

Forschungsgruppe Altern und Lebenslauf (FALL)

Forschungsbericht 60

Mai 1997

Waltraud Lay

**Die Vercodung der Angaben zu beruflichen Tätigkeiten
im Alters-Survey**

FREIE UNIVERSITÄT BERLIN

Forschungsgruppe Altern und Lebenslauf (FALL)

Leitung: Prof. Dr. Martin Kohli

FREIE UNIVERSITÄT BERLIN

Institut für Soziologie

Garystr. 55, D-14195 Berlin

Tel. +49 30 838 576 51, Fax +49 30 838 576 52

e-mail: fall@fall-berlin.de

home-page: www.fall-berlin.de

Waltraud Lay (1997): Die Vercodung der Angaben zu beruflichen Tätigkeiten im Alters-Survey. Forschungsgruppe Altern und Lebenslauf (FALL), Forschungsbericht 60. Berlin: Freie Universität.

Forschungsgruppe Altern und Lebenslauf (FALL)

Forschungsbericht 60

Mai 1997

Waltraud Lay

Die Vercodung der Angaben zu beruflichen Tätigkeiten im Alters-Survey

Zusammenfassung	3
1. Die Fragen zur beruflichen Tätigkeit im Alters-Survey	4
2. Das Klassifikationsschema	5
3. Die Konstruktion des Dictionärs und die Codierregeln	6
4. Die Pretestvercodung	8
5. Die Vercodung der Hauptstudie	11
6. Die Vercodungsleistung	14
Literatur	16

Zusammenfassung

Die berufliche Tätigkeit ist eine der wichtigsten und am häufigsten erfragten Hintergrundvariablen quantitativer sozialwissenschaftlicher Untersuchungen. Aufgrund der Vielzahl möglicher Angaben ist eine geschlossene Fragestellung oder eine Feldvercodung nicht praktikabel. Daher wird in Bevölkerungsumfragen üblicherweise eine offene Frage in der Formulierung der "Demographischen Standards" (Statistisches Bundesamt 1993) verwendet. Dies bietet gleichzeitig die Möglichkeit, die Angaben gemäß der Internationalen Standardklassifikation der Berufe (ISCO) vercoden zu können und – darauf aufbauend – vorliegende Klassifikationen und Skalen z.B. zur beruflichen Qualifikation und zum Berufsprestige einsetzen zu können.

Für die Verschlüsselung der Antworttexte bieten sich grundsätzlich zwei Möglichkeiten an: Zum einen die konventionelle inhaltsanalytische Vorgehensweise, bei der speziell geschulte Verocoder die Angaben lesen, die enthaltenen Informationen identifizieren und verschlüsseln. Zum anderen kann auf spezielle Computerprogramme zur Vercodung der Texteinheiten zurückgegriffen werden.

In Anbetracht der Datenmengen, die sich im Rahmen des Alters-Survey in Bezug auf die Fragen zur beruflichen Tätigkeit ergaben, wurde zunächst eine Entscheidung für eine maschinelle Vercodung getroffen, die nach einer Überprüfung der Vercodung der Pretest-Daten revidiert werden mußte. Die Vorgehensweise der automatischen Vercodung offener Angaben zur Berufstätigkeit erwies sich als zu fehleranfällig und zu zeitintensiv. Selbst bei den relativ kurzen und – im Vergleich zu anderen Texten – relativ eindeutigen Angaben kam es ohne die Interpretations- und Selektionsleistung von Vercodern zu einer kaum überschaubaren Fülle von Nicht- bzw. Fehlvercodungen. Eine eingehende Fehleranalyse führte zu dem Fazit, daß das Ausmaß der Fehlvercodungen mit den gewählten Instrumenten der standardisierten computerunterstützten Inhaltsanalyse (cui) nicht hinreichend reduziert werden konnte. Da aber nicht auf den zeitsparenden Einsatz des Computers verzichtet werden sollte, wurde in der Folge ein Verfahren zur computerunterstützten manuellen Vercodung entwickelt, welches die Vorteile der konventionellen Vercodung mit denen des Computereinsatzes verbindet.

Der Arbeitsbericht dokumentiert die Konstruktion des Basisdictionärs und des Regelwerkes, die Vercodung des Pretests, die Entwicklung des Vercodungsprogrammes, die Vercodung der Hauptstudie sowie den Abgleich der Vercodungen und die Reliabilitätsmessung. Im Ergebnis führte die computerunterstützte manuelle Vercodung durch zwei unabhängige Codierer zu einer guten Reliabilität von $Z=.88$ bei einer zufriedenstellenden Vercodungsleistung der 26.141 Angaben zur beruflichen Tätigkeit in der Haupterhebung. Dabei wurde lediglich für sieben Prozent der Angaben Restkategorien für unzureichend detaillierte Angaben verwendet, die für die Verwendung von Klassifikationen und Skalen nur teilweise verwertbar sind.

1. Die Fragen zur beruflichen Tätigkeit im Alters-Survey

Zielsetzung des interdisziplinär angelegten Alters-Survey ist eine Erhebung von Lebenszusammenhängen und Lebensentwürfen der gegenwärtigen und zukünftigen älteren Bevölkerung.¹ Neben der Erhebung der üblichen sozio-demographischen und sozialstrukturellen Daten beziehen sich die inhaltlichen Schwerpunkte des soziologischen Teils der Untersuchung auf die Bereiche Einkommen und finanzielle Absicherung, Wohnen, Gesundheit, soziale Beziehungen und Netzwerke, Tätigkeitsformen und gesellschaftliches Engagement, intergenerationelle Transfers und auf den Bereich Erwerbstätigkeit und Übergang in den Ruhestand. Dabei wurde auf eine ausführliche retrospektive Erhebung der Erwerbsbiographie zugunsten einer detaillierteren Erfassung aktueller Lebenszusammenhänge verzichtet. Es wurden folgende Daten zur Berufstätigkeit aller Befragten (Erwerbstätige, Nichterwerbstätige, Rentner, Pensionäre) sowie derjenigen ihrer Eltern, (Ehe-)Partner und Kinder erhoben (vgl. Dittmann-Kohli et al. 1997):

Befragte

- die erste berufliche Tätigkeit (Frage 105)
- die letzte hauptberufliche Erwerbstätigkeit bei Altersrentnern und Pensionären (Frage 126)
- die derzeitige berufliche Tätigkeit bei Altersrentnern und Pensionären, die einer Erwerbstätigkeit nachgehen (Frage 111)
- die berufliche Tätigkeit zum Zeitpunkt der Befragung bei Erwerbstätigen (Frage 141)
- die letzte hauptberufliche Erwerbstätigkeit bei zum Befragungszeitpunkt Nicht-Erwerbstätigen (Frage 166)

Herkunftsfamilie

- derzeitige bzw. letzte berufliche Tätigkeit (Fragen 18 und 19)

(Ehe-)Partner

- die berufliche Tätigkeit zum Zeitpunkt der Befragung bei erwerbstätigen (Ehe-)Partnern (Frage 213)
- die letzte hauptberufliche Erwerbstätigkeit bei nicht-erwerbstätigen (Ehe-)Partnern (Frage 217)
- die letzte berufliche Tätigkeit während der Ehe bei seinerzeit erwerbstätigen Ehepartnern (Frage 223)

¹ Das Forschungsprojekt "Alters-Survey" wird im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) unter der Leitung von Martin Kohli und Freya Dittmann-Kohli in Kooperation der Forschungsgruppe Altern und Lebenslauf (Berlin), der Forschungsgruppe Psychogerontologie (Nijmegen) und infas-Sozialforschung (Bonn) durchgeführt. Befragt wird eine repräsentative Stichprobe der Wohnbevölkerung der Bundesrepublik Deutschland mit deutscher Staatsbürgerschaft von 5000 Personen der Geburtsjahrgänge 1911 bis 1956 (vgl. Dittmann-Kohli et al. 1995).

- die letzte berufliche Tätigkeit bei seinerzeit nicht erwerbstätigen Ehepartnern (Frage 226)

Kinder

- derzeitige oder letzte berufliche Tätigkeit der ersten vier Kinder (Frage 306)

Die Frage nach der beruflichen Tätigkeit gliedert sich entsprechend der Empfehlungen der demographischen Standards in zwei Teile: Erfragt wurde zuerst direkt die berufliche Tätigkeit, anschließend die jeweilige Berufsbezeichnung (vgl. Statistisches Bundesamt 1993). Ausgehend von der Annahme, daß Eltern möglicherweise nur ungenau über die Tätigkeiten ihrer Kinder informiert sind und ältere Befragte sich eventuell nur schwer an die Berufstätigkeit ihrer Eltern erinnern können, wurde bei den Fragen zu den Eltern und Kindern diese Reihenfolge umgekehrt. Dabei sollte die Umkehrung der Fragereihenfolge die Funktion einer Erinnerungshilfe haben und Zeit zur Konkretisierung der Antwort lassen. Allen Fragen ging jeweils eine Frage nach der beruflichen Stellung voraus, die dort, wo potentiell Tätigkeiten in der DDR genannt werden konnten, um eine Frage nach den Leitungsfunktionen ergänzt wurde. Sie kann zur Differenzierung innerhalb des Systems beruflicher Stellung der DDR verwendet werden (vgl. hierzu Solga 1993).

2. Das Klassifikationsschema

Für die Vercodung von Berufsangaben stehen mehrere Klassifikationsschemata zur Verfügung: die vom Internationalen Arbeitsamt in Genf erarbeitete Internationale Standardklassifikation der Berufe in der Version von 1968 (ISCO-68, vgl. Statistisches Bundesamt 1968), aktualisierte und umstrukturierte Versionen dieses Klassifikationsschemas von 1988 und 1993 (ISCO-88 und ISCO-COM, vgl. Elias/Birch 1993) sowie die nationale Klassifikation des Statistischen Bundesamtes (Statistisches Bundesamt 1992). Bis heute wird in der sozialwissenschaftlichen Forschung zumeist ISCO-68 verwendet. Dies liegt einerseits an fehlenden bzw. unvollständigen Übersetzungen der neueren Versionen, andererseits an der Verfügbarkeit von auf diesem Schema aufsetzenden Skalen und Klassifikationen (z.B. Blossfeld 1985; Treiman 1977; Wegener 1985). Entsprechend wurde auch im Alters-Survey dieses Klassifikationsschema verwendet.

Dieses Klassifikationsschema klassifiziert nach der *Art* der beruflichen Tätigkeiten. Systematisiert werden diese hinsichtlich der Ähnlichkeit der verrichteten Arbeiten in Bezug auf Tätigkeitsbereiche, (implizit) den Qualifikationsanforderungen der verrichteten Tätigkeiten² und dem Ausüben von Leitungsfunktionen. Es weist dabei vier Ebenen auf: Berufshauptgruppen (8), Berufsuntergruppen (83), Berufsgattungen

² Die Qualifikationsanforderungen der Tätigkeit ("skill level") werden in der ISCO-68 nicht explizit als Kriterium angeführt, aber angewandt. Es bezeichnet den Grad der Komplexität der entsprechenden Aufgaben, die entweder eine entsprechende Qualifikation oder eine spezifische Einarbeitung voraussetzen. Explizit als Kriterium eingeführt wird es erst in der ISCO-88 (vgl. Elias/Birch 1993).

(284) und Berufsfelder (1506). Während die Berufshaupt- und Untergruppen eine verhältnismäßig große Bandbreite von Berufstätigkeiten umfassen, differenzieren die Berufsgattungen sowohl nach Tätigkeitsbereich bzw. Branche als auch nach dem skill-level, wobei aber von der numerischen Reihenfolge der 3. Stelle der Kategorie nicht automatisch auf Unterschiede von Qualifikationsniveaus geschlossen werden kann. Die Kriterien der Klassifizierung der Berufsgattung beziehen sich auf das Ausführen ähnlicher Arbeitsverfahren, das Bedienen ähnlicher Einrichtungen, das Leisten ähnlicher Dienste und das Herstellen ähnlicher Gegenstände. Die kleinste berufliche Einheit der ISCO-68 stellen die Berufsfelder dar, die die übergeordneten Berufsgattungen in Berufsfunktionstypen aufteilen. Da jedoch die auf dem ISCO-68 aufbauenden Klassifikationen und Skalen lediglich den Dreisteller voraussetzen, wurde auf die numerische Vercodung der Berufsfelder verzichtet.

Mit der Entscheidung für die ISCO-68 mußten allerdings auch Einschränkungen in Kauf genommen werden. Neuere Berufsfelder, insbesondere im Bereich der Dienstleistungsberufe, sind dort unterrepräsentiert bzw. nicht ausreichend differenziert. Teilweise sind Berufsbezeichnungen wie auch Tätigkeitsbeschreibungen veraltet.³ Und obgleich es sich um eine internationale Klassifikation handelt, ist die Zuordnung von Berufstätigkeiten in der ehemaligen DDR problematisch (vgl. Solga 1993).

3. Die Konstruktion des Diktionärs und die Codierregeln

Als Grundlage für den Diktionär, anhand dessen eine maschinelle Codierung der beruflichen Tätigkeiten mit dem Programm TEXTPACK (Mohler/Züll 1995)⁴ vorgenommen werden sollte, diente das 'Erweiterte alphabetische Verzeichnis der Berufsbenennungen' im Anhang der ISCO-68. Dieses wurde in einem ersten Schritt computerlesbar erfaßt. Anschließend wurde eine Disambiguierung der Begriffe vorgenommen – sämtliche Bezeichnungen der Berufstätigkeiten, die nicht eindeutig, d.h. ohne weitere Angabe, einer einzigen Kategorie zuordbar waren, wurden als 'multiple Kategorien' speziell gekennzeichnet. Auf diese Weise konnten nicht nur harte und weiche Indikatoren⁵ voneinander getrennt werden, sondern es wurde sichergestellt,

³ Hierbei handelt es sich nicht in allen Fällen um einen Nachteil, da die erfragten beruflichen Tätigkeiten – insbesondere bei den älteren Befragten – ebenfalls einige Zeit zurückliegen. Es erscheint aber in jedem Fall wünschenswert, die Skalen für die Verwendung mit aktuelleren Klassifikationen zu überarbeiten.

⁴ Für eine Beschreibung des Aufbau und der Funktionsweise dieses Programmsystems vgl. auch Züll et al. (1991).

⁵ Die Unterscheidung zwischen „harten“ und „weichen“ Indikatoren ist bei einer automatischen Codierung essentiell, da die Verschlüsselung des Textes auf einem Wort-zu-Wort-matching basiert. Als „harte Indikatoren“ werden Begriffe bezeichnet, „deren Auftreten allein ausreicht, um sie eindeutig einer Kategorie zuweisen zu können, unabhängig von Kontext, Kontext-Information oder sonstigen Bedeutungen des Wortes“ (Geis 1992: 21). „Weiche Indikatoren“ dagegen haben kontextabhängige Bedeutungen. Zu weichen Indikatoren sind u.a. Homographen (Wörter, die zwar gleich geschrieben werden, deren unterschiedliche Betonung bei der Aussprache aber zu unterschiedlichen Bedeutungen führt) und Homonyme (gleiche Wörter, die unterschiedliche Bedeutungen haben; vgl. Mohler 1992).

daß uneindeutige Vercodungen jederzeit identifizierbar und überprüfbar sind. In einem weiteren Schritt wurden den mehrdeutigen Begriffen spezifizierende Erläuterungen beigefügt, um eine spätere Überprüfung mittels manueller Vercodung zu erleichtern.⁶

Auf die Verwendung von Wortstämmen wurde angesichts einer befürchteten Erhöhung der Fehlvercodungsrate verzichtet. Die Verwendung von Wortstämmen bietet den Vorteil eines kleineren Diktionärs, erhöht aber auch in nicht absehbarem Umfang die Zahl der Falschvercodungen.⁷ Auf Mehrwortverbindungen wurde nur in eingeschränktem Maß zurückgegriffen, da die Befragten zumeist stichwortartig auf Fragen zur Berufstätigkeit antworten und gleiche Wörter empirisch relativ selten in einer festen Reihenfolge auftreten. Die Vercodungsleistung solcher Indikatoren ist sehr gering.

Auch bei der computerunterstützten Inhaltsanalyse kann aus mehreren Gründen nicht auf die manuelle Vercodung verzichtet werden. Zum einen hinterläßt die automatische Verschlüsselung sowohl eine Restliste der nicht verschlüsselten als auch eine Liste der nur vorläufig und mehrfach codierten Angaben ('multiplen Kategorien'). Beide müssen von speziell geschulten Vercodern nach den gleichen Regeln unabhängig voneinander vercodet werden. Darauf folgt ein Abgleich der manuellen Verschlüsselungen und die Diskussion der nicht übereinstimmenden Fälle, um zu einer gültigen Vercodung zu gelangen.⁸ Zum anderen kann eine Überprüfung des Diktionärs nur anhand eines Abgleichs der Ergebnisse der automatischen Vercodung mit einer konventionell durchgeführten Vercodung durchgeführt werden. Hinsichtlich der Kriterien der intersubjektiven Nachvollziehbarkeit und der Reliabilität der Codierung hat daher neben der Konstruktion des Diktionärs die Formulierung von Codierregeln eine zentrale Bedeutung.

Das Regelwerk des Alters-Survey für die Vercodung der Angaben zur beruflichen Tätigkeit besteht aus zwei Teilen. Ein übergeordneter, allgemeiner Teil, der an schon erprobte Vercodungsregeln in diesem Bereich (vgl. Klingemann/Schönbach 1984) angelehnt ist, formuliert die Basisregeln, während ihre spezifische Anwendung im zweiten Teil expliziert wird. Hier werden Regeln für das Vorgehen spezifischer Problemfälle gegeben, auf Besonderheiten der ISCO hingewiesen und für die Vercodung relevante Ergebnisse der Diskussion innerhalb der Forschungsgruppe zu Zwecken

⁶ Beispielsweise ist die Berufsbenennung „Dreher“ nicht eindeutig einer Kategorie zuzuordnen, da diese Bezeichnung für eine ähnliche Tätigkeit, aber in verschiedenen Bereichen und mit unterschiedlichen Materialien verwendet wird. Es kann sich u.a. sowohl um einen Metall- als auch um einen Steindreher handeln. Schließlich fließt bei dem Metalldreher auch das skill-level mit ein: Es wird unterschieden, ob die jeweilige Person die Maschinen einrichtet oder sie lediglich bedient. Das Wort „Dreher“ kann daher nicht als harter Indikator verwendet werden. Folglich wurde „Dreher“ als 'multiple Kategorie' in den Diktionär aufgenommen.

⁷ Dies hätte zur Folge, daß beispielsweise mit dem Eintrag „Winzer“ sowohl der Winzer als auch die Winzerin vercodet werden, aber auch der Winzergehilfe, was eine falsche Codezuweisung wäre.

⁸ Das heißt nicht, daß die Vercodungsrate als das ausschlaggebende Kriterium betrachtet wird. Selbstverständlich kann auch zu dem Schluß gelangt werden, daß Angaben nicht codierbar sind.

der intersubjektiven Nachvollziehbarkeit festgehalten. Im folgenden werden die allgemeinen Regeln kurz dargestellt:

- M1: Widersprechen sich die beiden Angaben (Tätigkeit und Beruf) nicht, so gilt der Code der spezifischeren Bezeichnung.
- M2: Sind beide Angaben gleich spezifisch, so erfolgt die Codezuweisung nach der Tätigkeitsangabe.
- M3a: Widersprechen sich die beiden Angaben, so muß unter der Berücksichtigung des gesamten Kontextes (inklusive beruflicher Stellung, evtl. dem Rückgriff auf das Qualifikationsniveau) und nach logischer Schlüssigkeit über die jeweilige Codezuweisung entschieden werden. Diese Fälle sind in der Gruppe zu diskutieren. Das Primat der Codierung liegt auf der beruflichen Tätigkeit.
- M3b. Dasselbe gilt für uneindeutige Angaben.
- M4: Allgemeine Angaben werden auf der untersten Ebene vercodet (Bsp.: „Lehrerin“ ohne weitere Angabe erhält den Code 133), d.h. es wird von der Stufe der einfachsten und/oder am häufigsten vorkommenden Tätigkeit des Berufsfeldes ausgegangen und von eventuellen Leitungs- oder Aufsichtsfunktionen abgesehen. Dies geschieht einerseits aus Gründen der Vercodbarkeit, andererseits im Hinblick auf die spätere Zuweisung von Prestigewerten. Unsystematische Fehlzuweisungen sollen hiermit vermieden werden. Dabei wird ein gewisser Umfang an zu niedrig angesetzten Prestigewerten in Kauf genommen.

4. Die Pretestvercodung

Die Vercodung der Pretest-Daten erfolgte in mehreren Schritten gemäß der geplanten Vorgehensweise: Nach dem Einlesen der Rohdaten und ihrer Edition wurden diese mit Hilfe von TEXTPACK verschlüsselt.⁹ Es entstanden drei je nach der Art ihrer Bearbeitung voneinander unterscheidbare Kategorien von Daten. Neben der Datei der vercodeten Daten lagen diejenige mit den multiplen Kategorien, also uneindeutigen Codes, sowie eine Resteliste nicht vercodeter Daten vor. Alle drei wurden einer konventionellen Bearbeitung übereignet. Die Ergebnisse der manuellen Zuordnungen wurden miteinander verglichen, die Intercoder-Reliabilität ermittelt, strittige Fälle diskutiert und ggf. weitere Informationen über die jeweiligen Berufsfelder eingeholt. Auf diese Weise entstand eine vollständige Verschlüsselung des gesamten Datensatzes.

Zur Validierung und zur Erweiterung des erstellten Dictionärs sollten die Vercodung der Pretest-Daten mit Hilfe von TEXTPACK einerseits, der konventionellen, manuellen Verschlüsselung andererseits und der anschließende Abgleich der jewei-

⁹ Die Arbeit mit TEXTPACK stellt einige Anforderungen an das Organisationstalent des Anwenders. Zwecks Reduktion der Dateienmenge erwies es sich als sinnvoll, hauptsächlich mit der Liste der vercodeten Fälle und der Resteliste zu arbeiten und diese für weitere Bearbeitung in andere Programme zu importieren. Auch auf die Anwendung der TEXTPACK-Prozeduren zur Disambiguierung (KWIC etc.) wurde aufgrund der einfacheren Handhabung zugunsten anderer Programme verzichtet.

ligen Ergebnisse dienen. Zugleich sollten damit das Ausmaß und die Güte der automatischen Vercodung sowie der benötigte Zeitaufwand untersucht werden.

Im Ergebnis wurden von den 582 zur Disposition stehenden Berufsangaben 322 maschinell vercodet, 260 Fälle blieben auf der Resteliste. Das entspricht einer Vercodungsleistung von 55%. Darin sind allerdings auch jene Angaben enthalten, die mehrfach, d.h. uneindeutig verschlüsselt wurden (zum Vergleich: nach der manuellen Vercodung eines Vercoders verblieben 44 nicht verschlüsselbare Fälle übrig). Da die Codezuweisung immer eindeutig sein muß, müssen alle Angaben, für die mehr als ein Code vergeben wurde, ebenfalls konventionell vercodet werden. Der Anteil dieser Fälle beträgt 46% aller automatisch vercodeten, also 26% aller zu vercodenden Angaben. Gemeinsam mit der Resteliste ergaben sich somit 408 Fälle, die prinzipiell einer manuellen Vercodung übereignet werden mußten; das sind rund 70% aller Zählleinheiten. Da potentielle Fehler selbst dann nicht ausgeschlossen werden können, wenn qua automatischer Verschlüsselung nur ein Code zugewiesen wird, wurden auch die automatischen Vercodungen auf Übereinstimmung mit den Ergebnissen der manuellen Vercodung untersucht. Im Vergleich zur konventionellen Verschlüsselung wurden demzufolge 159 Zählleinheiten korrekt vercodet, das entspricht 49% aller automatischen Vercodungen. Fehlvercodungen fanden sich in 15 Fällen (5% aller automatischen Vercodungen).

Darüber hinaus wurden in Anlehnung an die Vorgehensweise von Klingemann und Schönbach (1984) die bei der Vercodung möglicherweise auftretenden Fehler in verschiedene Fehlertypen aufgeteilt und die Häufigkeit ihres Auftretens ausgezählt. Die maschinelle Pretest-Vercodung wurde hinsichtlich des Auftretens spezifischer Fehlergruppen und -arten untersucht. Die Fehlergruppe I umfaßt dabei jene Fälle, in denen nur ein Code auf der Basis 'weicher Indikatoren' oder durch Nicht-Erfassung des Kontextes vergeben wurde. Die Fehlergruppe II beinhaltet Fälle, in denen zwei oder mehr Codes auf der Basis 'weicher Indikatoren' oder durch Nicht-Erfassung des Kontextes vergeben wurden. Selbstverständlich sind sämtliche Fehler, die bei der Vergabe eines Codes auftreten können, auch hier möglich. Damit die Kriterien der Relevanz, der Eindimensionalität, der Vollständigkeit und der Trennschärfe (vgl. Geis 1992) erfüllt werden können, mußte in jedem Einzelfall über eine eindeutige Zuweisung entschieden werden.

In dieser Auflistung möglicher Fehlvercodungen (vgl. Tabelle 1) treffen drei Problemfelder aufeinander.

- Die Frageformulierung: In der Alltagssprache wird selten deutlich zwischen Berufsbezeichnung und Berufstätigkeit unterschieden. Insbesondere wenn nach der Tätigkeit Dritter gefragt wird, über die der Befragte nicht gut informiert ist, dienen Angaben der Berufsbezeichnung häufig als Antwort auf die Berufstätigkeit. Andererseits werden Fragen zur Berufstätigkeit mit Termini beschrieben, die nicht alle in den Diktionär aufgenommen werden können - sie können oftmals nur interpretativ erschlossen und konventionell verschlüsselt werden.
- Der Diktionär: Die Komplexität der Kriterien der ISCO spiegelt sich weniger in einzelnen Begriffen als in Satzketten, bzw. Begriffsreihen der Antworten wider. Wie oben schon erwähnt, führt das zu einer Fülle an multiplen Kategorien, also weichen Indikatoren, die nicht als alleiniges Kriterium zur Codevergabe hinreichen.

Gerade aber die weichen Indikatoren, wie z.B. Lehrerin stellen sich als häufig verwendete Antworten dar. Es wäre daher nicht sinnvoll, sie aus dem Diktionär zu entfernen.

- Die Wortbezogenheit der computerunterstützten Inhaltsanalyse: Die Verschlüsselung erfolgt durch kontextfreies Wort-zu-Wort-matching auf der Basis „harter Indikatoren“, d.h. Worte, Wortstämme und Wortkombinationen werden nach dem vorab definierten Kategorienschema als Indikatoren für die kategoriale Zugehörigkeit ohne Berücksichtigung des Kontextes direkt in numerische Daten überführt. Die Wortbezogenheit, mit anderen Worten das "Nicht-Erkennen-von Sinn", ist ein Charakteristikum und eine „entscheidende Schwäche“ (Geis 1992: 21, vgl. auch Kuckartz 1988) der automatischen Vercodung. Diese Eigenart verweist auf die Bedeutung des Diktionärs, dessen Einträge eine eindeutige und kontextfreie Bedeutung haben müssen, um Fehlvercodungen zu vermeiden. Disambiguierungstechniken aber verlagern die Problematik auf eine zusätzliche manuelle Vercodung.

Tabelle 1: Ergebnisse der Fehleranalyse

	Gruppe I: Vergabe eines Codes	Gruppe II: Vergabe von zwei oder mehr Codes	Fehlertypus
1. Angabe einer oder mehrer Berufsbezeichnung/en ohne nähere Beschreibung der Tätigkeit	29 Fälle	70 Fälle	potentiell
2. Spezifikation notwendig, aber nicht im Text vorhanden	6 Fälle	3 Fälle	eindeutig
3. Spezifikation durch Begriffe oder Umschreibungen im Text vorhanden, wird aber nicht vom Diktionär erfaßt	11 Fälle	41 Fälle	potentiell
4. Beschreibung einer früheren Tätigkeit, die nicht erfragt werden sollte	3 Fälle	0 Fälle	eindeutig
5. Beschreibung der Tätigkeit eines Dritten	0 Fälle	0 Fälle	eindeutig
6. Die Angaben spezifizieren einander, wobei beide Angaben erfaßt werden		22 Fälle	potentiell
7. Die Angaben widersprechen einander	8 Fälle	12 Fälle	potentiell

Im einzelnen wurden folgende Fehlerarten unterschieden:

1. Die Antwort enthält eine oder mehrere allgemeine Berufsbezeichnungen, aus denen eine nähere Beschreibung der Tätigkeit nicht hervorgeht. Es wurde die einfache Berufsbezeichnung verschlüsselt.
2. Eine Spezifikation wäre für eine eindeutige Vercodung notwendig, ist aber nicht aus dem Antworttext ersichtlich. Der zugewiesene Code ist mit hoher Wahrscheinlichkeit falsch.
3. Die Spezifikation ist im Antworttext vorhanden, besteht aber aus Begriffen oder Umschreibungen von Tätigkeiten mit Begriffen, die im Diktionär nicht enthalten sind. Die Vercodung ist gemäß der Regel M1 falsch.
4. Das vercodete Stichwort beschreibt eine frühere Tätigkeit, die nicht vercodet werden sollte.

5. Das vercodete Stichwort beschreibt die Tätigkeit eines Dritten.
6. Bei einander widersprechenden Angaben wurde nur eine verschlüsselt.
7. Es werden zwei Codes zugewiesen, wobei die Angaben einander spezifizieren bzw. ergänzen.
8. Es werden zwei oder mehr Codes zugewiesen, wobei die Angaben einander widersprechen.

Die Häufigkeit der potentiellen Fehlvercodungen könnte durch Erweiterung und Disambiguierung des Diktionsärs und ein umfangreiches Regelwerk zwar vermindert werden. Sie sind aber nicht grundsätzlich zu verhindern. Um eine valide Vercodung zu erhalten, müßten also die Ergebnisse der automatischen Vercodung generell manuell überprüft werden. Dies aber bedeutet, daß die maschinelle Vercodung kaum noch Vorzüge gegenüber einer manuellen Vercodung hat.

5. Die Vercodung der Hauptstudie

Für die Vorgehensweise im Alters-Survey stellte sich die automatische Verschlüsselung der offenen Fragen zur Berufstätigkeit als zu zeitaufwendig und fehleranfällig dar. Gleichwohl sollte aufgrund des erheblichen Umfangs der im Alters-Survey zu bearbeitenden Daten nicht auf eine Computerunterstützung der Vercodung verzichtet werden. Aus diesem Grund wurde ein neues Vercodungsprogramm entwickelt, das als eine computerunterstützte manuelle Vercodung verstanden werden kann. Das Programm wurde als „Add-In“ für Word for Windows mit der programminternen Sprache entwickelt.¹⁰ Dies hat den Vorteil, während der Vercodung alle Funktionen des Textverarbeitungsprogrammes nutzen zu können. Auch hier stellt ein austauschbarer Diktionsär den Kern des Programms dar, allerdings werden die Zählleinheiten nicht automatisch vercodet. Der Vercoder hat zunächst gemäß der Vercodungsregeln zu entscheiden, welche Worte oder Wortbestandteile für die Verschlüsselung relevant sind. Der Diktionsär wird daraufhin nach dem jeweiligen Begriff oder Wortbestandteil durchsucht. Das Ergebnis der Suche erscheint in einem Fenster neben dem Text. In einem zweiten Schritt ist es die Aufgabe des Vercoders, zu entscheiden, ob die vorgeschlagene Verschlüsselung (bzw. welche im Falle mehrerer Fundstellen) zutreffend ist. Dabei leisten eingefügte Kommentare und Verweise auf ähnliche Begriffe im Diktionsär Hilfestellung. Nötigenfalls kann die Suche mit Spezifikationen oder "weicheren" Begriffen wiederholt werden.¹¹ Mit der Bestätigung eines vorgeschlagenen Codes wird dieser Code in den Datensatz aufgenommen. Vorläufig nicht codierbare Angaben können speziell gekennzeichnet und bei einer weiteren Bearbeitung des

¹⁰ Dabei wurde einerseits auf Kompatibilität mit TEXTPACK geachtet, so daß der Diktionsär weiterverwendet werden konnte, andererseits wurde das Programm so gestaltet, daß es auch problemlos für die Vercodung anderer offen erhobener Daten eingesetzt werden kann. Das Programm wurde von Harald Künemund entwickelt und ist auf Anfrage von der Forschungsgruppe erhältlich.

¹¹ Umgekehrt kann auch mittels Suche nach einem Code eine Liste der beruflichen Tätigkeiten erstellt werden, die diesen Code erhalten würden. Weiterhin besteht die Möglichkeit, die Kommentare für die Suche ein- oder auszuschließen.

Textes identifiziert werden – das Programm enthält die Option, nur die jeweils nicht oder nur vorläufig codierten Angaben zu bearbeiten. Im Prozeß des Vercodens besteht zudem die Möglichkeit der Dokumentation von erklärungsbedürftigen Codezuweisungen und der individuellen iterativen Erweiterung des Diktionärs. Findet sich z.B. ein Begriff wie „Arzthelferin“ nicht im Diktionär, so kann der zugewiesene Code (ggf. mit einem einschränkenden Kommentar) in den Diktionär aufgenommen werden. Solche neuen Einträge werden für die weitere Vercodung sofort verwendet, können aber jederzeit als Neueintrag identifiziert werden. Der Vercoder steht somit in ständiger Kommunikation sowohl mit dem Text als auch mit dem Diktionär, insofern ihm die Möglichkeit der Suche und des Vergleiches, der Interpretation, der Überprüfung seiner Interpretation und der Korrektur seiner Deutungen gleichermaßen gegeben ist.

Grundsätzlich verwendeten die Vercoder im Alters-Survey ihren eigenen Diktionär. Regelwerk und Diktionär in der Ausgangsfassung wurden – wie bereits dargestellt – im Rahmen der Vercodung der Pretest-Daten ausgearbeitet und diskutiert. Die erste Vercodung der Daten erfolgte unabhängig voneinander. Im Anschluß wurden die Ergebnisse der Vercodung wie auch die jeweiligen Erweiterungen des Diktionärs verglichen. Strittige und nur vorläufige Verschlüsselungen und Diktionäreinträge wurden in der gesamten Arbeitsgruppe zur Diskussion gestellt oder zu weiteren Nachforschungen über das Tätigkeitsprofil der Angaben in gesonderten Listen festgehalten. Als „gültig vercodet“ galt eine Angabe erst dann, wenn die Codes der unterschiedlichen Vercoder übereinstimmten und/oder die Diskussion der strittigen Fälle zu Übereinstimmungen führte. Gleiches galt auch für gültige, d.h. verbindliche Erweiterungen des Diktionärs und Ergänzungen der speziellen Vercodungsregeln. Nicht codierbare Daten erhielten einen speziellen Code. Abschließend wurde eine Konsistenzprüfung der Vercodung in bezug auf diese Ergänzungen und Erweiterungen durchgeführt.

Für nach den Kriterien der ISCO-68 nicht vercodbare, unspezifische Angaben wurden die unten aufgeführten 11 Restkategorien konstruiert. Diese ordnen die jeweiligen Fälle in 'grobe' Tätigkeitsgruppen ein, sind aber nicht Bestandteil des ISCO-Kategoriensystems. Diese Grobklassifikation hat den Vorteil, die Daten teilweise für andere Klassifikationen und Skalen verwertbar zu machen anstatt sie überhaupt nicht vercoden zu können. Es ist beabsichtigt, beispielsweise für diese Gruppen jeweils mittlere Prestigescores zu errechnen oder theoretisch begründet zu vergeben. Die Zuweisung zu diesen Restkategorien erfolgte nur, wenn der jeweilige Fall unter Beachtung der spezifischen Regelformulierung im Diktionär nicht vercodbar war.

Die folgende Liste gibt die bei der Grobklassifikation verwendeten Codes und die darunter subsumierten Tätigkeiten wieder:

- **0001:** Berufssoldaten und Wehrdienstleistende
- **0002:** Offiziere
- **0003:** Angestellte ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0004:** Kaufmännische Angestellte und Kaufleute ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind

- **0005:** Arbeiter ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0006:** Beamte ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0007:** selbständige Handwerker, Facharbeiter, Meister und Lehrausbilder im Handwerk bzw. in der Produktion ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0008:** Technische Angestellte ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0009:** Angestellte mit leitenden Tätigkeiten ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind
- **0010:** Geschäftsführer, tätige Inhaber und Selbständige ohne weitere Angaben oder ohne ausreichende Spezifizierung ihrer Tätigkeit. (Nicht eingeschlossen sind selbständige Kaufleute sowie selbständige Handwerker)
- **0101:** Zivildienstleistende ohne weitere Spezifizierung ihrer Tätigkeit oder mit Angaben, die nicht klassifizierbar sind

Ein besonderes Problem stellte die Klassifikation von Berufstätigkeiten in der ehemaligen DDR mittels der Kategorien der ISCO-68 dar. Sie erweist sich aus vielfältigen Gründen als problematisch. Eine eindeutige und präzise Abbildung der DDR-Realität ist mit Hilfe der ISCO nicht zu leisten (vgl. dazu Solga 1993). Dies gilt insbesondere dann, wenn es um berufliche Tätigkeiten geht, die auch politische Funktionsausübungen beinhalten, wie z.B. bei Kaderleitern. Grundsätzlich wurde, wie auch bei der Vercodung der West-Berufe, bei Angaben zu Berufstätigkeiten in der DDR in erster Linie der primäre Tätigkeitsbereich berücksichtigt. Leitende und politische Funktionen werden in den Fällen bei der Vercodung berücksichtigt, in denen sie das überwiegende Tätigkeitsmerkmal darstellen. Das betrifft Leitungstätigkeiten im Staats- und Parteiapparat (SED - Sekretäre und leitende Mitarbeiter der SED, des FDGB und der FDJ, Vorsitzende und leitende Mitarbeiter der Räte der Bezirke, Kreise, Gemeinden und Städte) sowie die Leitungstätigkeiten in den Bereichen außerhalb des Staats- und Parteiapparates. Bei letzteren wird von den Spezifika der staatssozialistischen Warenproduktion abstrahiert: Leiter von großen Betrieben und Kombinat werden somit der Berufsuntergruppe 2.1. der Internationalen Standardklassifikation der Berufe zugeordnet. Diese Kategorie der ISCO faßt Führungskräfte in der Privatwirtschaft zu einer Gruppe zusammen, eine Kategorienbezeichnung, die für DDR-Produktionsverhältnisse offensichtlich unsinnig ist. Da die überwiegenden Tätigkeitsmerkmale dennoch vergleichbar sind, erfolgt hier eine Analogisierung, die eine gewisse Unschärfe der Abbildung in Kauf nimmt. Die Frage der Ausübung von Leitungsfunktionen kann darüber hinaus anhand einer zusätzlich erhobenen Variable beantwortet werden, die jedoch aus systematischen Überlegungen keinen Eingang in die Vercodung der Tätigkeiten finden konnte.

Im Zuge des Vercodungsprozesses wurde unter Mitarbeit der Codierer ein „Handout für Besonderheiten und spezielle Vercodungsregeln“ erarbeitet, in dem spezifische Entscheidungsregeln dokumentiert werden. Ziel war es hierbei, ein verbindliches Regelwerk für Fallentscheidungen zu erarbeiten. Sie beinhalten Ausführungs-

vorgaben zur oben beschriebenen Vercodung von Tätigkeiten in der DDR, zu den bereits erwähnten Restkategorien und spezielle Vercodungsregeln für Angaben, die sich als gemäß des ISCO nicht eindeutig codierbar erwiesen haben.

6. Die Vercodungsleistung

Jede Vercodung offener Angaben ist auf Angaben ihrer Qualität angewiesen. Insbesondere stellt sich die Frage nach dem Anteil valide vercodeter Angaben und der Reliabilität der Vercodung. Insgesamt wurden in der Hauptstudie 1847 Angaben den Restkategorien zugeordnet. Dies sind rund 7% aller Vercodungen. Die Verteilung auf die Restkategorien) stellt sich folgendermaßen dar:

Tabelle 2: Restkategorien

Kategorie	0003	0004	0005	0006	0007	0008	0009	0010	0101
Anzahl	281	383	655	167	94	50	137	62	18

Nach der Internationalen Standardklassifikation der Berufe von 1968 gültig vercodet wurden 92,6% aller Angaben zur beruflichen Tätigkeit. Der Diktionär wurde in diesem Zusammenhang um insgesamt 1350 neue Einträge ergänzt.

Die Reliabilität ist das Maß für die Reproduzierbarkeit der inhaltsanalytischen Kategorisierung und läßt auf „die Präzision und unmißverständliche Beschreibung des methodischen Instrumentariums einerseits sowie dessen korrekte Anwendung andererseits“ (Früh 1989: 106) schließen. Validität und Reliabilität stehen in einem Spannungsverhältnis zueinander: Eine ausschließliche Verwendung von harten Indikatoren für die jeweiligen Kategorien führt beim Vercodungsvorgang zu einem einfachen Wort-zu-Wort-matching, vergleichbar einer automatischen Verschlüsselung der Texte. In diesem Fall geht die Kontextinformation verloren, was zu fehlerhafter Erfassung und Kategorisierung der tatsächlich gegebenen Information führt. Da die Interpretationsleistung des Vercoders als Fehlerquelle auf ein Minimum reduziert wird, ist die Reliabilität solcher Vercodungen nahe 1.

Je größer das Ausmaß der geleisteten Interpretation ist und je mehr weiche Indikatoren ein Kategorienschema enthält, um so höher ist die Wahrscheinlichkeit, daß die gleichen Textangaben entweder von dem gleichen Vercoder zu unterschiedlichen Zeitpunkten (Intracoder-Reliabilität) oder von unterschiedlichen Vercodern (Inter-coder-Reliabilität) unterschiedlich kategorisiert werden. Andererseits ist gerade die menschliche Interpretation als Erkennen von Sinn konstitutiv für eine valide Kategorisierung.

Im Rahmen der Vercodung der offenen Angaben zu beruflichen Tätigkeiten im Alters-Survey wurde diesem Spannungsverhältnis mit den weiter oben schon ausführlicher geschilderten Maßnahmen begegnet:

- Der Entscheidung für eine manuelle Vercodung,
- der Disambiguierung von weichen Indikatoren durch verbindliche einschränkende Erläuterungen,

- umfangreichen Vercodungsanleitungen und
- die Dokumentation des Vercodungsvorganges, mit Hilfe derer die Entscheidungen der einzelnen Verocoder rekonstruierbar blieben.

Die Intercoder-Reliabilität wurde nach dem Abgleich der Vercodungen konkret nach der Formel

$$Z = \frac{2\ddot{U}}{C_1 + C_2}$$

ermittelt (vgl. Lisch/Kriz 1978: 90). Die Reliabilität (Z) errechnet sich also aus dem Quotienten der Anzahl übereinstimmender Codierungen (\ddot{U}) und der Summe der insgesamt von beiden Codierern verschlüsselten Angaben (C). Nicht vercodete Angaben, die aber vom anderen Vercoder gültig verschlüsselt werden konnten, gehen also als ungleiche Vercodung in die Berechnung ein. Die Intercoder-Reliabilität bei der Verschlüsselung der Angaben zur beruflichen Tätigkeit im Alters-Survey betrug $Z=0.88$; sie kann damit – in Anbetracht der Komplexität – als ausgesprochen hoch bewertet werden.

Insgesamt waren 26141 Angaben zur beruflichen Tätigkeit zu verschlüsseln. Davon wurden 3247 Fälle nicht übereinstimmend vercodet und der Diskussion in der Codierergruppe übereignet. 84 dieser Fälle konnten weder nach der ISCO klassifiziert noch den Restkategorien zugeordnet werden.

Literaturverzeichnis

- Blossfeld, Hans-P. (1985): Berufseintritt und Berufsverlauf. Eine Kohortenanalyse über die Bedeutung des ersten Berufs in der Erwerbsbiographie. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 18, 177-197.
- Dittmann-Kohli, Freya, Martin Kohli und Harald Künemund (1995): Lebenszusammenhänge, Selbstkonzepte und Lebensentwürfe. Die Konzeption des Deutschen Alters-Surveys. Berlin: Freie Universität (Forschungsbericht 47 der Forschungsgruppe Altern und Lebenslauf).
- Dittmann-Kohli, Freya, Martin Kohli, Harald Künemund, Andreas Motel, Christina Steinleitner, Gerben Westerhof unter Mitarbeit von infas-Sozialforschung (1997): Lebenszusammenhänge und Lebensentwürfe – Erhebungsdesign und Instrumente des Alters-Survey. Berlin: Freie Universität (Forschungsbericht 61 der Forschungsgruppe Altern und Lebenslauf).
- Elias Peter & Margaret Birch (1993): ISCO-88 (COM). Internationale Standardklassifikation der Berufe 1988 zur Verwendung innerhalb der Europäischen Gesellschaft. Universität Warwick: Institut für Beschäftigungsforschung.
- Früh, Werner (1989): Inhaltsanalyse. Theorie und Praxis. München: Ötschläger (2. Auflage).
- Geis, Alfons (1992): Computerunterstützte Inhaltsanalyse - Hilfe oder Hinterhalt? In: Züll, Cornelia & Peter Ph. Mohler (Hrsg.): Textanalyse. Anwendungen der computerunterstützten Inhaltsanalyse. Beiträge zur 1. Textpack-Anwenderkonferenz. Opladen: Westdeutscher Verlag.
- Klingemann, Hans D. & Schönbach, Klaus (1984): Computerunterstützte Inhaltsanalyse als Instrument zur Vercodung offener Fragen in der Umfrageforschung. In: Klingemann, Hans D. (Hrsg.): Computerunterstützte Inhaltsanalyse in der empirischen Sozialforschung. Frankfurt/M.: Campus, 227-275.
- Kuckartz, Udo (1988): Computer und verbale Daten. Chancen zur Innovation sozialwissenschaftlicher Forschungstechniken. Frankfurt/M.: Lang.
- Lisch, Ralf & Jürgen Kriz (1978): Grundlagen und Modelle der Inhaltsanalyse. Reinbek: Rowohlt.
- Mohler, Peter Ph. (1992): cui bono - Computerunterstützte Inhaltsanalyse für die qualitative Sozialforschung. In: Hoffmeyer - Zlotnik, Jürgen H.P. (Hrsg.): Analyse verbaler Daten. Über den Umgang mit qualitativen Daten. Opladen: Westdeutscher Verlag, 389-401.
- Mohler, Peter Ph. & Cornelia Züll (1995): TEXTPACK PC. Release 5.0. Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA).
- Solga, Heike (1993): Systematik der beruflichen Tätigkeiten und Ausbildungen in der DDR. Berlin: Max-Planck-Institut für Bildungsforschung (Forschungsbereich Bildung, Arbeit und Gesellschaftliche Entwicklung, Projekt Lebensverläufe und historischer Wandel in der ehemaligen DDR, Arbeitsbericht 2/1993).
- Statistisches Bundesamt (1968): Internationale Standardklassifikation der Berufe. Übersetzung der „International Standard Classification of Occupation“ des Internationalen Arbeitsamtes, Genf, 1968. Stuttgart: Kohlhammer.
- Statistisches Bundesamt (Hrsg.) (1992): Klassifizierung der Berufe - Systematisches und alphabetisches Verzeichnis der Berufsbenennungen - Ausgabe 1992. Stuttgart: Metzler-Poeschl.
- Statistisches Bundesamt (Hrsg.) (1993): Demographische Standards - Eine gemeinsame Empfehlung des Arbeitskreises Deutscher Marktforschungsinstitute, der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute und des Statistischen Bundesamtes. Wiesbaden: Statistisches Bundesamt.
- Treiman, D.J. (1977): Occupational prestige in comparative perspective. New York: Academic Press.
- Wegener, B. (1985): Gibt es Sozialprestige? In: Zeitschrift für Soziologie, 14, 209-235.
- Züll, Cornelia, Peter Ph. Mohler & Alfons Geis (1991): Computerunterstützte Inhaltsanalyse mit TEXTPACK PC: Release 4.0 für IBM XT/AT und Kompatible unter MS/DOS ab Version 3.0. Stuttgart: G. Fischer.