

## Microsoft Access in der Analyse von Fragebögen und Interviews mit offenen Antwortformaten

*Thorsten Meyer, Harald Gruppe & Michael Franz*

### Keywords:

Inhaltsanalyse,  
offenes Antwort-  
format, MS  
Access, Visual  
Basic, SPSS,  
Intercoderrelia-  
bilität, Lebens-  
qualität, Patienten-  
zufriedenheit,  
Schizophrenie,  
psychiatrische  
Patienten

**Zusammenfassung:** Strukturierte Interviews oder Fragebögen beinhalten häufig offene Fragen, insbesondere um zusätzlich zu den standardisierten Fragen vom Instrument nicht abgedeckte Informationen von den Befragten zu erhalten. Die Analyse dieser offenen Fragen kann in der Regel nicht im gleichen Auswertungsprogramm durchgeführt werden, in dem die quantitativen Daten eingegeben und verwaltet werden, was die Kombination dieser Datentypen erschwert.

Das Datenbanksystem Microsoft Access ist weit verbreitet. Es lassen sich damit flexibel quantitative Daten eingeben, verwalten und wieder exportieren. Die Eingabe von Textmaterial ist durch Memo-Felder für die ökonomische Aufnahme von Daten fast beliebig großen Umfangs gut handhabbar. Vor dem Hintergrund der Möglichkeit, die gleiche Software für die Verwaltung häufig gemeinsam auftretender quantitativer und qualitativer Daten zu verwenden, wollten wir die Flexibilität von MS Access und seine Programmereigenschaften (Visual Basic mit MS-Access-Erweiterungen) für die inhaltsanalytische Auswertung nutzbar machen.

Dieser Beitrag umfasst die Beschreibung der Erstellung von Datentabellen, Abfragen, Kategorienlisten und von Oberflächen zur Eingabe von Rohdaten. Ebenso wird die Erstellung von Kodiereinheiten (Paraphrasierung des Materials), die Kodierung und die Umstrukturierung der Daten für die statistische Auswertung dargestellt. Zudem wird der Daten-Export in das Statistikprogramm SPSS sowie die Analyse der Intercoderreliabilität behandelt. Datengrundlage sind ein strukturiertes, offenes Lebensqualitätsinterview mit chronisch schizophren Erkrankten und eine schriftliche Beurteilung des stationären Aufenthalts durch psychiatrische Patienten.

### Inhaltsverzeichnis

- [1. Hintergrund](#)
- [2. Anwendung von MS Access am Beispiel eigener Forschungsprojekte](#)
  - [2.1 Projekt Patientenzufriedenheit: einführendes Beispiel zur Anwendung von MS Access bei offenem Antwortformat](#)
  - [2.2 Datentabellen](#)
  - [2.3 Formulare zur Eingabe von Rohdaten](#)
  - [2.4 Paraphrasierungen der Rohdaten](#)
  - [2.5 Kategoriensystem](#)
  - [2.6 Kodierung des Materials](#)
  - [2.7 Transformation von "ungeordneten" Tabellen in rechteckige Datenmatrizen](#)
  - [2.8 Export der Daten in SPSS](#)
  - [2.9 Überprüfung der Intercoderreliabilität](#)
- [3. Fazit](#)

[Literatur](#)

[Zu den Autoren](#)

[Zitation](#)

## 1. Hintergrund

In der empirischen Sozialforschung finden nicht selten sowohl geschlossene als auch offene Antwortformate Verwendung. Fragebögen beinhalten häufig offene Fragen, um zusätzlich zu den standardisierten Fragen vom Instrument nicht abgedeckte Informationen von den Befragten zu erhalten. Eine Kombination offener und geschlossener Antwortformate findet sich insbesondere in strukturierten Interviews. Eine angemessene Würdigung der Antworten auf diese offenen Fragen wird – auch unabhängig von forschungsparadigmatischen Vorlieben der jeweiligen Forscherinnen und Forscher – durch vorhandene Eingabe- und Analyseprogramme erschwert. Auf der einen Seite steht der primär an den standardisierten Fragen interessierte Forscher. Antworten auf offene Fragen werden in Textfeldern eingegeben. Auf der Grundlage einer Sichtung des Textmaterials werden ad hoc Kategorien entworfen, in die das Textmaterial im Nachhinein im Auswertungsprogramm kodiert wird – nicht selten eine willkommene Aufgabe für studentische Hilfskräfte. Im Extremfall unterbleibt sogar eine Analyse der Textstellen. Auf der anderen Seite stehen Forscherinnen und Forscher, die die Reichhaltigkeit und Tiefe des Textmaterials für die Überprüfung und Weiterentwicklung ihrer Forschungsfrage ausschöpfen möchten. Die Analyse der offenen Fragen kann in der Regel nicht im gleichen Programm durchgeführt werden, in dem die quantitativen Daten eingegeben und verwaltet werden. Damit wird auch die Kombination dieser Datentypen erschwert. Es bedarf dazu entweder einer aufwendigen Analyse "per Hand", oder es wird ein computergestütztes Verfahren zur Entwicklung und Kodierung der Aussagen angewendet. Forschungspraktisch ergeben sich dadurch jedoch Probleme:

1. Programme zur Dateneingabe haben in der Regel einen Schwerpunkt entweder auf quantitativen Daten oder auf Textdaten mit den entsprechenden Möglichkeiten zur Datenverwaltung. Ziel wäre es aus unserer Sicht, dass eine angelernte Hilfskraft Daten eines Fragebogens bzw. Interviewprotokolls in *eine* Eingabemaske eingeben kann, wodurch verschiedene Fehlerquellen bei der Eingabe von Daten besser kontrolliert werden können.
2. Zwischenanalysen, die die Antworten auf die offenen Fragen in direkte Beziehung zu anderen Daten (sowohl auf andere standardisierte als auch offene Fragen) stellen, können nur mit vergleichsweise hohem Aufwand erstellt werden. Wünschenswert wäre bspw. eine Möglichkeit, bei der Erstellung von Codes auch Informationen zu Antworten auf andere Fragen schnell abrufen zu können.
3. Zur weiteren Auswertung müssen die unterschiedlichen Datenquellen wieder miteinander kombiniert werden. Verschiedene computergestützte textanalytische Programme bieten die Möglichkeit des Exports von Daten. Beispielsweise lässt sich in ATLAS.ti (z.Zt. im Release 4.2) eine SPSS-Datentabelle über eine SPSS-Syntax-Datei erstellen. So steht in einer durch

ATLAS.ti erstellten SPSS-Datentabelle nicht eine Person, sondern eine Textstelle in einer Datenzeile. Auch aus dieser Perspektive wäre ein Programm wünschenswert, dass diesen Schritt unnötig macht. [1]

Das Datenbanksystem Microsoft Access bietet sich für die Lösung dieser Probleme an. Es lassen sich damit flexibel quantitative Daten eingeben, verwalten und wieder exportieren. Die Eingabe von Textmaterial ist durch sogenannte Memo-Felder für Daten fast beliebigen Ausmaßes gut handhabbar. Zudem hat es durch die Integration in das MS Office Paket den Vorteil einer weiten Verbreitung. Damit bietet es die Möglichkeit, verschiedenen Anwendern die Kategorisierung von vergleichbar strukturiertem Textmaterial ohne die Anschaffung und Einarbeitung in zusätzliche inhaltsanalytische Software zu ermöglichen. Vor dem Hintergrund der Möglichkeit, die gleiche Software für die Verwaltung quantitativer und qualitativer Daten zu verwenden, haben wir den Versuch unternommen, die Menü- und Programmierereigenschaften (Visual Basic mit MS-Access-Erweiterungen) von MS Access für die inhaltsanalytische Auswertung nutzbar zu machen. Auch wenn andere qualitative Forscher MS Access sogar als Ersatz für spezielle textanalytische Software betrachten (NATKIN 2000), stellt dies nicht unser primäres Anliegen dar. Vielmehr liegen unsere Interessen a) in der Kombination von geschlossenen und offenen Antwortformaten, b) in der Analyse strukturierter Befragungsmethoden mit offenem Antwortformat. Die strukturierten, offenen Aussagen zeichnen sich dadurch aus, dass sie eine Antwort auf *eine bestimmte* Frage darstellen, die jeder / jedem Befragten / Befragtem in gleicher Weise vorgelegt wird (strukturiertes Interview bzw. Fragebogen). Das Element des Rekursiven (d.h. Veränderungen der Fragen aufgrund der vorgenannten Antworten), das charakteristisch für sogenannte qualitative Forschungsmethoden ist, bleibt bei diesem Ansatz unberücksichtigt. Der Vorteil der strukturierten, offenen Erhebungsform besteht in der Realisierung hoher Fallzahlen – einer Voraussetzung für die Repräsentativität der Stichproben – bei gleichzeitig größerer inhaltlicher Tiefe im Vergleich zum standardisierten Fragebogen. [2]

Unsere Entscheidung für den Einsatz von MS Access hatte folgenden Hintergrund: Ein thematischer Schwerpunkt der Arbeitsgruppe Sozialpsychiatrie an der Psychiatrischen Klinik des Universitätsklinikums Gießen besteht in der Untersuchung der subjektiven Lebensqualität von Menschen mit einer schizophrenen Erkrankung. Fragen nach der subjektiven Lebensqualität von schizophrenen Menschen haben zwar eine große Popularität in der Evaluation psychiatrischer Versorgungsmaßnahmen und in klinischen Studien gewonnen (FRANZ, MEYER & GALLHOFER 2000), dennoch lag bis vor kurzem kein Instrument zur Erfassung der Lebensqualität vor, das auf der spezifischen Erfahrungs- und Lebenswelt chronisch schizophrener Menschen basiert (FRANZ, MEYER & GALLHOFER 2002). Aus diesem Grunde wurden in Lebensqualitätsstudien der Arbeitsgruppe Sozialpsychiatrie neben der konventionellen, psychometrischen Lebensqualitätsmessung auch offene Interviews mit den Betroffenen durchgeführt, um als Ausgangsbasis für weitere Lebensqualitätsuntersuchungen deren Vorstellungen zur Lebensqualität und relevante Komponenten bzw. den Bedeutungsraum von Lebensqualität aus Sicht

der Betroffenen zu erfassen. Da insbesondere bei den schwer chronifizierten Patienten mit häufig einhergehenden ausgeprägten kognitiven Funktionseinschränkungen längere, vertiefende Interviews nicht durchführbar sind, wurde ein strukturiertes offenes Lebensqualitätsinterview in Anlehnung an die Arbeiten von Monika LUDWIG (1991) entwickelt und mittlerweile an nahezu 500 schizophren Erkrankten aus unterschiedlichsten Versorgungssettings angewendet. Die ersten 268 Interviews wurden einer zusammenfassenden qualitativen Inhaltsanalyse nach Philipp MAYRING (1990) unterzogen (Darstellung der durchgeführten Inhaltsanalyse in BACKES 1999), die u.a. mit Hilfe des Programms Atlas/ti (Version 1.1E; MUHR 1993) durchgeführt wurde. Da in dieser Programmversion, die allerdings mittlerweile vollkommen überarbeitet wurde, jedes Interview in einer eigenen Textdatei abgelegt wird, stieg die Zugriffszeit auf die einzelnen Interviewtexte mit wachsender Fallzahl deutlich an. Zudem ergaben sich Schwierigkeiten in der differenzierten Analyse der Antworten auf einzelne Fragen sowie in der Paraphrasierung von Textstellen. Die Umstellung auf Access erfolgte in erster Linie, um die Eingabe, Paraphrasierung und Codierung dieser und neu hinzukommender Interviews zu vereinfachen, die Analyse auf der Ebene einzelner Fragen zu erleichtern, sowie diese Daten für Zwischenauswertungen routinemäßig mit anderen Daten verknüpfen zu können. Es sei allerdings erwähnt, dass sich die gleichen hier vorgestellten Analysen ebenfalls mit der aktuellen Version von ATLAS.ti (<http://www.atlasti.de>) durchführen lassen. [3]

In einem weiteren Projekt wurde im Rahmen der Einführung eines Beschwerdemanagements an der Psychiatrischen Klinik ein Fragebogen zur Patientenzufriedenheit entwickelt, der in einer ersten Entwicklungsphase offen nach positiven und negativen Kritikpunkten seitens der Patienten sowie nach Verbesserungsvorschlägen fragte. Aus diesem Projekt liegen Fragebögen von 232 psychiatrischen Patienten vor. Die Kombination der Eingabe und Kodierung der Antworten in Access ermöglicht vergleichsweise schnell und ökonomisch die Identifikation häufig genannter Kritikpunkte und die Analyse von Zusammenhängen mit einer allgemeinen Beurteilung des Klinikaufenthalts sowie weiterer Hintergrundvariablen. [4]

Der vorliegende Artikel soll die Erfahrungen mit der Anwendung von MS Access für Erstellung, Kodierung und Datenhandling im Kontext inhaltsanalytischer Auswertung offener, strukturierter Interviews oder Fragebogenitems anhand der Untersuchung zur Lebensqualität schizophrener Menschen und zur Patientenzufriedenheit darstellen. Die Beschreibung der einzelnen Vorgehensschritte orientiert sich eng an den Anforderungen der einzelnen Projekte. Sie stellt nicht primär eine detaillierte Anleitung zum Erstellen einer entsprechenden Datenbank dar, sondern soll anhand der Beispiele von unserer Arbeitsgruppe realisierte Problemlösungen aufzeigen. Eingegangen wird auf die Erstellung von Datentabellen, Oberflächen zur Eingabe von Rohdaten (sog. "Formularen"), Kategorienlisten, auf die Paraphrasierung des Materials und seine Kodierung, sowie auf die Transformation von Datentabellen mittels sog. "Abfragen". Zudem werden die Möglichkeiten und Schwierigkeiten des Exports der Daten in das Statistikprogramm SPSS behandelt und einige Anmerkungen

zur Überprüfung der InterCoderreliabilität gemacht. Die folgenden Ausführungen setzen ein rudimentäres Wissen im Umgang mit MS Access und SPSS voraus, z.B. wie eine Tabelle, eine Abfrage oder ein Formular erstellt wird. Bei der Definition der Variablen sollte darauf geachtet werden, dass sie kompatibel mit der statistischen Auswertungssoftware sind. Für das Programmpaket SPSS bedeutet dies bspw., dass Variablennamen aus nicht mehr als acht Zeichen bestehen dürfen, das erste Zeichen keine Zahl sein darf, Umlaute nicht verwendet werden dürfen, etc. [5]

Zur Einführung soll die praktische Anwendung von MS Access am Beispiel der Daten aus dem Patientenzufriedenheitsprojekt kurz dargestellt werden. [6]

## **2. Anwendung von MS Access am Beispiel eigener Forschungsprojekte**

### **2.1 Projekt Patientenzufriedenheit: einführendes Beispiel zur Anwendung von MS Access bei offenem Antwortformat**

In einem 1-seitigen Fragebogen werden Patienten der hiesigen Klinik routinemäßig am Ende ihrer Behandlung mit vier offenen Fragen gebeten, positive und negative Kritikpunkte bzgl. ihres Aufenthaltes in der Klinik aufzuführen. Die Daten werden regelmäßig von einer Hilfskraft in eine Eingabemaske (Access-Formular) eingegeben, die optisch an die Papier-Version des Fragebogens angelehnt ist. In einer weiteren Maske (Access-Formular) werden von einem Projektmitarbeiter die Antworten mittels Auswahllisten in definierte Kategorien eingeordnet. Dieses einfache Formular zeigt Abb. 1. Beispielhaft sind Kodierungen zu den Fragen nach Zufriedenheit mit dem Aufenthalt bzw. nach Veränderungswünschen zu sehen. Die erste Ziffer der Kodierung kennzeichnet dabei den inhaltlichen Bereich (z.B. umfasst Bereich 5 Aspekte der Versorgung und Verpflegung, Bereich 1 Aspekte der Betreuung und Behandlung durch das Personal), die dritte Stelle kennzeichnet, ob es sich um eine positive (=1) oder negative (=2) inhaltliche Aussage handelt, die beiden letzten Ziffern beziehen sich auf eine konkrete inhaltliche Kategorie (z.B. 04 im Bereich 1 für die Vielfältigkeit von Therapieangeboten). Eine –8 wird für alle nicht kodierten Variablen per Voreinstellung vergeben (=fehlender Wert). Die Antwort auf Frage 4 ("Würden Sie wieder in unsere Klinik kommen, wenn Sie entsprechende Hilfe und Unterstützung benötigen?") wird nicht einer inhaltlichen Kategorie zugeordnet, sondern nach einem vorgegebenen Auswertungsschema geratet im Sinne einer globalen Beurteilung des Klinikaufenthaltes. In diesem Projekt wurde allerdings noch nicht auf die Möglichkeiten der Datenaufbereitung durch Access zurückgegriffen. Diese Datenaufbereitung findet weiterhin in SPSS statt, das mittels einer SPSS-Syntaxdatei auf die Datentabelle mit den Kodierungen zugreift, die Daten in eine "rechteckige" Datenmatrix überführt mit den Patienten/innen als Fällen und den Kategorien als Variablen, weiterhin Variablen Label und fehlende Werte definiert, und die Daten über Häufigkeitsverteilungen darstellt. Auf diese Weise können die Daten schnell und effizient nutzbar gemacht werden – allerdings nur von einem/r Mitarbeiter/in mit hinreichenden Kompetenzen bzgl. SPSS.

Microsoft Access - [Eingabe\_Codierungen]

Projekt Patientenzufriedenheit - Kodierungen der schriftlichen Patientenaussagen

Frage 2: Zufriedenheit  
Sehr zufrieden war ich mit der Versorgung; gut gefallen hat mir die Gliederung der Stationen in einzelnen Häusern und die Möglichkeit der Bewegung und Begegnung in der Parkanlage. Sehr vielfältig und gut fand ich die Therapieangebote. Die Möglichkeit, die Station zu verlassen um Aktivitäten in der Stadt nachzugehen war sehr positiv.

Kod. 1: 50101 Kod. 2: 40101 Kod. 3: 10104 Kod. 4: 40101 Kod. 5: 40152 Kod. 6: -8

Frage 3: Veränderungswünsche  
Angesichts der Mehrheit von Rauchern auf Station - mich eingeschlossen - würde ich mir im Haus einen Raum für Raucher wünschen, da es doch gelegentlich sehr ungemütlich kalt war draußen vor der Tür.

Kod. 1: 30204 Kod. 2: -8 Kod. 3: -8 Kod. 4: -8 Kod. 5: -8 Kod. 6: -8

Frage 4: Wiederkommen  
Ich würde wiederkommen, wenn die "Umgebung" bei akuter Erkrankung, aber auch wegen des medizinischen Erfolges...

Kodierung:  
2 eindeutig negativ  
3 eher negativ  
4 weder noch  
5 eher positiv  
6 eindeutig positiv  
7 extrem positiv  
9 sowohl positiv als auch negativ  
-9 keine Angabe

Datensatz: 55 von 238

Formularansicht

Abb. 1: Formular zur Kodierung von offenen Antworten im Rahmen einer Patientenbefragung [7]

An die Daten des o.g. Lebensqualitätsprojektes wurden wesentlich höhere Anforderungen an die Auswertung gestellt. Z.B. sollten die Aussagen der Erkrankten in einzelne, inhaltstragende Textteile umformuliert werden (Kodiereinheiten) und diese "Paraphrasierungen" wiederum kodiert werden. Beispielsweise hat ein Patient auf die Frage, was ihn in der letzten Woche besonders erfreut hatte, geantwortet: "der Besuch von Verwandten und Freunden, sowie meine gesundheitlichen Fortschritte". Als eigene inhaltstragende Kodiereinheiten wurde die Antwort umformuliert in 1. der Besuch von Verwandten, 2. der Besuch von Freunden, 3. meine gesundheitlichen Fortschritte. Dieser Schritt dient dazu, dass Aussagen für sich stehen können, um die Zuordnung eines Ratings zu einer Aussage eindeutig und damit nachvollziehbar zu machen. Es handelt sich somit nicht um eine Paraphrasierung i.S. einer Zusammenfassung mit eigenen Worten. Eine Auswertung der Antworten auf *einzelne* Fragen sollte ermöglicht werden. Ebenso sollte die Paraphrasierung und Kodierung von verschiedenen Personen durchgeführt werden können, um Informationen zur Intercoderreliabilität erhalten zu können. Daher soll im folgenden am Beispiel der offenen Lebensqualitätsdaten detaillierter aufgezeigt werden, wie eine entsprechende MS Access-Datenbank aussehen kann und wie wir die Möglichkeiten von Access für unsere Fragestellungen nutzbar machen konnten. [8]



## 2.2 Datentabellen

Für die Rohdaten (=Antworttext) wird eine Tabelle erstellt, in der für jeden Fragebogen bzw. für jedes Interview eine neue Datenzeile vorgesehen ist. Neben eindeutigen Identifizierungsfeldern (z.B. ID und Datum der Erhebung) enthält diese Tabelle so viele Memo-Felder, wie offene Fragen vorliegen. D.h. analog der Erstellung eines numerischen Feldes pro Item eines standardisierten Fragebogens werden in diesem Schritt Felder im Memo-Format definiert. Memo-Felder zeichnen sich im Gegensatz zu Textfeldern dadurch aus, dass fast unbegrenzt viel Text eingefügt werden kann. Natürlich können in dieser Tabelle auch Zahlenfelder definiert werden, somit Text- und Zahlenfelder kombiniert werden. Je nach Umfang und Forschungsfrage oder auch aus Gründen der Übersichtlichkeit ist es möglich, verschiedene Tabellen zur Erfassung der Rohdaten vorzusehen. In Abb. 2 findet sich eine Rohdatentabelle aus dem Projekt "Lebensqualität bei schizophrenen Menschen" in der Entwurfsansicht.

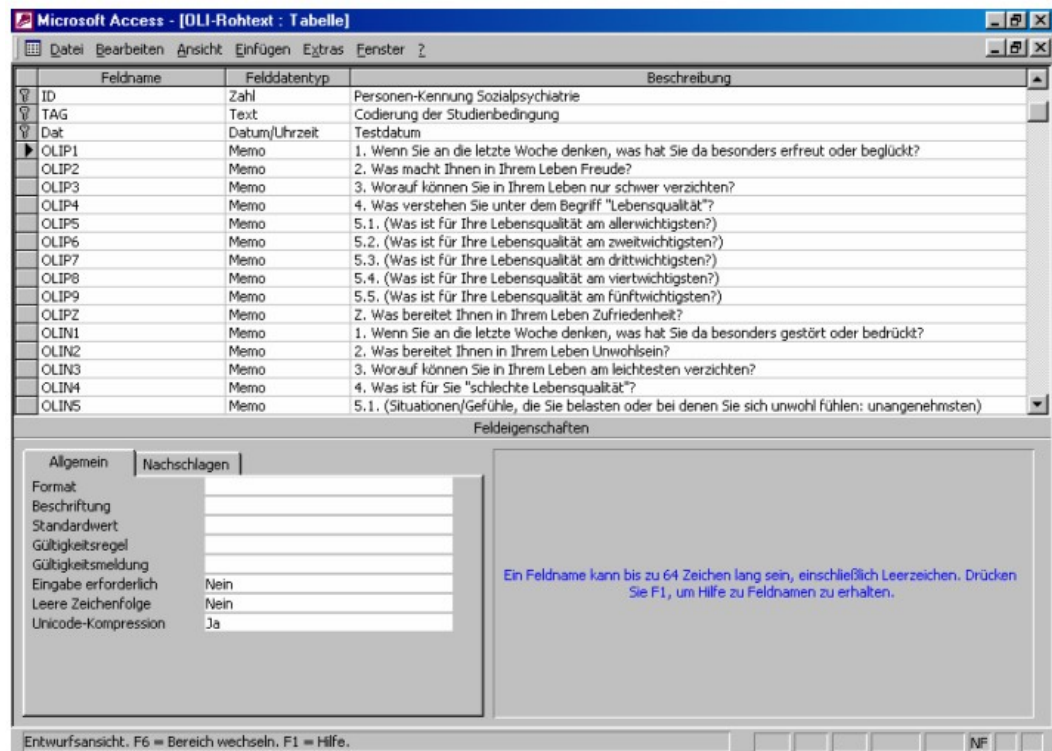


Abb. 2: Entwurfsansicht einer Tabelle zur Speicherung offener Antworten (Lebensqualitätsprojekt) [9]

## 2.3 Formulare zur Eingabe von Rohdaten

Mit Hilfe von sog. "Formularen" können flexibel Oberflächen gestaltet werden, die eine benutzerfreundliche Eingabe von Daten in die definierten Tabellen erlauben. Die Oberflächen können bspw. so gestaltet werden, dass sie dem verwendeten Fragebogen sehr ähnlich sind, womit sich bestimmte Eingabefehler vermeiden lassen. Formulare können nicht nur auf einer Tabelle basieren, sondern auch auf einer "Abfrage". Eine Abfrage kann u.a. mehrere Tabellen miteinander

kombinieren oder auch bestimmte Fälle aus einer Tabelle anhand von selber festzulegenden Kriterien auswählen. [10]

## 2.4 Paraphrasierungen der Rohdaten

Je nach Forschungsfrage ist es notwendig, aus dem Ausgangstextmaterial einzelne Analyseeinheiten zu formulieren. Damit werden inhaltstragende sprachliche Analyseeinheiten gebildet, die im Sinne einer Kodiereinheit (vgl. MAYRING 1990, S. 49) für sich stehen und damit eindeutig einer Kategorie zugeordnet werden können. Diesen Vorgang bezeichnen wir als Paraphrasierung des Ausgangstextes. Die Paraphrasierung dient dabei *nicht* der Zusammenfassung, sondern nur der Umformulierung des Materials. Da aus der Antwort auf eine Frage häufig mehrere Paraphrasierungen i.S. inhaltstragender Einheiten abgeleitet werden können, sollte datenbanktechnisch für jede Frage eine genügend große Anzahl von Feldern gebildet werden, in die der paraphrasierte Text eingefügt werden kann (vgl. Abb. 3). Da dieser Auswertungsschritt auf den Interpretationen der/des Paraphrasierers/in beruht, sollte diese Person über ein eigenes Feld identifiziert werden können, wenn mehrere Personen denselben Text paraphrasieren (s. Abb. 3, Feld PA\_NAME). Durch diesen Schritt besteht die Datenzeile nicht mehr aus *einem Fragebogen*, sondern aus *einer Paraphrasierung eines Fragebogens*. Ein Fragebogen kann somit in verschiedenen Datenzeilen erscheinen, wenn er von verschiedenen Personen paraphrasiert wird. Eine Liste von beteiligten Personen bei der Analyse der Daten kann wiederum in Form einer Tabelle erstellt werden. Ein zusätzliches Memo-Feld für jede Paraphrasierung erlaubt die Erstellung von Notizen, was gerade für komplexere Datensätze sinnvoll sein kann.



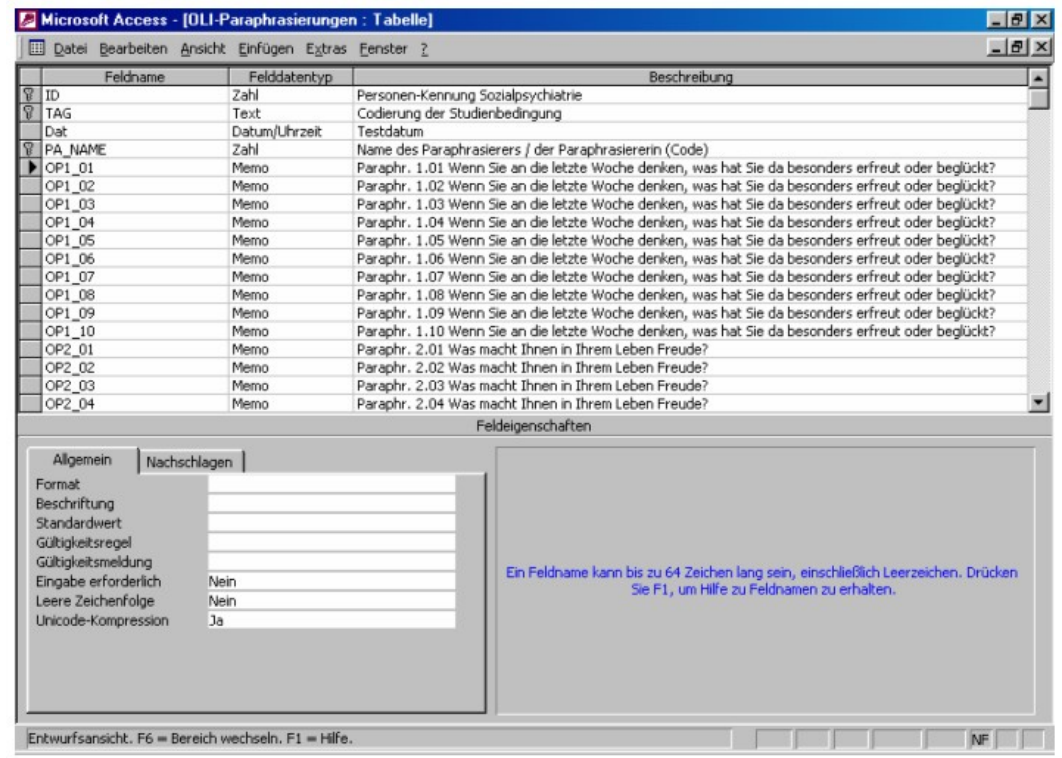


Abb. 3: Entwurfsansicht einer Tabelle zur Speicherung der Kodiereinheiten (Paraphrasierungen) [11]

Für den eigentlichen Prozess der Paraphrasierung haben wir ein Formular erstellt, das auf der linken Seite die Antworten auf eine Frage aufführt und auf der rechten Seite die Felder für die Paraphrasierungen (Abb. 4). Ggf. können in einem weiteren dazugehörigen Feld Notizen niedergelegt werden. Der "Rohtext" kann über einfache Kopier- und Einfügefunktionen von den linken Originalantwortfeldern in die Paraphrasierungsfelder auf der rechten Seite kopiert und ggf. sprachlich modifiziert werden. Der Vorgang des Paraphrasierens ist auf diese Weise arbeitstechnisch sehr komfortabel durchzuführen und bleibt zudem transparent und rekonstruierbar.

The screenshot shows a Microsoft Access form window titled "Microsoft Access - [e\_roh\_para\_positiv\_1 : Formular]". The form is titled "Paraphrasierung und Codierung der Offenen Lebensqualitäts-Interviews (OLI-P)" and contains the question "2. Was macht Ihnen in Ihrem Leben Freude?". On the left, there is a text box with the red text "Musik hören, Kaffee trinken, rauchen, Radio hören (Nachrichten)". To the right, there are several rows of input fields. Each row contains a text field for an activity and a dropdown menu for a numerical code. The visible data is as follows:

Activity	Code
Musik hören	7.7.1
Kaffee trinken	3.2.1.
rauchen	3.3.
Radio hören	7.1.1
Nachrichten hören	12.3

At the bottom of the form, there is a status bar showing "Datensatz: 16 von 939" and "Formularansicht".

Abb. 4: Formular zur Erstellung von Kodiereinheiten (Paraphrasierung) und zur Kodierung mittels Auswahllisten [12]

## 2.5 Kategoriensystem

Das Kategoriensystem wird in Form einer Tabelle erstellt. Dabei entspricht eine Datenzeile einer Kategorie. Die Tabelle sollte ein Feld beinhalten, das die Kategorie kurz bezeichnet, ein weiteres Feld, das diese Kategorie ausführlicher beschreibt, sowie eine Zahlenfeld, das einer Kategorie genau einem Zahlenwert (Code) zuordnet. In beiden Beispielfällen lag das jeweilige Kategoriensystem bei Erstellung der MS Access-Datenbank weitestgehend vor. Es waren nur noch kleinere Modifikationen am Kategoriensystem notwendig. Jerry NATKIN (2000) beschreibt ausführlicher, wie MS Access bei der Entwicklung eines Kategoriensystems verwendet werden kann bzw. inwieweit vorhandene Kategorien im Verlauf der Entwicklung eines Kategoriensystems sinnvoll angepasst werden können. [13]

## 2.6 Kodierung des Materials

Das Ergebnis der Analyse, d.h. die Zuordnung von Antworten zu Kategorien (Codes), wird in einer eigenen Tabelle festgehalten, deren Entwurfsansicht Abb. 5 zeigt. Diese Tabelle ist identisch aufgebaut wie die Tabelle mit den Paraphrasierungen, nur dass anstelle der Paraphrasierungen jeweils ein numerisches Feld definiert wird, das den Code, also das Resultat der Kodierung, beinhaltet. Wie in der Paraphrasierungstabelle beinhaltet eine Datenzeile die Kodierung eines einzelnen Fragebogens durch eine/n Kodierer/in. Wir haben aus

praktischen Gründen die beiden Schritte *Paraphrasierung* und *Kodierung* zusammengefasst, d.h. aus den Originalaussagen wird ein inhaltstragender Aspekt in das Paraphrasierungsfeld kopiert und ggf. paraphrasiert, anschließend wird von *derselben* Person der Inhalt des Paraphrasierungsfeldes kodiert (vgl. Abb. 5). Theoretisch ist es möglich, wiederum Paraphrasier/in und Kodierer/in voneinander zu trennen, was jedoch zu einer höheren Komplexität der Datenstruktur führen würde. [14]

Die Kodierung des Materials erfolgt in unserem Beispiel im selben Formular, in dem auch die Paraphrasierungen erfolgen. Es enthält damit zusätzlich zur Kontrolle auch die Original-Antwort des / der Befragten. Die numerischen Codes der Kategorien müssen nicht per Hand eingetippt werden, was eine deutliche Fehlerquelle darstellen kann. Die Kodierung erfolgt vielmehr über Auswahllisten, d.h. das Feld, in das der Code eingetragen werden soll, wird mit der Tabelle verknüpft, die die Kategorien auflistet (vgl. Abb. 1 und 4). Sie erscheint beim Anklicken mit einer "Laufleiste" als Auswahlliste, in der die entsprechende Kategorie nur noch markiert werden braucht. Auch wenn diese Form der Eingabe durch ungenaue Bedienung der Maus nicht absolut fehlerfrei ist, ist sie gegenüber der manuellen Eingabe der bis zu 8-stelligen Codes sicherlich in puncto Eingabeobjektivität als vorteilhafter zu bewerten.

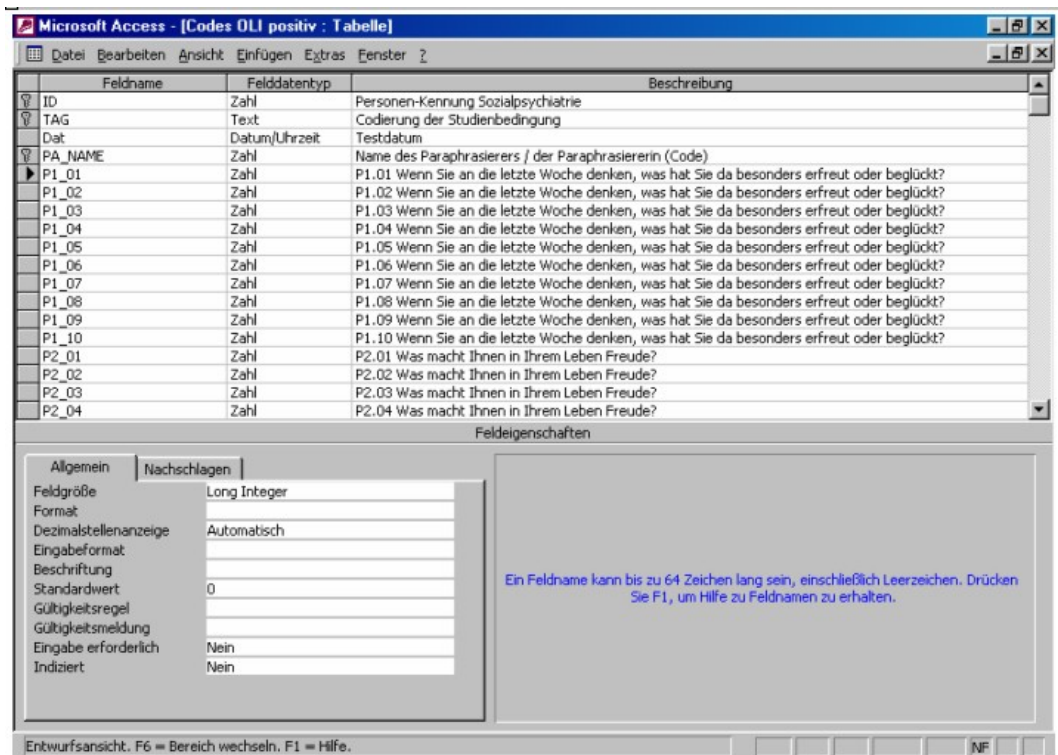


Abb. 5: Entwurfsansicht einer Tabelle zur Speicherung der Codes der kodierten Paraphrasierungen [15]

## 2.7 Transformation von "ungeordneten" Tabellen in rechteckige Datenmatrizen

Unter der Voraussetzung, dass eine Quantifizierung des Textmaterials anhand des Kategoriensystems inhaltlich sinnvoll ist, stellt sich die Frage, welche Datenstruktur als Ausgangsbasis einer statistischen Analyse dienen kann. Im vorliegenden Fall der Aussagen schizophrener erkrankter Menschen zu ihrer Lebensqualität sollte die *Häufigkeit der Nennung* eines bestimmten inhaltlichen Aspektes ein Indikator für die Bedeutung dieses Aspektes für die Lebensqualität dieser Personengruppe darstellen. Zum anderen sollten Zusammenhangsanalysen (d.h. welche Kategorien bzw. inhaltlichen Bereiche werden besonders häufig mit bestimmten anderen Kategorien genannt) ermöglicht werden. In der Ausgangsdatei für die Auswertungen sollte daher eine Person eine Zeile darstellen und jede Kategorie jeweils eine Variable ("rechteckige" Datenmatrix). Neben der Häufigkeit der Nennung eines inhaltlichen Aspektes hat uns interessiert, ob ein bestimmter Inhalt überhaupt im Interview angesprochen wurde oder nicht. Daher sollten die Variablen mit den jeweiligen Kategorien sowohl die *Anzahl der Nennungen* als auch *dichotomisiert* (Inhalt wurde genannt vs. nicht-genannt) vorliegen. [16]

Eine solche Datenstruktur entspricht allerdings nicht der Datenstruktur, in der die Daten nach der Kodierung der Paraphrasen vorliegen. Im nächsten Verarbeitungsschritt muss daher die stark von den qualitativen Aspekten des Ausgangsmaterials abhängige und daher flexible Datenstruktur in eine Datenstruktur überführt werden, die den formalen Kriterien der quantitativen Datenanalyse entspricht. Daher soll im folgenden erläutert werden, wie sich eine solche Umstrukturierung der Daten mit Hilfe der innerhalb von MS Access zur Verfügung stehenden Programmiersprache Visual Basic, einschließlich ihrer Access-spezifischen Erweiterungen, realisieren lässt. [17]

Nach der Kodierung der Paraphrasen liegen für jede befragte Person Listen von Kodierungen vor, die im Sinne der statistischen Analyse "ungeordnet" sind. Sie bestehen aus einer Aneinanderreihung von Kodierungen, bei der die Abfolge der Listenelemente durch die Aufeinanderfolge von inhaltstragenden Einheiten im ursprünglichen Textmaterial oder zusätzlich durch die Paraphrasierung bestimmt wird. Die Listen weisen einen nur geringen Strukturierungsgrad auf, da sie a) für die einzelnen Fälle und Texteinheiten eine nur kleine und zudem wechselnde Anzahl von Elementen verglichen mit der Gesamtanzahl möglicher Kodierungen enthalten, b) für die verschiedenen Fälle und Texteinheiten in der Regel eine unterschiedliche Auswahl von Kodierungen aus dem Gesamt aller Kodierungen umfassen und bei denen c) die Reihenfolge der Kodierungen nicht notwendigerweise der theorieabhängigen Struktur des Kategoriensystems entspricht, das den Kodierungen zugrunde liegt. Demgegenüber setzt die angestrebte statistische Auswertung die Anordnung der Daten in "rechteckigen" Datenmatrizen voraus, bei denen die Zeilen der Matrix den befragten Personen / Fällen und die Spalten den bei den einzelnen Personen erfassten Merkmalen / Variablen entsprechen. Eine einzelne Variable enthält dabei die Maßzahlen der Fälle für ein einzelnes, bei allen Fällen gleiches Merkmal. Bei allen Fällen müssen

dieselben Variablen in gleicher Anordnung vorliegen. Das Gesamt der Variablen entspricht allen Merkmalen der Personen, die im Rahmen der untersuchten Fragestellung erfasst wurden. Bezogen auf die Kodierung von Textmaterial entsprechen die Variablen (erfassten Merkmale) eines Falles (einer Person) der Information darüber, ob für jede einzelne Kategorie des Kategoriensystems eine Zuordnung zu jeder einzelnen Texteinheit dieses Falles erfolgte oder nicht. Die Anzahl notwendiger Variablen entspricht daher dem Produkt aus der Anzahl der Texteinheiten und der Anzahl der Kategorien. [18]

Das offene Lebensqualitätsinterview für schizophrene Menschen umfasst 10 offene Fragen in fester Reihenfolge. Die inhaltstragenden Aspekte der freien Antworten auf diese offenen Fragen werden einem Kategoriensystem zugeordnet, das 230 hierarchisch geordnete Kategorien enthält. Das Auswertungsschema sieht vor, dass bei der Paraphrasierung bzw. Kodierung bis zu zehn inhaltstragende Aspekte aus den Einzel-Antworten zu jeder der zehn Fragen extrahiert werden können (vgl. Abb. 3 und 4). Jeder der tatsächlich extrahierten Paraphrasen wird ein Kode (Kategorie) zugewiesen. Nach Abschluss dieses Schrittes liegen für das Textmaterial jeder Person maximal 100 Kodierungen vor, für die in der Datenbank entsprechende Datenfelder vorgehalten werden müssen. In der konkreten Auswertung wurden allerdings in den allermeisten Fällen von einer Frage nur ein bis drei Paraphrasen abgeleitet. Es wird deutlich, dass sich die Kodierungslisten für die einzelnen Personen und offenen Fragen in Abhängigkeit von Inhalt und Umfang der Antworten erheblich in Anzahl und Inhalt der Kodierungen unterscheiden. [19]

Für eine sinnvolle Weiterverarbeitung in einem Statistikprogramm müssen diese ungeordneten Kategorienlisten so umstrukturiert werden, dass bei jeder der zehn Fragen jede der 230 Kategorien als einzelne Variable in einer rechteckigen Datenmatrix repräsentiert ist. Dazu sind demnach insgesamt 2300 Variablen notwendig. Bei der Umstrukturierung der Daten muss berücksichtigt werden, dass sich die Kodierungen nicht auf die einzelnen Fragen selbst, sondern auf die aus ihnen abgeleiteten bis zu zehn Paraphrasen beziehen. Eine solche komplexe Umstrukturierung der Daten ist mit den Möglichkeiten, die das Datenbanksystem MS Access über die interaktive menü- oder makro-basierte Programmsteuerung anbietet, nicht möglich. Eine Automatisierung der Umstrukturierung kann nur mit Hilfe eines Algorithmus erfolgen, der die spezifischen Ein- und Ausgabedatenstrukturen berücksichtigt. Ein solcher Algorithmus lässt sich mit der in Access integrierten Programmiersprache Visual Basic einschließlich ihrer Access-spezifischen Erweiterungen realisieren. Die Grundzüge des Algorithmus zur Datenumstrukturierung sollen im folgenden erläutert werden. [20]

Der Datenumstrukturierungsalgorithmus greift auf Abfragen zurück, die getrennt für jede der zehn offenen Interview-Fragen mit Hilfe des Abfrageneditors von MS Access erstellt wurden. In einer derartigen Abfrage werden die "ungeordneten" Felder der Paraphrasierungstabelle, die die Kodierungen für die Paraphrasen einer bestimmten Interview-Frage enthalten, mit der "rechteckigen" Datentabelle verbunden, die für die betreffende Interview-Frage die Felder für die einzelnen Kategorien des Kategoriensystems enthält. Nachdem eine Abfrage innerhalb

einer Visual-Basic-Prozedur geöffnet worden ist, lässt sich mit Hilfe Access-spezifischer Programmiersprachenbefehle auf die einzelnen Datenfelder der Datensätze der beiden verbundenen Tabellen zurückgreifen. Die Umstrukturierung erfolgt in mehreren geschachtelten Programmschleifen. In der äußersten Schleife werden nacheinander die zehn Abfragen für die verschiedenen Interview-Fragen geöffnet. In den einzelnen Datensätzen einer Abfrage werden in den "ungeordneten" Feldern der Paraphrasierungstabelle die tatsächlich für eine Interview-Frage vergebenen Paraphrasierungskodes identifiziert. Anschließend wird abgezählt, für wie viele Paraphrasen einer Interview-Frage ein bestimmter Kode vergeben wurde. Das Ergebnis der Zählung wird in das Feld der zweiten Tabelle geschrieben, das der Kategorie mit dem jeweiligen Paraphrasierungskode entspricht. Am Ende dieses Verarbeitungsschrittes liegt für jede der zehn Interview-Fragen eine Tabelle vor, aus der für jede Person entnommen werden kann, wie oft eine bestimmte Kategorie des Kategoriensystems für die Paraphrasen der zugehörigen Interview-Frage vergeben wurde. Diese Datentabellen entsprechen "rechteckigen" Datenmatrizen, wie sie von Statistikprogrammen erwartet werden. [21]

Die im letzten Auswertungsschritt erzeugten Datenmatrizen enthalten die vollständigen Informationen über alle kodierten Antworten. Für die Identifizierung charakteristischer Antwortmuster oder die Untersuchung von Zusammenhängen mit anderen Variablen muss diese Information aber zunächst verdichtet werden. Auch diese Verdichtung wurde in dem beschriebenen Projekt innerhalb der Access-Datenbank mit Hilfe von Visual-Basic-Prozeduren durchgeführt. Obwohl eine Datenverdichtung auch innerhalb der Statistikprogramme möglich ist, schien es uns sinnvoll zu sein, die primäre Datenaufbereitung auf Ebene der eigentlichen Datenbank durchzuführen und dem Statistikprogramm einen Datensatz zu übergeben, der ohne weitere Vorverarbeitung statistisch ausgewertet werden kann. [22]

In einem ersten Schritt zur Informationsreduktion wurden die Ergebnisse der Kategorienzählung bei den verschiedenen Fragen dichotomisiert (zur inhaltlichen Bedeutung der Dichotomisierung s.o.). Die Datentabellen für die dichotomisierten Daten haben dieselbe Felderstruktur, wie die zuvor beschriebenen Matrizen mit den Zählergebnissen. Die Felder enthalten allerdings nicht das Ergebnis der Zählung, sondern nur noch den Wert 1, wenn die betreffende Kategorie mindestens einer Paraphrase der jeweiligen Interview-Frage zugeordnet wurde, oder 0, wenn diese Kategorie der Interview-Frage überhaupt nicht zugeordnet wurde. In einem abschließenden Schritt der Datenvorverarbeitung wurden die Daten einer Person zu den zehn verschiedenen Interview-Fragen zusammengefasst und in einer einzelnen Tabelle aggregiert. Auch dieser Schritt wurde mit Hilfe von Visual-Basic-Prozeduren realisiert, in denen je eine Access-Tabelle für die aggregierten Häufigkeits- bzw. dichotomisierten Daten erzeugt wurde. Die zuletzt erzeugten Tabellen enthalten die Information darüber, wie oft bzw. ob die einzelnen Kategorien des Kategoriensystems dem Textmaterial einer Person insgesamt zugeordnet wurden. [23]



## 2.8 Export der Daten in SPSS

Über den sog. ODBC-Treiber können MS Access Tabellen und Abfragen in SPSS eingelesen werden (im SPSS Menü Datei → Datenbank öffnen → Neue Abfrage). Der Befehl zum Einlesen einzelner Tabellen bzw. Abfragen kann in Syntax-Editor eingefügt werden, so dass auch routinemäßig schnell auf eingegebene Daten zurückgegriffen werden kann. Ein aufgrund der zumeist hohen Anzahl von Variablen bedeutendes Problem besteht dabei darin, dass die in Access vergebenen Bezeichnungen der Variablen in SPSS nicht als Variable Labels eingelesen werden. Sie müssen daher komplett noch einmal in SPSS-Syntax erstellt werden. [24]

## 2.9 Überprüfung der Intercoderreliabilität

Mittels der oben beschriebenen Visual Basic Applikationen ließen sich innerhalb von Access auch Datenaufbereitungen zur weiteren Analyse der Intercoderreliabilität realisieren. Beispielsweise könnten im Falle von zwei Kodierern die Übereinstimmungen und Abweichungen der Kodierungen auf der Ebene der einzelnen Fragen und auch der einzelnen Kategorien identifiziert werden. Damit würde eine detaillierte Untersuchung dazu ermöglicht, welche Kategorien besonders uneindeutig sind bzw. ob möglicherweise bestimmte Frage eine geringere Übereinstimmung zur Folge haben. Im vorliegenden Beispiel wurde die Aufbereitung und Analyse der Daten zur Intercoderreliabilität allerdings erst in SPSS mit Hilfe des AGGREGATE-Befehls vorgenommen. [25]

## 3. Fazit

Prinzipiell sind die hier durchgeführten Verarbeitungsschritte und -analysen auch mit einer textanalytischen Software wie ATLAS.ti (<http://www.atlasti.de>) oder WinMax (<http://www.maxqda.de/>) möglich. Spezielle textanalytische Software ist für die *Erstellung* eines Kategoriensystems sogar deutlich besser geeignet – wenn auch in Access mit größerem Aufwand entsprechend programmierbar. Der Vorteil von Access für die inhaltsanalytische Auswertung liegt in seiner Verbreitung und Flexibilität. Vorhandene Kategoriensysteme können routinemäßig auf neu anfallendes Textmaterial angewendet werden, ohne Einarbeitung in spezielle textanalytische Software. In unserer Forschungsgruppe erleichtert die Verarbeitung der Text- und quantitativen Daten in einem Datenbanksystem den Forschungsalltag. Unser Eingangsbeispiel zur Patientenbefragung sowie der von Jerry NATKIN (2000) vorgestellte Ansatz verweisen darauf, dass *auch ohne Programmierkenntnisse* mittels Access ein eigenes Analyseprogramm erstellt werden kann, das flexibel den eigenen Ansprüchen und Anforderungen angepasst werden kann und sogar die Möglichkeit von einfachen quantitativen Auswertungen bietet. Allerdings sind für komplexere Anwendungen wie in unserem Beispiel der Inhaltsanalyse von Lebensqualitätsinterviews mit schizophren erkrankten Menschen Programmierkenntnisse bzgl. Visual Basic unverzichtbar. Damit konnten wir sehr spezifischen Anforderungen an die Aufbereitung und Auswertung der Daten

genügen. Dieser Aufwand erscheint uns gerade dann lohnenswert, wenn offene Antworten in größerer Fallzahl routinemäßig ausgewertet werden sollen. [26]

Weitere Möglichkeiten für inhaltsanalytische Auswertungen erschließen sich für uns aus zukünftigen forschungspraktischen Anforderungen. Beispielsweise eröffnen Auswahlabfragen die Möglichkeit, schnell auf bestimmte Textpassagen zurückzugreifen. Damit ließe sich ein auf einer Auswahlabfrage basierendes Formular erstellen, das alle Textpassagen auflistet, für die ein bestimmter Code vergeben wurde. [27]

Insbesondere für strukturierte, offene Interviews oder Fragebögen erscheint die Datenverarbeitung in MS Access sehr geeignet. Die Übertragung dieses Ansatzes auf notwendige Kodierungen im Rahmen von Untersuchungen mit primär geschlossenen Antwortformaten liegt sehr nahe. [28]

## Literatur

- Backes, Sandra (1999). Komponenten der subjektiven Lebensqualität schizophrener Patienten. Eine qualitative Inhaltsanalyse offener Interviews. *Diplomarbeit am Fachbereich Psychologie der Justus-Liebig-Universität Gießen*.
- Franz, Michael; Meyer, Thorsten & Gallhofer, Bernd (2001). Lebensqualität und Neuroleptikabehandlung in der Schizophrenie – aktueller Stand der Forschung. In Dieter Naber & Franz Müller-Spahn (Hrsg.), *Leponex, Pharmakologie und Klinik eines atypischen Neuroleptikums, Aspekte der Lebensqualität in der Schizophrenietherapie* (S.72-83). Berlin: Springer.
- Franz, Michael; Meyer, Thorsten & Gallhofer, Bernd (2002/im Druck). Subjektive Lebensqualität schwer chronifizierter schizophrener Langzeitpatienten. Teil 3 der Hessischen Enthospitalisierungsstudie. *Psychiatrische Praxis*.
- Ludwig, Monika (1991). Lebensqualität auf der Basis subjektiver Theoriebildung. In Monika Bullinger, Monika Ludwig & Nicole von Steinbüchel (Hrsg), *Lebensqualität bei kardiovaskulären Erkrankungen* (S.24-34). Göttingen: Hogrefe.
- [Mayring, Phillip](#) (1990). *Qualitative Inhaltsanalyse* (2. Aufl.). Weinheim: Deutscher Studien Verlag.
- Muhr, Thomas (1993). *Atlas/ti. Computer aided text interpretation & theory building – Release 1.1E. User's Manual* (2. Aufl.), Berlin.
- Natkin, Jerry (2000). Using Microsoft Access for qualitative data analysis: an introduction. *Paper presented at the Annual Meeting of the American Evaluation Association, Honolulu, November 2000*.

## Zu den Autoren

*Thorsten MEYER*, Dipl.-Psych., Jg. 1968, 1996  
Diplom in Psychologie an der Christian-Albrechts  
Universität Kiel, seit 1996 wissenschaftlicher  
Mitarbeiter in der Arbeitsgruppe Sozialpsychiatrie,  
Psychiatrische Klinik am Universitätsklinikum der  
Justus-Liebig Universität Gießen

*Arbeitsschwerpunkte*: Entwicklung und Bewertung  
von Indikatoren psychiatrischer Versorgung,  
Evaluation (sozial-) psychiatrischer Interventionen.

Kontakt:

Thorsten Meyer

Arbeitsgruppe Sozialpsychiatrie  
Psychiatrische Klinik am Universitätsklinikum  
Justus-Liebig Universität  
Am Steg 24  
D-35385 Gießen

E-Mail: [thorsten.meyer@psychiat.med.uni-giessen.de](mailto:thorsten.meyer@psychiat.med.uni-giessen.de)

*Harald GRUPPE*, Dr. phil., Dipl.-Psych., Jg. 1954,  
1984 Diplom in Psychologie an der Justus-Liebig-  
Universität Gießen, 1991 dort Promotion zum Dr.  
phil., seit 1993 wissenschaftlicher Mitarbeiter in  
der Psychiatrischen Klinik des  
Universitätsklinikums Gießen

*Arbeitsschwerpunkte*: Biometrische, datenbank-  
und netzwerktechnische Betreuung von  
Forschungsvorhaben, kognitive  
Psychophysiologie.

*Michael FRANZ*, Dr. med., Jg. 1958, Studium von  
Medizin und Soziologie in Gießen, 1985  
medizinisches Staatsexamen, bis 1989  
Assistenzarzt psychiatrisches Krankenhaus  
Marburg, seitdem Universitätsklinikum Gießen,  
1995 Promotion zum Dr. med., seit 1996 Aufbau  
und Leitung der Arbeitsgruppe Sozialpsychiatrie  
an der JLU Gießen, Leitung des Akutbereiches  
und eines gemeindebasierten, integrierten  
Versorgungsmodells für psychisch Kranke

*Arbeitsschwerpunkte*: Sozialpsychiatrie in  
Versorgung und Forschung, Grundlagenforschung  
zu psychiatrischen Indikatoren

## Zitation

Meyer, Thorsten; Gruppe, Harald & Franz, Michael (2002). Microsoft Access in der Analyse von Fragebögen und Interviews mit offenen Antwortformaten [28 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(2), Art. 20, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0202206>.

Revised 2/2007