

Texterfassung für sozialwissenschaftliche Auswertung

Alfons Geis

Zentrum für Umfragen, Methoden und Analysen, Mannheim

Zusammenfassung

Der Stellenwert der "Texterfassung" im Forschungsprozess soll hervorgehoben und das Problembewusstsein für den Umgang mit Texten in sozialwissenschaftlichen Untersuchungen geschärft werden.

Es werden Hinweise gegeben, wie Texte entsprechend wissenschaftlichen Kriterien erfasst und für eine weitere Verwendung aufbereitet werden und wie ein dazu notwendiges Regelwerk erstellt wird.

ZUMA How-to-Reihe Nr. 13

2004

Inhaltsverzeichnis

<i>1 Grundsätze und allgemeine Hinweise</i>	3
<i>1.1 Warum Texterfassung problematisieren?</i>	3
<i>1.2 Forschungsfrage und Regelungsbedarf</i>	3
<i>1.3 Originalvorlage und Regelungsbedarf</i>	4
<i>1.3.1 Textübernahme</i>	4
<i>1.3.2 Einscannen</i>	5
<i>1.3.3 Abschreiben</i>	5
<i>1.3.4 Transkribieren von Tonaufzeichnungen</i>	6
<i>2 Regeln zur Texterfassung</i>	6
<i>2.1 Was muss geregelt werden</i>	7
<i>2.1.1 Umfang und Auswahl</i>	7
<i>2.1.2 Einheiten und Identifikatoren</i>	7
<i>2.1.3 Dateiformat</i>	9
<i>2.1.4 Kontrolle und Nachbearbeitung</i>	10
<i>2.2 Was soll geregelt werden</i>	10
<i>2.2.1 Nicht lesbar/unverständlich</i>	10
<i>2.2.2 Fehlender Wert/Redeunterbrechung</i>	10
<i>2.2.3 Rechtschreibung/Sprache</i>	11
<i>2.2.4 Arbeitsorganisation</i>	12
<i>2.3 Was kann geregelt werden</i>	12
<i>2.3.1 Textgestaltungsmerkmale</i>	12
<i>2.3.2 Groß-/Kleinschreibung</i>	12
<i>2.3.3 Sprechpausen</i>	13
<i>2.3.4 Nonverbale Äußerungen</i>	13
<i>2.3.5 Lautstärke</i>	13
<i>2.3.6 Zitate</i>	13
<i>2.3.7 Syntaktische Informationen</i>	13
<i>2.3.8 Anmerkungen, Kommentare</i>	14
<i>3 Beispiele für Regeln und Vorgehensweisen</i>	14
<i>3.1 Textübernahme</i>	14
<i>3.2 Abschreiben</i>	17
<i>3.3 Transkription</i>	21
<i>4 Schlussbemerkung</i>	22
<i>Literaturhinweise</i>	23

Die Verwendung von Texten in sozialwissenschaftlichen Untersuchungen gewinnt an Bedeutung, vor allem durch die zunehmende Anzahl von Texten, die als Dateien vorliegen oder die im Zuge von Untersuchungen entstehen, angefangen bei offenen Fragen im standardisierten Interview über Leitfadengespräche oder Gruppendiskussionen bis zum therapeutischen Gespräch. Um daraus einen analysierbaren Text zu erhalten, sind einige Arbeitsschritte notwendig, die eindeutig zu regeln sind.

Der Leser soll die Problematik dieser Untersuchungsphase erkennen und in die Lage versetzt werden, ein Regelwerk für Texterfassung zu erstellen, das den Erfordernissen der Wissenschaftlichkeit wie der Fragestellung gerecht wird.

Nach allgemeinen Hinweisen zur Texterfassung werden die Punkte aufgeführt, die geregelt werden müssen bzw. können und was dabei zu beachten ist. Im dritten Teil werden exemplarische Regelwerke vorgestellt und kommentiert.

1 Grundsätze und allgemeine Hinweise

1.1 Warum Texterfassung problematisieren?

Das große Textangebot und die leichte Verfügbarkeit – der Text liegt vor und muss nicht erst „produziert“ werden – lässt leicht darüber hinwegsehen, dass durch die Verwendung und Weiterbearbeitung des Textes etwas mit ihm geschieht. Er wird verändert, und wenn es nur Farbe, Größe, Schrifttyp oder Layout sind, die im Original anders aussehen als im analysierten Text. Noch mehr wird verändert, wenn der Text neu erstellt oder abgeschrieben wird. Und eine gänzlich neue Form entsteht bei der Transkription von Audiovorlagen.

Soll eine Vorgehensweise wissenschaftlich sein, müssen u.a. die Kriterien Systematik und Intersubjektivität/Transparenz gewährleistet sein. Bezogen auf die Texterfassung bedeutet dies, dass jede Veränderung des Originaltextes systematisch - nicht ad hoc oder willkürlich - vorgenommen werden muss und dieser Prozess genau protokolliert und dokumentiert werden muss (nicht: „... ich weiß doch, was ich gemacht habe.“).

Sozialwissenschaftliche empirische Projekte bestehen aus vielen Arbeitsschritten, die aufeinander aufbauen. Die Qualität des Projektes kann nur so gut sein, wie der schlechteste Arbeitsschritt. Sind die Antworten auf eine offene Frage oder ein Intensivinterview unzureichend verschriftet worden, kann die darauf aufbauende Auswertung auch nur unzureichende Ergebnisse hervorbringen.

Gesprochene Sprache als Tonaufzeichnung ist zwar noch kein Text, kann aber in einen solchen überführt werden, indem das Gehörte niedergeschrieben wird, wonach es wie Text zu behandeln ist. Insofern wird Transkription zur Texterfassung gezählt.

1.2 Forschungsfrage und Regelungsbedarf

Wenn Texte ausgewertet werden, handelt es sich im weitesten Sinne um Textanalyse. Der Breite des methodischen Instrumentariums (vgl. Alexa 1997) entspricht die Vielfältigkeit des Verwendungszweckes: Erstellung einer Stichwort- oder Themensammlung, Festlegung von Deskriptoren einer Bibliothek, Suche nach Vorgaben für eine geschlossene Frage, Dokumentation oder Archivierung von Texten, Bestimmung der Verbalisierungsfähigkeit des Befragten oder gar Erfolgskontrolle einer psychotherapeutischen Behandlung. Je nach Zweck erfordert die dazugehörige Texterfassung und -aufbereitung einen unterschiedlichen Aufwand. Für das letzte Beispiel, die Erfassung von psychotherapeutischen Gesprächen, ist der Regelungsbedarf am höchsten. Wenn ein vorliegender Text lediglich aufbewahrt werden

soll, ist der Aufwand deutlich geringer, und wenn ich nur wissen will, welche Themen in einem Zeitungsbeitrag oder nach einer offen gestellten Frage angesprochen werden, ist er noch geringer.

Der Verwendungszweck im engeren Sinn wird beschrieben durch die Forschungsfrage. Um beispielsweise (in einem Pretest) die Motive für die Teilnahme an einer Online-Befragung zu erforschen, ist der Zweck eine textanalytische Auswertung, sei es um Vorgaben für eine spätere geschlossene Frage zu ermitteln oder um die Antworten zu strukturieren. Im ersten Fall reicht vielleicht eine Auflistung der möglichen Motive für die Teilnahme, ohne die gesamte Antwort zu erfassen. Für eine strukturierte Darstellung aller Befragtenaussagen muss alles Gesagte notiert werden; sonst würde die Darstellung des Befragungsergebnisses unvollständig. Soweit es die finanziellen Ressourcen zulassen, sollte ein minimalistisches Vorgehen vermieden werden, weil sich im Laufe einer Untersuchung sehr oft weitere Fragestellungen ergeben, die zu Beginn nicht absehbar waren, bei vollständiger Texterfassung aber beantwortet werden können.

So wie hier die Frage der Vollständigkeit anhand des Ziels zu beantworten ist, sind auch alle anderen Regulierungsmöglichkeiten, wie sie im Abschnitt 2 (Regeln zur Texterfassung) beschrieben werden, mehr oder weniger zu berücksichtigen. Maßgeblich ist immer die Forschungsfrage.

1.3 Originalvorlage und Regelungsbedarf

Ob der Regelungsbedarf im Sinne der Wissenschaftlichkeit groß oder klein ist, richtet sich weiterhin nach dem Maß der Veränderung, die der Originaltext erfährt. Es gibt vier Formen von Textquellen, die sich bezüglich des erforderlichen Regelungs- und Arbeitsaufwandes für die Weiterverarbeitung unterscheiden: Übernahme, Scannen, Abschreiben und Transkribieren.

1.3.1 Textübernahme

Der geringste Aufwand entsteht und die geringste Veränderung erfolgt, wenn der Text bereits als Datei vorliegt und maschinenlesbar ist. Er muss nur in das erforderliche Textformat gebracht und ggf. gegliedert, d.h. mit Identifikatoren (vgl. Punkt 2.1.3) versehen werden. Der Originaltext erfährt gegenüber den anderen Texterfassungsformen die geringste Veränderung. In der Forschung steigt der Anteil dieser Textart kontinuierlich: es gibt kaum eine Notiz, die nicht ohnehin am Computer erstellt wird, kaum ein aktuell produziertes Papier, das nicht auch über das Netz verfügbar wäre. Neben diesen aus der Alltagsarbeit entstandenen Texten werden Textkorpora und Medientexte auch systematisch gesammelt und zur Verfügung gestellt (meist gegen Gebühr): Institut für Deutsche Sprache, GBI (Gesellschaft für Betriebswirtschaftliche Informationen mbH), Projekt Gutenberg (frei beziehbare deutschsprachige Literatur im Internet), Lexis-Nexis, CD-ROMs mit kompletten Jahrgängen von Zeitungen und Zeitschriften u.a.m. - die aktuellen Links sind mit den entsprechenden Stichworten über Suchmaschinen sehr leicht zu finden.

Wird ein Text für die eigene Untersuchung übernommen, muss festgelegt werden, was übernommen wird, wie der Text gegliedert wird und in welchem Textformat er gebraucht wird - in Abschnitt 2.1 (Muss geregelt werden) beschrieben.

1.3.2 Einscannen

Der Text liegt bereits in schriftlicher Form vor, muss aber noch mittels Scanner und Texterkennungssoftware in eine bearbeitbare Datei überführt werden.

Ein Scanner erfasst zunächst nur ein Bild, dessen einzelne Punkte (Pixel) durch Farbe und Helligkeit definiert sind. Die separate Software fasst zusammengehörige Bildpunkte zu Buchstaben zusammen, denn nur diese können von den Textverarbeitungsprogrammen erkannt werden. Der Vorgang, die Bildpunkte als Buchstaben zu interpretieren, gelingt unterschiedlich gut. Er ist abhängig von der Qualität der Vorlage und der Software, die technischen Merkmalen des Scanners (Auflösungsvermögen, Lesegeschwindigkeit u.ä.) spielen kaum noch eine Rolle, weil der Minimalstandard heute schon ausreichend ist. Für große Mengen (regelmäßig mehrere Tausend Seiten) ist ein professionelles Gerät mit automatischem Einzug unerlässlich; dessen Kosten liegen nicht bei ca. 100 € und aufwärts wie für ein einfaches Gerät, sondern beim Zehnfachen. Die Fähigkeiten der Software zur Texterkennung sind in einem ständigen Wandel begriffen. Den aktuellen Stand findet man in gängigen PC-Zeitschriften (z.B. PC Professional). Für gute Vorlagen und wenige Vorlagen reicht die zusammen mit dem Scanner gelieferte Software meist schon aus. Neben der Software spielt die Qualität der Vorlage selbst eine große Rolle. Je kontrastreicher die Flächen und schärfer die Linien sind, um so geringer die Lesefehler. Deutliche Verbesserungen lassen sich oft durch Erstellen einer Zwischenkopie mit höherem Kontrast erreichen. Bei schlechten Vorlagen ist es nicht ausgeschlossen, dass ein Abschreiben der Texte finanziell und zeitlich günstiger ist als Einscannen und Nachbearbeitung. Klarheit schafft nur ein Test und Hochrechnen von Kosten und Zeit.

Neben den für Textübernahme notwendigen Regelungen, muss darauf geachtet werden, den Text bereits beim Scannen in einer Form abzulegen, wie es die spätere Anwendung (Programm oder Verwendungszweck) erfordert - vgl. Abschnitt 2.1.3 (Dateiformat).

1.3.3 Abschreiben

Vom Abschreiben wird auch in Zeiten der fortschreitenden Digitalisierung häufig Gebrauch gemacht, beispielsweise dann, wenn die Vorlage eine zu hohe Fehlerquote beim Einscannen ergibt oder wenn es sich um handschriftliche Vorlagen handelt. Ältere oder gar historische Zeugnisse bereiten schon einem geübten und kompetenten Leser Probleme, sind also für das Einscannen erst recht ungeeignet. Manchmal werden nur Bruchteile des Textkorpus gebraucht, so dass eine spätere Selektion des eingescannten Textes aufwendiger wäre als das Eintippen gezielt ausgewählter Textpassagen. Auch wenn die Vorlage eine Form aufweist, die eingescannt die Textteile falsch zuordnet, z.B. in komplexen Tabellen oder bei der Verwendung von Pfeilen oder anderen Sonderzeichen, die eine Texterkennung nicht verarbeitet, sind Scanner und Software überfordert und es muss abgeschrieben werden. Trotz CAPI (Computer Aided Personal Interview) und CATI (Computer Aided Telephone Interview) und sonstigen Computereinsatzes ist der Anteil der schriftlichen Interviews noch sehr hoch, besonders dort, wo einzelne Forscher ohne große technische Ausstattung Untersuchungen durchführen müssen oder bei postalischen Umfragen mit Selbstausfüllern, so dass auch da eine Verschriftung stattfinden muss.

Die Frage, ob professionelle und damit teure Schreibkräfte oder weniger geübte und billigere studentische Hilfskräfte eingesetzt werden sollen, ist nicht generell zu beantworten. Die i.d.R. geringere Schreibleistung von Studenten kann durch die geringere Bezahlung gegenüber

Fachkräften kompensiert werden. Die Fertigkeiten einer geübten Sekretärin können auch nur da voll zum Tragen kommen, wo längere zusammenhängende Texte zu erfassen sind, wirken sich aber kaum aus, wenn die Arbeitsvorbereitung groß ist, z.B. viel geblättert werden muss, und jeweils nur kurze Texte zu erfassen sind. Die sicherste Grundlage für eine Kalkulation von Zeit und Kosten besteht in einer Probe-Verschriftung. Die dabei gewonnenen Erfahrungswerte werden sich aufgrund des Übungseffektes eher verbessern.

Bei der Texterfassung schriftlicher Vorlagen ist der Regelungsbedarf und der Veränderungsgrad schon deutlich höher als bei Textübernahme und Einscannen: Anweisungen zur Rechtschreibung, Lesbarkeit u.a., wie in den Abschnitten 2.2 und 2.3 behandelt, müssen gegeben werden.

1.3.4 Transkribieren von Tonaufzeichnungen

Die Transkription, auch Umschrift genannt, "überträgt" Inhalte aus dem akustischen in das visuelle Medium. Es ist ein Übersetzungsprozess, der nie vollständig und befriedigend und immer nur in Hinsicht auf die Erfordernisse der Fragestellung und des Untersuchungsziels gelingen kann. Während die Verschriftung von Vorlagen noch mit wenigen Regeln auskommen kann, verlangt die Übertragung von gehörten Wörtern (und anderen akustischen Äußerungen) in Schrift mehr und exaktere Richtlinien für den Transkribenten, als bei anderen Erfassungsformen erforderlich sind, weil das Ausmaß der Veränderung wesentlich größer ist. Ohne Regelwerk ist für Außenstehende gar nicht und für die unmittelbar Beteiligten schon wenige Wochen später nicht mehr rekonstruierbar, wie in dieser oder jener Situation verfahren wurde, z.B. bei gleichzeitigem Sprechen, Wiederholungen oder unverständlichen Passagen. Eine Äquivalenz von Original und Text kann es nicht geben. Selbst eine "Partitur" der Bandaufnahme, wie sie von Ehlich (1992) entwickelt wurde, kann diese nicht ersetzen, sondern stellt immer eine Selektion der akustischen Situation dar, die ihrerseits schon ein eng begrenzter Teilbereich der tatsächlichen sozialen Gesamtsituation ist. Weil so viele Merkmale erfasst werden können, aber nicht alle gleichzeitig in die Transkription eingehen können oder von Bedeutung sind, kommt der Fragestellung und dem theoretischen Hintergrund eine um so größere Bedeutung zu. Er bestimmt, was geregelt werden muss und welcher Aufwand gerechtfertigt ist.

2 Regeln zur Texterfassung

Zu jedem der nachfolgend genannten Punkte könnte der einschränkende Hinweis gesetzt werden, dass es letztlich auf die Forschungsfrage ankommt, ob und in welchem Ausmaß eine Regelung vorgenommen werden muss. Aber selbst wenn ein Gesichtspunkt für das Forschungsziel keine Rolle spielt, sollte dies explizit gesagt werden, schon um eventuelle unnötige Arbeit zu ersparen. In jedem Fall müssen die Regeln schriftlich und vor Beginn der Arbeiten vorliegen. Wenn sich im Laufe der Texterfassung neue Fragen ergeben, können die Regeln ergänzt werden, es muss aber sichergestellt sein, dass der neu geregelte Sachverhalt noch nicht aufgetreten ist oder das entsprechende Material nachbearbeitet wird.

Die Anweisungen müssen möglichst konkret und einfach formuliert sein. Damit wird zum einen das eigene Vorgehen klar dokumentiert, zum anderen vermeidet es Missverständnisse zwischen Forscher und Mitarbeitern.

Es gibt Punkte, die in jedem Fall geregelt sein müssen, andere, die je nach Bedarf und Fragestellung mehr oder weniger von Bedeutung sind.

2.1 Was muss geregelt werden

Was im Folgenden aufgeführt ist, stellt das Minimum an Bearbeitung dar, das notwendig ist, um einen Text für weitere Zwecke verwenden zu können. Es gilt für alle Textformen und Verwendungszwecke. Ausreichend ist dieser Minimalaufwand in der Regel nur bei Texten, die bereits als Datei vorliegen.

2.1.1 Umfang und Auswahl

Zunächst wird gesagt, um welche Art von Texten es sich handelt, wo sie zu finden sind, welche Textteile für die eigene Untersuchung relevant ist und wie die Auswahl für die Verschriftung zu erfolgen hat. Soll die Grundgesamtheit der Texte systematisch reduziert, also eine Stichprobe gezogen werden, muss das Auswahlverfahren beschrieben werden. Nicht selbst produzierte Texte sind meist einem größeren Zusammenhang entnommen, der für die aktuelle Fragestellung eingegrenzt werden muss. Beispielsweise stellt sich für E-Mails die Frage, ob wirklich alle Informationen erhalten bleiben müssen, wie ausführliche Header-Daten, Anhänge, Bilder, Graphiken oder die Kopie des Schreibens, auf das Bezug genommen wurde.

Was soll verschriftet werden? Sind in Umfragen neben den Antworten der Befragten auch die vorgedruckten Frageformulierungen abzuschreiben? Handelt es sich z.B. um immer gleich formulierte Fragen, wird es ausreichen, den Text des Interviewers nur dann zu transkribieren, wenn er von der Vorlage abweicht. Was geschieht mit Randnotizen, Fußnoten? Schreiben oder weglassen?

Muss wortwörtlich abgeschrieben werden oder reichen die sinntragenden wichtigen Wörter? Um während der Texterfassung entscheiden zu können, was "wichtig" ist, wird theoretisches Hintergrundwissen vorausgesetzt, oder es muss eine ausführliche Einweisung stattfinden. Wenn eine wortgetreue Wiedergabe der Antworten wichtig ist, kann dies nicht deutlich genug formuliert sein. Nur so lässt sich ausschließen, dass "gutmeinende" VerschrifteterInnen den Forschern "helfen" wollen, die eine oder andere Aussage "klarer zu machen", und dabei unter Umständen genau die Informationen verfälschen, auf die es ankommt.

Im Zweifelsfall und im Sinne einer umfassenden Dokumentation sollte die Textvorlage komplett abgeschrieben oder jedes gehörte Wort transkribiert werden. Dann kann eine Auswahl auch später noch erfolgen, und die Vorgehensweise bleibt nachvollziehbar.

2.1.2 Einheiten und Identifikatoren

Mit Identifikatoren wird festgelegt, in welchen "Teilstücken" der Text später verarbeitet werden kann. Nur was definiert ist, lässt sich zählen und beschreiben. Die wenigsten Programme erlauben nach dem ersten Einlesen noch eine Umorganisation des Textes. Daher ist es sinnvoll, sich bereits in der Phase der Texterfassung vorzustellen, wie die Ergebnisse einmal präsentiert werden sollen, welche Informationen wie differenziert eine Tabelle enthalten soll. Die Texteinheiten, auf die darin Bezug genommen wird, über die eine Aussage gemacht werden soll, müssen eindeutig beschreibbar sein und für das Programm ansprechbar sein. Kleinere Einheiten lassen sich später wieder zusammenfassen, größere selten nach inhaltlichen Gesichtspunkten wieder aufteilen; deshalb ist im Zweifelsfall differenzierter zu gliedern. Ist z.B. das Werk eines Schriftstellers lediglich nach Titeln untergliedert, können nur über ein Buch als ganzes Aussagen gemacht werden, nicht aber über einzelne Kapitel, Anfang oder Ende innerhalb des Buches, beispielsweise über die Abfolge der angesprochenen Themen. Dazu müsste weiter untergliedert werden: Kapitel, Seite, Absatz, Satz, Zeile oder gar

Wort. Einige Textanalyseprogramme nehmen von sich aus eine Aufgliederung in Zeilen vor, andere, wie TEXTPACK (Mohler & Züll 2002), bieten die Möglichkeit, aufgrund von sogenannten Markern (z.B. Satzzeichen) Untergliederungen automatisch vorzunehmen.

Die Form, wie die Untergliederung vorgenommen wird, ist abhängig vom verwendeten Programm. Eine vielgenutzte Möglichkeit ist die Kennzeichnung jeder Zeile durch vorangestellte Identifikatoren, was einer fixen Datenstruktur (fixed format) bzw. einer tabellarischen Darstellung entspricht (vgl. Tabelle 1).

Tabelle 1: Identifikatoren bei fixer Datenstruktur

01	03	001	Die Wintersonne stand nur als armer Schein, milchig und matt hinter Wolkenschichten über der engen Stadt. Naß und zugig war's in den giebeligen Gassen, und manchmal fiel eine Art von weichem Hagel, nicht Eis, nicht Schnee.
		002	Die Schule war aus. Über den gepflasterten Hof und heraus aus der Gatterpforte strömten die Scharen der Befreiten, teilten sich und enteilten nach rechts und links ...
...			
	276	...	Blonden und Blauäugigen, den hellen Lebendigen, den Glücklichen, Liebenswürdigen und Gewöhnlichen.
	277		Schelten Sie diese Liebe nicht, Lisaweta; sie ist gut und fruchtbar. Sehnsucht ist darin und schwermütiger Neid und ein klein wenig Verachtung und eine ganze keusche Seligkeit.
04	001	...	
...			
02	01	001	...

Beispiel aus "Tonio Kröger" von Thomas Mann

In diesem Beispiel bestehen die Gliederungselemente aus Autor, Titel und Absatz. Das jeweilige Textelement befindet sich immer auf ein- und derselben (fixen) Position einer Zeile; in dem Beispiel ist die Autorenkennung immer auf der Spalte 1 bis 2 ("01" für Thomas Mann, "02" für einen weiteren Schriftsteller), Spalte 3 bleibt zur besseren Lesbarkeit leer. Der Code für den Titel steht auf der Spalte 4 bis 5 ("03" für Tonio Kröger, anderer Ziffern für andere Werke) und die laufende Nummer für die Absätze immer auf den Spalten 7 bis 9), während der Text konstant in den Spalten 11 bis 75 zu finden ist - 6 und 10 bleiben wieder leer.

Dies ist auch das Format, wie es die sogenannten portablen Dateien meist haben, die benutzt werden, um Daten aus einem Programm in ein anderes zu überführen. Anstelle der Leerzeichen, die hier für die Spaltenstruktur sorgen, werden auch andere Trennzeichen wie Tabulatorzeichen oder Semikolon verwendet.

Eine andere Form, Texteinheiten zu beschreiben, ist die mit Präfixen, die den Kennziffern vorangestellt werden; TEXTPACK sieht diese Möglichkeit vor.

Tabelle 2: Identifikatoren mit Präfix und Fließtext

\$1
\$\$\$3
\$\$\$1
Die Wintersonne stand nur als armer Schein, milchig und matt hinter Wolkenschichten über der engen Stadt. Naß und zugig war's in den giebeligen Gassen, und manchmal fiel eine Art von weichem Hagel, nicht Eis, nicht Schnee.
\$\$\$2
Die Schule war aus. Über den gepflasterten Hof und heraus aus der Gatterpforte strömten die Scharen der Befreiten, teilten sich und enteilten nach rechts und links ...
...
\$\$\$276
... Blonden und Blauäugigen, den hellen Lebendigen, den Glücklichen, Liebenswürdigen und Gewöhnlichen.
\$\$\$277 Schelten Sie diese Liebe nicht, Lisaweta; sie ist gut und fruchtbar. Sehnsucht ist darin und schwermütiger Neid und ein klein wenig Verachtung und eine ganze keusche Seligkeit.
\$2
\$\$\$1
\$\$\$1
...

Erläuterung zum Beispiel: Das Präfix „,\$“ besagt, dass die folgende Ziffer „1“ die Kennzeichnung eines bestimmten Autors (Thomas Mann) ist, das Präfix „,\$\$“ mit "3" bezeichnet den Titel (Tonio Kröger), und „\$\$\$“ steht vor der laufenden Nummerierung der Absätze. Ab der darauffolgenden Zeile steht fortlaufend der Text, bis der Text eines neuen Absatzes, eines neuen Titels oder eines neuen Autors beginnt. Dann muss eine neue Identifikatorenzeile für Absatz, Titel oder Autor eingefügt werden.

Dieses TEXTPACK-Format erweist sich als praktisch, wenn Texte per Hand eingegeben werden, weil es ein Schreiben ohne Einrückung oder sonstige Formatierung ermöglicht, während sich das fixe Format meist dann anbietet oder ergibt, wenn Texte aus Tabellen oder anderen Programmen übernommen werden. Im übrigen ist auf die Vorgaben der verwendeten Software zu achten.

2.1.3 Dateiformat

Es reicht nicht zu wissen, dass ein Text verfügbar ist, er muss auch lesbar sein, und zwar nicht nur am Bildschirm, sondern auch vom Programm, mit dem die Auswertung vorgenommen werden soll. Dass z.B. die Internet-Texte nur eine Bildschirmdarstellung sind und als Daten ganz anders aussehen, lässt sich leicht feststellen, wenn die Seite als sogenannte Textdatei abgespeichert wird. Hinter dem im Netz verwendeten Format (HTML) verbirgt sich ein System von Schlüsselbegriffen, Spitzklammern und sonstigen Sonderzeichen. Sie beschreiben die Darstellung des Textes auf dem Bildschirm. Die eigentliche Textinformation macht meist nur einen Bruchteil des Dateiumfangaus. Ähnliche Erfahrungen können mit allen programmspezifischen Dateien gemacht werden (Word, SPSS, PDF). Noch immer gibt es kaum Textanalyseprogramme, die diese Formate verarbeiten. Es ist vorab zu klären, welches Datenformat das Auswertungsprogramm erfordert. Was alle Programme lesen und verarbeiten können, ist das ASCII- bzw. ANSI-Format. Es speichert keinerlei Formatierungen, sondern nur den reinen Text; deshalb wird dieses Format in MS-Word "Nur Text" genannt und hat die Dateierweiterung ".txt".

Weitere Restriktionen für die Verwendbarkeit der Texte in Programmen können sich aus der maximal zulässigen Zeilenlänge, dem zulässigen Umfang der Texteinheiten oder der Behandlung von Sonderzeichen oder Umlauten ergeben. Bei Scannererfassung, Verschriftung und Transkription lassen sich diese Bedingungen schon bei der Texterfassung

berücksichtigen, bei der Übernahme von Texten muss ggf. eine Umformatierung vorgenommen werden.

2.1.4 Kontrolle und Nachbearbeitung

Der Forscher selbst muss sich vergewissern, ob die von ihm aufgestellten Regeln auch richtig verstanden und umgesetzt werden, indem er selbst an den Erfassungsarbeiten teilnimmt und/oder stichprobenweise überprüft. Darüber hinausgehende, regelrechte Reliabilitätstests (mehreren Personen verschriften ein und dieselbe Textstichprobe und die Textdateien werden miteinander verglichen) werden in der Praxis selten durchgeführt, zumal es für die Messung noch kein standardisiertes Verfahren gibt. Fehler passieren besonders leicht bei Transkriptionen von Tonaufnahmen, selbst bei großer Sorgfalt und Konzentration. Eine einfache und effiziente Kontrolle besteht darin, dass eine unabhängige Person das Band abhört und mit dem Text vergleicht. Gemessen an der Transkriptionszeit macht dies vielleicht nur ein Zehntel aus, vermeidet aber grobe Fehler. Über alle Texte (übernommen, eingescannt, abgeschrieben und transkribiert) sollte eine Rechtschreibprüfung laufen (natürlich nicht, wenn Rechtschreibung von Probanden, Interviewern oder Transkribenten Gegenstand der Forschungsfrage ist). Damit wird die korrekte Bearbeitung von übernommenen oder eingescannten Texten überprüft, ebenso wie die manuelle Erfassung, bei der Tipp- und Rechtschreibfehler nie auszuschließen sind.

2.2 Was soll geregelt werden

2.2.1 Nicht lesbar/unverständlich

Was ist zu tun, wenn der Text nicht lesbar ist, fehlt oder eine Angabe explizit verweigert wird? – Sonderzeichen oder unverwechselbare Schlüsselwörter/Codes (z.B. „xyz“) für ein nicht lesbares Wort oder andere Unklarheiten erleichtern die Nachbearbeitung oder Interpretation des Textes.

Was geschieht mit unverständlichen Redebeiträgen? Wie spezifisch werden nicht verständliche Teile gekennzeichnet? – Die Stelle könnte nur markiert werden (mit Code/Sonderzeichen) oder es könnten auch Gründe und Umfang der Unverständlichkeit festgehalten werden. Eine Regelung könnte folgendermaßen aussehen: *Wird ein Text auch nach mehrmaligem Hören nicht verstanden, dann soll eine zweite oder dritte Person hinzugezogen werden. Kann der Text auch von drei Personen nicht identifiziert werden, wird die Stelle mit zwei Fragezeichen („??“) erfasst, wenn es sich um einzelne Wörter oder Satzteile handelt; sind dagegen größere Textpassagen betroffen, wird diese Stelle mit drei Fragezeichen („???“) markiert. Sind keine geeigneten Personen zum Kontrollhören vorhanden, ist diese Stelle möglichst genau zu notieren (Texteinheit, Band-Zählerstand u.ä.), um später diesen Abschnitt überprüfen zu können. Die Zahl der Abhörversuche soll (aus arbeitsökonomischen Gründen) auf drei beschränkt bleiben.*

Zwei oder drei Fragezeichen direkt hintereinander kommen in einem normalen Text nicht vor. Somit können die unklaren Stellen über den Suchbefehl leicht angesteuert werden.

2.2.2 Fehlender Wert/Redeunterbrechung

Wenn eine Redeunterbrechung für die Fragestellung nicht von Bedeutung ist (weil es z.B. nur auf die Sammlung von Gesichtspunkten ankommt) und die Abfolge der Redebeiträge nicht interessiert, dann werden die Aussagen einfach hintereinander gesetzt. Wenn ein Sprecher dem anderen ins Wort fällt, gibt es die Möglichkeit, dass der Unterbrochene schweigt oder

weiterredet. In dem Transkript kann dies dadurch festgehalten werden, dass im ersten Fall (Schweigen des Erstredners) die Rede mit einem zu vereinbarenden Zeichen (z.B. drei Punkte) abgeschlossen und so verfahren wird wie bei einem Rednerwechsel. Reden dagegen mehrere Sprecher gleichzeitig, kann dies durch ein besonderes Zeichen (in Tabelle 3 das "+") zu Beginn und Ende der Gleichzeitigkeit angedeutet werden.

Tabelle 3: Beispiel für Markierung von Redeunterbrechungen/Gleichzeitigkeit

<p>Die Markierung der Gleichzeitigkeit erfolgt immer auf Wortgrenze, auch wenn die Gleichzeitigkeit nur auf Wortteile beschränkt ist. In den folgenden Beispielen wird „T/“ und „P/“ als Sprecherkennung verwendet. Zunächst ein Fall, in dem der Unterbrecher das Wort behält (oder an sich reit):</p> <p><i>P/ ich habe da eine + Venenentzndung</i> <i>T/ja ja. + das sagten Sie schon mal</i></p> <p>Im nchsten Beispiel bleibt der erste Sprecher am Wort:</p> <p><i>P/ ich habe da eine + Venenentzndung</i> <i>T/ ja ja.+</i> <i>P/ die ich einfach nicht loswerde</i></p> <p>zitiert aus: Mergenthaler, E.,: Die Transkription von Gesprchen, Ulm 1992, S. 6. Hier findet sich eine sehr umfangreiche und praktisch erprobte Sammlung von Transkriptionsregeln.</p>
--

Erluterung zum Beispiel: "P" bedeutet Patient und "T" Therapeut, "+" kennzeichnet Beginn und Ende der gleichzeitigen Rede.

Andere Mglichkeiten sind Anmerkungen in Kommentarzeilen, die Behandlung wie nacheinander gesprochene Beitrge oder schlichtes Weglassen der Zwischenreden.

2.2.3 Rechtschreibung/Sprache

Bezglich der Rechtschreibung muss klar geregelt sein, was verndert werden darf. Soweit nicht die Orthographie, der Dialekt oder die Sprachfrbung des Befragten oder des Interviewers Forschungsgegenstand sind, sollten die korrekte Schreibweise und die Hochsprache verwendet werden. Eine Standardsprache erleichtert jede Form von Recherche, Textidentifizierung und Zhlung, weil der Suchende mit einer klar berschaubaren Variationsbreite der Ausdrucksformen rechnen kann. Andererseits gilt fr die computeruntersttzte Inhaltsanalyse mit Wrterbchern, dass eine korrekte Schreibweise nicht immer notwendig ist. Die Buchstabenfolge der zu analysierenden Wrter muss sich im Wrterbuch wiederfinden, gleichgltig ob richtig und/oder falsch geschrieben. Ein hufig falsch geschriebenes Wort wrde man in der korrekten wie in der falschen Schreibweise in das Wrterbuch aufnehmen.¹

hnlich wie beim Abschreiben festgelegt wird, was am Text verndert werden darf, muss fr die gesprochene Sprache klar sein, ob der phonetische oder der semantische Text zu transkribieren ist. Im letzten Fall bedeutet dies, dass ausschlielich in Hochsprache zu

¹ In der automatischen Berufsvercodung bei ZUMA hat ein Test gezeigt, da eine Rechtschreibprfung zwar den Anteil der automatisch codierten Flle erhht (um 5%), die manuelle Nachbearbeitung (Codierung) der Angaben aber nur ein Fnftel der Zeit kostet wie die der Rechtschreibprfung (5 Stunden). Aufwand und Erfolg mssen in einem angemessenen Verhltnis stehen.

übertragen ist. Aber was soll mit nicht "übersetzbaren", regionaltypischen Wörtern und Ausdrucksweisen geschehen? Eine phonetische Transkription würde die computergestützte, automatische Auswertung erheblich erschweren, kann aber für manche Fragestellungen (z.B. bzgl. Dialektgrenzen und -entwicklungen) unumgänglich sein. Meist reicht es, lediglich eine Übertragung von Umgangssprache/Dialekt vorzunehmen, d.h. die Wörter werden nicht ersetzt, sondern nur der lexikalischen Schreibweise angeglichen. "Woast" wird z.B. zu "weiß", "kriegste" wird zu "kriegst du". Wörter, Wortstellungen, Satzkonstruktionen sollen im Sinne einer unverfälschten Dokumentation unverändert bleiben. Für Zweifelsfälle bietet sich die Regel an: "Jedes Wort, das im Duden zu finden ist, darf geschrieben werden". Dies gilt besonders für Ausdrücke der Umgangssprache.

2.2.4 Arbeitsorganisation

Umfangreiche Erfassungen werden nicht von Einzelpersonen vorgenommen, sondern erfolgen arbeitsteilig. Neben der Tatsache, dass gerade dann die Regeln sehr ausführlich und unmissverständlich formuliert sein müssen, ist auch ein Mindestmaß an Organisation erforderlich, um Arbeitsvorbereitung, -teilung und -kontrolle zu gewährleisten. Dazu können eine einfache Arbeitsliste und das Abzeichnen erfasster Vorlagen dienen oder auch zusätzlich die Vergabe von Personenkennziffern bei den Identifikatoren, die den Erfasser jedes Textes ausweist. U.a. wird damit vermieden, dass ein Text mehrfach von verschiedenen Personen eingegeben wird.

2.3 Was kann geregelt werden

2.3.1 Textgestaltungsmerkmale

Die Übernahme von maschinenlesbaren Texten bedeutet zwar, dass alle Wörter übernommen werden, weitergehende Merkmale des Textes gehen aber in der Regel bei der Verwendung von Textanalyseprogrammen verloren. Es bleibt lediglich der einfache Text. Das bedeutet, dass z.B. nur noch bekannt ist, dass es sich um den Buchstaben "F" in einem Wort handelt, nicht aber, wie das "F" im Originaltext ausgesehen hat. Überschriften und Absatzstruktur können durch entsprechende Vergabe von Identifikatoren erhalten bleiben, Schrifttypen, Schriftgröße, Farbe, Unterstreichungen oder andere Hervorhebungen sind nicht mehr erkennbar, soweit keine Vorkehrungen getroffen werden. Sind diese Merkmale für die Forschungsfrage relevant, können sie durch Kennzeichnung mit Sonderzeichen oder Buchstabenkombinationen markiert werden. "#alles" könnte dann beispielsweise eine Hervorhebung von "alles" bedeuten. Formatierung, Layout und sonstige Gestaltungsmerkmale bleiben nur dann erhalten, wenn sie durch ASCII-Zeichen dargestellt werden, also bloße Zeichen ohne weitere Merkmale wie Schrifttyp, Farbe, Unterstreichung o.ä. In Umfragen werden bei Aufzählungen z.B. gerne Spiegelstriche ("-") verwendet oder neue Zeilen. Um falsche Zusammenhänge zu vermeiden und diese Darstellung zu erhalten, können neue Aufzählungen/Zeilen z.B. mit einem vorangesetzten "<" markiert werden. Ähnlich kann man mit Unterstreichung, Fettdruck, Farbe, Kursivschrift, Schriftgröße usw. verfahren. Es muss nur sichergestellt sein, dass das verwendete Zeichen sonst nicht als Text vorkommt. Diese Darstellungsweise ist vergleichbar mit den Formatierbefehlen in HTML-Texten, wo der Umlaut "ö" z.B. mit den ASCII-Zeichen "ö" dargestellt wird.

2.3.2 Groß-/Kleinschreibung

Die bei ZUMA seit Jahren durchgeführte automatische Berufsvercodung greift auf ein Wörterbuch zurück, in dem nur Großbuchstaben verwendet werden. Eine Unterscheidung

groß/klein ist bei der Verschriftung der Berufsangaben also nicht notwendig, weil die Information groß/klein nicht gebraucht wird. Gleichzeitig erspart man sich, in dem Wörterbuch alle möglichen Groß- und Kleinschreibungen aufzuführen. Grundsätzlich in Großbuchstaben zu verschriften ist eher mühsam, vor allem aber erschwert es das Lesen des Textes. Es kann so geschrieben werden, wie es den geringsten Aufwand macht. Mit einem einzigen Klick (in Word z.B. über die Befehle Format/Zeichen/Großbuchstaben) ist der gesamte Text vor der Analyse in Großbuchstaben umgesetzt.

Kommt es bei der Analyse der Texte aber darauf an, Substantive identifizieren zu können und auch von Nicht-Substantiven am Satzanfang unterscheiden zu können, so kann die Groß-/Kleinschreibung so geregelt werden, dass alle Wörter so zu schreiben sind, als stünden sie mitten im Satz, es beginnt dann also nicht jeder Satz automatisch mit Großschreibung.

2.3.3 Sprechpausen

Für die Transkription von Interviews, besonders im Bereich der Psychologie, wo auch nichtverbale Informationen wichtig sind, kann es eine Rolle spielen, wie viel Zeit sich der Proband für die Antwort lässt (spontan, zögerlich, nach langer Überlegung). Dann muss dafür eine Notationsform festgelegt werden. Beispielsweise könnten je 10 Sekunden Sprechpause durch einen Bindestrich dargestellt werden. In diesem Fall sollte sichergestellt sein, dass der Bindestrich sonst nicht verwendet wird.

2.3.4 Nonverbale Äußerungen

Dazu gehören u.a. Räuspern, Husten, Stottern, Lachen, Kindergeschrei, Hintergrundgeräusche u.a.m. Je mehr der Text selbst nicht ausschließlicher Untersuchungsgegenstand ist, sondern eher als Indikator für nichtverbalisierbare Eigenschaften gilt, um so mehr müssen neben dem Text auch andere Merkmale festgehalten werden. Dies trifft vor allem für psychotherapeutische Gespräche oder für Test- und Experimentalsituationen zu.

2.3.5 Lautstärke

Für die Untersuchung gruppenspezifischer Prozesse ist nicht nur die Aussage selbst von Bedeutung, sondern u.a. auch die Art und Weise, wie der eigene Standpunkt vorgetragen wird. Mit Buchstaben-Zeichen-Zahlenkombinationen - z.B. "<l+>" für lauter und "<l->" für leiser als gewöhnlich - zu Beginn der Rede oder der Veränderungen lassen sich auch diese Merkmale in dem Transkript der Gruppendiskussion dokumentieren.

2.3.6 Zitate

Werden Reden oder Ansprachen transkribiert, sollte Vorsorge getroffen werden, dass zitierte Beiträge als solche formal abgegrenzt werden und nicht nur durch die Nennung des Autorennamens im Text davor oder dahinter identifizierbar werden. Besonders bei computerunterstützter Auswertung, wo lediglich gezählt wird, wer sagt was wie oft, können sich verfälschte Häufigkeiten ergeben. Der sicherste Weg, falsche Zuschreibungen zu vermeiden, ist die Vergabe einer eigenen "Sprecher"-Identifikation für Zitate.

2.3.7 Syntaktische Informationen

Gewisse syntaktische Informationen werden in jeder Transkription - auch durch den Interviewer beim Interview - ohne weiteres festgelegt, nämlich die Setzung von Punkt und Komma; dabei dürfte der Sprecher kaum die Satzzeichen gesprochen haben, sie werden also durch die verschriftende Person nach deren Gefühl vergeben. Solange nur Themen oder Aussagen betrachtet werden, spielt dies keine Rolle. Rücken aber Redestil, Grammatik, Satz-

oder Sprachstrukturen in den Mittelpunkt der Betrachtung, werden genaue Regeln für die Zeichensetzung erforderlich. Die Veränderung der ursprünglich satzzeichenlosen Rede muss zumindest immer in der gleichen Weise erfolgen, um willkürliche oder zufällige Abweichungen zu vermeiden.

2.3.8 Anmerkungen, Kommentare

Während der Verschriftung können Fragen, Unsicherheiten und Probleme auftreten, die nicht in den Regeln aufgeführt sind, oder es gibt wichtige Hinweise für das Projekt, z.B. deutliche Hinweise auf Interviewfälschungen (in Form von sich wiederholende Allgemeinangaben bzw. Verweigerungsbegründungen) oder offensichtlich falsch verstandene Fragen. In Transkriptionen sind Hinweise auf eine besonders schlechte Aufnahmequalität hilfreich, ebenso die Information, dass Umgebungslärm zu hören ist, der die Konzentrationsfähigkeit des Probanden möglicherweise beeinträchtigt. Mit einer eigenen Kennung (Identifikator) für Anmerkungen, die wie Text einer eigenen Frage behandelt werden, lässt sich die Information zielgenau festhalten.

Natürlich erhöht sich mit immer komplexeren Regeln der Verschriftungsaufwand, und mit der Elaboriertheit eines Regelwerkes erhöht sich auch die Anfälligkeit bezüglich Genauigkeit und Reliabilität der Umsetzung. Deshalb, die Liste der möglichen Regeln abschließend, der Hinweis: So viel festlegen wie nötig, so wenig reglementieren wie möglich - immer in Hinsicht auf die Forschungsfrage.

3 Beispiele für Regeln und Vorgehensweisen

In Folgendem werden drei Beispiele für Regelwerke vorgestellt: für die Textübernahme, für das Abschreiben und für die Transkription. Ein Beispiel für das Einscannen erübrigt sich: Der Text selbst muss hier genauso bearbeitet werden wie bei der Übernahme. Lediglich beim Speichern des eingescannten Textes ist darauf zu achten, dass ein Format gewählt wird, das von dem gewählten Textanalyseprogramm erkannt oder leicht entsprechend angepasst werden kann.

3.1 Textübernahme

In dem Beispiel handelt es sich um das Kategorienschemas zur ISCO-88 (International Standard Classification of Occupations). Die Texte der Dokumentation sollten daraufhin untersucht werden, in wieweit Verben zur Definition der Berufskategorien benutzt werden. Diese Fragestellung spielt insofern eine Rolle als die Texteinheiten den Kategorien/Codes (Ziffern im Text) genau zuzuordnen sein mussten.

Tabelle 4: Vorgegebener Text im Original

MAJOR GROUP 1: LEGISLATORS, SENIOR OFFICIALS AND MANAGERS

Legislators, senior officials and managers determine, formulate, direct ... other workers may be included.

- 12 Corporate managers (This group is intended ... requiring a total of three or more managers.)
- 13 General managers (This group is intended ... classified according to their specific tasks.)

Note

In some case where specific professional, technical or operational skills and knowledge ... major group.

SUB-MAJOR GROUP 11 LEGISLATORS AND SENIOR OFFICIALS

Legislators and senior officials determine, formulate or advise on and direct government policies, make, ratify, amend ...

Tasks performed by workers in this sub-major group usually include: determining and formulating, or advising on and ...

Occupations in this sub-major group are classified into the following minor groups:

- 111 Legislators
- 112 Senior government officials
- 113 Traditional chiefs and heads of villages
- 114 Senior officials of special-interest organisations

Note

In some cases where specific professional, technical or operational skills and knowledge may be required of workers at ... is to allocate research and development funds from a government budget, on the basis of medical knowledge, then the job belongs in this sub-major group.

MINOR GROUP 111 LEGISLATORS

Legislators determine, formulate, and direct policies ... Tasks performed usually include: presiding over or participating in the proceedings of legislative assemblies and administrative councils of national, state, regional or local ...

Occupations in this minor group are classified into the following unit group:

1110 LEGISLATORS

Legislators determine, formulate, and direct policies of national, state, regional or local governments, and make, ratify, amend or repeal laws, public rules and regulations.

- Tasks include -
- (a) presiding over or participating in the proceedings ...
 - (b) determining, formulating, and directing policies ...

- ... (h) supervising other workers.

MINOR GROUP 112 SENIOR GOVERNMENT OFFICIALS

Senior government officials advise governments on ... of other workers may be included.

Occupations in this minor group are classified into the following unit group:

- 1120 Senior government officials

Note

Senior government officials who manage ...

1141 SENIOR OFFICIALS OF POLITICAL-PARTY ORGANISATIONS

Senior officials of ... their behalf.

- Tasks include -
- (a) determining and formulating the ...

- ... (g) supervising other workers.

Zur Gliederung des Textes für eine computerunterstützte Inhaltsanalyse werden folgende Regeln aufgestellt:

1. Die Textstruktur bleibt erhalten. Zeilenwechsel, Absätze, Einrückungen, Aufzählungszeichen, Leerzeichen usw. werden nicht verändert.
2. Als Kennung der Texteinheit dient der ISCO-Code, ggf. mit Nullen auf vier Stellen erweitert, mit dem Präfix "\$" (ohne Leerzeichen dazwischen).
3. Die Kennung wird vor dem zugehörigen Text auf einer eigenen Zeile eingefügt.
4. Es kann vorkommen, dass sich ein Kommentar auf eine allgemeine (zusammenfassende) Kategorie bezieht, die an anderer Stelle aufgeführt ist. Dann ist vor diesen Kommentar der Code mit Präfix (ID-Kennungen) zu setzen, auf den sich der Kommentar bezieht.
5. Reine Titel oder Berufsbenennungen (ohne weiteren Text oder Erläuterungen), die an anderer Stelle noch einmal aufgeführt und dort erläutert werden, sind zu streichen.
6. Wiederholungen, die direkt hintereinander stehen, löschen.

Demnach verändert sich der Originaltext folgendermaßen:

Tabelle 5: Vorgegebener Text, der verändert wurde

(Der veränderte Text ist fett wiedergegeben.)

Text mit Veränderungen	Kommentar
<p>\$1000 MAJOR GROUP 1: LEGISLATORS, SENIOR OFFICIALS AND MANAGERS</p> <p>Legislators, senior officials and managers determine, formulate, direct ... other workers may be included.</p> <p>\$1200 12 Corporate managers (This group is intended ... requiring a total of three or more managers.)</p> <p>\$1300 13 General managers (This group is intended ... classified according to their specific tasks.)</p> <p>\$1000 Note In some case where specific professional, technical or operational skills and knowledge ... major group.</p> <p>\$1100 SUB-MAJOR GROUP 11 LEGISLATORS AND SENIOR OFFICIALS</p> <p>Legislators and senior officials determine, formulate or advise on and direct government policies, make, ratify, amend ...</p> <p>Tasks performed by workers in this sub-major group usually include: determining and formulating, or advising on and ...</p> <p>Occupations in this sub-major group are classified into the following minor groups:</p> <p>111 Legislators 112 Senior government officials 113 Traditional chiefs and heads of villages 114 Senior officials of special-interest organisations</p> <p>Note In some cases where specific professional, technical or</p>	<p>Auf einer eigenen Zeile (Regel 3) wird vor den Code 1 ein "\$" als Kennung für die Einheit gesetzt. Der Code wird um drei Nullen ergänzt (Regel 2).</p> <p>Diese Kategorien werden zwar unten ausführlich erläutert, aber auch hier finden sich Definitionen - deshalb kommt Regel 5 nicht in Betracht.</p> <p>Was unter "Note" aufgeführt ist, bezieht sich auf die allgemeinere Hauptgruppe 1 - deshalb wird, abweichend von der Reihenfolge, "\$1000" eingefügt.</p> <p>Auf einer eigenen Zeile (Regel 3) wird vor den Code 11 ein "\$" als Kennung für die Einheit gesetzt. Der Code wird um zwei Nullen ergänzt (Regel 2).</p> <p>Es ist nur die Übersicht der Kategorien, die unten erläutert werden - deshalb gestrichen (Regel 5).</p>

Text mit Veränderungen	Kommentar
<p>operational skills and knowledge may be required of workers at ...</p> <p>... is to allocate research and development funds from a government budget, on the basis of medical knowledge, then the job belongs in this sub-major group.</p> <p>§1110 MINOR GROUP 111 LEGISLATORS</p> <p>Legislators determine, formulate, and direct policies ... Tasks performed usually include: presiding over or participating in the proceedings of legislative assemblies and administrative councils of national, state, regional or local ...</p> <p>Occupations in this minor group are classified into the following unit group:</p> <p>1110 LEGISLATORS Legislators determine, formulate, and direct policies of national, state, regional or local governments, and make, ratify, amend or repeal laws, public rules and regulations. Tasks include - (a) presiding over or participating in the proceedings ... (b) determining, formulating, and directing policies (h) supervising other workers. ...</p> <p>§1120 MINOR GROUP 112 SENIOR GOVERNMENT OFFICIALS</p> <p>Senior government officials advise governments on ... of other workers may be included. Occupations in this minor group are classified into the following unit group: 1120 Senior government officials</p> <p>Note Senior government officials who manage</p> <p>§1141 1141 SENIOR OFFICIALS OF POLITICAL-PARTY ORGANISATIONS Senior officials of ... their behalf. Tasks include - (a) determining and formulating the ... (g) supervising other workers.</p>	<p>Auf einer eigenen Zeile (Regel 3) wird vor den Code 111 ein "\$" als Kennung für die Einheit gesetzt. Der Code wird um eine Null ergänzt (Regel 2).</p> <p>Auf einer eigenen Zeile (Regel 3) wird vor den Code 112 ein "\$" als Kennung für die Einheit gesetzt. Der Code wird um eine Null ergänzt (Regel 2).</p> <p>Auf einer eigenen Zeile (Regel 3) wird vor den Code 1141 ein "\$" als Kennung für die Einheit gesetzt (Regel 2).</p>

Durch das Einfügen der (fett wiedergegebenen) Identifikatoren wird festgelegt, welche Textteile jeweils zusammengehören: von Dollarzeichen bis Dollarzeichen. Das konnte deshalb nicht automatisch erfolgen (nach der Regel: "Immer wenn sich in den ersten eine Ziffer befindet, nimm diese als Identifikationskennziffer."), weil auch inhaltliche Gesichtspunkte zu berücksichtigen waren - wenn Kommentare nicht bei den Texten standen, auf die sie sich bezogen.

3.2 Abschreiben

Das Projekt, aus dem das Beispiel stammt, untersuchte, mit welchen Stellen und Institutionen die Befragten Kontakt hatten. Dabei war eine Grobgliederung, wie zum Beispiel "Presse", "Radio, Fernsehen" usw. vorgegeben. Die Befragten sollten ihre Kontakte jeweils im einzelnen dem Interviewer nennen. Die vom Interviewer im Fragebogen notierten Kontaktstellen wurden zum Zweck einer computerunterstützten Inhaltsanalyse verschriftet.

Tabelle 6: Beispiel für Verschriftungsregeln (TEXTPACK-Konventionen)

Verschriftungsregeln zum Projekt „Kontakte“	Erläuterungen
1. Zu verschriften sind die vom Interviewer notierten Antworten zur Frage 74, die nach Kennziffern unterteilt ist. Der Antworttext wird fortlaufend mit automatischem Zeilenumbruch geschrieben.	<i>Umfang, Teil des zu verschriftenden Fragebogens. Innerhalb der Texte keine CRs (Absatzmarken) notwendig.</i>
2. Von jedem Fragebogen muss zumindest die Kennziffer 01 verschriftet werden. Ist überhaupt keine Angabe gemacht worden, so wird hinter der Kennziffer 01 als Text "9" verschriftet, der Code für fehlende Werte.	<i>Damit wird gewährleistet, dass kein Fall vergessen wird, bzw. eine Endkontrolle anhand der ID-Liste möglich ist.</i>
3. Es wird nur verschriftet, wenn eine Antwort gegeben wurde. Grundsätzliche Anmerkungen zur Frage oder zum Fragebogen bleiben unberücksichtigt.	<i>Beschränkung auf die reinen Antworten. Unnötige Schreibearbeit wird vermieden.</i>
4. Es werden bis zu drei Identifikatoren vergeben: 1. Befragten-Nr./Fragebogen-Nr.: die Paginierstempel-Ziffern auf dem "Zusatzblatt zu Frage 74". Unmittelbar vor die Fragebogen-Nr. wird immer "+" gesetzt; jeder Identifikator muss immer auf einer eigenen Zeile stehen. 2. Kennziffer: sie steht in den Kästchen am linken Rand. Unmittelbar vor die Kennziffer-Nr. wird immer "#" gesetzt; jeder Identifikator muss immer auf einer eigenen Zeile stehen. Angaben ohne Kennziffer erhalten die Kennziffer 80. 3. Nennung: Für jede Angabe innerhalb der Kennziffer wird eine laufende Nummer vergeben, ohne dass der Text dadurch verändert werden darf. Eine neue Angabe ist meist durch eine Komma o.ä. abgetrennt; jeder Kontakt (= genannte Institution) stellt eine eigene Angabe/Nennung dar. Unmittelbar vor die lfd. Nr. der Angabe wird immer "<" gesetzt.	<i>Auf einem FRAGEBOGEN befinden sich meist viele Ziffern, deshalb die Beschreibungen. Das „+“ ist die Kennung für den Identifikator der 1. Ebene, das „#“ die Kennung für die 2. Ebene. Die >80< regelt das Fehlen von Identifikatoren, die handschriftlich eingetragen wurden Die Nennungen innerhalb der Kennziffer werden beim Verschriften mit dem Präfix „<“ für die 3. Identifikatorebene versehen (die Durchnummerierung kann in TEXTPACK automatisch erfolgen).</i>
5. Die Identifikatoren müssen nur geschrieben zu werden, wenn sie sich ändern. Beispiel: jede Fragebogen-Nr. braucht nur einmal geschrieben zu werden.	<i>Solange die übergeordnete Ebene gleich bleibt, muss sie nicht wiederholt werden.</i>
6. Die Texte sind nach ID sortiert einzugeben. Ein nicht in der Sortierfolge liegender Fragebogen sollte an der entsprechenden Stelle eingefügt werden. Ähnlich ist mit den Kennziffern zu verfahren: häufiger wird ein und dieselbe Kennziffer ein zweites Mal genannt; dann sind die dazugehörigen Angaben zu der bereits vorhandenen Kennziffer zu schreiben.	<i>Bei einer Präfix-Notation kann TEXTPACK nicht sortieren. Zur Kontrolle, bes. bei Fehlern im Identifikationsfeld, muss eine feste und übereinstimmende Reihenfolge von Fragebögen und Texten vorliegen.</i>
7. Der Text ist so abzuschreiben, wie er zu den einzelnen Fragen vorgefunden wird, also z.B. auch "wie gesagt", "weiß ich nicht", "sage ich nicht", "kann ich nichts sagen", "kenne mich nicht so gut aus", "keine Meinung", "fällt mir nichts dazu ein", "keine", "ebenfalls", "?", "kein Urteil", "keine Nennung", "" (Gänsefüßchen).	<i>Hier muss der Forscher später selbst entscheiden, wie die Verweise zu werten sind, und der Verschrifterin bleiben zeitraubende Entscheidungszwänge erspart.</i>
8. Offensichtliche Fehler (Rechtschreibung, Zeichensetzung) sind zu korrigieren, nicht aber Satzbau, Wortstellung oder Wörter selbst.	<i>Die Grenzen der Veränderbarkeit werden beschrieben.</i>
9. Soweit verwendete Abkürzungen eindeutig und bekannt sind, sollten sie ausgeschrieben werden.	<i>Nur bekannte - vermeidet Fehldeutungen.</i>

Verschriftungsregeln zum Projekt „Kontakte“	Erläuterungen
10. Angaben, die einen eindeutigen Hinweis enthalten, dass sie als Antwort auf eine andere Kennziffer zu werten sind (Pfeil o.ä.), werden unter der Kennziffer verschriftet, auf die verwiesen wurde.	<i>Nur eindeutige Verweise sind gemeint, um Überforderung bei Verschriftung zu vermeiden.</i>
11. Kann ein Wort nicht gelesen werden, sollten weitere "Schriftkundige" herangezogen werden. Ist keine Klärung möglich, wird anstelle des nicht lesbaren Wortes "xyz" geschrieben.	<i>Verfahren bei Nichtlesbarkeit</i>
12. Der verschriftete Text ist von Zeit zu Zeit, spätestens nach 1 Stunde, zu sichern	<i>Selbst bei einer automatischen Sicherung können Teile verloren gehen.</i>
13. In der vorbereiteten Liste ist genau festzuhalten, wer, was, wann verschriftet, wo und unter welchem Dateinamen abgespeichert hat.	<i>Betrifft die Arbeitsorganisation. So können systematische (personenbezogene) Fehler leichter korrigiert werden</i>
14. Jeder bearbeitete Fragebogen wird an einer gleichbleibenden Stelle (rechts oben auf dem Deckblatt) mit dem Namenskürzel gekennzeichnet.	<i>Arbeitsorganisation: keine Doppelverschriftung, zuständige Person für Rückfragen.</i>
15. Die Dateien sind unter j:\users\temev\projekte\elite\... zu sichern, und zwar mit folgenden Dateinamen: ID (von ... bis)/Dateiname/Person, die verschriftet/Abschluss-Datum 0001-0100 EL95-01.TXT N.N: 12.03.2003 0101-0200 EL95-02.TXT N.N: 12.03.2003 0201-0300 EL95-03.TXT N.N: 12.03.2003 0301-0400 EL95-04.TXT N.N: 12.03.2003 0401-0500 EL95-05.TXT N.N: 12.03.2003 0501-0600 EL95-06.TXT N.N: 12.03.2003 ...	<i>Die Projektleitung kann jederzeit den Stand der Arbeiten sehen. Liste für spätere Aufbereitung (Zusammenfassung u.ä.).</i>

Die Textfassung nach obigen Regeln sieht folgendermaßen aus:

Tabelle 7: Beispiel für den verschrifteten Text

Text	Kommentar
<p>+301 #13 Kultur, <Senat Berlin #17 Siemens #18 Konzertagenturen #24 Berliner Tageszeitungen (alle), <Zitty, <Tip #25 ARD #26 Berliner FU, <TU, <Humboldt Universität #27 Staatstheater, <Staatsoper, <Deutsche Oper, <Konzerthaus, <Deutsches Theater, <Gorki- Theater, <Museen (staatliche in Berlin) #29 ev. und kath. Kirche in Berlin #27 Philharmonie, <Konzerthaus, <Akademie der Künste</p>	<p>← Befragtennummer "301", mit einem "+" davor gekennzeichnet ← Kennziffer "13", mit einem "#" davor gekennzeichnet - zu den Kennziffern 1 bis 12 lag keine Text vor (vgl. Regel 4, Abs. 1 und 2). Innerhalb der Kennziffer "26" wurden drei verschiedene Kontaktstellen genannt - vor jeder neuen Nennung (semantische Einheit) wird ein "<" eingefügt (vgl. Regel 4, Abs. 3).</p>
<p>+302 #06 Kanzleramt #07 Umweltministerium</p>	<p>← Es beginnt der Text eines neuen Fragebogens mit der Nummer 302.</p>
<p>+306 #01 9</p>	<p>← Im Fragebogen Nummer 306 findet sich zur Frage 74 kein Text: Es wird der Code für fehlende Werte "9" als Text geschrieben (vgl. Regel 2).</p>
<p>+503 #23 Gewerkschaften in xyz #24 alle #25 alle #80 Vorsitzender des Anwaltsvereins</p>	<p>← Der Ortsname (vermutlich) konnte trotz mehrfacher Versuche nicht entziffert werden - anstelle des nicht lesbaren Wortes wird "xyz" geschrieben (vgl. Regel 11). ← Dieser Text war keiner Kennziffer zugeordnet worden, sondern vom Befragten an den Rand geschrieben worden; er wird mit "#80" als sonstige Angaben gekennzeichnet (vgl. Regel 4, Abs. 2)..</p>

3.3 Transkription

In dem Beispiel für die Transkriptionsregeln eines Leitfadengesprächs werden nur die Punkte aufgeführt, die über den Regelungsbedarf beim Abschreiben hinausgehen.

Tabelle 8: Beispiel für Transkriptionsregeln (TEXTPACK-Konventionen)

Regeln	Erläuterungen
<ul style="list-style-type: none"> Es muss sichergestellt sein, dass die Kopie der Aufnahme und nicht das Original bearbeitet wird. 	<p><i>Durch intensive Nutzung beim Transkribieren könnte der Datenträger (Band, CD) leicht beschädigt werden</i></p>
<ul style="list-style-type: none"> Jede verbale Äußerung sollte festgehalten werden, auch Verlegenheitswörter wie "nun", "na ja", "was soll ich sagen" usw. 	<p><i>Damit können Rückschlüsse auf die Bestimmtheit der Aussage gezogen werden.</i></p>
<ul style="list-style-type: none"> Es werden drei Identifikatoren vergeben, die jeweils auf einer eigenen Zeile stehen müssen: Interview-Nummer, Frage-Nummer, Sprecher-Nummer. <u>Interview-Nummer:</u> sie befindet sich auf der Kassette bzw. Mini-Disk. Präfix ist "*". <u>Frage-Nummer:</u> sie entspricht der Nummer im Leitfaden (1 bis 28). Sollte nicht ersichtlich sein, zu welchem Themenbereich gesprochen wird, wird die Frage-ID 30 für sonstiges Thema vergeben. Präfix ist "+". <u>Sprecher-Nummer:</u> Präfix ist "#"; es gilt folgende Zuordnung: 1 = Interviewer: Stimmt die Frageformulierung mit der des Leitfadens überein, so kann anstelle des Textes "=" und Nummer der Frage geschrieben werden. 2 = Befragter 3 = sonstige Person, z.B. Ehepartner 4 = Anmerkung beim Transkribieren Die Identifikatoren müssen immer dann geschrieben werden, wenn sie sich ändern. 	<p><i>Festlegung der Texteinheiten, die für die Auswertung eine Rolle spielen.</i></p> <p><i>So kann über die Interviews hinweg Arbeitszeit eingespart werden. Der Fragetext selbst kann später mit Suchen und Ersetzen leicht wieder eingefügt werden.</i></p> <p><i>Hier kann all das festgehalten werden, was nicht Text des Interviews ist.</i></p>
<ul style="list-style-type: none"> Es wird in Hochsprache verschriftet, wobei eine Übertragung von umgangssprachlichen oder Dialektausdrücken stattfindet; "haste" wird z.B. zu "hast du". Wortstellung und Grammatik bleiben unverändert. Jedes Wort, das sich im Wörterbuch finden lässt, darf auch geschrieben werden. 	<p><i>Damit wird auch eine computerunterstützte Auswertung leicht möglich sein, und es kommt nicht auf das Wie, sondern das Was der Aussagen an..</i></p>
<ul style="list-style-type: none"> Gleichzeitiges Sprechen wird so aufgelöst, dass die zuerst begonnen Rede zu Ende geführt wird und dann der Text des Unterbrechenden geschrieben wird. 	<p><i>Die Interaktion während des Interviews ist ohne Bedeutung, wichtig sind die Aussagen. Der Aufwand für die Erfassung der Gleichzeitigkeit kann eingespart werden.</i></p>
<ul style="list-style-type: none"> Sprechpausen, die länger dauern als 2 Sekunden (Zähledauer für "21" und "22") werden festgehalten: 3 - 5 Sekunden = "/" (1 Schrägstrich) 6 - 10 Sekunden = "//" (2 Schrägstriche) 11 - 20 Sekunden = "///" 21 - 60 Sekunden = "////" mehr als 1 Minute = "/////" 	<p><i>Wenn der Befragte länger braucht, bis er antwortet, ist dies ein Hinweis auf eine schwierige Fragestellung; der Antwort geht eine längere Überlegungszeit voraus, so dass es sich nicht um eine spontane Antwort handelt.</i></p>

Regeln	Erläuterungen
<ul style="list-style-type: none"> • Unverständliche Redepassagen werden nicht mehr als drei Mal abgehört. Ist man sich dann immer noch nicht sicher, was es heißt, hat aber eine Vermutung, wird die Stelle (Nummer des Bandzählwerkes bzw. die Zeit der Mini-Disk) und der angenommene Text in Klammern geschrieben. Weiterhin unverständliche Textteile werden mit "xyz" und der Stelle festgehalten. 	<p><i>Nach Abschluss der Verschriftung oder beim Kontrollhören können die unverständlichen Stellen gezielt gesucht und überprüft werden. Klammern oder "xyz" kommen im normalen Text nicht vor, so dass diese Stellen eindeutig zu identifizieren sind.</i></p>
<ul style="list-style-type: none"> • Hinweise, die sich nicht direkt aus dem transkribierten Text ergeben, z.B. störende Hintergrundgeräusche oder sonstige Beeinträchtigungen, die aber für die Interpretation der Aussagen oder die Qualität der Transkription wichtig sein können, werden unter der Sprecher-ID "4" festgehalten. 	<p><i>Die Anmerkungen erfolgen punktgenau und können nach Erledigung leicht entfernt werden.</i></p>

4 Schlussbemerkung

Das Ziel dieses Beitrags ist erreicht, wenn die Texterfassung im Forschungsprozess als eine Phase gesehen wird, an die der gleiche Maßstab anzulegen ist wie an Datenerhebung und Auswertung:

- alle Entscheidungen werden begründet und sind ausgerichtet auf die Forschungsfrage - Relevanz
- es wird nach vorab festgelegten Regeln gearbeitet - Systematik
- alle Arbeitsschritte werden dokumentiert - Intersubjektivität und Nachvollziehbarkeit

Literaturhinweise

Alexa, M. & Geis, A. & Schmidt, I. (1999). TECA-Projekt: Mehrschichtig annotierte Texte für interpretative Textanalyse. In: Gippert, Jost (Hrsg.): *Multilinguale Corpora. Codierung, Strukturierung, Analyse. 11. Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung.* (S. 65-74) Prag: Enigma Corporation.

Alexa, M. (1997): *Computer-assisted text analysis methodology in the social sciences. Arbeitsbericht 97/07.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Alexa, M. & Züll, C. (1999). *A Review of Software for Text Analysis, ZUMA-Nachrichten Spezial, Bd. 5.* Mannheim: Zentrum für Umfragen, Methoden und Analysen:

Dittmar, N. (2001). *Transkribieren: ein Leitfaden für Forscher und Laien.* Opladen: Leske + Budrich.

Ehlich, K. (1992). Computergestütztes Transkribieren - das Verfahren HIAT-DOS. In: Richter, G. (Hrsg.), *Methodische Grundfragen der Erforschung gesprochener Sprache.* (S. 47-59). Frankfurt a. M.: P. Lang.

Früh, W. (2002). *Inhaltsanalyse: Theorie und Praxis* (5. überarbeitete Auflage). Konstanz: UVK Verlagsgesellschaft.

Fuchs, W. (1984). *Biographische Forschung: eine Einführung in Praxis und Methoden.* Opladen: Westdeutscher Verlag.

Mergenthaler, E. (1992). *Die Transkription von Gesprächen : eine Zusammenstellung von Regeln mit einem Beispieltranskript* (3. neu überarbeitete Auflage). Ulm: Ulmer Textbank.

Merten, K. (1995). *Inhaltsanalyse: Einführung in Theorie, Methode und Praxis* (2. verb. Auflage). Opladen : Westdeutscher Verlag.

Mohler, P. Ph. & Züll, C. (2002): *TEXTPACK User's Guide.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Neubauer, M. & Hub, I. & Thimm, C. (1994). *Transkribieren mit LATEX: Transkriptionsregeln, Eingabeverfahren und Auswertungsmöglichkeiten.* Heidelberg: Psychologisches Institut der Universität Heidelberg.

Schmidt, S. J. (1976). *Pragmatik/ Pragmatics II: zur Grundlegung einer expliziten Pragmatik.* München: Fink.

Schmidt, S. J. (1976). *Texttheorie: Probleme einer Linguistik der sprachlichen Kommunikation* (2. verbesserte und ergänzte Auflage). München: Fink.