

## Die Auswertung inhaltsanalytischer Kategorien mit Latent-Class Modellen

Tarnai, Christian; Rost, Jürgen

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:

Tarnai, C., & Rost, J. (1991). Die Auswertung inhaltsanalytischer Kategorien mit Latent-Class Modellen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 28, 75-87. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-202484>

### 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.

Die Auswertung inhaltsanalytischer Kategorien mit Latent-Class Modellen  
von Christian Tarnai und Jürgen Rost<sup>1</sup>

### 1. Einleitung

Innerhalb der Anwendung von Methoden der Inhaltsanalyse ist mit der validen und reliablen Zuordnung von Inhalten zu Kategorien der erste Schritt zu einer Überprüfung von Forschungshypothesen geleistet. In vielen Fällen ist es auch von Interesse, welche Zusammenhänge zwischen den Kategorien bestehen und inwieweit diese eine spezifische (latente) Struktur von Inhalten repräsentieren. Eine Auswertungsstrategie ist die Zerlegung der vorliegenden mehrdimensionalen Kontingenztabelle. Diese wird aber sehr schnell unübersichtlich. Es werden daher komplexere Verfahren zur Abbildung der Zusammenhänge von Inhaltskategorien benötigt. Die Latent-Class-Analyse (LCA) von *Lazarsfeld* (1950) ist hier eine geeignete Methode. Die LCA steht für eine Vielzahl von Modellen (*Lazarsfeld/Henry* 1968), die durch die Entwicklung von Rechenprogrammen auch praktisch anwendbar geworden sind (vgl. *Formann* 1984; *Langeheine* 1988). Die Erweiterung der LCA durch *Rost* (1988) gestattet die Einbeziehung ordinal abgestufter Variablen in die Analysen. Mit LACORD (Latent Class Analysis for Ordinal Variables; *Rost* 1990a) steht ein Rechenprogramm zur Verfügung, das die gleichzeitige Berücksichtigung von Variablen mit dichotomen, polytomen und ordinalen Abstufungen erlaubt, wodurch auch für die Auswertung von Inhaltsanalysen neuartige Anwendungen ermöglicht werden.

Generell werden mittels Latent-Class-Analyse nicht direkt beobachtbare (latente) Klassen von Objekten (Personen) anhand von manifesten Variablen abgebildet. Die einzelnen Klassen sind die verschiedenen Stufen der latenten Variablen, die für die Zusammenhänge der beobachtbaren (manifesten) Variablen verantwortlich sind. Die LCA berücksichtigt nicht nur die paarweisen Assoziationen der manifesten Variablen, sondern auch die Gesamtheit der Zusammenhänge höherer Ordnung. Hierin unterscheidet sie sich von Verfahren, die häufig mit derselben Zielrichtung angewandt werden, wie beispielsweise Clusteranalyse und Faktorenanalyse.

Im folgenden sollen die Grundzüge der LCA für ordinale Daten dargestellt werden.<sup>2</sup> Dies geschieht im Zusammenhang mit einer speziellen Fragestellung der Untersuchung von

1 Anschriften der Autoren: Dr. Christian *Tarnai*, Institut für Sozialpädagogik, Weiterbildung und Empirische Pädagogik, Universität Münster, Georgskommende 33,4400 Münster.  
PD Dr. Jürgen *Rost*, Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel, Ols-  
hausenstraße 62,2300 Kiel.

2 Vgl. auch die Darstellung des Modells im Beitrag von *Georg/Rost* (1991) in diesem Heft.

*Tarnai/Bos* (1989, 1990), in der die Definitionen und Operationalisierungen des Begriffs Emanzipation durch Studierende der Pädagogik analysiert werden.

**2. Empirisches Beispiel**

In der Untersuchung von *Tarnai/Bos* (1989, 1990) wurde Studierenden der Pädagogik (N = 120) folgende Aufgabe vorgelegt: "Versuchen Sie eine Definition des Begriffes 'Emanzipation' nach Ihrem eigenen Verständnis und geben Sie dann Indikatoren an, um die Begriffsinhalte empirisch bestimmbar machen zu können".

Neben der Analyse der durch die Studierenden gegebenen Definitionen und Operationalisierungen auf der Grundlage der von *Tarnai* (1989a) entwickelten Inhaltskategorien ist eine interessierende Frage, inwieweit bei den Angaben explizit *Frauen, Geschlechter* oder *Männer* genannt werden.<sup>3</sup>

Eine dichotome Betrachtung der drei Inhaltskategorien nach 'genannt' und 'nicht genannt' bedeutet einen Verzicht auf Information, da gerade das mehrmalige Ansprechen der Kategorien Ausdruck verschiedener Sichtweisen (Gewichtungen) des Begriffes Emanzipation sein können. Es liegt somit ein typischer Anwendungsfall der LCA für ordinale Daten vor. Obwohl man Häufigkeiten generell als absolut skaliert auffassen kann, interessieren hier die Übergänge von ein, zwei und mehr Nennungen im Sinne einer Ordinalskala. Für die beabsichtigte Analyse werden vier Stufen gebildet, von keiner bis drei (und mehr) Nennungen. Die Häufigkeitsverteilung der Nennungen in diesen drei Inhaltskategorien ist in Tabelle 1 in Form von prozentualen Nennungshäufigkeiten wiedergegeben.

**Tabelle 1:** Prozentuale Nennungshäufigkeiten in den drei Inhaltskategorien: Frauen, Geschlechter, Männer (N = 120)

		Frauen	Geschlechter	Männer
Nennungs-	0	48	88	60
häufigkeiten	1	17	10	18
	2	9	1	8
	3	26	1	14

<sup>3</sup> Diese Kategorien beinhalten alle eindeutig zuordenbaren Begriffe wie beispielsweise weiblich, männlich, Mädchen oder Junge.

Es zeigt sich, daß bei den beiden Inhaltskategorien *Frauen* und *Männer* jeweils eine bimodale Häufigkeitsverteilung vorliegt, d.h. die häufigste Nennungsanzahl ist in jedem Fall 0 (keine Nennung), jedoch tritt die Nennungshäufigkeit 3 häufiger auf als die Nennungshäufigkeit 2. Eine solche bimodale Häufigkeitsverteilung ist bei der Annahme einer ordinal skalierten Variablen, wie es Nennungshäufigkeiten eigentlich sein müßten, ungewöhnlich. Eine naheliegende Annahme ist, daß die beobachtete Häufigkeitsverteilung, wie sie sich in Tabelle 1 darstellt, aufgrund einer heterogenen Personenstichprobe zustande gekommen ist. Die Methode der Latent-Class-Analyse versucht nun, die beobachteten Kontingenzen durch die Annahme zu erklären, daß die beobachteten Daten eine Mischung aus verschiedenen Teilstichproben darstellen. Im ersten Schritt der Analyse muß daher festgestellt werden, welche Anzahl von latenten Klassen von Personen den Daten am besten entspricht. Dies geschieht durch Berechnung von Statistiken für verschiedene Anzahlen von Klassen, hier z.B. für 1 bis 5 latente Klassen.

Eine inferenzstatistische Beurteilung, welche Klassenanzahl den Daten am angemessensten ist, kann im Prinzip mit Hilfe von Likelihoodquotienten festgestellt werden. Praktisch scheidet dieses Vorgehen jedoch oft daran, daß die asymptotischen Eigenschaften von Likelihoodquotiententests aufgrund zu geringer Häufigkeiten in der Kontingenztabelle nicht erfüllt sind. So hat z.B. die hier vorliegende Kontingenztabelle  $4^3 = 64$  Zellen, von denen jedoch 44 Zellen beobachtete Häufigkeiten von 0 aufweisen. Damit ist eine inferenzstatistische Absicherung der Klassenanzahl mit einem Likelihoodquotiententest nicht möglich. Einen Ausweg bietet der sogenannte AIC-Index (Akaiikes Information Criterion), der eine Funktion von Likelihood und Anzahl der Parameter eines Modells ist (vgl. Bozdogan 1987)<sup>4</sup>

$$AIC = -2 \log L + 2N_{\text{parm}} \quad (1)$$

Der AIC-Index setzt also für eine bestimmte Klassenanzahl die Likelihood der Daten mit der Anzahl der Parameter in Beziehung, die für die Erreichung dieses Likelihoodwertes benötigt wird. Während die Likelihood mit zunehmender Klassenanzahl zwangsläufig ansteigt, verändert sich der AIC-Index nicht unbedingt in einer Richtung. Er wird vielmehr immer dann kleiner, wenn der Anstieg der Likelihood größer ist als die zusätzliche Anzahl von Parametern. Steigt die Likelihood im Vergleich zur aufgewendeten Parameteranzahl unverhältnismäßig langsam, so wird der AIC-Index wiederum größer. Das bedeutet, daß der kleinste AIC-Index das am relativ besten passende Modell ausweist. Für die vorliegenden Daten sind die AIC-Indizes in Tabelle 2 wiedergegeben.

<sup>4</sup> Für einen Vergleich des AIC-Index mit anderen Indizes siehe *Read/Cressie* (1988).

Tabelle 2: Ergebnis der Latent-Class-Analyse für unterschiedliche Anzahl von Klassen: log-Likelihood und AIC-Index

	log-Likelihood	Nparm	AIC-Index
1	-327.526	5	672.051
2	-277.599	19	593.197
3	-266.415	29	590.829
4	-263.813	39	605.626
5	-263.138	49	624.276

Anzahl der Klassen

log-Likelihood des saturierten Modells: -258.447

Es zeigt sich, daß die 3-Klassenlösung den geringsten AIC-Index hat und daher die beobachteten Daten am besten und sparsamsten beschreibt. Die Ergebnisse der 3-Klassenlösung sind in Tabelle 3 dargestellt.

Tabelle 3: Bedingte Wahrscheinlichkeiten der Antwortkategorien für die 3-Klassenlösung der Latent-Class-Analyse

Verteilung der Befragten auf:		Inhaltskategorien:			
		Frauen	Geschlechter	Männer	
1.Klasse 22,2%	Anzahl	*0*	0.000	0.847	0.010
	der	*1*	0.000	0.153	0.016
	Nennungen	*2*	0.155	0.000	0.337
		*3*	0.845	0.000	0.637
2.Klasse 15,9%	Anzahl	*0*	0.003	0.743	0.002
	der	*1*	0.453	0.152	0.998
	Nennungen	*2*	0.199	0.052	0.000
		*3*	0.346	0.052	0.000
3.Klasse 61,8%	Anzahl	*0*	0.781	0.933	0.966
	der	*1*	0.153	0.067	0.034
	Nennungen	*2*	0.041	0.000	0.000
		*3*	0.025	0.000	0.000



Die erste Klasse wird von 22,2% aller Personen ( $N = 120$ ) gebildet, welche die Inhaltskategorie *Frauen* in 84,5% aller Fälle mindestens dreimal ansprechen und die Kategorie *Männer* in 63,7% der Fälle mindestens dreimal nennen. Die Personen der zweiten Klasse (15,9% von 120 Studierenden) sprechen die Kategorie *Frauen* auf jeden Fall an, jedoch mit unterschiedlicher Häufigkeit, d.h. einmal, zweimal oder dreimal mit Wahrscheinlichkeiten zwischen  $p=0,453$  und  $p=0,199$ . Anders sieht es in dieser Klasse mit der Inhaltskategorie *Männer* aus, die fast ausschließlich einmal angesprochen wird ( $p=0,998$ ). Die dritte Klasse setzt sich aus der Mehrheit der Studierenden (61,8%) zusammen, die sich dadurch auszeichnen, daß sie in allen drei Inhaltskategorien überwiegend keine Nennungen aufweisen. Hier ist auch bei der Inhaltskategorie *Geschlechter*, die in den ersten beiden Klassen zumindest in 15% bzw. 25% der Fälle genannt wurde, die geringste Auftretenshäufigkeit zu sehen.

Zusammenfassend lassen sich diese drei Klassen etwa dahingehend interpretieren, daß die erste Personenklasse den Begriff Emanzipation sehr stark unter dem Blickwinkel der Gemeinsamkeit oder der Abgrenzung von Mann und Frau betrachtet, was sich in den hohen Nennungshäufigkeiten dieser Inhaltskategorien ausdrückt. Die Personen der zweiten Klasse sprechen Aspekte der Emanzipation mehr im Zusammenhang mit *Frauen* als mit *Männer* an. Im Gegensatz zu den beiden ersten Klassen stehen die Personen der dritten Klasse, die möglicherweise den Emanzipationsbegriff gänzlich unabhängig von der Geschlechtsrollenproblematik zu definieren versuchen. Interessant ist es daher, in einem nächsten Analyseschritt das Geschlecht der befragten Studierenden einzubeziehen. Dieses Merkmal wird als vierte Variable berücksichtigt. Ebenfalls anhand des AIC-Indexes wird festgestellt, daß die 3-Klassenlösung den Daten am angemessensten ist (AIC-Index = 735,8). Tabelle 4 gibt die entsprechenden Ergebnisse wieder.

Das Ergebnis für die Zusammensetzung der latenten Klassen ist mit der Lösung ohne Berücksichtigung des Personengeschlechtes annähernd gleich. Anhand der Auftretenswahrscheinlichkeiten für das Personengeschlecht zeigt sich, daß in der ersten Klasse, in der der Emanzipationsbegriff mit extrem häufigen Nennungen der Inhaltskategorien *Frauen* und *Männer* definiert wurde, der Anteil weiblicher Personen überwiegt. Diese Klasse umfaßt 88% Studentinnen und 12% Studenten. In beiden anderen Klassen beträgt das Verhältnis in etwa 2:1, was bei der gegebenen Zusammensetzung der Stichprobe von 71% Studentinnen und im Vergleich zur ersten Klasse als eine ausgeglichene Zusammensetzung zu bezeichnen ist. Es sind somit überproportional viele Studentinnen, die im Zusammenhang mit dem Begriff der Emanzipation *Frauen* und *Männer* erwähnen. Neben diesem Gewinn für die Interpretation der Daten durch die Einbeziehung des Geschlechts der Personen fällt vor allem die Stabilität der 3-Klassenlösung auf. Die Einbeziehung einer vierten Variablen in die Analyse hat somit nichts an der markanten Struktur der drei latenten Klassen und ihrer relativen Größen zueinander geändert.



**Tabelle 4:** Bedingte Wahrscheinlichkeiten der Antwortkategorien für die 3-Klassenlösung der Latent-Class-Analyse mit dem Personengeschlecht

Verteilung der Befragten auf:			Inhaltskategorien:			Personengeschlecht
			Frauen	Geschlechter	Männer	
1.Klasse 21,8%	Anzahl	*0*	0.000	0.847	0.000	0.115 männl.
		*1*	0.000	0.153	0.005	0.885 weibl.
	Nennungen	*2*	0.154	0.000	0.344	
		*3*	0.846	0.000	0.650	
2.Klasse 20,4%	Anzahl	*0*	0.001	0.774	0.203	0.319 männl.
		*1*	0.427	0.144	0.796	0.681 weibl.
	Nennungen	*2*	0.209	0.041	0.000	
		*3*	0.363	0.041	0.000	
3.Klasse 57,8%	Anzahl	*0*	0.836	0.936	0.966	0.334 männl.
		*1*	0.138	0.064	0.034	0.666 weibl.
	Nennungen	*2*	0.027	0.000	0.000	
		*3*	0.000	0.000	0.000	

### 3. Restriktion der LCA für ordinale Daten

Die bisherigen Ergebnisse wurden mit Hilfe der normalen Latent-Class-Analyse gewonnen und beinhalten keinerlei Restriktionen oder Annahmen bezüglich der Ordnung der vier Antwortkategorien, die hier Nennungshäufigkeiten darstellen. Dadurch enthält das Modell sehr viele Parameter, nämlich im gegebenen Beispiel 10 unabhängige Kategorienwahrscheinlichkeiten für jede Klasse und 2 unabhängige Klassengrößenparameter, zusammen 32 unabhängige Modellparameter.

Die Idee von LACORD (Latent-Class-Analyse für ordinale Daten, Rost 1990a) besteht nun darin, daß man durch die Berücksichtigung der Skalenqualität der Antwortvariablen Parameter spart. Der Schlüssel für die Einführung geeigneter Restriktionen liegt im Konzept der Schwellenwahrscheinlichkeiten. Diese lassen sich nur bei ordinal skalierten Variablen definieren, denn sie bezeichnen die Übergangswahrscheinlichkeit von einer Antwortkategorie zur darauffolgenden:



$$Sw = p_x / (p_x + p_{x-1}) \quad (2)$$

wobei  $p_x$  die Wahrscheinlichkeit ist, mit der Antwortstufe  $x$  zu antworten, d.h. die Nennungshäufigkeit  $x$  zu haben, und  $p_{x-1}$  die Antwortwahrscheinlichkeit der vorhergehenden Stufe. Das Ergebnis der Analyse könnte für das vorliegende Datenbeispiel folgendermaßen aussehen:

$p_x$	0.2	0.5	0.2	0.1
$x$	0	1	2	3
$Sw$	.71	.29	.33	

In diesem Beispiel ist der Übergang von Stufe 0 zu Stufe 1 wahrscheinlicher als der von Stufe 1 zu 2 ( $Sw_{01} = 0,71$  und  $Sw_{12} = 0,29$ ). Im Kontext der Inhaltsanalyse heißt dies, daß der durch die Kategorie repräsentierte Inhalt mit hoher Wahrscheinlichkeit einmal genannt wird, aber eine zweite Nennung weniger wahrscheinlich ist. Die weitere Schwellenwahrscheinlichkeit ( $Sw_{23} = 0,33$ ) offenbart, daß es etwas wahrscheinlicher ist, bei zwei Nennungen noch eine dritte zu machen als bei nur einer Nennung eine zweite ( $Sw_{12} = 0,29$ ).

Das Zustandekommen der Schwellenwahrscheinlichkeiten wird auf zusammenwirkende Komponenten zurückgeführt, die - je nach Modell - von der Inhaltskategorie  $i$ , der latenten Klasse  $g$  und der Antwortstufe  $x$  abhängen.<sup>5</sup> Die additive Zerlegung der  $Sw$  wird wie beim Rasch-Modell nicht für die Wahrscheinlichkeiten, sondern für die logistisch transformierten Parameter durchgeführt:

$$Sw = \exp(\beta_{gix}) / (1 + \exp(\beta_{gix})) \quad (3)$$

Die Art der Zerlegung von  $\beta_{gix}$  ergibt die acht verschiedenen Modelle von LACORD (vgl. Rost 1990a, Georg/Rost 1991 in diesem Heft). Im vorliegenden Fall kommt nur eine Art der Zerlegung infrage, da die verwendeten Variablen unterschiedliche Stufenzahl haben (Modell 4 von LACORD). In diesem Modell wird angenommen, daß sich jede Schwellenwahrscheinlichkeit aus zwei Größen zusammensetzt, nämlich zum einen aus der Leichtigkeit, mit der eine Inhaltskategorie in einer latenten Klasse genannt wird. Dieser Parameter wird als *Lokationsparameter*  $\mu$  bezeichnet und charakterisiert jeweils eine Inhaltskategorie in einer latenten Klasse. Die zweite Komponente, die zur Schwellenwahrscheinlichkeit beiträgt, ist die *Leichtigkeit einer bestimmten Schwelle*  $\tau$  (tau) bei einem Item, welche unabhängig von der betreffenden latenten Klasse ist. Diese Parameter werden Schwellenparameter genannt und charakterisieren jede Schwelle bei jedem Item unabhängig von der Klasse. Die Parametrisierung der Schwellenwahrscheinlichkeiten sieht für dieses Modell wie folgt aus:

<sup>5</sup> Bei den meisten anderen Anwendungen des Modells sind die beobachteten Variablen  $i$  die Items eines Fragebogens oder ähnliches.



$$\beta_{gix} = \mu_{ig} + \tau_{ix} \text{ und } \sum_x \tau_{ix} = 0 \quad (4)$$

Der Lokationsparameter  $\mu_{ig}$  ist infolge der Restriktion  $\sum_x \tau_{ix} = 0$  die mittlere Leichtigkeit aller Schwellen einer Klasse und daher als Leichtigkeitsparameter der Inhaltskategorie  $i$  für die Personen der latenten Klasse  $g$  interpretierbar. Der Parameter  $\tau_{ix}$  ist der Leichtigkeitsparameter der Antwortstufe  $x$  in der Inhaltskategorie  $i$ .

Wendet man dieses Modell auf das gegebene Datenbeispiel an, so zeigt sich zunächst, daß der AIC-Index niedriger ist als für das unrestringierte Latent-Class Modell (716,0 gegenüber 735,8). Einem geringfügigen Abfall der Likelihoodfunktion steht die drastische Ersparnis von immerhin 12 Modellparametern gegenüber, was zu diesem Resultat führt. Die 20 Modellparameter setzen sich wie folgt zusammen: es gibt für jede Inhaltskategorie und das Personengeschlecht in jeder latenten Klasse eine Leichtigkeit (12 Parameter) und einen Parameter für jede Schwelle bei jeder Inhaltskategorie (6 Parameter, da jeweils eine der Schwellen vom Lokationsparameter abhängig ist). Hinzu kommen 2 unabhängige Klassengrößenparameter, welche die Verteilung der Befragten auf die Klassen bestimmen.

Die inhaltliche Interpretation der drei latenten Klassen unterscheidet sich nicht wesentlich von dem Ergebnis der unrestringierten LCA, d.h. die Antwortwahrscheinlichkeiten und relativen Klassengrößen sind nahezu dieselben geblieben (vgl. Tabelle 5). Geht man bereits bei der ersten Analyse der drei Inhaltskategorien von restringierteren Modellen aus, wie dies in der Arbeit von *Tarnai* (1989b) geschieht, so ergibt sich mit und ohne Berücksichtigung des Personengeschlechts als Kriteriumsvariable eine noch stabilere Struktur der Inhaltskategorien als bisher aufgezeigt wurde.

**Tabelle 5:** Bedingte Wahrscheinlichkeiten der Antwortkategorien für die 3-Klassenlösung der Latent-Class-Analyse, Modell 4

Verteilung der Befragten auf:		Inhaltskategorien:			Personengeschlecht	
		Frauen	Geschlechter	Männer		
1.Klasse 23,3%	Anzahl	*0*	0.000	0.867	0.001	0.118 männl.
	der	*1*	0.014	0.119	0.094	0.882 weibl.
	Nennungen	*2*	0.107	0.008	0.296	
		*3*	0.878	0.008	0.610	
2.Klasse 21,2%	Anzahl	*0*	0.004	0.746	0.305	0.337 männl.
	der	*1*	0.427	0.195	0.665	0.663 weibl.
	Nennungen	*2*	0.314	0.026	0.029	
		*3*	0.254	0.032	0.001	
3.Klasse 55,5%	Anzahl	*0*	0.869	0.942	0.963	0.332 männl.
	der	*1*	0.13	0.056	0.037	0.668 weibl.
	Nennungen	*2*	0.000	0.002	0.000	
		*3*	0.000	0.000	0.000	

Für die Interpretation der in Tabelle 5 vorgestellten Lösung sind die in der restringierten LCA eingeführten Schwellenparameter von Bedeutung (vgl. Tabelle 6). Diese können folgendermaßen interpretiert werden. Die extrem niedrigen (logistischen) Lokationsparameter der dritten, größten latenten Klasse für die drei Inhaltskategorien geben an, daß eine Nennung in einer dieser drei Kategorien extrem unwahrscheinlich ist. Die erste Klasse hat dem gegenüber extrem hohe Leichtigkeiten für die Inhaltskategorie *Frauen* und *Männer*, jedoch eine geringe Leichtigkeit für die Inhaltskategorie *Geschlechter*. Dies bedeutet, daß die Personen der ersten Klasse bei ihrer Definition des Begriffs *Emanzipation* die beiden Inhaltskategorien *Frauen* und *Männer* oft gemeinsam ansprechen, während die Personen der dritten Klasse überhaupt selten entsprechende Nennungen machen. In der zweiten Klasse ist die Inhaltskategorie *Frauen* überwiegend alleine bei der Verkodung der Nennungen maßgebend.

**Tabelle 6:** Schwellenparameter für die 3-Klassenlösung der Latent-Class-Analyse, Modell 4

		Lokationsparameter $\mu$			
		Inhaltskategorien:			Personen- geschlecht
		Frauen	Geschlechter	Männer	
1.Klasse	*	3.669	-1.693	2.308	2.009
2.Klasse	*	1.354	-1.048	-1.971	0.677
3.Klasse	*	-5.116	-2.538	-6.000	0.699
		Leichtigkeitsparameter der Schwellen			
		0:1	1:2	2:3	
Inhaltskategorie:					
Frauen		3.227	-1.660	-1.567	
Geschlechter		-0.292	-0.973	1.266	
Männer		2.749	-1.165	-1.584	

Die Leichtigkeiten der Schwellen<sup>6</sup> sollten bei ordinalen Variablen in absteigender Reihenfolge über die Antwortstufen geordnet sein, d.h. von Schwelle zu Schwelle sollte es schwerer werden, diese zu überschreiten (vgl. Rost 1988, S 157ff). Dies ist im gegebenen Beispiel nur für die jeweils ersten beiden Schwellen jeder Inhaltskategorie gegeben, während es bei der ersten und zweiten Inhaltskategorie genauso leicht (Kategorie *Frauen*) oder sogar leichter ist (Kategorie *Geschlechter*), die Schwelle von 2 zu 3 Nennungen zu überschreiten. Bei der Inhaltskategorie ist es bereits sehr schwierig überhaupt die erste Schwelle zu überschreiten, d.h. eine Nennung in dieser Kategorie aufzuweisen. Dies bedeutet, daß die Inhaltskategorie *Geschlechter* zu selten angesprochen wird, um Unterschiede zwischen den Personen anzuzeigen. Differenzierungen der Definition des Begriffs *Emanzipation* anhand der drei vorliegenden Inhaltskategorien ergeben sich alleine durch die beiden Kategorien *Frauen* und *Männer*.

Das vorliegende Beispiel zeigt, daß die Mehrzahl der Studierenden (Klasse 3: 55,5%) bei der Definition des Begriffs *Emanzipation* ohne die drei Inhaltskategorien *Frauen*, *Geschlechter* und *Männer* auskommt. Werden entsprechende Nennungen öfter gemacht, dann lassen sich zwei Gruppen (Klasse 1 und Klasse 2) unterscheiden. Die Personen der ersten Klasse (23,3%

<sup>6</sup> Die Variable Personengeschlecht hat keinerlei Schwellenparameter, weil bei einer dichotomen Variablen die einzig vorhandene Schwelle bereits durch den Lokationsparameter parametrisiert ist.



der Befragten) erwähnen *Frauen* und *Männer* häufig *und* gemeinsam, wobei in dieser Klasse in auffälliger Weise die Studentinnen überwiegen. Die Personen der zweiten Klasse (21,2% der Befragten) sprechen bei ihrer Definition von *Emanzipation* überwiegend die Kategorien *Frauen* an. In dieser Klasse entspricht die Zusammensetzung von Studentinnen und Studenten in etwa der der Stichprobe. Aufgrund dieses Ergebnisses kann gefolgert werden, daß es eher die Studentinnen sind, die in ihrer Definition von *Emanzipation* die weiblichen und männlichen Aspekte in gleicher Weise betonen. Dies ist ein Ergebnis, das im Zusammenhang mit weiteren Auswertungen mit Hilfe der LCA ein differenziertes Bild des Emanzipationsverständnisses von Studierenden der Pädagogik ergibt (vgl. *Tarnai/Bos* 1989, 1990).

#### 4. Diskussion

Das empirische Beispiel gibt einen Einblick in die Anwendungsmöglichkeiten der Latent-Class-Analyse für ordinale Daten (LACORD) von *Rost* (1990a). Gezeigt wurde hier die gleichzeitige Berücksichtigung von dichotomen und ordinalen Variablen, die Auswertungen erlaubt, die bisher nur unter Verzicht auf vorliegende Informationen möglich waren. Die Modelle von LACORD schließen die Latent-Class-Analyse (LCA) ohne Nebenbedingungen für dichotome Daten ein. Diese Anwendung bei Inhaltsanalysen ist immer dann angezeigt, wenn alleine die Auftretenswahrscheinlichkeit von Inhaltskategorien von Interesse ist (vgl. *Tarnai* 1989b). Die Erweiterung, ähnlich dem vorgestellten Beispiel, kann weitergehende Hypothesen testen, indem Kriteriumsvariablen in die Analysen einbezogen werden.

Mit LACORD sind noch weitergehendere Anwendungen in den Sozialwissenschaften möglich als im vorliegenden Zusammenhang aufgezeigt werden kann. Ein Beispiel für die Verwendung bei Ratingskalen geben *Giegler/Rost* (1990) anhand von Daten der Allgemeinen Bevölkerungsbefragung (ALLBUS) 1986. Es wird dort deutlich, wie LACORD zur multivariaten Analyse von Fragebögen eingesetzt werden kann. Die Verbindung von Komponenten der Einstellungsmessung mit Schwellenwertmodellen von *Thurstone* zeigt *Rost* (1990b).

Bei allen Einsatzmöglichkeiten von LACORD ergibt sich eine wesentliche Anwendungsbeschränkung aus der Anzahl der Variablen und der Zahl ihrer Antwortstufen. Dies ist weniger eine Beschränkung, die aus der Rechenkapazität- bzw. Speicherbegrenzung folgt, sondern vielmehr eine Frage der Generalisierbarkeit der Ergebnisse, da zur statistischen Absicherung eine entsprechend große Anzahl von Personen benötigt wird. Sind diese Bedingungen erfüllt, so liegt mit LACORD ein Verfahren vor, das sich bei vielen Fragestellungen nicht nur der Inhaltsanalyse als geeignete Methode empfiehlt.



### Hinweis zum Programm LACORD

Das Programm LACORD kann für DM 50.- beim Zweitautor bezogen werden. Es handelt sich um ein PC-Programm (MS-DOS) mit selbsterklärender Benutzeroberfläche und ausführlichem Manual. Empfehlenswert, aber nicht notwendig, ist ein (schneller) Rechner mit Coprocessor.

### Literatur

- Bozdogan, H.*, 1987:  
Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52: 345-370.
- Formann, A.K.*, 1984:  
Die Latent-Class-Analyse. Weinheim: Beltz.
- Giegler, H./J. Rost*, 1990:  
Ordinale manifeste Variablen - Nominale latente Variablen - Latent Class Analyse für ordinale Daten (S. 461-470). In F. Faulbaum/R. Haux/K.-H. Jöckel (Hrsg.), *SoftStat '89. Fortschritte der Statistik-Software 2*. Stuttgart: Gustav Fischer.
- Georg, W./J. Rost*, 1991:  
Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus". In *ZA-Information* 28: 52-74.
- Langeheine, R.*, 1988:  
New developments in Latent Class Theory. In R. Langeheine /J. Rost (Eds.), *Latent Trait and Latent Class Models* (pp. 77-108). New York: Plenum.
- Lazarsfeld, R.R.*, 1950:  
The logical and mathematical foundation of latent structure analysis. In S.A. Stouffer/L. Guttman/E.A. Suchman/P.F. Lazarsfeld/S.A. Star/J.A. Clausen (Eds.), *Studies in social psychology in World War II, Vol. IV*: [zitiert nach Formann, 1984].
- Lazarsfeld, P.F./N.W. Henry*, 1968:  
*Latent structure analysis*. Boston: Houghton Mifflin.
- Read, T.R.C./N.A. Cressie*, 1988:  
*Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Rost, J.*, 1988:  
*Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.
- Rost, J.*, 1990a:  
LACORD. Latent class analysis for ordinal variables. A FORTRAN program. 2nd edition. Kiel: IPN.
- Rost, J.*, 1990b:  
Einstellungsmessung in der Tradition von Thurstone's Skalierungsverfahren. *Empirische Pädagogik* 4: 83-92.
- Tamai, Ch.*, 1989a:  
Eine empirisch-semantische Analyse des Begriffs Emanzipation anhand der Enzyklopädien Brockhaus und Meyer von 1830 bis 1988. In W. Bos/Ch. Tamai (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie* (S. 137-154). Münster: Waxmann.



**Tarnai, Ch.**, 1989b:

Abbildung der Struktur von Inhaltskategorien mittels Latent-Class-Analyse für ordinale Daten.

In *W. Bos/Ch. Tarnai* (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie* (S. 303-315). Münster: Waxmann.

**Tarnai, Ch./W. Bos.**, 1989:

Emanzipation und Emanzipatoren: Definition des Begriffs Emanzipation durch Studierende der Pädagogik.

In *W. Bos/Ch. Tarnai* (Hrsg.), *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie* (S. 155-172). Münster: Waxmann.

**Tarnai, Ch./W. Bos.**, 1990:

Was Studierende der Pädagogik unter Emanzipation verstehen. Sozialwissenschaftliche Forschungsdokumentationen. Münster: Institut für sozialwissenschaftliche Forschung e.V.