

2.0 Maschinelle Texthaltung und Textverarbeitung als Voraussetzung computer-unterstützter Prozessforschung

2.1 Der Ablauf der Datenerfassung

2.2 „Elektronische Verbalanalyse“ Ein Programmsystem zur maschinellen Inhaltsanalyse von Texten (EVA)

2.3 Der Ablauf der Datenauswertung

2. Maschinelle Texthaltung und Textverarbeitung als Voraussetzung computer-unterstützter Prozessforschung

Die psychotherapeutische Forschung geriete spätestens dann, wenn sie Verbatim-Protokolle, die wichtigste methodische Neuerung der letzten Jahre, zur Grundlage der Forschung macht, in kaum zu bewältigende Schwierigkeiten, eine adäquate Verarbeitung der verfügbaren Daten zu erreichen. Ganz abgesehen von den immer noch bestehenden Widerständen der meisten Therapeuten, über Ton oder Filmaufzeichnungen ihrer Behandlungen eine breite Basis für vergleichende Forschung zu ermöglichen, sind die bestehenden Aufzeichnungs- und Auswertungsprozeduren sehr zeitraubend. So schreiben BECKMANN et al. (1974):

„Als praktisches Problem ist der hohe zeitliche Aufwand zu erwähnen: für eine Therapiestunde werden etwa 17 Stunden für die Verschlüsselung benötigt“ (S. 192).

Da bei diesem großen zeitliche Aufwand eine systematische Verlaufsforschung nur unter großen Schwierigkeiten realisierbar wäre, haben wir in den letzten Jahren ein Datenverarbeitungssystem zur elektronischen Inhaltsanalyse aufgebaut. Die Teamarbeit von Psychoanalytikern, Psychologen und Informatikern war hierzu eine unerlässliche Bedingung, um die vielfältigen Probleme zu lösen, die sich beim Aufbau einer zeitlich und ökonomisch vertretbaren Methodik stellen.

2.1 Der Ablauf der Datenerfassung

„In der psychoanalytischen Behandlung geht nichts anderes vor als ein Austausch von Worten zwischen dem Analysierten und dem Arzt“ (FREUD, S. 1917, S. 98).

Auch wenn die Verbalisierung in der psychoanalytischen Situation nicht das Gesamt der Interaktion umfasst, können wir FREUD's Bemerkung sicherlich dahingehend verstehen, dass der Gedankenaustausch in Form eines Dialoges in der Psychoanalyse ohne Zweifel einen zentralen Platz einnimmt. Wir haben an anderer Stelle uns mit den besonderen Möglichkeiten auseinandergesetzt, die eine Aufzeichnung des therapeutischen Gespräches für die psychoanalytische Prozessforschung bietet (KÄCHELE, SCHAUMBURG u. THOMÄ, 1973). Ohne hiermit ausschließen zu wollen, dass andere Fragestellungen andere Beobachtungsmethoden verlangen mögen, so haben wir uns in diesem Projekt mit der Analyse von Verbatimprotokollen auf eine bestimmte Ebene des wissenschaftlichen Fragens eingelassen und stehen somit wie eine Reihe anderer Kollegen vor den methodischen Problemen, die sich aus der Verwendung von solchen sehr umfangreichen Protokollen ergeben. Wie das folgende Schema darstellt, waren bisher die Möglichkeiten, solche Protokolle zu sammeln, zu archivieren und später dann auszuwerten, durch die Anwendung „einfachster“ Methoden gekennzeichnet. Das Gespräch wird von einem Tonband aufgezeichnet. Darauf wird es von einer Sekretärin verbatim geschrieben und wandert in Aktenordner, die über die Jahre immer mehr an Umfang zunehmen und dabei ständig an Zugänglichkeit im Sinne von Verfügbarkeit über das Archivierte abnehmen.

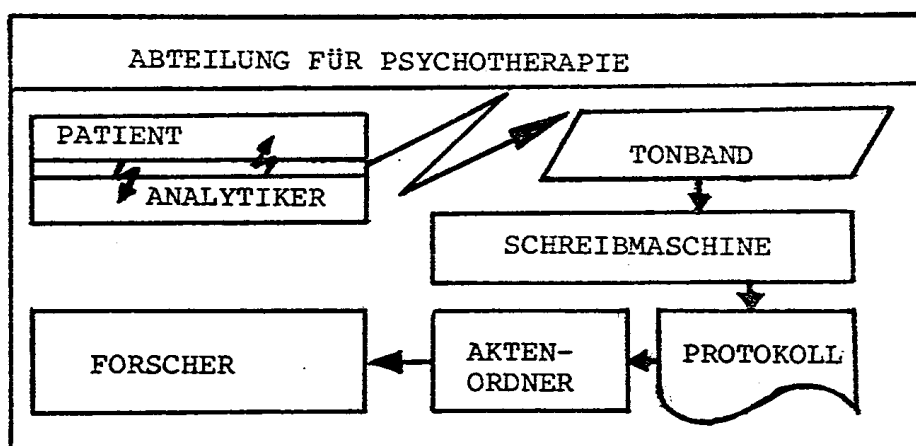


Abbildung 6

Betrachtet man diesen Ablauf als typisch für die bisherige Forschungspraxis, so lassen sich die Klagen über den „enormen Zeit- und Kostenaufwand“, die sich in allen einschlägigen Arbeiten finden (z.B. WALLERSTEIN und SAMPSON,

1971, S. 26) leicht verstehen. COHEN und COHEN (1961), die sich ebenfalls mit einer großen Zahl transkribierter Interviews beschäftigt haben (N = 700 von drei Patienten) resümieren ihren Verzicht folgendermaßen:

„Aber die Ergebnisse dieses unglaublichen Aufwandes an Zeit, gedanklicher Arbeit und Gefühl haben ihren Niederschlag nur in unseren beruflichen Karrieren gefunden, aber nicht direkt in wissenschaftlichen Publikationen. Obwohl wir einige Arbeiten veröffentlicht haben, sind wir am Ende in der Masse der Daten ertrunken“ (S. 47).

Die hier angeschnittenen Probleme sind ubiquitär und so ist es nicht verwunderlich, dass trotz einer nun über 20-jährigen Forschungstradition die psychoanalytische Prozessforschung noch immer vom Exemplarischen lebt, d.h. es gibt noch keine „primäre Datenbank“, wie dies LUBORSKY und SPENCE (1971, S. 426) genannt haben. Für eine solche Datenbank, die entsprechend den Archiven in der Geschichtswissenschaft, den Bibliotheken etc. Forschungsmaterial zugänglich machen würde, stellen sie folgende Bedingungen auf:

„Idealerweise sollten zwei Bedingungen erfüllt sein: der Fall sollte eindeutig als psychoanalytischer Fall gekennzeichnet sein, was immer auch eine Gruppe von Experten an Kriterien anlegen wird; und die Daten sollten mit Tonband aufgezeichnet sein, verschriftet und so indexiert sein, dass die Zugänglichkeit und die Transparenz des Materials maximiert ist“ (S. 426).

Dieser Stand der Dinge erscheint verwunderlich, um so mehr als die maschinelle Datenverarbeitung ähnliche Probleme, nämlich Textverarbeitung größeren Umfanges, für andere Bereiche der Sozialwissenschaften längst in Angriff genommen hatte und eine Reihe von Lösungsmöglichkeiten bereits vorliegen oder z.Zt. entwickelt werden.

Der amerikanische Psychologe Neal MILLER hatte schon früh die Bedeutung des Computers für die Sozialwissenschaften erkannt, als er 1957 „in seltenen Momenten der Begeisterung“ die Entwicklung der Computertechnologie mit der Bedeutung der Entdeckung des Mikroskopes für die Biologie verglich (1957, S. 44).

In der Tat brachte die Einführung der maschinellen Datenverarbeitung die Möglichkeit, in einer Art und Weise mit Daten umzugehen, die dem neuen Verständnis der Beziehung von Beobachtung und Theorie zu entsprechen scheint. Die Verfügbarkeit über einmal akkumulierte Daten erlaubt ein wiederholtes Erproben verschiedener theoretischer Ansätze. Beobachtungsdaten werden somit zur Verfügung des Forschers gestellt; anstatt durch sie erdrückt zu werden, kann er sie zur Überprüfung manipulieren.

Wir wollen anhand unserer eigenen Erfahrung den Aufbau einer Datenbank und eines Forschungssystems schildern, welches es erlaubt, einer Vielzahl von

Fragestellungen nachzugehen, die nur mit Hilfe von Verbatimprotokollen gelöst werden können.

Im folgenden gehen wir nun auf den Ablauf des Datenerfassungsprozesses ein, wie er in Zusammenarbeit mit dem Rechenzentrum der Universität Ulm realisiert werden konnte. Das folgende Datenflussdiagramm zeigt, dass die entscheidende Neuerung im Bereich der Abteilung für Psychotherapie darin besteht, dass die Sekretärin die Protokolle nicht mehr auf normales Papier schreibt, sondern sogenannte Belege verwendet.

Bei den bisherigen Ansätzen, die Möglichkeiten der maschinellen Textverarbeitung zu nutzen, wurden als Eingabeträger in die Rechenmaschine Lochkarten verwendet. Das zog eine doppelte Schreiarbeit nach sich. Zuerst musste ein normales Schreibmaschinenprotokoll erstellt werden und dann erfolgte die Umschreibung auf Lochkarten. Diese doppelte Prozedur kann nun mit Hilfe des sogenannten Klarschriftlesersystems entscheidend vereinfacht werden.

Das Klarschriftlesersystem CDC SC 1700/CDC 955

Seit längerer Zeit werden sogenannte Belegleser wie z.B. das Klarschriftlesersystem CDC SC 1700/CDC 955 angeboten. Dieses System ist in der Lage, genormte Schrifttypen – OCR-A-Schrift, Sonderzeichen und handschriftliche Druckerzeichen – zu lesen. Die Formulare (Belege) für den Belegleser können formatfrei oder formatgebunden sein. Formatgebundene Belege sind z.B. Standardfragebögen, wo die Daten an genau festgelegten Stellen eingetragen werden, formatfreie dagegen sind solche, auf denen nur ein pastellfarbener Rand gekennzeichnet ist und in die z.B. mit schwarzer Schrift freier Text eingetragen werden kann. Normalerweise wird mit Hilfe eines Standardleseprogramms der Text optisch über eine Spiegelplatte ‚abgelesen‘ in der Form, dass das System nur Kontraste zwischen hellem Untergrund und schwarzgefärbten Zeichen ‚sieht‘. Anschließend wird der Text auf Magnetband gespeichert.

Für den Zweck, eine Datenbank aufzubauen, die den verschiedenen Bedürfnissen gerecht werden kann, war es jedoch nötig, ein spezielles Leseprogramm zu entwickeln^{*)}. Es ist nämlich von größter Bedeutung, die verbale Interaktion zwischen Patient und Analytiker erforschen zu können. Hierzu ist es notwendig, dass die Speicherung der Texte getrennt erfolgt, d.h. dass der Text für den Patient und Analytiker jeweils getrennt zugreifbar ist, wobei gleichzeitig die Abfolge der Texteinheiten erhalten bleiben musste. Von

^{*)} Für die Entwicklung dieses Leseprogramms danken wir Herrn Tschandl vom Universitätsrechenzentrum Ulm.

daher stellte sich die Aufgabe, dass jede einzelne Äußerung eine numerische Kodierung erhalten musste.

Diese maschinelle Verwaltung und Organisation der Texte erfolgt nun über eine vierzehnstellige Zahl, die zu Beginn jeder Belegseite aufgeführt wird:

Die Textkennziffer

Zunächst wird jedes Formular durch die sogenannte Therapeutennummer (dreistellig) und die Patientennummer (dreistellig) gekennzeichnet. Die folgende Ziffer (einstellig) gibt an, um welche Art von Text es sich handelt. Dabei können folgende Angaben gemacht werden: Textsorte (z.B. Psychoanalyse, Psychotherapie, Erstinterview); dann folgt die Angabe der Gesprächsart (z.B. Dialog, Gruppe, etc.).

Abschließend folgen die laufende Nummer der Therapiestunden und die laufende Nummer des Beleges für die betreffende Stunde. Durch diese letzte Nummer werden die Blätterseiten der Protokolle zu einem einzigen Fall zusammengefasst.

Abschließend folgt nochmals die Nummer zur Übersicht:

001	001	1	9	0040	03
Therapeut	Patient	Textsorte	Gespräch	Stunde	Blattnummer

Diese Kennziffer wird jeweils in die erste Spalte des Beleges eingetragen; in der zweiten Spalte folgt dann der Text^{*)}.

Diese Textkennziffer wird aber durch die Leistung des Belegleseprogrammes noch ergänzt. Wird nämlich ein Dialog oder ein Gruppengespräch eingelesen, so wird auf dem Beleg jeder einzelne Sprecher durch einen Buchstaben (A1 – A99 = Therapeuten; P1 – P99 = Patienten) mit nachfolgender Nummer kenntlich gemacht. Das Leseprogramm legt diese Nummern in einer zwanzigstelligen Kennzahl auf dem Band ab, wobei die zeitliche Reihenfolge der einzelnen Äußerungen verschiedener Sprecher erhalten bleibt. Bei dem bisherigen Stand ist also die einzelne Äußerung die elementare Texteinheit, die vom Leseprogramm verarbeitet werden kann.

Die Bequemlichkeit des Beleglesesystems besteht nun besonders darin, dass die Belege bei uns im Hause auf einer IBM-Schreibmaschine mit einem speziellen Kugelkopf geschrieben werden können. Damit kann die Sekretärin einen

^{*)} Wir haben eine ausführliche Schreibanweisung erstellt, die es jedem Interessenten ermöglicht, auf einer IBM-Schreibmaschine seine Protokolle in dieser speicherkompatiblen Form abzufassen.

Arbeitsgang einsparen. Sie kann nun direkt vom Tonband die computerlesbaren Protokolle schreiben, womit erheblich Zeit gespart werden kann. Allerdings erfordert das Schreiben von Belegen eine besondere Aufmerksamkeit, weil das System Schreibfehler nicht erkennt, sondern das Wort als eigenes Wort behandelt, schlecht lesbare Buchstaben aber durch andere ersetzen würde. Hier hilft die Verwendung eines ebenfalls von IBM entwickelten Schreibautomaten MB 72.

Die MB 72 bietet die Möglichkeit, Texte, die auf ihr geschrieben werden, gleichzeitig über eine eingebaute Magnetkassettenstation auf speziellen Magnetbändern zu speichern. Werden Korrekturen nötig, so können die Verbesserungen auch direkt auf diese Magnetbänder aufgezeichnet und die Texte anschließend fehlerfrei wiedergegeben werden. Dies ist besonders beim Schreiben von Verbatimprotokollen eine große Hilfe, weil der Text nicht immer gleich richtig verstanden wird.

Das Klarschriftlesersystem überspielt die auf Belegen angelieferten Texte auf einen Datenträger, der als Eingabeeinheit für das im folgenden dargestellte Programmsystem dienen kann. Wir benutzen hierzu einen Plattenspeicher. Umfangreichere Texte können allerdings auf Magnetbänder überspielt werden, die als externe Speicher wesentlich günstiger und billiger sind. Diese können jederzeit wieder in den Plattenspeicher zurückgespielt werden und von dort aus durch entsprechende EDITOR-Programme zugreifbar gemacht werden.

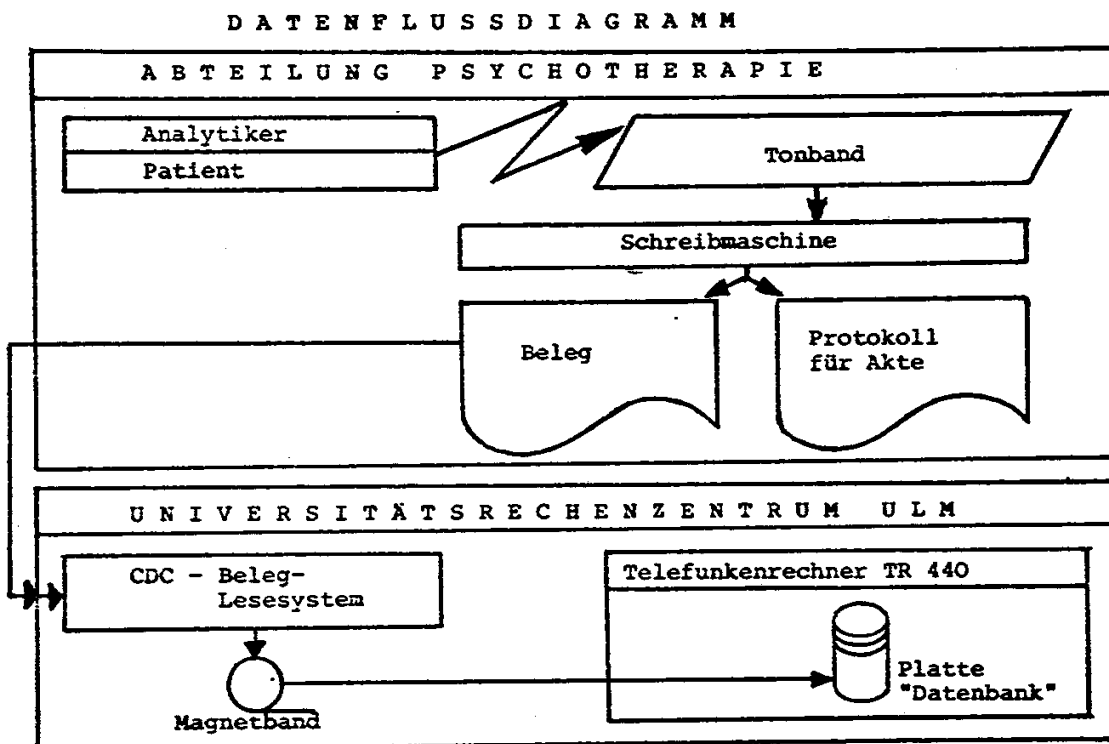


Abbildung 7

2.2 „Elektronische Verbalanalyse“ - ein Programmsystem zur maschinellen Inhaltsanalyse von Texten (EVA)

Bevor wir im folgenden das EVA-System (Elektronische Verbal-Analyse) vorstellen, soll anhand eines Schemas der allgemeine Aufbau eines solchen Analysesystems erläutert werden, wie es allen bisher entwickelten Programmsystemen eigen ist:

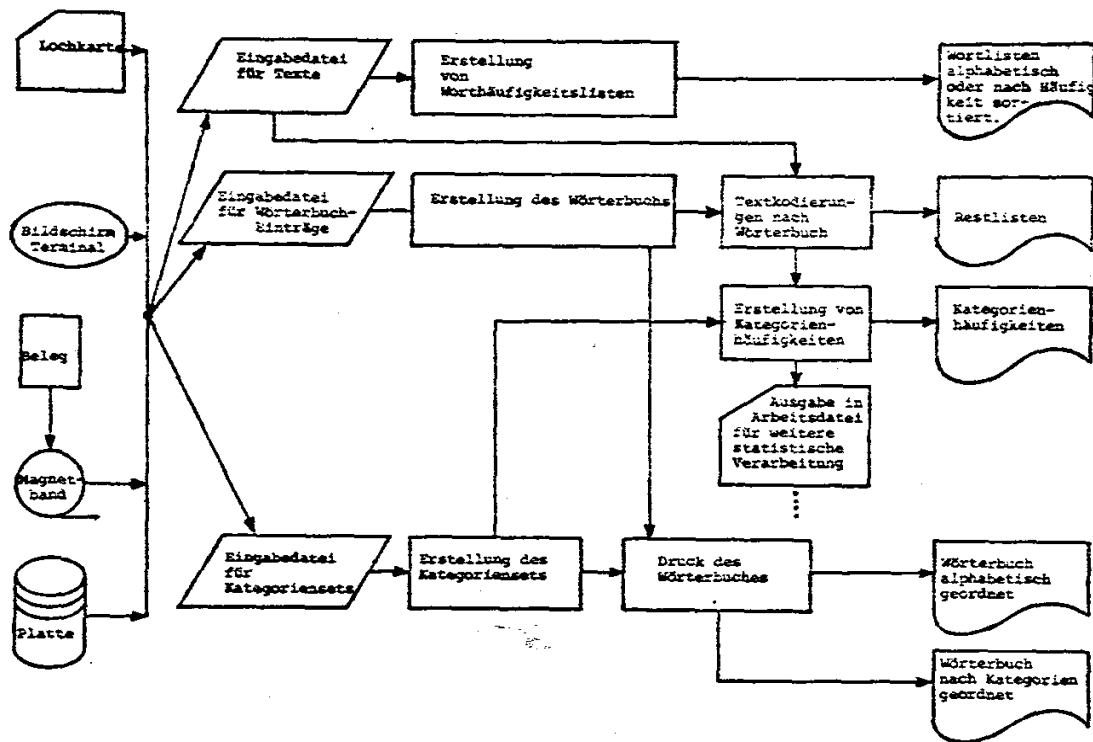


Abbildung 8

Von verschiedenen Eingabemöglichkeiten aus (Lochkarten, Terminal, Belegen, Datenbanksystem) wird ein Text, der analysiert werden soll, zunächst in eine sogenannte Texthaltungsdatei gebracht. Die erste Aufgabe zur Vorbereitung einer maschinellen Inhaltsanalyse besteht darin, sich über den Wortschatz zu orientieren, der analysiert werden soll. Die Erstellung von Worthäufigkeitslisten – wobei verschiedene Kennzeichen der Wörter zugleich ermittelt werden – stellt die im Abbild oben dargestellte Analyse-Ebene dar.

Die Aufgabe des Forschers besteht nun im wesentlichen darin, nach der Inspektion des Wortschatzes die ihn interessierenden Wörter einer oder mehreren semantischen Kategorien zuzuordnen. Damit wird die Inhaltsanalyse zu einem Spezialfall der automatischen Sprachübersetzung (SCHEUCH, 1967, S. 667). Die Wörter, die nach Ansicht des Forschers ein bestimmtes theoretisches Konzept repräsentieren, werden dann durch Bildung einer Kategorie als semantische Äquivalente zusammengefasst.

Metasprache

Konzept KOMMUNIKATION

Objektsprache

Stichwörter: antworten, diskutieren,
berichten, erzählen, fragen,
informieren, etc.

Abbildung 9 (nach MOCHMANN, 1974, S. 193)

Die den Forscher interessierenden Konzepte werden somit extensional definiert, was eine ständige Überprüfbarkeit und intersubjektive Verständigung über den Kodiervorgang ermöglicht. Die Erstellung des inhaltsanalytischen Wörterbuches erfolgt technisch dadurch, dass die kodierten Wörter mit den numerischen Zuweisungen für die Kategorien in einer Eingabedatei für Wörterbucheinträge abgelegt werden und mit dem semantischen Kategoriennamen, die aus einer gesonderten Eingabedatei kommen, zusammengebracht werden müssen. Dieses Wörterbuch liegt maschinenintern in alphabetischer Wortfolge vor, kann aber durch ein spezielles Programm in Form von semantisch geordneten Begriffslisten ausgedruckt werden.

Mit der Erstellung des inhaltsanalytischen Wörterbuches ist das Beobachtungsinstrument der maschinellen Inhaltsanalyse erstellt. Wie für jedes andere Instrument der Wissenschaften gilt es auch hier Reliabilitäts- und Validitätsaspekte zu berücksichtigen. Vor allem aber muss die theoriegebundene, leitende Funktion des Wörterbuches unterstrichen werden. Diese primäre Funktion des Wörterbuches, der inhaltsanalytischen Beobachtung eine theoretische Orientierung zu geben, vollzieht sich auf drei Ebenen (nach GOLDHAMER, 1969, S. 349):

- a) Die erste Auswirkung der Wörterbuch-vermittelten Theorie besteht in der Selektion der Inhaltskategorien, die der Forscher aus der unendlichen Vielzahl der Bedeutungskategorien seiner Kultur aussondert. Damit wird festgelegt, welche Bedeutungen als Ergebnis (im Programmdruck) festgehalten werden sollen. Hierbei handelt es sich um einen statischen Einfluss, bei dem außerdem viele Bedeutungsaspekte nicht berücksichtigt werden, die bei dem Vorgang der Textkodierung intermediär benötigt werden.
- b) Eine zweite Einflussebene vollzieht sich in der Wort-zu-Kategorie-Zuordnung, welche den Gehalt des Wörterbuches ausmacht. Dieser Einfluss ist relativ dynamisch, aber er führt doch zu einer fixierten Strategie, Kategorien auf der empirischen Ebene des Textes zu identifizieren.
- c) Die dritte (und am meisten dynamische und wirksamste) Einflussebene vollzieht sich durch die zusammenfassenden Operationen, die über die

Einheiten der Untersuchung gemacht werden. Auf dieser Ebene vollzieht sich die eigentliche Abstraktionsleistung; hier werden Themen und Trends entdeckt, hier werden Strategien der Kategorienidentifizierung angepasst und modifiziert, und hier kann der Forscher einen Überblick gewinnen und sich fragen, was hier eigentlich vorgeht.

Alle drei Ebenen des theorievermittelten Einflusses sind notwendig um die inhaltsanalytischen Prozeduren für die zu prüfende Theorie relevant zu gestalten. Dies führt in der Forschungspraxis zur Entwicklung verschiedener Wörterbücher, die jeweils den zugrundeliegenden Fragestellungen entsprechend gestaltet werden.

So führt STONE (1966) bereits eine Liste von 17 Wörterbüchern an, die ein breites Spektrum sozialwissenschaftlicher, anthropologischer und literaturwissenschaftlicher Fragestellungen umfassen (S. 140).

Zusammenfassend können wir sagen, dass neben den noch zu diskutierenden technischen Problemen der Realisierung das Hauptproblem der maschinellen Inhaltsanalyse die richtige, d.h. forschungsrelevante Vorgabe der Übersetzungsanweisung von der Objekt- in die Metasprache darstellt. So bezeichnet HOLSTI (1968) das Wörterbuch als das vitale Band zwischen der Theorie des Forschers und der Mechanik der Computertechnologie (S. 665).

Das EVA-System, das wir nun beschreiben^{*)}, wurde am Soziologischen Seminar der Universität Hamburg von R. THIEMANN und K. HOLZSCHECK entwickelt und stellt ein vielseitiges, ökonomisches und zuverlässiges Forschungsinstrument für den Inhaltsanalytiker dar; die Adaption von EVA an die TR440-Anlage des Rechenzentrums der Universität Ulm wurde ebenfalls von K. HOLZSCHECK besorgt.

EVA besteht aus einer Reihe von Programmen, die in ALGOL 60 und Telefunken Assembler (TAS) programmiert sind und aus einigen Kommandoprozeduren, welche die Benutzung des Systems sehr vereinfachen. Weiterhin enthält EVA fest vereinbarte Dateien, die bei dem Aufruf des Systems erzeugt werden; über diese Dateien ist die Kommunikation der einzelnen Programme sichergestellt, d.h. die Ausgabe eines Programms kann zugleich Eingabe für ein anderes Programm sein.

Für die praktische Erstellung des Wörterbuches, des eigentlichen Analyseinstrumentes jeder maschinellen Inhaltsanalyse, werden auf der mittleren Ablauebene die Wörterbucheinträge in eine Datei eingegeben. Hier ist

^{*)} Die Beschreibung stützt sich auf GRÜNZIG, HOLZSCHECK u. KÄCHELE, 1976; detailliertere Beschreibungen finden sich bei TIEMANN, 1973, HOLZSCHECK, 1975 und HOLZSCHECK und MERGENTHALER, 1976).

ebenfalls die Eingabe über die verschiedenen Eingabeeinheiten möglich, wie auch die Korrekturen an vorliegenden Eingaben über den Terminal direkt durchführbar sind. Das hier operierende Programm LINWAB wandelt die alphabetisch geordneten Wörterbucheinträge mit numerischer Kategorieninformation in ein Wörterbuch nach EVA-Konvention um und legt dieses in einer Zwischendatei ab. Das Wörterbuch wird aber erst durch die Hinzunahme der Bezeichnungen für die Kategorien vervollständigt. Diese werden auf der dritten Ablaufebene eingebracht und mit Hilfe des Programmes TIVOLI auf einer permanenten Datei EVASYS gespeichert. Die Erstellung des Wörterbuches, mit kategorialer Ordnung der Einträge, wird nun durch ein Zusammenspiel der zweiten und dritten Ablaufebene ermöglicht. Das Programm BINLIN holt sich die alphabetisch gespeicherten Wörterbucheinträge und ersetzt die numerischen Zuweisungen durch die Namen der Kategorien. Die inhaltsanalytische Analyseprozedur im engeren Sinne, d.h. die Zuweisung der Wörter zu semantischen Faktoren, den Kategorien, erfolgt durch ein Zusammenwirken aller drei Ablaufebenen. Der Text wird aus der Eingabedatei in der oberen Ebene geholt; jedes im Text vorkommende Wort wird darauf hin geprüft, ob es in der alphabetisch sortierten Liste der Wörterbucheinträge vorkommt. Dieser Vorgang wird von dem leistungsstarken Programm LINKOD durchgeführt: ist eine Wortform im Wörterbuch vorhanden, so wird das Wort unter den entsprechend zugeordneten Kategorien verrechnet. Kann das Wort nicht identifiziert werden, so wird nach dem längsten Wörterbucheintrag gesucht, welcher mit dem zu kodierenden Wort vom ersten Zeichen ab übereinstimmt. Wird eine solche Zeichenfolge gefunden, so wird diese vom Gesamtwort abgespalten und unter den entsprechenden Kategorien verrechnet; mit dem noch nicht identifizierten Restwort wird dieselbe Identifizierungsprozedur wiederholt; es sind insgesamt maximal 10 derartige Kodierläufe pro Wortform vorgesehen. Kann irgendein Teil des Wortes nicht identifiziert werden, so werden die bereits vorgenommenen Kategorienzuweisungen rückgängig gemacht und das gesamte Wort in die sogenannte Restliste (RLD) aufgenommen. Die Restliste enthält alle Wörter, die sich nicht restlos aus den vorhandenen Wörterbucheinträgen zusammensetzen lassen. Hier werden dann die absolute Häufigkeit, die relative Häufigkeit und die aufsummierte Häufigkeit der Wörter angegeben um somit in der Restliste die häufigen Wörter leicht identifizieren zu können.

Der Druck der gesamten alphabetisch sortierten Restliste wird durch BINLIT2 besorgt.

Zur Veranschaulichung der Wortidentifizierungsprozedur diene folgendes Beispiel: Im Text trete das Wort „gesehen“ auf; das Wörterbuch enthalte u.a. die Wörter „ge“, „seh“ und „en“. Da das gesamte Wort nicht im Wörterbuch vorhanden ist, wird zunächst die Vorsilbe „ge“ identifiziert, in den nächsten zwei Suchläufen auch die restlichen Wortteile, so dass schließlich das gesamte

Wort kodiert werden kann. Das Wort „geseht“ kann bei den gegebenen Wörterbucheinträgen nicht in dieser Weise aufgelöst werden, da die Zeichenfolge „nt“ nicht im Wörterbuch enthalten ist; folglich werden die bereits vorgenommenen Kodierungen „ge“ und „seh“ rückgängig gemacht und das gesamte Wort in die Restliste aufgenommen.

Das Ergebnis dieser Analyse kann man sich so veranschaulichen, dass die Textwörter in die Stammwörter^{*)} des Wörterbuchs zerlegt sind; für jedes Stammwort enthält der Datensatz die „Bedeutungen“, die über die Kategorienzuweisungen den Einträgen des Wörterbuches zugeordnet sind. Mit diesem Datensatz kann nun in zwei Weisen weiterverfahren werden: einmal kann dieser Datensatz direkt zur Erzeugung einer Häufigkeitsverteilung über die Kategorien des Wörterbuches herangezogen werden. Zum anderen bildet diese Darstellung der Information die Grundlage für weitere Analyseschritte, welche die Bedeutung der Textwörter in einem jeweiligen Kontext ihrer Benutzung genauer eingrenzen sollen. Voraussetzung hierfür ist, dass den Stammwörtern des Wörterbuchs neben Kategorien (gewissermaßen als invarianten Bedeutungszuschreibungen) auch Regeln zugeordnet werden können, die Aussagen darüber enthalten, unter welchen Kontextbedingungen einem Textwort oder einem Stammwort eine vorgegebene Bedeutung zugeschrieben bzw. eine im allgemeinen Fall gültige Bedeutung gestrichen werden soll. Diese Regeln werden vom Benutzer in einer vereinbarten formalen Sprache formuliert. Das Ergebnis der Anwendung dieser Regeln besteht also darin, dass die Zuweisung von Bedeutungen (Kategorien) in Abhängigkeit vom Kontext korrigiert werden kann. Die Anwendung solcher Regeln auf den Datensatz erfolgt durch das Programm SENSEVA, das sich zur Zeit noch in der Testphase befindet. Weiterhin können diejenigen frei wählbaren Wortformen, die im Text sehr häufig auftreten und/oder als nicht sehr informationshaltig angesehen werden – z.B. und, der, die, das etc. –, mit LINKOD von vorneherein aus der Analyse eliminiert werden; diese Wortformen erscheinen dann auch nicht in der Restliste. Mit der Eliminierung der bedeutungsleeren, zugleich aber sehr häufig auftretenden Wortformen gelingt es in vielen Fällen, den Textumfang auf bis zu 50 Prozent zu reduzieren.

Mit QUSU schließlich wird die Berechnung und Ausgabe verschiedener Kennwerte durchgeführt. Es können absolute oder zu verschiedenen Werten relativierte Häufigkeiten pro Fall ausgegeben werden. Mehrere Möglichkeiten der Definition eines Falles sind vorgesehen. Einmal können Zeichen vereinbart werden, die das Ende eines Satzes kennzeichnen; dieser Satz ist nicht notwendigerweise ein grammatikalischer Satz. Hierdurch wird das Textdoku-

^{*)} In Anlehnung an den Begriff ‚Wortstamm‘ bezeichnen wir einen Eintrag im Wörterbuch als ‚Stammwort‘. Als Stammwort wählt man sinnvollerweise eine Buchstabenfolge, die zur Identifizierung einer möglichst großen Zahl von Flexionen dienen kann, vorausgesetzt, dass alle diese Flexionsformen in dieselbe Inhaltskategorie fallen. Eine solche Buchstabenfolge ist häufig mit dem Wortstamm identisch.

ment in mehrere Einzelfälle aufgelöst, über welche getrennte Kategorienhäufigkeiten berechnet werden. Zum anderen können verschiedene Textdokumente zu einem einzigen Fall zusammengezogen werden; Kriterium für die Fallkombination ist die Übereinstimmung der ursprünglichen Fallnummern hinsichtlich der frei wählbaren Spalten innerhalb der Fallnummer. Das heißt, alle diejenigen Dokumente des Gesamttexts, die in den angegebenen Fallnummernspalten übereinstimmen, werden vor der Auswertung zu einem Fall zusammengefasst.

Die folgende Tabelle verdeutlicht die Form des EVA-Ausdruckes für eine inhaltsanalytische Prozedur: (Anwendung des Dogmatismus-Wörterbuches auf eine Äußerung des Patienten).

Text: (0000201)

Ja, ich habe Angst, irgendetwas falsch zu machen, ich mein, die Angst, bei irgendetwas ertappt zu werden oder eingeschätzt zu werden, den falschen Eindruck zu erwecken, nichts zu wissen, unhöflich zu sein, jemand weh zu tun, eigene Schwächen, Fehler usw. entweder zu zeigen oder von außen darauf hingewiesen zu werden.

Ergebnis: (0000201)

Wörter:	53	kodierte Wörter:	6	Zuweisungen	6
Kat.-Nr.:	2	(Name: Anzahl, Mengen [positiv])	1	18,9	166,7
Kat.-Nr.:	9	(Name: Grad, Maß [negativ])	2	37,7	333,3
Kat.-Nr.:	11	(Name: Ausschluss, Einbeziehung)	3	56,6	500,0

Das Ergebnis besagt, dass der zu untersuchende Text 53 Wörter enthält, von denen insgesamt 6 Wörter durch das Dogmatismus-Wörterbuch (ERTEL, 1975) kodiert wurden. Diese Wörter verteilen sich auf drei Kategorien, für die dann jeweils die absolute Häufigkeit und prozentual relativierte Werte angegeben werden.

Diese hier illustrierten Ergebnisse sind durch das Programm QUSU einerseits ausdrückbar, andererseits aber über spezielle Arbeitsdateien gleich indirekt in statistische Programme (z.B. BMD) überzuführen. Damit ist eine weitere Bequemlichkeit des EVA-Systems angesprochen, dass nämlich die inhaltsanalytischen und statistischen Prozeduren nahtlos miteinander verbunden sind.

Über die skizzierten Programme mit ihrem festen Leistungsumfang hinaus enthält EVA eine interne Sprache (FOSKOX, formale Sprache zur Kontextbeschreibung), mit deren Hilfe weiterführende Fragestellungen verfolgt werden können. So können mit FOSKOX beispielsweise sogenannte KWIC-

Listen^{*)} ausgegeben oder alle Fälle herausgesucht und gedruckt werden, in denen vom Benutzer vorher definierte Kategorienkombinationen auftreten. Darüber hinaus kann mit FOSKOX auch eine entsprechende Modifikation des bereits von LINKOD erstellten Datensatzes erfolgen, welcher sodann von QUSU, wie oben beschrieben, ausgewertet wird. Die in FOSKOX formulierten Anweisungen werden von dem Operator BINKOX übersetzt und auf syntaktische Richtigkeit geprüft.

Das folgende Diagramm fasst noch einmal die Abläufe zusammen, wie wir sie beschrieben haben; statt der umgangssprachlichen Beschreibung der Abläufe wie wir sie im oben aufgeführten Diagramm gegeben haben, stehen hier nur die Namen der speziellen Programmroutinen und die Namen der verschiedenen Speicher, wie sie auch im Text erwähnt wurden.

Abschließend soll aber betont werden, dass die hier gegebene Darstellung des EVA-Programmsystems (S. 33, Abb. 10) benutzerorientiert ist. Alle darüber hinausgehenden Angaben zu Einzelheiten des Aufbaus und der Programmierprobleme solcher umfangreicher datenverarbeitender Systeme finden sich bei HOLZSCHECK und MERGENTHALER (1976).

^{*)} KEY-WORD in CONTEXT; es werden alle Textabschnitte ausgedruckt, die ein jeweils definiertes Suchwort enthalten.

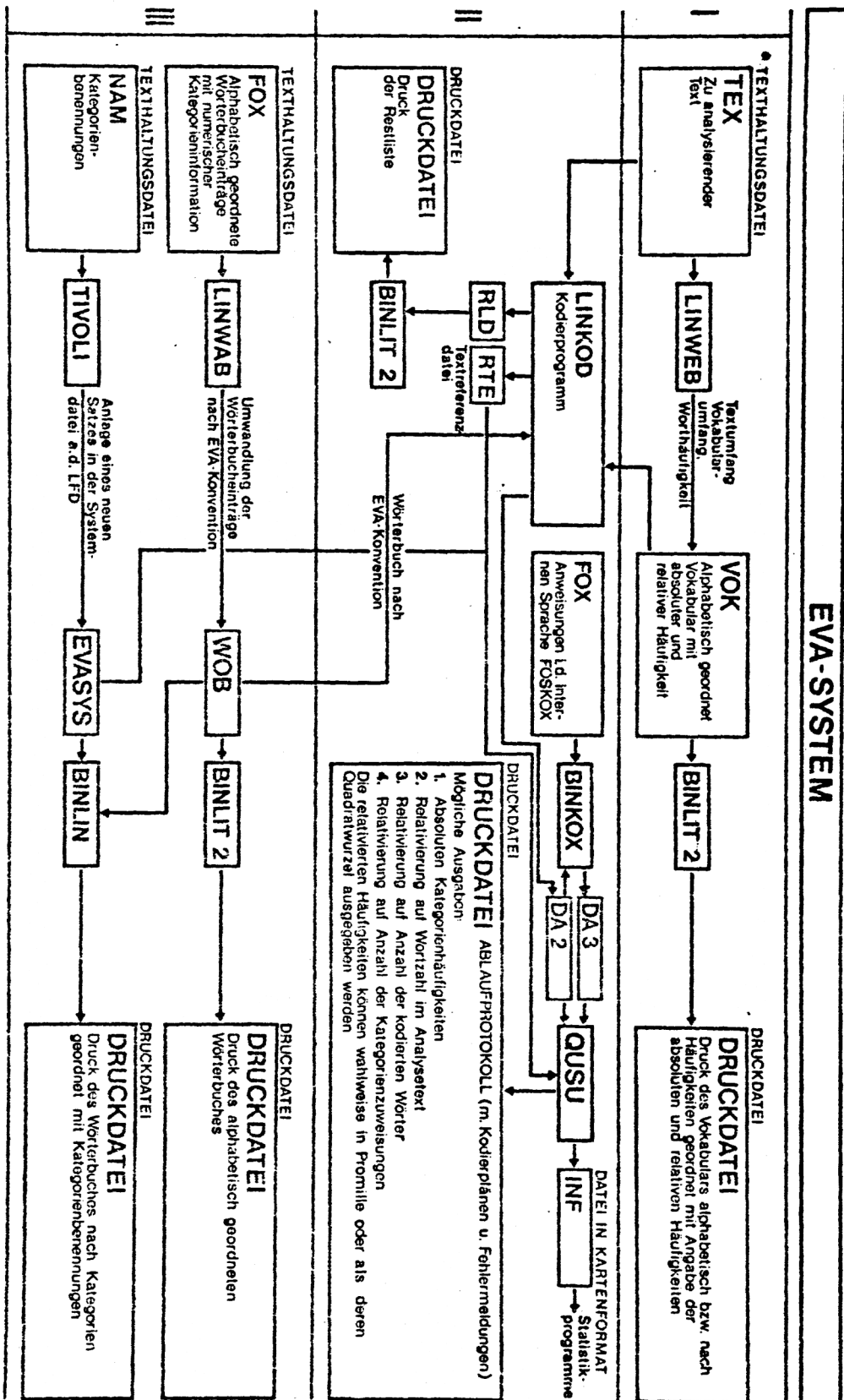


Abbildung 10

2.3 Der Ablauf der Datenauswertung

Nachdem wir nun das inhaltsanalytische Programmsystem kurz beschrieben haben, mit dem wir die maschinelle Textanalyse durchführen, können wir zur Beschreibung der Forschungssituation zurückkehren. Wir hatten bei der kurzen einleitenden Charakterisierung der manuellen Inhaltsanalysen besonders auf den Aspekt kritisch hingewiesen, dass der Aufwand eine Wiederholung von Auswertungsprozeduren von sich aus verbietet. Deshalb wollen wir bei der folgenden Beschreibung der Prozeduren zur Datenauswertung besonders auf den Aspekt der instrumentalen Leichtigkeit abheben, mit dem inhaltsanalytische Fragen dann verfolgt werden können, wenn bestimmte technische Möglichkeiten realisiert sind.

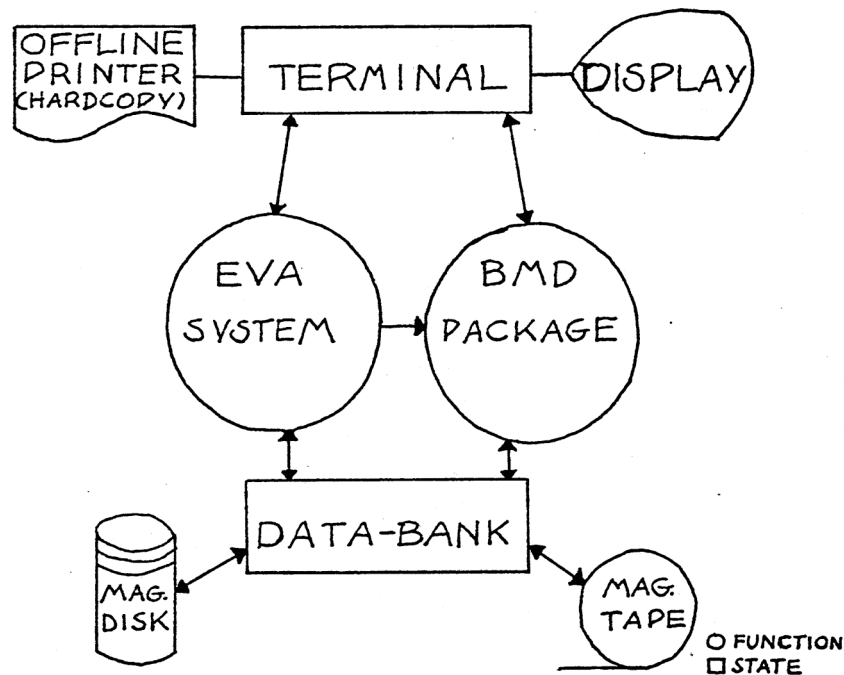


Abbildung 11

Nachdem der Vorgang der Datenspeicherung, wie er unter 2.1 beschrieben wurde, abgeschlossen ist, steht die flexible Beziehung zwischen dem Forscher, seinem Arbeitsinstrument und den Daten im Mittelpunkt der Forschungsprozedur. Diese Beziehung wird durch eine direkte Verbindung zwischen dem Rechenzentrum und dem Arbeitsraum des Benutzers hergestellt, indem ein Bildschirm-Terminal über eine telefonische Standleitung direkt mit dem Rechenzentrum Verbindung herstellt. Von dem Sichtgerät, welches sowohl

eine Eingabeeinheit wie auch eine Ausgabeeinheit darstellt (optimalerweise wird noch ein Drucker angeschlossen) können die Daten direkt mit Hilfe des Programmsystems bearbeitet werden. Damit ist eine wichtige Eigenschaft der elektronischen Inhaltsanalyse gewährleistet; ein lebendiges Wechselspiel zwischen Fragestellungen des Forschers und seinen Ergebnissen in sich wiederholenden Schleifen ist möglich:

Von der Konsole des Terminals aus wird das Programmsystem aktiviert. Dann werden die benötigten Texte von den Datenträgern (Magnetband, Plattenspeicher) geholt und in die Arbeitsspeicher des Programmpakets eingetragen. Die Ergebnisse werden zunächst auf dem Bildschirm des Terminals ausgegeben. Zur schriftlichen Dokumentation können sie über den Drucker ausgegeben werden. Für die statistische Weiterverarbeitung ist es aber auch möglich, die Daten im Rechenzentrum auf Lochkarten ausstanzen zu lassen oder auch direkt die Weiterverarbeitung mit statistischen Programmpaketen anzusteuern.¹⁾

Werden graphische Darstellungen der Verlaufskurven gewünscht, so werden die Ergebnisse über Lochkarten in ein Zeichengerät, den ‚Plotter‘ eingegeben, der dann programmgesteuert die Graphiken zeichnet.

Dieses System gewinnt einen interaktiven Charakter besonders dadurch, dass der Forscher nicht an einmal gefundene Ergebnisse nur deshalb fixiert bleibt, weil es große Mühe machte, diese Ergebnisse überhaupt zu bekommen, sondern er kann ohne große Mühe das Wörterbuch verändern, Kategorienstrukturen systematisch variieren und erneut die Inhaltsanalyse durchführen. Damit wird überhaupt erst der Vergleich verschiedener Wörterbücher am gleichen Text durchführbar, der bisher nur ein Desideratum war.

Fazit

Zusammenfassend können wir sagen, dass von technischer Seite her die Probleme einer psychotherapeutischen Datenbank als lösbar erscheinen und dass damit die vergleichende Erforschung einer größeren Zahl von Behandlungen ermöglicht wird. Das hier beschriebene System erlaubt die Einbeziehung einer Vielzahl von Behandlungen, die auch an anderen Orten transkribiert werden können. Die minimale Bedingung ist die Einhaltung der Schreibkonvention für die transkribierten Texte; Speicherung dieser Texte und ihre inhaltsanalytische Aufarbeitung sind mit dem hier beschriebenen System im Prinzip gelöst. Damit sind die technischen Voraussetzungen erfüllt um psychotherapeutische

¹⁾ Wir benutzen hierzu die in Ulm implementierten Biomedical Programs (DIXON, 1973).

Behandlungsprotokolle zu speichern und auszuwerten und der Aufbau einer Datenbank kann begonnen werden²⁾.

²⁾ Wir bereiten z. Zt. Eine derartige Kooperation mit einer Reihe von Forschungsgruppen vor: mit Herrn Dr. rer. biol. Hum. E. Brähler u. Mitarbeiter vom SFB 32 in Giessen, Herrn Prof. Dr. med. Enke u. Mitarbeiter von der Forschungsstelle für Psychotherapie in Stuttgart, Herrn Prof. Dr. med. Meyer u. Mitarbeiter von der Abt. für Psychosomatik in Hamburg, Herrn PD. Dr. med. Köhle von der Abteilung für Psychosomatik in Ulm und Herrn Prof. Dr. med. P. Hahn von der Abt. für allg. klin. Medizin in Heidelberg.