

## Optimal-Matching-Technik: ein Analyseverfahren zur Vergleichbarkeit und Ordnung individuell differenter Lebensverläufe

Erzberger, Christian; Prein, Gerald

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

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

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Erzberger, C., & Prein, G. (1997). Optimal-Matching-Technik: ein Analyseverfahren zur Vergleichbarkeit und Ordnung individuell differenter Lebensverläufe. *ZUMA Nachrichten*, 21(40), 52-80. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208443>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

# OPTIMAL-MATCHING-TECHNIK: EIN ANALYSEVERFAHREN ZUR VERGLEICHBARKEIT UND ORDNUNG INDIVIDUELL DIFFERENTER LEBENSVERLÄUFE

*CHRISTIAN ERZBERGER UND GERALD PREIN*

Die zunehmende Heterogenität von Lebens- bzw. Erwerbsverläufen führt bei quantitativ erhobenen, am Längsschnitt orientierten Daten zu dem Problem, mögliche, hinter diesen Verläufen liegende Ordnungsmuster erkennen zu können. Mit 'Optimal-Matching' - einem Verfahren der Mustererkennung, das etwa in der Gentechnik angewandt wird - steht ein Instrument zur Verfügung, das in der Lage ist, Unterschiede zwischen Personen hinsichtlich des Musters von Verläufen, das heißt der Lage und Abfolge von Erwerbssequenzen zu ermitteln. Auf dieser Grundlage können Personen mit ähnlichen Mustern zu Gruppen zusammengeführt werden. Anhand eines empirischen Beispiels werden mit dem Verfahren des 'Optimal-Matching' Erwerbsverläufe von Männern und Frauen in Abhängigkeit von familialen Verpflichtungen betrachtet. Dabei zeigt sich zum einen, daß generell noch immer die Geburt von Kindern der entscheidende Faktor für die geschlechtsspezifisch unterschiedliche Erwerbsbeteiligung ist, zum anderen aber können gemischtgeschlechtliche Gruppen mit je spezifischen Verlaufsmustern identifiziert werden, die diese traditionelle Trennung sprengen.

Due to the increasing heterogeneity of peoples' life courses and occupational biographies it is more and more difficult to analyse life course patterns in quantitative longitudinal datasets. 'Optimal-Matching' - a technique employed in biological research - is an analytical instrument that facilitates the discovery of life course patterns by comparing the position and order of sequences of life course events. Occupational biographies of women and men were compared with 'Optimal-Matching' by means of an empirical example with respect to their family obligations. It becomes obvious that the birth of children is still the dividing element between men and women in respect of their labour market participation. On the other hand, it is possible to find groups with non-traditional life course patterns, groups of people trying to practice new life course arrangements.

## 1. Einleitung

Unter einer quantitativen, längsschnittorientierten Perspektive lassen sich Lebensverläufe als auf einer Zeitachse chronologisch aneinandergereihte Ereignisse betrachten, wobei die Zusammenschau dieser Ereignisse den Verlauf des Lebens einer Person über den betrachteten Zeitraum wiedergibt. Als Ereignisse stellen sich dabei zum Beispiel unterschiedliche Erwerbsarbeitsformen, Zeiten von Arbeitslosigkeit, Geburten, Heiraten, Scheidungen usw. dar. Aus der Struktur solcher Verläufe, das heißt aus der Lage, Abfolge und Dauer von Ereignissequenzen, ist es zum Teil möglich, Rückschlüsse über die Bedeutung einzelner Ereignisse für den weiteren Lebensverlauf der untersuchten Population zu ziehen oder sogar gesellschaftlichen Wandel zu erfassen, wenn der Bedeutungsgehalt von Ereignissen unterschiedlicher Kohorten miteinander verglichen wird.

Abweichend von dieser ereigniszentrierten Betrachtungsweise wird auch versucht, Lebensverläufe nach dem Muster der Lage und Abfolge von Sequenzen zu ordnen. Unter diesem Gesichtspunkt kann die Entwicklung von Phasenmodellen des Lebens betrachtet werden. So wurde für Frauen der Übergang von einem Zwei- zu einem Drei-Phasen-Modell festgestellt, wobei das Drei-Phasen-Modell durch eine Berufsarbeit vor und nach einer erwerbsarbeitsfreien Familienphase und das Zwei-Phasen-Modell durch einen Ausstieg aus der Erwerbsarbeit ohne weiteren Wiedereinstieg charakterisiert wird (vgl. Myrdal/Klein 1956). Ist man an einer Ordnung von Lebens- bzw. Erwerbsverläufen interessiert, ohne bereits intervenierende Ereignisse im Blick zu haben, so sperren sich komplexe und heterogene Anordnungen von Phasensequenzen gegen die Aufdeckung von Strukturmustern, wie ja auch die immer wieder festgestellte geringe empirische Relevanz der Phasenmodelle zeigt (Erzberger 1993; Krüger u.a. 1989; Lauterbach 1991, 1994).

Diese Schwierigkeiten bestehen vor allen Dingen bei Personen, die gezwungen sind, die beiden Lebensbereiche Familie und Erwerbsarbeit miteinander zu vereinen. Das trifft auf familial gebundene Frauen zu, aber möglicherweise auch auf Männer der jüngeren Generation, deren verstärktes Engagement im Familienbereich für Brüche in der bislang für sie geltenden kontinuierlichen Beteiligung am Erwerbsprozeß sorgen könnte. Eindeutig feststellbare Muster in der Anordnung von Erwerbsphasen im Verlaufe des Lebens sind unter diesen Bedingungen nur noch schwer zu entdecken. Die Lebens- bzw. Erwerbsverläufe entziehen sich daher umso mehr der Aufdeckung von Strukturmustern, je weniger die in ihnen erfaßten Ereignisse klar umgrenzten Bereichen zuzuordnen sind. Auch hier sind Phasenmodelle vorstellbar, die aber neben einer Zwei- und Dreiphasigkeit eine Vielzahl von Strukturmustern erkennen lassen können. Allerdings

bedarf es dazu eines Instrumentes, das in der Lage ist, komplexe Verläufe nach Strukturähnlichkeiten zu ordnen.

Um diesen Ähnlichkeiten - und auch Unterschiedlichkeiten - auf die Spur zu kommen, ist in diesem Beitrag das Verfahren der 'Optimal-Matching-Technik' angewendet worden. Diese Technik erlaubt eine Exploration von Daten innerhalb eines Längsschnittdesigns. Durch den Vergleich zwischen unterschiedlichen Erwerbs- und Lebensverläufen können so Muster, die sich durch Lage, Länge und Abfolge von Ereignissequenzen auszeichnen, bestimmt werden. Diese Muster wiederum bilden den Ausgangspunkt für die Erstellung bestimmter Typologien, die dann in einem weiteren Analyseschritt mit anderen Variablen in Beziehung gesetzt werden können. Bislang sind solche Verfahren der Mustererkennung in der Sozialwissenschaft äußerst selten angewendet worden, und Buchmann/Sacchi, die die 'Optimal-Matching-Technik' in einem Artikel in der KZfSS erwähnen, kommen entsprechend zu dem Schluß, daß das Verfahren nur bei einer eingeschränkten Anzahl von Elementen funktioniere und darüber hinaus noch keine entsprechende Software zur Verfügung stünde (vgl. ebenda 1995). In anderen Wissenschaftszweigen dagegen finden Mustererkennungsverfahren schon seit geraumer Zeit ihre Anwendung. Ausgehend von Erfahrungen in der Gentechnik und der Informatik sowie den wenigen Beispielen aus den Sozialwissenschaften wird am Sonderforschungsbereich 186 der Universität Bremen versucht, die Mustererkennung für die Analyse von Längsschnittdaten fruchtbar zu machen. In diesem Beitrag wird eine erste Anwendung vorgestellt, die helfen soll, die Analysemöglichkeiten zu verdeutlichen.

Das hier verwendete Beispiel stammt aus der Lebensverlaufsforchung und beschäftigt sich mit Ein- und Ausstiegen aus der Erwerbs- und Familienarbeit von Frauen und Männern. Ausgehend von der Beschreibung der besonderen Heterogenität gerade weiblicher Ereignissequenzen und der Begründung der Anwendung dieses Verfahrens gerade bei offensichtlich nicht eindeutig zu qualifizierenden Verläufen, folgt eine allgemeine Beschreibung der Technik des 'Optimal-Matching'. Anschließend werden die Lebensverläufe von Frauen und Männern hinsichtlich ihrer Musterunterschiedlichkeiten und -ähnlichkeiten untersucht.

## **2. Lebensverläufe zwischen Arbeitsmarkt und Familie**

Ausgehend von der Annahme, daß "der Lebenslauf (...) in den modernen Gesellschaften um das Erwerbssystem herum organisiert" ist und sich in "die Dreiteilung in Vorbereitungs-, Aktivitäts- und Ruhephase gliedert" (Kohli 1985: 3), bedeutet dieses zum einen "die Regelung des sequentiellen Ablaufs des Lebens, zum anderen die Strukturierung

der lebensweltlichen Horizonte bzw. Wissensbestände, innerhalb derer die Individuen sich orientieren und ihre Handlungen planen" (Kohli 1985: 3; vgl. auch Blossfeld 1990: 125; Krüger u.a. 1993). Folglich lassen sich Lebensverläufe von Personen am besten nachzeichnen, wenn deren Leben stark mit dem Erwerbssystem verschränkt ist. Das Augenmerk der Analyse liegt dann auf der Institutionalisierung dieser Phasen und "der biographischen Bedeutungszuweisung der Statuspassagen zwischen den Phasen, etwa der Abhängigkeit der Aktivitätsphase im Erwerbssystem von im Bildungssystem erworbenen Zertifikaten und der Position im Ruhestand von vorher per Erwerbsarbeit/Lohn erworbenen Rentenberechtigungen" (Krüger 1995: 5). Mit diesem Betrachtungsfokus aber lassen sich lediglich Verläufe von Personen betrachten, die 'Spuren' im Erwerbssystem bzw. in den dieses System umgebenden Institutionen hinterlassen. Ausgeklammert sind all diejenigen, die entweder im Erwerbssystem kaum 'Spuren' hinterlassen, weil sie daran nicht teilhaben, oder deren 'Spuren' nicht zu interpretieren sind, da deren Leben in weiten Teilen durch eine Vielzahl von nicht an das Erwerbssystem gekoppelten Ereignissen und Entscheidungen bestimmt wird.<sup>1)</sup>

Dieses trifft vor allen Dingen auf Personen zu, die nicht bzw. nicht ausschließlich im Erwerbssystem verankert sind, das heißt Betreuungsaufgaben im Familienbereich übernommen haben. In der Regel handelt es sich dabei um Mütter - aber auch um Väter, soweit diese die Betreuungsarbeiten für ihre Kinder mit allen Konsequenzen leisten. Das Leben dieser Betreuungspersonen zerfällt (spätestens) ab Geburt des Kindes in zwei das weitere Leben beeinflussende unterschiedliche Bereiche: Erwerbsarbeit und Familie. Dieses doppelte Eingebundensein schränkt die Planbarkeit des Lebens insgesamt ein und zeitigt Erwerbsbeteiligungen, die sich scheinbar unstrukturiert über den Lebensverlauf verteilen. Dieses ist bislang sehr eingehend für die Frauen bzw. Mütter beschrieben worden, da sie es üblicherweise sind, die die beiden Bereiche miteinander vereinen müssen (vgl. Jurczyk/Rerrich 1993; Krüger/Born 1987; Streckeisen 1991; Sørensen 1990). Die folgende Verdeutlichung der Konsequenzen dieser zweifachen Verankerung findet ihren Bezugspunkt daher in der Betrachtung des Lebens von Frauen.

Levy hat zur Analyse des Eingebundenseins von Personen in unterschiedliche Bereiche der Gesellschaft den Begriff der Statuskonfiguration entwickelt (Levy 1977). Der Lebenslauf erscheint dann als eine mehr oder weniger stark institutionalisierte Sequenz von Statuskonfigurationen (vgl. Levy 1977: 31). Statuskonfiguration meint dabei die Bestimmung der Stellung eines Mitglieds in der Struktur seiner Gesellschaft durch die Gesamtheit seiner Positionen in den Sektoren, an denen es partizipiert (vgl. Levy 1977: 28). Als Sektoren werden dabei institutionelle Ordnungen bezeichnet, die eine moderne Gesellschaft strukturell gliedern und voneinander abgegrenzte Status- und Interaktionsbereiche bilden, an denen die Individuen teilhaben. Die Unterscheidung

zwischen familiären und außerfamiliären Bereichen bezeichnet zum Beispiel eine solche Sektorengrenze.

Bei jedem Übergang von einem Sektor zu einem anderen können Veränderungen positionaler und/oder konfiguratoraler Art stattfinden. Positionsveränderungen bestehen aus Auf- und Abwärtsmobilität zum Beispiel in Berufs- und Bildungskarrieren. Konfigurationsveränderungen dagegen werden als Aufnahme oder Abgabe von Partizipationen an gesellschaftlichen Statusbereichen definiert. Dabei kann es zu Verschiebungen kommen (zum Beispiel weg von der Schule hinein ins Berufssystem), zu Erweiterungen (zum Beispiel Aufnahme einer Berufstätigkeit bei Weiterführung des Familienhaushalts) oder auch zu Reduktionen (zum Beispiel Aufgeben eines politischen Mandats bei Weiterführung der Berufstätigkeit) (vgl. Streckeisen 1991: 37f.).

Gerade Familienkarrieren von Frauen sind durch Abfolgen von Statuskonfigurationen gekennzeichnet: "Der Familienzyklus als normativer Orientierungsrahmen strukturiert den Ablauf des Lebens der Frau in erster Linie am Anfang: Heirat und Kinderbekommen sind weichenstellende Konfigurationsveränderungen. Was danach kommt, hat weniger den Charakter von Scheidewegstationen; eher geht es um Konsequenzen von bereits getroffenen Entscheidungen bzw. um Anpassungen an vorgegebene Entwicklungen" (Streckeisen 1991: 47). Arbeit in der Familie kann als eher funktional diffus betrachtet werden, da die Anforderungen starken Schwankungen unterliegen. So verlangen Kindergartenkinder nach einer auch zeitlich anderen Betreuung als Kleinkinder oder Schulkinder. Ebenso ist der Ehemann nicht ganz ohne Einfluß auf die Arbeitsorganisation der Ehefrau. Es fallen, je nach sozioökonomischer Stellung, unterschiedliche Unterstützungsleistungen an, seien es normative Vorgaben der Repräsentation oder die unterschiedlich verteilte Einflußnahme der Männer auf Erwerbsentscheidungen ihrer Ehefrauen (vgl. Blossfeld 1995; Erzberger 1993).

Gerade Frauen, die die Kinderphase als Unterbrechung der Erwerbsarbeit geplant haben und in den Erwerbsbereich zurückkehren wollen, sind gezwungen, zwischen raumzeitlich mehr und raumzeitlich weniger zergliederten Lebenszusammenhängen zu alternieren (vgl. Streckeisen 1991: 180).

Auch Hagestad/Neugarten konstatieren, "women's lives are characterized by discontinuity and form no 'orderly' patterns like those generally found in men's lives" (1985: 49). Sie führen dieses auf sogenannte 'counter transitions' zurück. Gemeint sind damit Statusveränderungen eines anderen Familienmitglieds, die die eigene Lebensführung zur Disposition stellen: zum Beispiel der Übergang zur Großmuttertschaft, wenn die eigenen Kinder erwerbstätige Eltern bleiben, oder der Übergang zur Pflegerin eines Familienmitglieds. Die 'counter transitions' "gestalten und standardisieren individuelle

Lebensverläufe auf eigenwillige Art, tragen den Charakter von 'passiven Entscheidungsprozessen', eben: Großmutterwerden, ohne vorher gefragt zu sein, oder Pflegerin des Ehegatten nach einem Schlaganfall. Sie strukturieren sich über die Abhängigkeit von den 'transitions' eines anderen Familienmitglieds, die sich der Planbarkeit, Kontinuierung und verlässlichen Vorhersagbarkeit in der Gestaltung bzw. der Selbstbestimmtheit der eigenen Lebensführung entziehen - und diese nachhaltig durcheinanderwirbeln" (Krüger 1995: 7f.).

Jedoch haben diese von Hagestad/Neugarten beschriebenen 'counter transitions' ihren Ursprung nicht nur im Familienbereich. Gerade die Verfaßtheit von Erziehungs- und Pflegeinstitutionen mit den Angeboten an Kindergartenplätzen und den Öffnungszeiten von Kindergärten, den unregelmäßigen Halbtagschulformen, den Hauspflegeformen nach der Gesundheitsstrukturreform oder der öffentlichen Organisation der Altenpflege zeigt, daß diese Institutionen ohne einen 'Puffer' in der Familie ihrem Auftrag allein nicht gerecht werden können. Um die Familie rankt sich also nun ein Bündel von Anliegerinstitutionen, die die Familie keineswegs ersetzen, sondern im Gegenteil von ihren Organisationsprinzipien her auf möglichst viel Pflege- und Versorgungsarbeiten ausführende Familien(mitglieder) bauen.

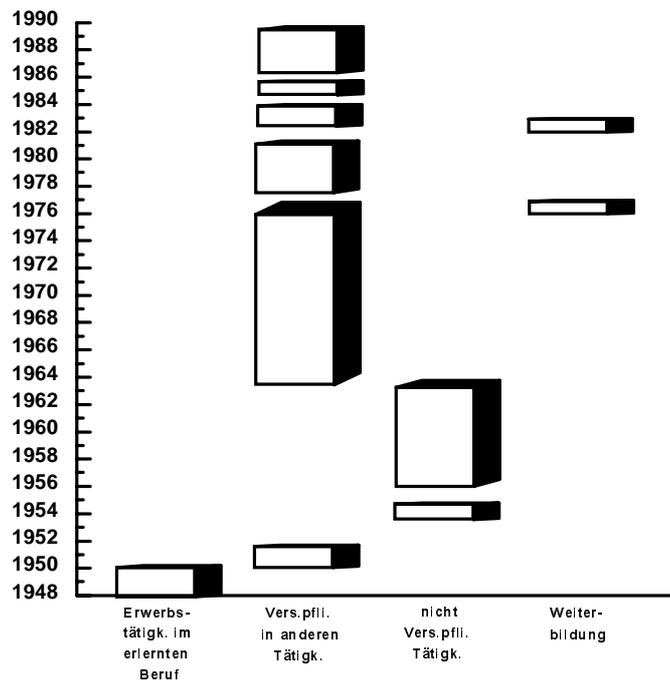
Born (1995) und Krüger (1995) nennen diese Anliegerorganisationen in Anlehnung an Hagestad/Neugarten 'counter institutions'. Da die Versorgungsleistungen innerhalb der Familie in der Regel von den Frauen organisiert werden, sind sie es auch, die dem Eigenleben dieser Institutionen ausgesetzt sind und ihr Leben nach deren Erfordernissen ausrichten müssen. Anliegerinstitutionen - in privater wie staatlicher Organisationsform - greifen in das Leben der Frauen ein und verlangen von ihnen eine permanente Ressourcenaufstellung zur Übernahme von Verpflichtungen. Die Entscheidungsgewalt über Annahme oder Ablehnung der Verpflichtungen liegt dabei in der Regel nicht auf Seiten der Frauen. Durch die Abhängigkeit von den Anliegerinstitutionen entzieht sich der Lebenslauf der eigenen Gestaltbarkeit und Selbstbestimmtheit. So kann die Änderung des Pflegegesetzes, welche eine vergütete Pflege von alten Menschen vorsieht, für die Frau bedeuten, daß neben die Erwerbstätigkeit und Haustätigkeit im eigenen Haushalt die dauerhafte Pflege eines Verwandten in dessen Haushalt hinzutritt und in der Folge der Umfang der eigenen Erwerbstätigkeit eingeschränkt werden muß.

Weibliche familiäre Lebensverläufe sind demnach gekennzeichnet durch eine große Heterogenität, die eben häufig nicht mehr auf einzelne Ereignisse zurückgeführt werden kann. Sicherlich spielt auch die in diesem Zusammenhang häufig fokussierte Geburt eines Kindes eine entscheidende Rolle für den weiteren Lebensverlauf. Es kann jedoch in der Folge zu einer Überlagerung von Ereignissen kommen, die, ausgelöst durch

externe Entscheidungen bzw. außerhalb der Familie auftretende Geschehnisse, zu eigenen Entscheidungen zwingen, die sich der Planbarkeit entziehen. Planungen sind damit an eine Vielzahl von Randbedingungen gekoppelt, die letztendlich die Vorhersagbarkeit des weiteren Lebensverlaufs verhindern, bzw. eigene Entscheidungen sich häufig nur noch im Rahmen von Reaktionen auf faktisch veränderte Situationen bewegen können. Dadurch werden weibliche Lebensverläufe sehr 'unübersichtlich', zumal dann, wenn die Frauen im Erwerbsbereich weiterhin einer Tätigkeit nachgehen. Das Alternieren zwischen dem Familienbereich mit seinen unplanbaren Anforderungen und den daraus folgenden unterschiedlichen Erwerbsarbeitsbeteiligungen führt häufig zu kaum entschlüsselbaren Verlaufsformen, wie sie in quantitativen längsschnitorientierten Daten beobachtet werden können.

Ein solcher Verlauf ist in Grafik 1 beispielhaft zu sehen. Sie zeigt den Erwerbsverlauf einer Kinderpflegerin, die sich heute bereits in Rente befindet.<sup>2)</sup>

**Grafik 1: Engagement im Erwerbsleben (Kinderpflegerin)**



Es ist ein vielfach zergliederter Erwerbsverlauf zu erkennen, der viel Erwerbstätigkeit in berufsfremden Tätigkeiten anzeigt, wobei diese Tätigkeiten jedoch häufig unterbrochen werden. In ihrem gelernten Beruf als Kinderpflegerin hat diese Frau nur kurz gearbeitet (zur Verlaufstypik von ausgesuchten Ausbildungsberufen: Krüger 1993; Krüger u.a. 1989; Krüger/Born 1991). Die hier zu beobachtenden 'Spuren', die sich in Wechseln und Unterbrechungen zu unterschiedlichen Zeitpunkten im Lebensverlauf niederschlagen, lassen sich durch die Betrachtung des Erwerbsverlaufes allein nicht erklären. Erst durch die Hinzunahme familienbiographischer Daten zeigt sich die Interferenz der beiden Dimensionen Erwerbsarbeit und Familie, die je eigenen Logiken folgen. Damit lassen sich die 'Spuren' in der Grafik 1 zumindest teilweise interpretieren:

1951	Heirat, Geburt des ersten Kindes
1953	Geburt des zweiten Kindes
1967	Scheidung
1969	2. Heirat
1975	2. Scheidung
1976	Geburt des 1. Enkelkindes
1981	Geburt des 2. Enkelkindes

Offensichtlich führt zum Beispiel die Geburt der Enkelkinder zeitweise zu einer Aufgabe der Erwerbsarbeit. Nach einer Betreuungsphase wird diese dann aber in beiden Fällen wieder fortgesetzt, was allerdings nicht zwangsläufig bedeutet, daß die Betreuung der Enkelkinder damit beendet wird.

Der Einfluß der Anliegerinstitutionen ist in dieser Grafik nicht dargestellt. Aufgezeichnet werden müßten wechselnde Betreuungsintensitäten je nach Alter der Kinder, Pflegeverpflichtungen und andere Unterstützungsleistungen. Gewollte oder erzwungene Veränderungen auf dieser Ebene haben häufig Auswirkungen auf Form und Umfang der Erwerbsarbeit. Diese Interdependenzen sind nun aber sehr schwierig zu erfassen, da die Anliegerinstitutionen zu zeithistorisch (Einführung des Erziehungsurlaubs) und lebensbiographisch (Großmutterschaft) unterschiedlichen Zeitpunkten in das Leben der Frauen eingreifen können. Auch ist der Eingriff nicht zwangsläufig. Die Einführung des Erziehungsurlaubs hat für Frauen mit 'alten' Kindern keine Auswirkungen mehr, und die Geburt von Enkelkindern bedeutet nicht automatisch die Verpflichtung zur Übernahme von Betreuungsarbeiten. Auf der quantitativen Ebene der Darstellung der Verläufe lassen sich daher zunächst nur Heterogenitäten feststellen, welche sich spontanen Ordnungsversuchen entziehen.

Grundlage dieser Ordnungsversuche sind häufig theoriegeleitete Annahmen über die Bildung von Lebensverlaufsmustern (zum Beispiel Phasenmodelle). Liegen aber keine solchen Annahmen über die möglichen Muster der Lebensverläufe vor, so muß auf eine

eher explorativ angelegte Datenanalysetechnik zurückgegriffen werden, deren Ziel es ist, diese Muster zu bestimmen und Verlaufstypen zu identifizieren. Es geht also um die Identifizierung der 'Spuren' auf der Grundlage von Datensätzen, die die für die Fragestellung ausreichenden Informationen zur Verfügung stellen.

### 3. Konstruktion von Ordnung: 'Optimal Matching'

Zur Identifizierung unterschiedlicher Verlaufstypen auf der Grundlage standardisierter Daten ist es notwendig, Verfahren anzuwenden, mit denen (1.) die Ähnlichkeit bzw. Unähnlichkeit von Lebensverläufen (das heißt von Ereignissequenzen) berechnet und (2.) die Fälle auf dieser Grundlage gruppiert werden können. Während letzteres mit klassifikatorischen Standardverfahren wie etwa der Clusteranalyse problemlos durchführbar ist, sobald eine Distanzmatrix vorliegt, ist die Berechnung von Distanzmaßen für Längsschnittdaten in den Sozialwissenschaften weitgehend Neuland. Allerdings sind andere Wissenschaftszweige seit geraumer Zeit mit ähnlichen Problemen beschäftigt und haben hierzu Lösungen entwickelt, die zum Teil auf sozialwissenschaftliche Problemstellungen übertragen werden können.<sup>3)</sup> Vordringlich zu nennen sind Anwendungen in der Biologie - und hier speziell in der Genforschung -, der Chromatographie, der Geologie, einzelnen Bereichen der Geschichts- und Kulturwissenschaft sowie der Informatik (Text- und Spracherkennung, dem Dateivergleich und der Bildverarbeitung) (vgl. dazu Kruskal 1983). Das allgemeine Problem all dieser Disziplinen besteht im Vergleich der Sequenzen oder Muster zwischen Wörtern, Aminosäuren, Gesteinsschichten, Bits oder Bytes, Tanzschritten etc.

Eine Sequenz setzt sich aus einer Aneinanderreihung von Merkmalen, Ereignissen oder Zuständen zusammen. Diese Merkmale können als Werte von Variablen beschrieben werden, wobei jede Variable  $V_i$  für das Merkmal an der Position  $i$  in der Sequenz steht ( $V_1$ - $V_n$ ). Diese Merkmale können hierbei sowohl als metrisch als auch als kategorial angesehen werden, wie zum Beispiel die Sequenzen 'M U N I C H' und 'M U N C H E N': Jede Sequenz kann damit als Zeilenvektor  $s'$  dargestellt werden, wobei  $s'_1$  im oben genannten Beispiel das Merkmal 'M' enthält,  $s'_2$  das Merkmal 'U',  $s'_4$  im einen Fall das Merkmal 'I', im anderen das Merkmal 'C', etc.

Die Bewertung der Distanz zwischen den Sequenzen 'M U N I C H' und 'M U N C H E N' auf der Grundlage eines direkten Vergleichs der Elemente  $s'_i$  führt zum Aufweis einer starken Unähnlichkeit: lediglich drei Elemente beider Vektoren stimmen überein (die drei Anfangsbuchstaben 'M U N'), da durch das eingeschobene 'I' in der ersten Sequenz sich die beiden Sequenzen gemeinsamen Elemente 'C' und 'H' an unterschiedlichen Positionen befinden. Somit sind Anwendungen vorstellbar, bei denen ein solcher

direkter Vergleich zu einer falschen Einschätzung von Differenzen führen kann, beispielsweise beim Dateivergleich, wenn zwei Dateien zwar identisch sind, zu Beginn der einen jedoch ein Byte fehlt. Um solche Artefakte zu umgehen, wurde die 'Optimal-Matching-Technik' entwickelt.

Bei dieser Vorgehensweise wird das Maß der Unterschiedlichkeit, die sogenannte Levenshteindistanz, zwischen zwei Mustern als minimale Summe von 'Kosten' berechnet, die entstehen, wenn das eine Muster in das andere überführt wird. Hierbei wird also ein Muster als 'Quellsequenz' definiert, ein anderes als 'Zielsequenz'. Die Quellsequenz muß nach der Transformation mit der Zielsequenz identisch sein. Die hierbei notwendige Überführung geschieht mittels dreier möglicher Aktionen: löschen, einsetzen und ersetzen (vgl. Levenshtein 1965; Kruskal 1983: 10ff.), für die jeweils die Kosten spezifiziert werden müssen.

Wenn das Muster 'MUNICH' in das Muster 'MUNCHEN' transformiert wird, indem zuerst das I an der vierten Position gelöscht wird und dann am Ende die Zeichen E und N eingefügt werden, sind damit drei Aktionen notwendig. Wenn - wie in allen folgenden Beispielen - für löschen, ersetzen und einfügen identische Kosten von '1' festgelegt werden, addieren sich die Kosten im vorliegenden Beispiel auf drei.<sup>4)</sup>

In der Regel bestehen die zu klassifizierenden Sequenzen aus wesentlich mehr Elementen und unterscheiden sich zudem in ihrer Länge. Da die Bedeutung der 'Kosten' aber von der Länge der Sequenzen abhängig ist - so ist ein Übergang in einer Sequenz von insgesamt fünf Ereignissen bedeutender als ein Übergang in einer Sequenz von 30 - ist es nötig, diese Kosten zu standardisieren. Das kann dadurch geschehen, daß die zunächst ermittelten 'Transformationskosten' durch die Länge der Sequenzen (bei unterschiedlicher Länge durch die Länge der längeren Sequenz) dividiert werden (vgl. Abbot/Hrycak 1990: 153ff.). Wir erhalten dann ein Maß für die Distanz zwischen zwei Sequenzen. Diese beträgt im oben dargestellten Beispiel zwischen 'MUNICH' und 'MUNCHEN'  $3/7 \approx 0,43$ . Die Distanz kann damit zwischen 0 (maximale Ähnlichkeit) und 1 (maximale Unähnlichkeit) variieren.

Auch Lebensverlaufsdaten können in der Form von Sequenzen abgebildet werden.<sup>5)</sup> Da wir im Falle des Vergleichs von solchen Lebenslaufsequenzen vor einem ähnlichen Problem stehen wie zum Beispiel beim Vergleich von DNA-Sequenzen, liegt also eine Anwendung sequenzanalytischer Techniken auch hier nahe: Auch beim Vergleich von Lebensverläufen finden sich selten die Ereignisse der untersuchten Personen auf den gleichen Positionen: Längere schulische Bildung bedingt einen späteren Eintritt in das Erwerbsleben, kindbedingte Erwerbsunterbrechungen finden je nach Alter der Person zu

unterschiedlichen Zeitpunkten statt. Dennoch können die Lebensverläufe sich sehr ähnlich sein, wenn auch dann gegeneinander verschoben.<sup>6)</sup>

In der Regel besteht das Problem darin, eine große Anzahl von Verläufen miteinander vergleichen zu wollen, die wesentlich mehr Zustände bzw. Ereignisse aufweisen als im Beispiel oben demonstriert. Es ist dann unerlässlich, diese Aufgabe einem Computer zu übertragen, der diesen Algorithmus auf die Daten anwendet. Es entsteht eine Matrix von Distanzwerten, in der dann über die Angabe von Reihe und Spalte der Distanzwert in der so definierten Zelle für die interessierende Kombination zweier Verläufe ermittelt werden kann.

Dieses ist der erste Schritt zum Auffinden von Ordnung, wobei nicht nur die Länge und Häufigkeit, sondern ebenso die Lage bzw. Abfolge von Ereignissen und Sequenzen in den Ordnungsprozeß einfließen. So lange das Interesse bei der Kombination zweier Verläufe verbleibt, wie zum Beispiel beim Vergleich von Lebensverläufen bei Paaren (zum Beispiel Mutter/Tochter oder Ehemann/Ehefrau), können die Distanzwerte der jeweiligen Kombination über ein Scattergramm ausgedruckt werden: Die nahe zusammenliegenden Verläufe weisen die niedrigsten Werte auf, während steigende Distanzwerte ein Zeichen sind für die Unterschiedlichkeit der Verläufe. Ein weiterer Analyseschritt aber ist nötig, wenn eine Vielzahl miteinander verglichener Verläufe geordnet werden soll, zum Beispiel um Unterschiede bzw. Ähnlichkeiten im Lebensverlauf zwischen Männern und Frauen oder zwischen unterschiedlichen Geburtskohorten zu ermitteln.

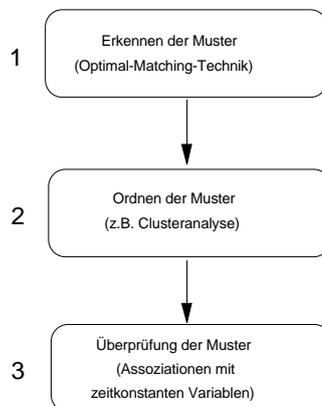
Diese in Form einer Distanzmatrix vorliegenden Daten können dazu in einem zweiten Schritt mit Hilfe statistischer Klassifikationsverfahren wie etwa Multidimensionaler Skalierung oder Clusteranalyse weiterbearbeitet werden. Während der erste Teil des Prozesses sich mit dem Vergleich der Verläufe beschäftigt, werden diese dann im zweiten Teil zu Gruppen mit ähnlichen Musterkombinationen gebündelt.

Die geordneten Muster stellen aber noch nicht unbedingt die eigentliche Analyse dar. Die ermittelten Gruppen können in einem dritten Schritt in bezug auf Assoziationen mit interessierenden, zeitkonstanten Variablen geprüft werden. Es ist zum Beispiel möglich, daß bestimmte Lebenslaufmuster mit bestimmten Berufsausbildungen korrelieren und sich hieraus gegebenenfalls Hypothesen über kausale Mechanismen erarbeiten lassen. Wir werden hierzu im folgenden Teil einige Beispiele geben.

Wichtig erscheint es uns an dieser Stelle, zumindest stichpunktartig auf zum Teil noch ungelöste Probleme der Anwendung von 'Optimal-Matching-Techniken' auf Lebensverlaufsdaten hinzuweisen.

1. Zunächst muß eine der Forschungsfrage angemessene Lösung des Problems ungleich langer Sequenzen gefunden werden. Hierbei sind historisch oder kulturwissenschaftlich orientierte Forschungsvorhaben von solchen zu unterscheiden, die im Bereich der Lebensverlaufsforchung stattfinden: Während beispielsweise Tanzschritte (vgl. Abbott/Forrest 1986) als abgeschlossene Sequenzen betrachtet werden können, sowie die unterschiedliche Länge ein inhaltlich relevantes Unterscheidungsmerkmal ist, kommen in der Lebensverlaufsforchung unterschiedlich lange Sequenzen zumeist durch Zensierungen zustande: Wenn zu einem bestimmten Zeitpunkt der Lebensverlauf einer 25jährigen Frau mit dem einer 35jährigen verglichen wird, so ist die Differenz in der Länge der Sequenz lediglich ein Effekt des Alters, zudem sind beide Sequenzen prinzipiell in die Zukunft hin offen. Eine im folgenden Beispiel dokumentierte, praktikable Lösung ist, daß bei unterschiedlicher Länge der Sequenzen die längere Sequenz nur bis zur Länge der kürzeren Sequenz in den Vergleich einbezogen wird. Eine Standardisierung der Distanz findet dann auf der Basis der Länge der kürzeren Sequenz statt. Welche Auswirkungen sehr stark divergierende Sequenzlängen auf die Stabilität des Verfahrens haben und wann hieraus Artefakte resultieren können, ist eine Frage, die noch durch weitere, eingehende Untersuchungen zu prüfen ist.

### Grafik 2: Überblick über die einzelnen Schritte der Analyse



2. Sodann müssen die Probleme untersucht werden, die aus der Kodierung von Lebensverläufen mit stetiger Zeit in Sequenzen mit diskreten Zustandsvariablen entstehen. Hier entstehen methodische Probleme und gegebenenfalls Artefakte nicht vor allem deshalb, weil ein Informationsverlust vorliegt - dies ließe sich durch entsprechend feine Ab-

stufung der betrachteten Intervalle lösen, wenn er überhaupt praktisch forschungsrelevant ist. Viel weitgehender sind Probleme, die daraus entstehen, daß längere Episoden - beispielsweise ein Zeitraum der Berufstätigkeit von drei Jahren - in voneinander getrennt betrachtete Subepisoden - beispielsweise in die drei Zustandsvariablen 'Berufstätigkeit' (Vollerwerb, Teilzeit, nicht versicherungspflichtig) - aufgeteilt werden. Bei der Vorgehensweise des 'Optimal-Matching' werden nun diese drei Subepisoden als voneinander unabhängig betrachtet. Diese Annahme der Unabhängigkeit von Elementen einer Sequenz ist natürlich unproblematisch im Falle anderer Anwendungen dieser Technik wie beispielsweise der DNA-Analyse: hier ist eine Base der DNA-Kette nicht von der vorherigen abhängig. Anders ist es bei Lebensläufen, denn die Wahrscheinlichkeit, in einem Jahr berufstätig zu sein, ist - bei allen Unwägbarkeiten - sicherlich höher bei jemandem, der im Vorjahr berufstätig war, als bei jemandem, der im Vorjahr Haus- und Familienarbeit geleistet hat. Prinzipiell kann dieses Problem durch zwei Modifikationen gelöst werden, die jedoch derzeit noch nicht anwendungsreif sind: a) durch die Kodierung der Lebensverläufe in stetiger Zeit und b) durch die stärkere Gewichtung von Zustandswechseln im Gegensatz zur Fortschreibung des vorherigen Zustands bei der Berechnung der Distanzmatrix. Durch ein so modifiziertes Verfahren könnte eine stärker dynamische Herangehensweise der Sequenzanalyse entwickelt werden.

3. Ein weiteres Problem betrifft die Gewichtung der jeweiligen Kosten generell (vgl. Abbot/Hrycak 1990: 178): Das hier vorgestellte Verfahren des 'Optimal Matching' geht von nominalskalierten Variablen aus. Es ist aber ebenso möglich und in den meisten Fällen auch sinnvoll, Stationen im Lebensverlauf über ordinal- bzw. intervallskalierte Variablen in dieses Verfahren einzubeziehen, wenn etwa die Variablenwerte der einzelnen Zustände bzw. Ereignisse des Verlaufs in einem hierarchischen Verhältnis zueinander stehen. Dieses ist zum Beispiel bei Karriereverläufen der Fall. Bei einem Karrieresprung vom Praktikanten zum Geschäftsführer wird eine größere Distanz überbrückt als bei einer Regelbeförderung innerhalb einer Beamtenlaufbahn. Die Ersetzungskosten können dann über einen anderen Algorithmus berechnet werden. Der Sprung vom Wert 1 zum Wert 3 verursacht dann andere Kosten als der Sprung von 1 nach 2. Die 'Kosten' sind es, die das Ausmaß an Distanz bzw. das Maß an Nähe zwischen den Verläufen bestimmen. Bei der Ermittlung des Algorithmus, der diesen Kosten zugrundeliegt, ist daher vorsichtig vorzugehen. Das anzuwendende Verfahren sollte sich nach den theoretischen Überlegungen richten, die der Untersuchung beziehungsweise der Fragestellung zugrunde liegen. Auch hinsichtlich des Verhältnisses von Ersetzkosten und Einfüge- bzw. Löschkosten müssen noch weitere Untersuchungen stattfinden: Wenn in einem Lebensverlauf ein Zustand wie

beispielsweise Arbeitslosigkeit vorkommt, im anderen jedoch nicht - das heißt wenn der Wert einer Zustandsvariablen hier ersetzt werden muß -, ist dies sicherlich ein stärker zu gewichtender Unterschied, als wenn in einem der beiden Lebensverläufe ein Übergang in die Arbeitslosigkeit nach dem zweiten Berufsjahr, im anderen allerdings erst nach dem dritten stattfindet.

Trotz dieser bestehenden Probleme soll im folgenden Teil gezeigt werden, daß bereits mit nur leicht modifizierten Ersetzungs-, Einfügungs- und Löschkosten sich mittels 'Optimal-Matching' theoretisch interessante Resultate erzeugen lassen.

#### **4. Frauen und Männer: Erwerbsarbeit oder Familie?**

Die Daten der im folgenden präsentierten Ergebnisse sind im Verlauf des Projektes 'Statuspassagengestaltung und intergeneracionales Erbe. Zum Wandel der Sequenzmuster zwischen Erwerbsverlauf und Familie im Generationentransfer' des Sonderforschungsbereichs 186 erhoben worden.<sup>7)</sup> Untersucht werden sollen die Erwerbs- und Familienverläufe von Frauen und Männern im Alter zwischen 30 und 41 Jahren. Ziel der Analyse ist zunächst die Ermittlung von Lebensverlaufsmustern, wobei diese Muster durch die Abfolge und Lage von Erwerbssequenzen charakterisiert werden. Die Erklärung der Unterschiede zwischen den Mustern erfolgt anschließend durch die Variablen 'Leben mit Kindern' und 'Leben in einer festen Partnerschaft'.

Es stehen retrospektiv erhobene Längsschnittdaten von 129 Personen (76 Frauen, 53 Männern) zur Verfügung. Die Abbildung der Verläufe geschieht durch Variablen, die die Stellung im Erwerbsleben jeder Person zwischen dem 17. und 41. Lebensjahr anzeigen. Die Verläufe weisen als Minimum 14 Variablen (17.-30. Lebensjahr) und als Maximum 25 Variablen (17.-41. Lebensjahr) auf. Die Variablen können dabei folgende Ausprägungen annehmen:

- V Vollerwerbstätigkeit
- T Teilzeiterwerbstätigkeit
- H Hausarbeitstätigkeit/geringfügige Erwerbstätigkeit
- A schulische und berufliche Ausbildung
- Z Bundeswehr/Zivildienst/Auslandsaufenthalte.

Die Arbeitsloskeitszeiten sind in diesem Sample sehr gering. Sie betragen, bezogen auf die Gesamtzahl der betrachteten Jahre, gerade 0,6%. Diese Zeiten wurden in diesem Fall aus der Analyse ausgeschlossen. Ein höherer Anteil von Arbeitslosenzeiten hätte zu einer eigenen Ausprägung in den Variablen geführt.

Die Bildungsbeteiligung ist, bezogen auf die allgemeine Schulbildung, sehr heterogen. Gymnasialabschlüsse sind im Sample jedoch deutlich überrepräsentiert. Männer und Frauen unterscheiden sich aber hinsichtlich der quantitativen Besetzung der Abschlüsse nicht. Alle Personen haben eine Berufsausbildung abgeschlossen.

Der Vergleich der Verläufe von zwei 32jährigen Personen stellt sich dann zum Beispiel folgendermaßen dar (Abbildung 1). Zunächst befinden sich beide Personen in Ausbildung (A), diese geht in eine lange Vollzeiterwerbstätigkeit (V) über und mündet in eine Hausarbeitstätigkeit (H). Allerdings sind die Verläufe nicht identisch. Die Person 2 verfügt über eine längere Ausbildungsphase, während die Person 1 etwas länger im Hausbereich tätig ist. Zur Transformation sind vier Aktionen erforderlich. Nach der zuvor verwendeten Terminologie kann im vorderen Teil des Verlaufes dreimal Vollzeittätigkeit (V) durch Ausbildung (A) - und am Ende an der Position  $V_{131}$  Hausarbeit (H) durch Vollzeittätigkeit (V) ersetzt werden. Die minimale Intersequenzdistanz beträgt bei der Zugrundelegung der Transformationskosten jeder Aktion von 1 hier 0,235.

**Abbildung 1: Beispiel für die Kodierung zweier Lebensverläufe**

	$V_{117}$	$V_{118}$	$V_{119}$	$V_{120}$	$V_{121}$	$V_{122}$	$V_{123}$	$V_{124}$	$V_{125}$	$V_{126}$	$V_{127}$	$V_{128}$	$V_{129}$	$V_{130}$	$V_{131}$	$V_{132}$	$V_{133}$
1	A	V	V	V	V	V	V	V	V	V	V	V	V	V	H	H	H
2	A	A	A	A	V	V	V	V	V	V	V	V	V	V	V	H	H

Für die Analyse der Daten werden Transformationskosten gewählt, die die Übergänge vom Ereignis des Quellverlaufs zum Ereignis des Zielverlaufs angemessener wiedergeben können. Es wird davon ausgegangen, daß zum Beispiel der Übergang von einer Vollzeittätigkeit zur Hausarbeit größere 'Kosten' verursacht als ein entsprechender Übergang zur Teilzeittätigkeit und vice versa. Die differenzierte Berechnung der 'Ersetzungskosten' ist in der folgenden Tabelle wiedergegeben.

Für das Einfügen bzw. Löschen eines Ereignisses werden jeweils 'Kosten' von 0,5 berechnet. Durch diese Regelung wird dem Vorgang des Löschens und Einfügens Vorrang vor dem des Ersetzens gegeben. Damit ist bei gegeneinander verschobenen Verläufen die Aktion des Verschiebens des Zielverlaufes 'billiger' als das Ersetzen von Ereignissen.

**Tabelle 1: Berechnung der 'Ersetzungskosten'**

Quellvariable		Zielvariable	Kosten
Vollzeit	↔	Hausarbeit	1

	↔	Teilzeit	0,8
	↔	Bund/Ausland	1
	↔	Ausbildung	1
Teilzeit	↔	Hausarbeit	1
	↔	Bund/Ausland	0,8
	↔	Ausbildung	0,8
Hausarbeit	↔	Bund/Ausland	1
	↔	Ausbildung	1
Bund/Ausland	↔	Ausbildung	0,4

Die jeweiligen Intersequenzdistanzen werden über ein C-Programm berechnet und in einer Matrix abgelegt.<sup>8)</sup> Die Distanzen in dieser Matrix werden - oben dargestellt als Teil 2 des Analyseprozesses - im Anschluß über eine Clusteranalyse im Rahmen des Programmpaketes SAS nach Verlaufsmustern geordnet.

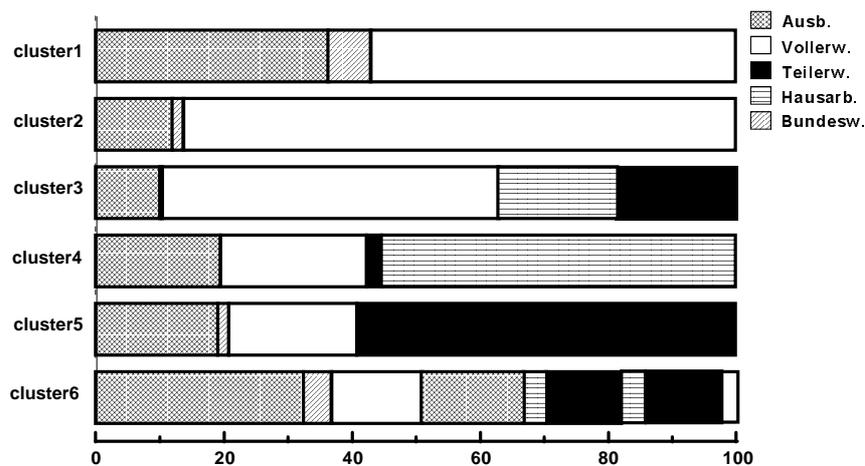
Die 'Optimal-Matching-Technik' erlaubt es, Männer und Frauen gemeinsam zu untersuchen. Es werden keine gesonderten Analysen vorgenommen, sondern die Lebensverläufe beider Geschlechter werden innerhalb eines Analyseschrittes miteinander verglichen. Sollten sich geschlechtsspezifische Unterschiede herausstellen, so müßten sich klar nach Geschlecht getrennte Gruppen unterschiedlicher Verläufe bilden. Sind sich dagegen Männer und Frauen hinsichtlich ihrer Verlaufsstruktur sehr ähnlich, so sollten sie auch eine gemeinsame Gruppe bilden, die sich durch eine charakteristische Verlaufsform auszeichnet.

Der Vergleich der 129 männlichen und weiblichen Verläufe erbrachte sechs Cluster mit sehr unterschiedlichen Phasenkombinationen.<sup>9)</sup> In Grafik 3 sind die Gruppen und deren spezifische Phasen abgebildet. Die Länge ist dargestellt als Anteil der einzelnen Phasen an der Gesamtlänge aller Phasen im jeweiligen Cluster. Die Phasen selbst sind hier bereits nach ihrem je nach Cluster unterschiedlich typischen Muster wiedergegeben. Diese Clustertypik wurde über eine Visualisierung der Lage und Anordnung der Phasen ermittelt. Da nicht immer alle Personen eines Clusters sich in das Ordnungsmuster einfügen, wurden Phasenverläufe dann als typisch charakterisiert, wenn mindestens 60% der Personen des jeweiligen Clusters diesem Muster folgten.<sup>10)</sup> Um darüber hinaus die Typik besser herausstellen zu können, sind Phasenanteile unter 5% nicht wiedergegeben.

Cluster 1 wird danach geprägt durch lange Ausbildungs- und Vollerwerbszeiten. In Cluster 2 finden sich geringe Ausbildungslängen und lange Zeiten der Vollerwerbstätigkeit.

Cluster 3 zeigt Personen, die lange voll- und teilzeiterwerbstätig sind, relativ lange den eigenen Haushalt versorgen, jedoch nur über kurze Ausbildungszeiten verfügen. Darüber hinaus ist zu erkennen, daß die Lage und Anordnung der Sequenzen ein Drei-Phasen-Modell anzeigt: Die Vollerwerbstätigkeit wird durch eine Haustätigkeit unterbrochen und später als Teilzeittätigkeit weitergeführt (vgl. Myrdal/Klein 1956). Cluster 4 dagegen zeigt ein Zwei-Phasen-Modell: Nach einer Vollzeit- bzw. Teilzeiterwerbstätigkeit steigen die Personen dieses Clusters ganz aus der Erwerbstätigkeit aus. In Cluster 5 mündet die Vollerwerbstätigkeit in eine Teilzeitarbeit. Dem Cluster 6 ist kein eindeutiges Muster zuzuordnen. Häufige Wechsel der Erwerbstätigkeitsformen, zweite Ausbildungen und Unterbrechungen durch Hausarbeitstätigkeiten sind hier zu finden. Das typische Muster dieses Clusters besteht in seiner 'Unruhe', in der nicht zu erkennenden eindeutigen Phasenspezifik.

**Grafik 3: Anordnung und Länge der unterschiedlichen Phasen in den einzelnen Clustern**



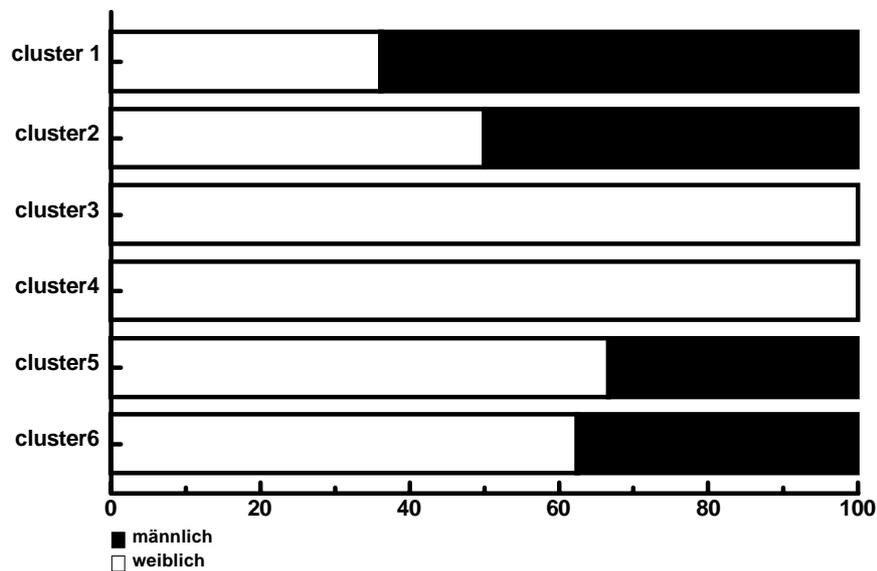
Die größte Clusterbesetzung findet sich in Cluster 1 mit 37% aller Personen, bis auf Cluster 6 mit 5% entfallen auf die restlichen Cluster die verbleibenden 58% zu etwa gleichen Teilen. Die Altersverteilung läßt keine großen Unterschiede zwischen den Clustern erkennen. Lediglich in Cluster 4 liegt der Altersdurchschnitt geringfügig über dem der anderen Cluster.

Grafik 4 zeigt nun die Verteilung von Männern und Frauen auf die ermittelten Cluster. Das unabhängig vom Geschlecht durchgeführte Ordnungsverfahren läßt Gruppen von

Personen entstehen, die sich durch charakteristische Muster der Lebensverläufe auszeichnen und zeigt ebenso, daß die Länge und Anordnung von Erwerbssequenzen im Lebensverlauf eine eindeutige Geschlechtertrennung aufweisen.<sup>11)</sup>

Es fällt auf, daß die Cluster, in denen das Zwei- bzw. Drei-Phasen-Modell vorherrscht (Cluster 3 und 4), ausschließlich mit Frauen besetzt sind. Viele Frauen befinden sich auch im Cluster 5, in dem Teilzeiterwerbstätigkeitsformen vorherrschen. Die Männer hingegen finden sich in dem Cluster, das durch lange Ausbildungs- und lange Vollerwerbstätigkeitszeiten charakterisiert ist. Aber wir finden immerhin mehr als 30% Männer im Cluster 5, welches durch lange Teilzeiterwerbstätigkeit gekennzeichnet ist. Offensichtlich gibt es Männer, die über Teilzeitarbeitsformen Verläufe aufweisen, wie sie in der Regel eher für Frauen typisch sind. Es ist aber auch zu erkennen, daß es zwei Verlaufsformen gibt, bei denen das Geschlecht keine so große Rolle zu spielen scheint: die Kombination von kurzer Ausbildung mit Einmündung in eine Vollerwerbstätigkeit (Cluster 2) und die unspezifische Verknüpfung von Ein-, Aus-, Um- und Wiedereinstiegen (Cluster 6). Hier finden sich etwa gleich viele Männer und Frauen in den Clustern.

**Grafik 4: Verteilung von Männern und Frauen auf die einzelnen Cluster**



Um die insgesamt geschlechtsdifferente Verteilung erklären zu können, müssen die Lebensformen in die Analyse eingeführt werden: Leben mit einem Partner und Leben mit Kindern.

Zunächst wird daher die Kombination aus 'Optimal Matching' und Clusteranalyse nur für die Personen wiederholt, die mit Kindern leben, bzw. in deren hier betrachteten Lebensverlauf Kinder hineingeboren wurden. Dieses war bei 78 von den 129 Personen der Fall. Grafik 5 zeigt die Länge und auch die Lage der einzelnen Phasen im Lebensverlauf ab dem 17. Lebensjahr. Der rechte Teil der Grafik gibt die Verteilung der Geschlechtsvariable in jedem Cluster wieder.

Es entstehen unter dieser Bedingung sechs Cluster mit ähnlicher Phasenordnung wie in dem vorangegangenen Ordnungsversuch. Dabei fällt auf, daß die Cluster mit Drei-Phasen- (Cluster 3) und Zwei-Phasenstruktur (Cluster 4) wiederum ausschließlich mit Frauen besetzt sind. Dieses sind auch die Cluster, in denen Bildung in nur geringem Maße akkumuliert worden ist. Während die Geburt des ersten Kindes in allen Clustern im Mittel etwa bei 37 Jahren liegt, sind die Mütter und Väter der Kinder im Cluster 6 wesentlich jünger (33,5 Jahre).

Die Männer dagegen finden sich auch hier in dem Cluster, in dem in Abhängigkeit von langer Bildungsbeteiligung viel Vollzeiterwerbstätigkeit akkumuliert wird (Cluster 1). Unterbrochen werden deren Lebensverläufe lediglich durch den Einsatz bei der Bundeswehr. Insgesamt verteilen sich daher Männer und Frauen wiederum sehr unterschiedlich auf die Cluster.<sup>12)</sup>

Bemerkenswert ist aber, daß in den Clustern 1 und 2 Frauen zu finden sind, obgleich sich hier keine Auswirkungen der Geburt von Kindern auf den Erwerbsverlauf zeigen.<sup>13)</sup> Gleichzeitig lassen sich Männer in weiblich dominierten Clustern auffinden (Cluster 5 und 6). Hier arbeiten sie vielfach unter Teilzeitbedingungen bzw. weisen eine große Anzahl von Aus-, Um- und Wiedereinstiegen in das Erwerbsleben auf.

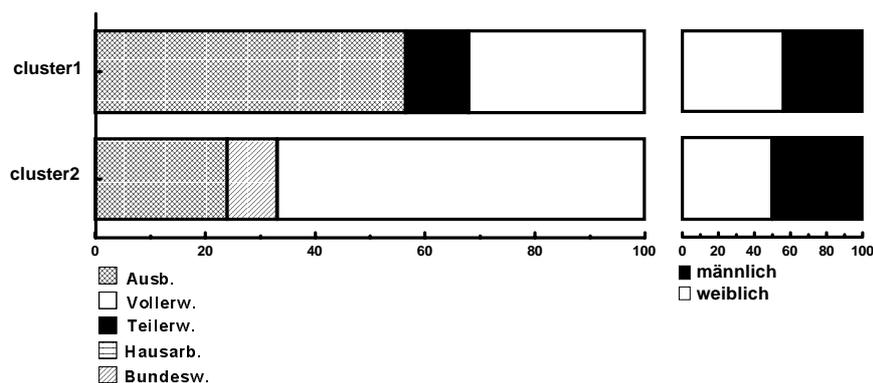
Im letzten Schritt ist das Verfahren auf Personen angewendet worden, die keine Kinder haben, jedoch mindestens sieben Jahre mit einem Partner zusammenleben bzw. gelebt haben. Es entstehen hierbei lediglich zwei Cluster. In Cluster 1 münden lange Ausbildungen über eine Teilzeitphase in Vollerwerbstätigkeit. In Cluster 2 folgt - abgesehen von kurzen Unterbrechungen durch Bundeswehrzeiten/Auslandsaufenthalt - nach einer Ausbildung die Tätigkeit als Vollzeiterwerb.

Wie im rechten Teil der Grafik 6 zu sehen, verteilen sich unter diesen Bedingungen Männer und Frauen absolut gleich auf die beiden Cluster. Es bestehen ganz

offensichtlich - trotz unterschiedlicher Phasentypik - keine Unterschiede zwischen Frauen und Männern hinsichtlich der quantitativen Besetzung dieser Cluster.<sup>14)</sup>

Fassen wir die Ergebnisse zusammen, so zeigt sich, daß Unterschiede zwischen den Erwerbsverläufen von Männern und Frauen im Falle langjähriger Partnerschaft (oder Ehe) nicht zu erkennen sind. Der Familienstand allein hat demnach keine Bedeutung für die Strukturierung dieser Verläufe. Erst mit der Geburt von Kindern treten geschlechtsspezifische Unterschiede auf. Das Kind wird nun zur bedeutendsten Weichenstellung für die Differenzierung der Berufsverläufe: Die Trennungslinie verläuft nun nicht mehr zwischen Männern und Frauen, sondern zwischen Frauen mit familiären Pflichten und Frauen und Männern ohne Familienbindung (vgl. dazu auch Lauterbach 1994; Schulze Buschhoff 1995).

**Grafik 6: Anordnung und Länge der Phasen bei Männern und Frauen ohne Kinder**



Das Zwei- bzw. Drei-Phasen-Modell existiert weiterhin, jedoch lediglich als eine mögliche Phasenanzordnung unter anderen. Quantitativ machen beide Modelle zusammen bei der Analyse aller Verläufe ca. 20% der Verläufe aus.

Neben diesen generellen Ergebnissen läßt das 'Optimal-Matching-Verfahren' aber noch differenziertere Betrachtungen zu. In den Grafiken 4 und 5 sind neben den starken Unterschieden zwischen den Geschlechtern ebenso gemischtgeschlechtliche Gruppen mit ähnlichen Verläufen zu erkennen. Grafik 5 zeigt, daß es Frauen mit Kindern gibt, die sich in eher erwerbsorientierten bzw. männlich dominierten Clustern finden (Cluster 1 und 2), wie es auch Männer gibt in eher weiblich dominierten Clustern (Cluster 5 und 6). Neben eher typisch männlichen und typisch weiblichen Verläufen gibt es Personen

beiderlei Geschlechts, deren Art, Länge und Anordnung der Erwerbssequenzen im betrachteten Lebensverlauf, trotz der Intervention einer Geburt, sich hinsichtlich der musterbildenden Abfolge der Sequenzen nicht voneinander unterscheiden. Häufig sind es gerade diese Personen, deren Verläufe den Erwartungen oder Hypothesen zunächst widersprechen. Diese Gruppe ist insgesamt nicht sehr groß, jedoch kann gerade hier Neues entdeckt werden. Vielleicht handelt es sich bei diesen Personen um Pioniere und Pionierinnen, die neue Lebensformen ausprobieren. Bei genügender Fallzahl können weitere quantitative Analysen folgen. Bei geringer Fallzahl können sie den Grundstock zu einem, mit qualitativen Methoden zu untersuchenden, Sample bilden. Letzteres auch dann, wenn sich in den quantitativen Daten keine Hinweise finden lassen, die die Anordnung bzw. Lage der Sequenzen erklären könnten.

## 5. Zusammenfassung

Es sollte deutlich geworden sein, daß die 'Optimal-Matching-Technik' sich sehr gut zum Vergleich von Verläufen eignet, da sie deren differenzierte Betrachtung ermöglicht. Dieses ist nötig vor allen Dingen bei Personen, deren Lebensverläufe sich als sehr heterogen darstellen, was in der Regel dann der Fall ist, wenn diese Personen in die zwei unterschiedlichen Bereiche Familie und Arbeitsmarkt eingebunden sind. Familien- und Erwerbsbereich können vielfach unterschiedliche Verbindungen aufweisen und führen daher häufig zu Anordnungen von quantitativ erhobenen Erwerbssequenzen, die sich nur schwer ordnen lassen. Gerade die Abfolge der Sequenzen aber ist es, die Aussagen über Chancen und Behinderungen von Personen oder Personengruppen ermöglicht. Während zum Beispiel Zwei-Phasen-Modelle den Abbruch der Erwerbskarriere anzeigen, kann ein Drei-Phasen-Modell ein Indikator für relativ gute Wiedereinstiegschancen sein. Verlaufsmuster mit vielen Aus-, Um- und Wiedereinstiegen mit häufigem Changieren zwischen Erwerbs- und Familienbereich können auf eine relativ neue - 'moderne' - Lebensführung verweisen, dieses umso mehr, je gleichberechtigter die Geschlechter in diesen Verlaufsformen vertreten sind. Welche Variablen für die unterschiedlichen Muster letztendlich verantwortlich gemacht werden können, ist durch einen eigenen Analyseschritt festzustellen.

Ein Vorteil der 'Optimal-Matching-Technik' zeigt sich auch darin, daß abweichend von quasiexperimentellen Designs, bei denen Versuchs- und Kontrollgruppe im Datensatz vorab definiert werden, hier zunächst keine derartigen Festlegungen getroffen werden müssen. Ob sich zum Beispiel Männer und Frauen hinsichtlich des untersuchten Erwerbsverlaufes voneinander unterscheiden, wird über den Ordnungsprozeß selber gesteuert. Große Unterschiede zwischen den Verläufen führen dann zu klar voneinander getrennten, geschlechtsspezifisch eindeutig festgelegten Gruppen mit charakteristischen

Verlaufsmustern. Gleichzeitig können auch die Personen bestimmt werden, die sich dieser Trennung widersetzen, seien es Männer in 'Frauenclustern' oder Frauen in 'Männerclustern'. Diese können den Ausgangspunkt für weitere Analysen bilden.

'Optimal-Matching' ist damit als ein exploratives Verfahren zu bezeichnen, welches in dem hier gezeigten Zusammenhang eingesetzt werden kann, um Strukturmuster von Erwerbsarbeitsbeteiligungen zu ermitteln, die in ihren konkreten Ausprägungen nicht ex ante hätten bestimmt werden können. Zu betonen ist allerdings, daß der Einsatz der oben beschriebenen Methode im Bereich kausaler Modellbildung, das heißt als konfirmatorisches Verfahren, problematisch ist, da Lebensverläufe hier retrospektiv in ihrer Gesamtheit betrachtet werden. Die Blickrichtung der Analyse folgt damit nicht der des handelnden Akteurs mit jeweils situativ unterschiedlichen Handlungsoptionen, sondern nimmt Lebensläufe von Akteuren nur über die realisierten Handlungsalternativen wahr. Dies führt zwangsläufig zu den bei Blossfeld und Rohwer (1995) diskutierten Problemen der Schätzung von Modellen, in denen die zeitliche Ordnung von Ursache und Effekten nicht mehr kontrolliert werden kann.

Am Schluß soll noch darauf verwiesen werden, daß das beschriebene Verfahren sehr kurze singuläre Sequenzen vernachlässigt. Kurze Sequenzen werden für den Ordnungsprozeß nur dann bedeutsam, wenn sie durch ihre Kombination Muster bilden, die mit anderen Mustern vergleichbar sind. So ist der Einfluß der etwa zweijährigen Bundeswehr- und Zivildienstzeit auf den Ordnungsprozeß sehr gering. Erst längere Zeiten schlagen sich als Ordnungskriterium nieder. Es war daher auch kein Problem, Männer und Frauen miteinander zu vergleichen, da diese ausschließlich Männer betreffenden Sonderzeiten den Ordnungsprozeß nicht nachhaltig beeinflussen. Allerdings ist dieser Prozeß abhängig von der Länge der Verläufe, das heißt die Frage der Relationierung spielt eine nicht zu vernachlässigende Rolle. Je kürzer ein Verlauf, desto eher können spezifische Ereignisse als Ordnungsgeber durchschlagen und damit den Vergleichsprozeß beeinflussen. Die Länge der Verläufe, die Auswahl der Ereignisse und die Berechnung der Transformationskosten verlangen daher viel Aufmerksamkeit.

## Anhang

Das Ziel des Vergleiches von Sequenzen ist, die *minimalen* Transformationskosten zu ermitteln. Es kann zum Beispiel sein, daß durch eine Kombination von löschen, einsetzen und ersetzen geringere Kosten entstehen als durch die Anwendung lediglich einer Aktion. Dieses zeigt sich im Vergleich der Sequenz 1 A B C D E F mit der Sequenz 2 B C D E F A. Werden die Elemente der Sequenz 1 vollständig durch die Elemente der Sequenz 2 ersetzt, so sind hierfür sechs Aktionen erforderlich. Offensichtlich aber sind die beiden Sequenzen sehr ähnlich, denn der Wert A in der

Sequenz 1 muß lediglich vorn gelöscht und dafür hinten eingesetzt werden. Ersetzungen finden nicht statt. Statt sechs Aktionen werden hier also nur 2 benötigt. 15)

Prinzipiell können zwei Sequenzen immer auf verschiedenen Wegen durch das Ersetzen, Löschen oder Einfügen von Elementen ineinander überführt werden: Betrachten wir drei verschiedene Möglichkeiten, die Sequenz GEMSE in die Sequenz BREMEN zu überführen: In den ersten beiden Fällen sind je vier Aktionen notwendig, im dritten Fall fünf. Welche Kosten sich hieraus ergeben, hängt von der Gewichtung der einzelnen Aktionen ab: Werden für alle drei Aktionen (Löschen, Ersetzen, Einfügen) jeweils Kosten von 1 berechnet, so ergeben sich für die Lösungen A und B Gesamtkosten von 4, für Lösung C Kosten von 5. Werden hingegen Kosten unterschiedlich definiert, beispielsweise Kosten von 1 für das Ersetzen und Kosten von 0,5 für Einfügen oder Löschen, so ergeben sich für Lösung A Kosten von 2,5, für Lösung B hingegen Kosten von 3,5 und für Lösung C Kosten von 4,5.

#### Lösung A

	G	E	M	S	E		
1.	<b>B</b>	G	E	M	S	E	füge B ein
2.	B	<b>R</b>	E	M	S	E	ersetze G durch R
3.	B	R	E	M	E		lösche S
4.	B	R	E	M	E	<b>N</b>	füge N ein

#### Lösung B

	G	E	M	S	E		
1.	<b>B</b>	E	M	S	E		ersetze G durch B
2.	B	<b>R</b>	E	M	S	E	füge R ein
3.	B	R	E	M	<b>E</b>	E	ersetze S durch E
4.	B	R	E	M	E	<b>N</b>	ersetze E durch N

#### Lösung C

	G	E	M	S	E		
1.	<b>B</b>	E	M	S	E		ersetze G durch B
2.	B	<b>R</b>	M	S	E		ersetze E durch R ein
3.	B	R	<b>E</b>	S	E		ersetze M durch E
4.	B	R	E	<b>M</b>	E		ersetze S durch M
5.	B	R	E	M	E	<b>N</b>	füge N ein

Zur rechnerischen Ermittlung der Levenshteindistanz schlagen Abbot und Hrycak in Anlehnung an Kruskal vor, eine iterative Prozedur auf eine Matrix anzuwenden, um den Transformationsweg mit den minimalen 'Kosten' zu ermitteln (s. Tabelle 2 sowie Abbot/Hrycak 1990: 178ff., Kruskal 1983: 23ff.).

Die Tabelle hat jeweils eine Zeile mehr als die Quellsequenz und eine Spalte mehr als die Zielsequenz. Die Quellsequenz (GEMSE) ist hierbei vor der ersten Spalte, die Zielsequenz (BREMEN) oberhalb der ersten Zeile der Tabelle eingetragen. Die minimale Distanz wird nun so berechnet, daß die Bewegungsrichtung in der Matrix von links oben nach rechts unten führt; das heißt in der letzten Zelle unten rechts finden wir die Minimalkosten für die Transformation. Ein Schritt nach rechts bedeutet dann das Einfügen eines Wertes, ein Schritt nach unten löschen eines Wertes und ein diagonalen Schritt (nach unten und rechts) 'ersetzen' eines Wertes. 'Einsetzen' ist immer assoziiert mit Aktionen in der Spalte und bedeutet, daß in die Zielsequenz Werte eingesetzt werden. 'Löschen' dagegen ist immer assoziiert mit Aktionen in der Zeile und bedeutet, daß in der Quellsequenz Werte gelöscht werden. 'Ersetzen' ist definiert als die Übernahme des Wertes der Zielsequenz in die Quellsequenz, die keine Kosten verursacht, wenn Quell- und Zielsequenz an dieser Stelle gleich sind.<sup>16)</sup> Die jeweiligen Kosten dieser Aktionen werden in den Zellen der Tabelle notiert.

Kosten für Ersetzen ↘	Kosten für Löschen ↓
Kosten für Einfügen →	minimale Gesamtkosten

Für die Positionen in den einzelnen Zellen gilt: Die Pfeile in der Tabelle zeigen die Wege an, die mit den geringsten Kosten durch die Matrix führen. Es wird hiermit deutlich, daß die oben beschriebenen Lösungen 1 und 2 zu den Transformationswegen gehören, die bei den vorgegebenen Gewichtungen die Kosten minimal halten: Werden für das Löschen, Ersetzen und Einfügen jeweils Kosten von 1 definiert, verursacht die Übertragung von GEMSE nach BREMEN Kosten in Höhe von 4.

Tabelle 2: Matrix zur Ermittlung der Levenshteindistanz

	Ziel- sequenz ⇨	B	R	E	M	E	N
Quell- sequenz ⇩	Start - 0 → ↓	-	-	-	-	-	-
G	- 1 - 1	1 1 1 1 →	1 1 1 2 ↘	1 1 1 3	1 1 1 4	1 1 1 5	1 1 1 6
E	- 1 - 2	1 1 1 2	1 1 1 2	0 1 1 2 ↘	1 1 1 3	0 1 1 4	1 1 1 5
M	- 1 - 3	1 1 1 3	1 1 1 3	1 1 1 3	0 1 1 2 ↘ ↓	1 1 1 3	1 1 1 5
S	- 1 - 4	1 1 1 4	1 1 1 4	1 1 1 4	1 1 1 3 ↘	1 1 1 3 ↘	1 1 1 4
E	- 1 - 5	1 1 1 5	1 1 1 5	1 1 1 5	1 1 1 4	0 1 1 3 →	1 1 1 4 Ziel

Zur Standardisierung der Kosten bei unterschiedlicher Länge der Sequenzen werden bei dem von Abbott und Hrycak vorgeschlagenen Verfahren die über die Tabelle ermittelten Kosten durch die Länge der längeren Sequenz dividiert. Dies ist im vorliegenden Fall die Sequenz BREMEN mit einer Länge von 6, so daß sich eine Levenshteindistanz von  $4/6 \approx 0,67$  ergibt.

### Korrespondenzadresse

Christian Erzberger, Gerald Prein  
 Universität Bremen, Sfb 186  
 Wiener Str. - FVG-West  
 Postfach 330440  
 D-28334 Bremen

### Anmerkungen

- 1) In den beiden großen repräsentativen Datensätzen des Bundesinstituts für Berufsbildung und des Instituts für Arbeitsmarkt- und Berufsforschung "Qualifikation und Be-

rufsverlauf" mit 26361 befragten Personen 1985 und 23476 befragten Personen 1991/92 sind die Zielpersonen über ihre Beteiligung am Erwerbsleben definiert. Personen, die zum Zeitpunkt der Befragung arbeitslos sind, geringverdienend arbeiten oder einer Haus- bzw. Familienarbeit nachgehen, werden in die Untersuchung nicht aufgenommen. Ebenso enthält die IAB-Beschäftigtenstichprobe (Meldungen der Arbeitgeber gegenüber der gesetzlichen Kranken- und Rentenversicherung und der Bundesanstalt für Arbeit) mit 427022 Personen keinerlei Hinweise über Tätigkeiten unterhalb der Versicherungspflichtgrenze und ebenso nur sehr wenige Hinweise über familiäre Ereignisse. In diesen Datensätzen finden sich zum einen keine Angaben von Personen, die nicht entsprechend der Zielgruppendefinition in den Arbeitsmarkt eingebunden sind, und zum anderen unterliegen die befragten Personen einem eingeschränkten Betrachtungsfokus, so daß sich für sie möglicherweise Muster von Aus- und Wiedereinstiegen - also 'Spuren' - ergeben, die nur sehr schwer zu interpretieren sind.

2) Das Beispiel stammt aus dem Teilprojekt "Statussequenzen von Frauen zwischen Erwerbsarbeit und Familie" des Sonderforschungsbereichs 186. Unter der Leitung von Helga Krüger wurden dort 220 Frauen in fünf typischen Frauenberufen über ihre Lebens- und Erwerbsbiographie befragt (vgl. Born/Krüger/Lorenz-Meyer 1996).

3) Auf die hierzu notwendigen Modifikationen wird weiter unten eingegangen; hier soll zunächst nur prinzipiell ein vielen Problemstellungen gemeinsames Verfahren der Mustererkennung dargestellt werden.

4) Der in der Literatur häufig vorgeschlagene Weg zur Berechnung der Minimalkosten wird ausführlich im Anhang dargestellt und an einem Beispiel erläutert.

5) So läßt sich der in Graphik 1 dargestellte Lebensverlauf in Form einer Sequenz kodieren. Definieren wir beispielsweise, daß Erwerbstätigkeit im erlernten Beruf den Kode E, andere versicherungspflichtige Tätigkeit den Kode V, nicht versicherungspflichtige Tätigkeit den Kode N und Weiterbildung den Kode W erhält. Wenn wir nur an der Abfolge der Zustände interessiert sind, läßt sich der Verlauf (ohne Berücksichtigung der Lücken) durch die Sequenz EVNNVWVWVV darstellen. Soll bei der Kodierung auch die Dauer der unterschiedlichen Phasen berücksichtigt werden, ist dies etwa dadurch möglich, daß für ein bestimmtes Zeitintervall (zum Beispiel ein Jahr) die vorherrschende Tätigkeit kodiert wird, was hier etwa durch die Sequenz EEVNNNNNNNEEEEEEEEEEEWEEEEEEEEEE dargestellt werden könnte. Letztere Art der Kodierung wird im folgenden benutzt, da wir davon ausgehen, daß nicht nur die Abfolge, sondern auch die zeitliche Bedeutung von Lebensphasen für unsere Analysen von großer Bedeutung ist.

6) Weitere Anwendungsgebiete beschreiben Buchmann/Sacchi (1995) für die Ordnung der Abfolge von ausgeübten Berufen, Abbot/Hrycak (1990) für das Auffinden von Mustern in Musikerkarrieren im Deutschland des 18. Jahrhunderts und Abbot/Forrest (1986) für die Veränderung von Tanzsequenzen bestimmter Englischer Tänze zwischen 1867 und 1945.

7) Dieses Teilprojekt des Sonderforschungsbereichs 186 wird seit 1994 unter der Leitung von Helga Krüger durchgeführt.

8) Das C-Programm kann bei den Autoren kostenfrei angefordert werden. Inzwischen ermöglicht allerdings auch das Programm TDA 6 von Götz Rohwer die Berechnung entsprechender Matrizen.

9) Es entstehen sieben Cluster, aber in diesem siebten Cluster befinden sich zwei Fälle, die als Ausreißer ausgeschlossen wurden. Verglichen wurden die hierarchischen Clusterverfahren 'Complete Linkage', 'Average Linkage', 'Ward' und 'Flexible Beta'. Auch wenn 'Flexible Beta' eine den Daten nicht angemessene Clustermethode darstellt, so wurde sie - aus Ermangelung adäquaterer Verfahren - dennoch gewählt, da sie die besten Ergebnisse erbrachte: die Gruppen waren in sich sehr homogen und gleichzeitig am schärfsten gegeneinander abgesetzt. Über die Anzahl der Cluster wurde nach den Werten der Inhomogenität der einzelnen Agglomerationschritte über die Inspektion des Distanzkoeffizienten entschieden.

10) Als Beispiel sind hier die Verläufe der Personen im Cluster 4 dargestellt:

```

AAAAAAVVVVVHHHHHHHHHH
AAAAAAVVVVVHHHHH
VVVVVVVVVHHHHHHHHHHH
VVVVVVVVVHHHHHHHHHHHHHHHH
AAAVVVVVVHHHHHHHHHHHHHHHH
AAAAAVVVVHHHHHHHHHHHH
AAHHHHHHHHHHHHHHHH
AAVVVHHHHHHHHHHHHHHHH
AAAAVVVVTTTTHHHHHHHHHHH
AAAAAAHHVVVHHHHHHHHHHHH
AAAAAAAAAAVTTTHHHH
AAVVVVVVVHHHHHHHHHHHHT

```

11) Zur Beschreibung der Stärke der Beziehung zwischen den nominalen Variablen wurde Cramer's V herangezogen. Der Koeffizient beträgt in diesem Fall bei  $\chi^2 = 32,08; V = 0,5$ .

12) Cramer's V weist bei einem  $\chi^2 = 33,7$  mit 0,65 auf eine relativ starke Assoziation zwischen den Clustern mit ihrer typischen Phasenstrukturierung und dem Geschlecht hin.

13) Es kommen geringfügige Zeiten von Hausarbeit in diesen Clustern vor, jedoch wird davon die Phasentypik nicht berührt (Anteil unter 5%).

14) Cramer's V weist mit 0,06 ( $\chi^2 = 1,86$ ) darauf hin, daß keine Beziehung zwischen Clustertypik und Geschlecht besteht.

15) Gängige Verfahren der Clusteranalyse sind zur Ermittlung solcher Distanzen nicht geeignet. Diese vergleichen die Ausprägungen von einzelnen Variablen miteinander und verstehen die Aneinanderreihung von Variablen nicht als zusammengehörige Zeitreihe. Das heißt, gegeneinander versetzte Reihen können von diesen Clusteranalysen nicht adäquat analysiert werden. Durch die Berechnung der Distanz zwischen diesen beiden Ver-

läufen nach dem Tanimoto-Koeffizienten für nominalskalierte Variablen erhalte man als Ergebnis einen maximalen Unterschied.

16) Die Bezeichnung der drei möglichen Aktionen beruht auf Konvention.

## Literatur

Abbot, A./Hrycak, A., 1990: Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers. *American Journal of Sociology* 1: 144-185.

Abbot, A./Forrest, J., 1986: Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History* XVI: 471-494.

Blossfeld, H.-P., 1990: Berufsverläufe und Arbeitsmarktprozesse. Ergebnisse sozialstruktureller Längsschnittuntersuchungen. S. 118-145 in: K.U. Mayer (Hrsg.), *Lebensverläufe und sozialer Wandel*. Sonderband der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Opladen: Westdeutscher Verlag.

Blossfeld, H.-P./Drobnic, S./Rohwer, G., 1995: Employment Patterns: A Crossroad Between Class and Gender. A Long-term Longitudinal Study of Spouses in West Germany. Vortrag im Rahmen des 4. Symposiums des Sfb 186: Lebenslaufpolitik - Institutionen und Statusmanagement am 21./22. September 1995 in Bremen.

Blossfeld, H.-P./Rohwer, G., 1995: *Techniques of Event History Modeling*. Mahwah. Lawrence Erlbaum Associates.

Born, C., 1995: Modernisierungsgap - Angleichung im Wandel geschlechtsspezifischer Lebensführungen? Vortrag im Rahmen des 4. Symposiums des Sfb 186: Lebenslaufpolitik - Institutionen und Statusmanagement am 21./22. September 1995 in Bremen.

Born, C./Krüger, H./Lorenz-Meyer, D., 1996: Der unentdeckte Wandel. Annäherung an das Verhältnis von Struktur und Norm im weiblichen Lebenslauf. Berlin: edition sigma.

Buchmann, M./Sacchi, S., 1995: Mehrdimensionale Klassifikation beruflicher Verlaufsdaten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 3: 413-442.

Erzberger, C., 1993: Erwerbsarbeit im Eheleben. Männlicher und weiblicher Erwerbsverlauf zwischen Dependenz und Unabhängigkeit. Sfb186, Arbeitspapier Nr. 16. Universität Bremen.

Hagestad, G./Neugarten, B., 1985: Age and the Life Course. p. 35-61 in: R. Binstock./E. Shanas (eds.), *Handbook of Aging and the Social Sciences*. New York: Van Nostrand Reinhold Company.

Jurczyk, K./Rerrich, M.S., 1993: *Die Arbeit des Alltags*. Beiträge zu einer Soziologie der täglichen Lebensführung. Freiburg: Lambertus.

Kohli, M., 1985: Die Institutionalisierung des Lebenslaufs. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 1: 1-29.

Krüger, H., 1993: Bilanz des Lebenslaufs: Zwischen sozialer Strukturiertheit und biographischer Selbstdeutung. *Soziale Welt* 3: 375-391.