innen oder Teilnehmer/innen von USENET-Newsgroups systematisch untersucht werden können (vgl. Galliker/Herman/Weimer 1997, 1998). Der Vorteil der Co-occurrence-Analyse gegenüber traditionellen computergestützten Textanalysen besteht darin, daß nicht nur je besondere Kategorien, sondern von vornherein Kombinationen von Kategorien und damit thematisch komplexere Geltungsbereiche einer quantitativen Analyse zugänglich werden. Im Unterschied zum gängigen Auswertungsmodus wird die Analyse auf verschiedene Zählereinheiten bezogen, womit auf der Basis von Massendaten auch sozialpsychologisch relevante Detailinformationen systematisch aufarbeitbar sind.

Die CD-ROMs vieler Tageszeitungen erlauben allerdings nur Kombinationen weniger Suchwörter. Zudem ist in der Regel die maximale Häufigkeit des Wortvorkommens begrenzt. Ferner sind mit den meisten Rechercheprogrammen nur die Co-occurrences pro Artikel auszählbar. Besteht keine Zugriffsoption auf kleinere Textteile wie Absätze oder Sätze, müssen die Häufigkeiten der Co-occurrences bezüglich dieser Zählheiten von Auswerter/innen bestimmt werden. In diesem Fall sind die aus einer Liste abrufbaren Artikel einzeln auf den Bildschirm zu holen und bezüglich der verschiedenen Texteinheiten auszuzählen. Dieses Verfahren erlaubt es, Bedeutungen von Wörtern in ihrem Kontext zu kontrollieren bzw. Validitätsprobleme zu minimieren (siehe unten). Die Auswertung ist sicherlich aufwendiger als bei einer automatischen Auszählung, doch gemessen an traditionellen Textanalysen ist der Zeitaufwand immer noch gering, liegt doch das zuvor mittels Keywords thematisch eingegrenzte Textmaterial in verdichteter, temporal geordneter und elektronisch leicht abrufbarer Form vor. So konnten vergleichsweise umfangreiche Untersuchungen zur verbalen Kontextualisierung von Migranten in der FAZ 1993-95 und in der NZZ 1993 sowie zur positiven und negativen latenten Evaluation verschiedener Altengruppen in der FAZ 1993-95 in relativ kurzer Zeit realisiert werden (Galliker 1996; Galliker/Klein 1997; Galliker/Weimer 1996).

Die anhand von CD-ROMs gewonnenen Daten sind exportierbar und können mittels Zusatzprogrammen weiter analysiert werden. Allerdings gestatten nicht alle CD-ROMs einen globalen Export der aussortierten Artikel. Bei einigen Printmedien sind ausgewählte Artikel nur einzeln auf die Festplatte übertragbar (z.B. FAZ 1993-97; NZZ 1994). Hingegen können beispielsweise bei der New York Times bis zu 3.000 Artikel miteinander exportiert und auf der eigenen Festplatte gespeichert werden.

Nur unter der Voraussetzung eines globalen Exports ist in absehbarer Zeit eine nachträgliche Recherche großer Textmengen mit einem externen Programm möglich (siehe unten). Dieser Auswertungsmodus ist z. B. bei Längs- und/oder Querschnittsvergleichen von Printmedien verschiedener Länder erforderlich. Hierzu notwendige Vergleiche sind mit relativ geringem Aufwand und ohne die bei Längsschnittstudien sonst üblichen Ausfälle durchführbar.

3.2 Analysemöglichkeiten mit Hilfe von Suchdiensten

Einige Online-Dienste bieten Recherchemöglichkeiten, die mit jenen der CD-ROMs durchaus vergleichbar sind. So werden beispielsweise bei CompuServe mit der Definition des Suchkriteriums AND-, OR und/oder NOT-Operatoren eingesetzt. Durch die Formulierung der Suchaufträge in Klammersausdrücken sind relativ komplexe Recherchen möglich.

Im Unterschied zu Untersuchungen, die mittels CD-ROMs durchgeführt werden, besteht bei der Analyse von Online-Datenbanken oft Unklarheit über die Population. Bei unvollständiger Online-Repräsentation der Druckausgabe sind die Auswahlkriterien unbekannt und/oder unausgewiesen. Das Problem der Repräsentativität bleibt hier auch bei Einzelanalysen bestehen. Zwar gibt es Online-Dienste, die Volltexte zur Verfügung stellen (z.B. bei *T-Online* die *dpa*-Weltnachrichten), doch ist bei diesen Diensten noch keine Stichwortsuche mit Booleschen Operatoren möglich und/oder die für weitergehende Analysen relevanten Exportmöglichkeiten sind limitiert.

3.3 Analysen mit Hilfe von WWW-Suchmaschinen

Die für Recherchen relevanten Operatoren von Suchmaschinen sind in der Regel einfach zu handhaben. Zu den wichtigsten methodischen Prinzipien gehören die Inklusion und Exklusion von Wörtern. Bei AltaVista wird auf dieselben wie folgt hingewiesen:

„To make sure that a specific word is always included in your search topic, place the plus (+) symbol before the keyword in the search box. To make sure that a specific word is always excluded from your search topic, place a minus (-) sign before the keyword in the search box. Example: To find recipes for cookies with oatmeal but without raisins, try recipe cookie +oatmeal -raisin.“ (<http://altavista.digital.com/av/content/help.htm>).

Mit dem Oder-Operator kann ein Bedeutungsfeld (etwa synonyme und familienähnliche Wörter) abgesteckt werden. Auf der Ebene der Buchstabenfolgen werden die Variationsmöglichkeiten mit einem Platzhalter (*) bestimmt (Beispiel: mit wish* wird auch wish, wishes, wishful, wishbone usw. registriert). Bei der Eingabe des Wortstammes werden mit Vorteil nur Kleinbuchstaben verwendet, was folgendes Beispiel verdeutlicht:

„Wenn you search for paris, you'll find Paris, paris, and PARIS in our result pages. However, when you search for Paris, you'll only see Paris in the result pages.“ (<http://altavista.digital.com/av/content/help.htm>)

Mit der LiveTopic-Software von AltaVista (Java-Oberfläche) werden zu einem Suchbegriff die am häufigsten miteinander auftretenden Begriffe in ihrer numerischen Relation dargestellt. Sozusagen durch alle zugänglichen Seiten des Web hindurch ergibt sich eine Art komprimierter verbaler Kontext des Suchbegriffs, wobei deren Elemente in ihrer Vernetzung erscheinen. Visualisiert wird nach Kriterien wie Standort, Nähe und Häufigkeit der einzelnen Occurrences. Die resultierenden Strukturbilder sind vor allem für heuristische Zwecke verwendbar (näheres in Galliker/Männel 1998).

Qualität und Differenziertheit einer WWW-Recherche bestimmt sich durch die Anzahl verwendbarer Suchbegriffe sowie durch die Komplexität der Kombination dieser Begriffe. AltaVista erlaubt begrenzte Kombinationsmöglichkeiten, die aber für die Beantwortung

einfacherer Fragestellungen genügen. So wurde die Delegitimierungstheorie u.a. anhand der Repräsentation älterer Menschen im Web überprüft und festgestellt, inwieweit die Konnotationen vergleichbar sind mit jenen, die in der Untersuchung der FAZ 1993-95 (vgl. Galliker/Klein 1997) eruiert wurden.

WWW-Recherchen sind in methodologischer Hinsicht nicht unproblematisch. Mit Ausnahme von *Hotbot* erreichen Suchmaschinen lediglich einen Teil des Web. Als Zähleinheit steht oft nur die relativ große Texttheinheit einer Web-Seite zur Verfügung (bei Alta-Vista sind über den NEAR-Operator Wortabstände bis zu maximal acht Wortheinheiten möglich). Ferner ist das WWW in Bewegung, was die Bestimmung von Reliabilität und Objektivität in Frage stellt. Allenfalls könnten Momentaufnahmen auf CD-ROMs übertragen und zu Reproduktionszwecken verwendet werden.

3.4 Analysen mit externen Programmen

Mit Hilfe von WWW-Suchmaschinen oder Suchmasken von CD-ROMs können nur einzelne Co-occurrences von zwei (allenfalls drei oder vier) Occurrences ausgezählt werden. Für weniger punktuelle Untersuchungen müssen externe Programme verwendet werden. Als Beispiel sei hier TEXTPACK angeführt. Dieses Programm erschien zuerst in einer DOS-Version (vgl. u.a. Züll/Mohler/Geis 1991) und steht inzwischen auch in einer Windows 95/NT-Version zur Verfügung (Mohler/Züll 1998), wobei die Funktionalität weitgehend gleich geblieben ist (nähere Informationen bei Cornelia Züll erhältlich; textpack@zuma-mannheim.de). Mit der für die Co-occurrence-Analyse wesentlichen Subroutine TAGCODER von TEXTPACK ist für Kategorien umfangreicher Lexika simultan feststellbar, ob bzw. wie häufig zwei oder mehr Kategorien pro Zähleinheit erscheinen, so daß abstrahierend von den sehr häufigen bedeutungsarmen Wörtern (Artikel, Hilfsverben, Modalverben usw.) der Textkorpus annäherungsweise flächendeckend registrierbar ist (vgl. u.a. Galliker/Herman/Wagner/Weimer 1996). Die Auszählung der Häufigkeiten verbaler Co-occurrences ist auf verschiedene Zähleinheiten beziehbar (u.a. auf Sätze, Absätze und Artikel). In der Folge wird berechnet, wie oft zwei oder mehrere Kategorien in Texteinheiten verschiedener Größen auftreten und welche Verteilungsmuster resultieren. Aus dem Vergleich der pro Zeitabschnitt (Jahrgang, Quartal usw.) eines Printmediums in einen multidimensionalen Raum projizierten verbalen Konstellationen lassen sich Bewegungen der Verteilungsmuster erfassen (Galliker/Herman/Weimer/Wagner 1996). Die relativen Häufigkeiten verbaler Besetzungen kleinerer respektive größerer Zähleinheiten verweisen auf intentionelle bzw. auf inzidentelle verbale Konstellationen (vgl. u.a. Galliker 1996). Aufgrund von Längsschnittstudien ist eruiert, inwieweit sich die Distributionen (z.B. Ansammlungen von unmittelbaren Attributionen)

verändern. Auf diese Weise eröffnen sich auch Perspektiven hinsichtlich der Untersuchung impliziter Diskriminierungen (näheres in Galliker/Herman/Imminger/Weimer 1998).

Zur Illustration sei ein vom Schweizerischen Nationalfonds finanziertes Forschungsprojekt zur Repräsentation der Schweiz in der New York Times angeführt. Ausgangspunkt der von Jan Herman (Mannheim) ausgewerteten Untersuchung ist die Frage, mit welchen Wörtern die früher in der New York Times fast ausschließlich mit ökonomischen und betriebswissenschaftlichen Kategorien kontextualisierte Schweiz vorzugsweise verbunden wird, nachdem das Verhalten maßgebender Kreise dieses Landes zur Zeit des Zweiten Weltkrieges allgemein bekannt wurde. In einem ersten Auswertungsschritt werden sämtliche Artikel, in denen der Signifikant Schweiz erscheint (Suchwörter: *Switzerland; swiss; switzerland; switzerland's*), von den CD-ROMs (Jahrgänge 1995-97) auf die Festplatte übertragen. In einem zweiten Schritt werden Worthäufigkeitslisten erstellt. Im weiteren werden Wörter gleicher Wortstämme gruppiert. Die Häufigkeiten sowie die Bedeutung der Wortgruppen bilden die Grundlage des Wörterbuches. Bei der Bildung des Lexikons werden sehr häufig vorkommende Wörter unspezifischer Bedeutung (*der, die, das, haben, sein, gehen, dürfen*, usw.) nicht berücksichtigt. Im weiteren werden relativ selten auftretende Einzelwörter aus der quantitativen Analyse ausgeschlossen. Im vorliegenden Fall werden Häufigkeiten > 300 berücksichtigt, was insgesamt 51 Kategorien ergibt (siehe Tabelle 1). Anschließend werden die Wortkategorien kombiniert und die Häufigkeiten sämtlicher Co-occurrences in einer Dreieckstabelle aufgelistet.

Tabelle 1: Ausschnitt aus einem Wörterbuch*

Kategorie	Wörter	Total
1	America; America's; American; Americans	1541
2	artist; artist's; artistic; artists; arts; arts	417
3	Bank; Banker; Bankers; Banking, bank; bank's, banker, bankers, banking; Bundesbank; Bundesbank's; bundesbank; bundesbanks; citibank; Citibank's, citybank; citybank's; Banque; banque	1827
4	business; Business; businesses; businessman; businessmen; Businessman; Businessmen	486
51	School; Schools; school; schools	359

* Konstitutionsbasis: New York Times (195-97). Keywords: Switzerland; swiss; switzerland; switzerland's.

Aus der Studie geht hervor, daß in der New York Times der letzten Jahre im Kontext der Schweiz am häufigsten Kombinationen der Wörter *Bank, Germany, Money, Gold, Holocaust* und *Jews* erschienen. Anhand der Anzahl Co-occurrences pro Quartal verdeutlicht sich die Entstehung der gegenwärtigen Holocaust-Kontroverse. Schon ein Jahr bevor die

Kombination *Gold-Holocaust* mit der Schweiz in Verbindung gebracht und die veränderte Einstellung vieler Amerikaner/innen gegenüber der Schweiz in diesem Land überhaupt zur Kenntnis genommen wurde, erschien in der New York Times u.a. die Kombination *Bank-Money* überzufällig häufig. In den entsprechenden Artikeln wird einerseits über fragwürdige aktuelle Geldtransfers der Schweizer Großbanken (Drogenhandel, Geldwäscherei) und andererseits über deren Konzentrationsbestrebungen berichtet, was hinsichtlich einer Interpretation des späteren Diskurses (US-Boycott der Schweizer Geldinstitute) aufschlußreich sein könnte.

4. Schlußbemerkungen

Die sozialwissenschaftliche Datenlage hat sich in den letzten Jahren erheblich verbessert und wird sich weiterhin verbessern. Die neuen Informationsquellen können allerdings nur gewinnbringend genutzt werden, wenn die bereits beachtlichen elektronischen Zugriffsmöglichkeiten und Textverarbeitungsverfahren mit der Entwicklung der Datenquellen Schritt halten. Die mit TEXTPACK und anderen Programmen ermöglichte systematische Analyse von horizontaler sowie vertikaler Co-occurrences eröffnet Perspektiven für die computergestützte Auswertung sehr großer Datenmengen, insbesondere für die Komprimierung der bei der Massenkommunikation transportierten redundanten Informationen. Detailstudien sind ebenfalls durchführbar, indem quasi trichterförmig vorgegangen wird und aussagekräftige Einzelbeispiele extrahiert werden (vgl. Wagner/Galliker/Weimer 1997). Nicht selten sind in der gleichen Studie qualitative sowie quantitative Aspekte auswertbar (vgl. u.a. Galliker/Imminger/Weimer/Bock 1997). Für computergestützte qualitative Datenanalysen im engeren Sinne stehen besondere Programme zur Verfügung (u.a. ATLAS/ti).

Die drei wichtigsten Voraussetzungen zur Bearbeitung von Massendaten mit elektronischen Mitteln sind bei den skizzierten elektronischen Informationsquellen meistens bereits erfüllt:

1. Die Suche erfolgt mit den Booleschen Operatoren.
2. Die Daten sind en bloc vom Lieferanten auf eigene Datenbanken übertragbar.
3. Programme sind beziehbar, mit denen eine Vielzahl von Suchbegriffen behandelt werden.

Gewichtiger als die zwar noch bestehenden, aber jährlich sich verringernden technischen Probleme, die kaum mehr einen Beleg für eine negative Einschätzung der neuen sozialwissenschaftlichen Möglichkeiten bieten, scheinen methodische Probleme zu sein. Neben statistischen Fragen (u.a. Gewichtung der Co-occurrences bei unterschiedlichen Häufig-

keiten der Occurrences), die zumindest prinzipiell beantwortbar sind (vgl. u.a. Galliker 1997), gibt es eine Reihe von Schwierigkeiten, die in Zukunft differenzierter bearbeitet werden müssen.

Die größten Probleme zeigen sich bei der *Bildung des Wörterbuches*. Sie sind teilweise eliminierbar (wenngleich nicht lösbar), indem entweder auf die Verwendung des ODER-Operators (und damit auf eine durch synonyme und familienähnliche Wörter angereicherte Wortkategorie) verzichtet wird (vgl. u.a. Galliker/Klein 1997), oder indem die Wortkategorien nach einer prototypischen Methode auf der Grundlage von Hege (1987) bestimmt werden (vgl. u.a. Galliker et al. 1996). Das erste Verfahren ist nur sinnvoll bei einem *a-priori-Ansatz*, der auf der Basis einer mehr oder weniger ausformulierten Theorie von relativ genauen Annahmen ausgeht. Die Zuordnung erfolgt, indem ein theoretisch bestimmtes Lexikon, aus wenigen Wortkategorien bestehend, mit dem Textkorpus abgeglichen wird. Demgegenüber wird bei einem *empirizistischen Ansatz* das Wörterbuch erst im nachhinein aufgrund der Häufigkeiten der im Textkorpus vorkommenden Wörter festgelegt. Im Prinzip kann jedes signifikante Wort zusammen mit seinen Variationen eine eigene Wortkategorie bilden (vgl. Tabelle 1). Wird eine möglichst flächendeckende Analyse auf der Ebene signifikanter Einzelwörter angestrebt, ist ein Programm mit großer Kapazität bezüglich des Lexikons erforderlich. Bei kleinen Textkorpora können Synonym- und Lemmatisierungslisten eingesetzt werden, um die für eine Co-occurrence-Analyse notwendigen Häufigkeiten zu erreichen.

Bei den skizzierten computergestützten Verfahren werden die Probleme der *Reliabilität* und *Objektivität* minimalisiert, während das Problem der *Validität* bestehen bleibt. Doch scheint die Frage, ob tatsächlich das gemessen wird, was man zu messen vorgibt, beantwortbar zu sein, wenn in theoretischer sowie in methodologischer Hinsicht vom Primat des Signifikanten, also von der (computerlesbaren) Buchstabenfolge, ausgegangen wird. Demnach wird eine Bedeutung nicht über das Signifikat, sondern über andere Signifikanten in dessen Kontext bestimmt (näheres in Galliker 1990). Das heißt, je kleiner die Zählleinheit ist bzw. je näher fragliche Occurrences beieinander liegen, desto größer ist die Wahrscheinlichkeit, daß sie tatsächlich in der angenommenen Bedeutung erscheinen. Mit Keyword-in-Context (KWIC) von TEXTPACK (näheres u.a. in Züll et al. 1991) ist eine eingehende, wenngleich relativ aufwendige Validitätsprüfung möglich: Sämtliche Textstellen, die ein fragliches Wort enthalten, werden in ihrem unmittelbaren Kontext untereinander aufgelistet, wodurch überprüfbar ist, ob der Eintrag im Wörterbuch semantisch unmißverständlich ist oder nicht. Eindeutige Zuordnungen bzw. trennscharfe Kategorien gehören zu den wichtigsten Bedingungen der Möglichkeit *interner Validität*. Eine Untersuchung gilt als intern valide, wenn es möglich ist, eine realisierte Veränderung der

abhängigen Variablen auf die Veränderung einer oder mehrerer unabhängigen Variablen und/oder auf deren Interaktionen zurückzuführen (vgl. Gadenne 1984).

Die Frage nach der Grundgesamtheit stellt sich bei Einzelanalysen von im Volltext vorliegenden Printmedien in der Regel nicht. Hingegen löst sich bei Vergleichsstudien das Problem der *externen Validität* bzw. der *Repräsentativität* keineswegs auf, sondern scheint sich mit den anwachsenden Selektionsmöglichkeiten eher noch zu verschärfen. Das Problem kann pragmatisch angegangen werden, indem auf eine ausgewogene Auswahl der Zeitungen geachtet wird. Damit soll eine möglichst homogene Basis konstituiert werden, auf der die relativen Häufigkeiten verglichen werden. Eine materialbedingte einseitige Auswahl ist vermeidbar, indem nur gespeicherte Tageszeitungen (vergleichbarer Auflage, Distribution usw.) eines bestimmten Gebietes berücksichtigt werden, die sich auf gemeinsamer Basis möglichst komplementieren, so daß die hypothetisch relevanten Variablen gegeneinander getestet werden können. Es stellt sich aber die Frage, ob Prognosen möglich sind, wenn die je besonderen Voraussetzungen der verfügbaren Medien auch bei einem Verzicht auf divergente Textkorpora in nicht annähernd variierbarer Form vorliegen. Eine Lösungsperspektive dieses Problems bietet die in Absetzung von der traditionellen Populationslogik entwickelte *Typikalitätslogik*, wonach sich der Geltungsbereich einer Aussage nicht über den tatsächlich untersuchten Bereich hinaus auf rein statistischem Wege erweitern, sondern nur mittels zusätzlicher Forschungsarbeiten von Fall zu Fall neu abstecken läßt (näheres in Foppa, 1986). Eine elektronische Version dieses Vorgehens läßt sich wie folgt skizzieren: IQuest von CIS erlaubt das 'Überfliegen' ganzer Datenbanken. Hierbei wird abgeklärt, in welchen Datenbanken ein bestimmter Suchbegriff vorkommt. Nach der Bezeichnung einer bestimmten Datenbankgruppe sowie der Eingabe des Suchkriteriums wird ermittelt, ob die Häufigkeit einer Co-occurrence im Verhältnis zur maßgebenden Occurrence im neuen Geltungsbereich (andere Typen von Zeitungen oder Zeitschriften) ungefähr gleich bleibt oder ob sich signifikante Abweichungen ergeben.

Korrespondenzanschrift

*PD Dr. Mark Galliker
Psychologisches Institut der Universität Heidelberg
Hauptstr. 47-51
69117 Heidelberg
Fax und Tel 06221/54 77 03
e-mail: ZBJ1@psi-sv1.psi.uni-heidelberg.de*

Literatur

- al Wadi, D., 1995: Comas Benutzerhandbuch. Mannheim: IDS.
- Bar-Tal, D., 1989: Delegitimization: The Extreme Case of Stereotyping and Prejudice. S.169-182 in: D. Bar-Tal/C.F. Graumann/A. Kruglanski/W. Stroebe (Hrsg.), Stereotyping and Prejudice: Changing Conceptions. New York: Springer.
- Blank, Th./Heinrich, H.-A., 1998: "Wienerwald auf Mallorca": Negative Emotionen von Deutschen gegenüber ihrer eigenen Nation. ZUMA-Nachrichten 42: 120-143.
- Bock, H./Zafirov, B./Prießhäuser, B./Gigler, B./Künzel-Jacob, H./Stelzer, M./Wallner, C./Nowak, W./Landersdorfer, U., 1992: Aids in der Presse. Regensburg: Roderer & Welz.
- Cornellissen, J., 1997: Tageszeitungen Online. Ziele, Strategien und Erfahrungen von Web-Zeitungen - Eine E-Mail-Befragung unter Online-Redaktionen. Zeitschrift für Sozialpsychologie 25: 81.
- Cronin, B./Davenport, E./Martinson, A., 1997: Women's studies: bibliometric and content analysis of the formative years. Journal of Documentation 53: 123-138.
- Davenport, E., 1995: Who cites women? Whom do women cite? An exploration of gender and scholarly citation in sociology. Journal of Documentation 51: 404-410.
- Foppa, K., 1986: 'Typische Fälle' und der Geltungsbereich empirischer Befunde. Schweizerische Zeitschrift für Psychologie 45: 151-163.
- Gadenne, V., 1984: Theorie und Erfahrung in der Psychologischen Forschung. Tübingen: Mohr.
- Galliker, M., 1990: Sprechen und Erinnern. Göttingen: Hogrefe.
- Galliker, M., 1996: Delegitimierung von Migranten im öffentlichen Diskurs. Co-occurrence-analysis von drei Jahrgängen einer Tageszeitung mit CD-ROMs als Datenquelle. Kölner Zeitschrift für Soziologie und Sozialpsychologie 48: 704-724.
- Galliker, M./Herman, J., 1994: Verbales Reflektieren im therapeutischen Diskurs. Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie 42: 42-59.
- Galliker, M./Herman, J./Weimer, D., 1997: Vertical and horizontal Co-occurrence-analysis. Empirical examples on full-text research using TEXTPACK. S. 243-250 in: W. Bandilla/F. Faulbaum (Eds.), Advances in Statistical Software 6. Stuttgart: Lucius & Lucius.
- Galliker, M./Herman, J./Weimer, D., 1998: Mutual Reference in Dialogue. Vertical Co-occurrence-analysis of a Conversation Series. S. 111-116 in: S. Cmejrková/Hoffmannová, J./Müllerová/Svetlá (Eds.), Dialoganalyse VI. Proceedings of the 6th Conference Prague 1996. Tübingen: Niemeyer.
- Galliker, M./Herman, J./Imminger, K./Weimer, D., 1998: The investigation of contiguity: Co-occurrence analysis of printmedia using CD-ROMs as a new data source, illustrated by a discussion on migrant delinquency in a daily newspaper. Journal of Language and Social Psychology 17: 200-219.

- Galliker, M./Herman, J./Wagner, F./Weimer, D., 1996: Latente Abwertung sozialer Gruppen im öffentlichen Diskurs. Zur Co-occurrence-Analyse von vier Jahrgängen einer Tageszeitung. *Medienpsychologie* 8: 3-20.
- Galliker, M./Imminger, K./Weimer, D./Bock, H., 1997: Intensivierung des Diskurses durch Verneinung: Quantitative und qualitative Analyse der Co-occurrences von 'Soldaten' und 'Mörder' in der FAZ im Vergleich mit der NZZ. S. 99-119 in: A. Disselnkötter/S. Jäger/H. Kellershohn/S. Slobodzian (Hrsg.), *Evidenzen im Fluss*. Duisburg: DISS.
- Galliker, M./Klein, M., 1997: Implizite positive und negative Bewertungen älterer Menschen. Eine Kontextanalyse der Personenkategorien 'Senioren', 'ältere Menschen', 'alte Menschen' und 'Greise' bei drei Jahrgängen einer Tageszeitung. *Zeitschrift für Gerontopsychologie und -psychiatrie* 10: 27-41.
- Galliker, M./O. Männel, 1998: Suchmaschinen als Datenerhebungsinstrument von WWW-Inhaltsanalysen. Möglichkeiten und Grenzen. In: B. Batinic/L. Graef/A. Werner/W. Bandilla (Hrsg.), *Online Research - Methoden, Anwendungen und Ergebnisse*. Göttingen: Hogrefe.
- Galliker, M./Wagner, F., 1995: Implizite Diskriminierungen und Antidiskriminierungen anderer Menschen im öffentlichen Diskurs. Zu den Aussiedlungs- und Übersiedlungsdebatten im deutschen Bundestag. *Zeitschrift für Politische Psychologie* 3: 69-86.
- Galliker, M./Weimer, D., 1996: Benachteiligung von Migranten im Mediendiskurs. Co-occurrence-Analyse einer auf CD-ROM vorliegenden Tageszeitung. *Zeitschrift für politische Psychologie* 4: 301-318.
- Graumann, C.F. , 1995: Discriminatory discourse. *Patterns of Prejudice*: 29-83.
- Grawe, K./Caspar, F.M./Ambühl, H., 1990: Die Berner Therapievergleichsstudie: Wirkungsvergleich und differentielle Indikation. *Zeitschrift für Klinische Psychologie* 19: 316-337.
- Hege, G., 1987: Einfache Bedeutungskategorien: Ein Beitrag zur computergestützten Textanalyse psychotherapeutischer Texte. Berlin: Springer.
- Herberg, D., 1993: Die Sprache der Wendezeit als Forschungsgegenstand. *Muttersprache*, 103: 264-266.
- Herberg, D./Stickel, G., 1992: Gesamtdeutsche Korpusinitiative. Ein Dokumentationsprojekt (1989/90). *Deutsche Sprache* 20: 182-192.
- Helbig, G., 1988: *Lexikon deutscher Partikel*. Leipzig: VEB Verlag Enzyklopädie.
- Jäger, S., 1996: Der Großregulator - oder: wie die "Bild"-Zeitung das Bewußtsein der Menschen reguliert. S. 478-500 in: F. Deppe/G. Fülberth/G. Rilling (Hrsg.), *Faschismus*. Heilbronn: Diestel.
- Iker, H.J./Harway, N.L., 1969: A Computer System Approach Toward the Recognition and Analysis of Content. S. 381- 406 in: G. Gerbner (Hrsg.), *The Analysis of Communication Content*. New York: University Press.

- Klingemann, H. D., 1984 (Hrsg.): Computerunterstützte Inhaltsanalyse in der empirischen Sozialforschung. Frankfurt a.M.: Campus.
- Kolvenbach, M., 1987: Kleine Chronologie der Arbeiten am Handbuch-Korpus. S. 4-9 in: R. Schmidt (Hrsg.), Informationsschrift der Arbeitsstelle Linguistische Datenverarbeitung 1986/87, LDV-Info 6. Mannheim: Institut für Deutsche Sprache.
- Mohler, P. Ph./Züll, C., 1998: TEXTPACK PC. User's Guide. ZUMA, Mannheim.
- Nissan, E./Schmidt, K., (Hrsg.), 1995: From Information to Knowledge. Conceptual and Content Analysis by Computer. Oxford: Cromwell.
- Popper, K., 1966: Logik der Forschung. Tübingen: Mohr.
- Riehm, U./Böhle, K./Gabel-Becker, I./Wingert, B., 1992: Elektronisches Publizieren. Eine kritische Bestandsaufnahme. Berlin, Heidelberg: Springer.
- Ring, P., 1997: Zeitung im Internet. Ein Vergleich der Olympiaberichterstattung 1996 von "Die Welt" und "Welt online". Transfer, 2/3, 77.
- Schütz, H./Six, B., 1994: Mehr als 7000 Seiten Sozialpsychologie: eine Zeitschrift in der Retrospektive. Zeitschrift für Sozialpsychologie 25: 5-17.
- Stone, P., 1966: The General Inquirer - A Computer Approach to Content Analysis. Cambridge: Mass.
- Van Dijk, T.A., 1991: Racism and the Press. London: Routledge.
- Wagner, F., 1996: CD-ROM und computerlesbare Zeitungskorpora als Datenquellen für Linguisten. S. 78-87 in: B. Ruschoff/U. Schmitz (Hrsg.), Kommunikation und Lernen mit alten und neuen Medien. Bern: Lang.
- Wagner, F./Huerkamp, M./Jockisch, H./Graumann, C. F., 1993: Sprachliche Diskriminierung. S. 281-288 in: H. Löffler (Hrsg.), Dialoganalyse IV. Referate der 4. Arbeitstagung. Tübingen: Niemeyer.
- Wagner, F./Galliker, M./Weimer, D., 1997: Implizite sprachliche Diskriminierung von Ausländern zur Zeit der Wende. S. 230-240 in M. Jung/M. Wengeler/K. Böke (Hrsg.), Die Sprache des Migrationsdiskurses. Opladen: Westdeutscher Verlag.
- Werner, A., 1997: Pole-Position. ZD Internet Professionell, Juni, 93-95.
- Züll, C./Mohler, P.Ph./Geis, A., 1991: Computerunterstützte Inhaltsanalyse mit TEXTPACK PC. Stuttgart: Fischer.