

Ein Versuch der Trennung deskriptiver und evaluativer Bedeutungskomponenten in Eigenschafts-beschreibenden Begriffen

Peter Borkenau und Fritz Ostendorf

Universität Bielefeld

Zusammenfassung. In vier aufeinander aufbauenden Studien wurden 20 deutschsprachige Quartette von Trait-Adjektiven etabliert, in denen deskriptive und evaluative Bedeutungskomponenten unabhängig voneinander variierten. In der ersten Studie wurde die soziale Erwünschtheit von 230 Adjektiven durch 26 Beurteiler eingeschätzt und es wurden 116 Eigenschaften selektiert, welche als sehr erwünscht oder sehr unerwünscht eingeschätzt wurden. In Studie 2 hatten 10 Beurteiler die deskriptive Ähnlichkeit solcher Begriffspaare einzuschätzen, welche entweder eine deskriptiv ähnliche und evaluativ gegensätzliche Bedeutung aufwiesen oder eine evaluativ ähnliche und deskriptiv gegensätzliche. Die beiden folgenden Studien dienten der Replikation der Befunde von Peabody (1967): Die Probanden wurden über eine Eigenschaft einer fiktiven Person informiert und mußten angeben, ob diese Person eher eine deskriptiv ähnliche Eigenschaft mit konträrer Bewertung aufweisen dürfte oder eine ähnlich bewertete Eigenschaft mit konträrem deskriptiven Gehalt. Es zeigte sich eine beträchtliche Konvergenz der Befunde: Bei 20 Adjektiven überwog durchgängig die Bedeutung der deskriptiven Bedeutungskomponente gegenüber der evaluativen Komponente. Verwendungsmöglichkeiten für Adjektivquartette der hier generierten Art werden diskutiert.

Schlüsselwörter: Bedeutungskomponenten, Deskription, Evaluation, Konnotation, Soziale Erwünschtheit

Summary. Four studies were conducted to establish sets of four trait-descriptive terms where descriptive and evaluative meaning components varied independently. In the first study, 26 judges rated the social desirability of 230 trait adjectives; those terms were selected that were rated as clearly socially desirable or socially undesirable, respectively. In the second study, 10 judges estimated the semantic similarity of pairs of trait-descriptive terms that were either descriptively similar and evaluatively dissimilar or evaluatively si-

milar and descriptively dissimilar. The last two studies replicated Peabody's (1967) findings. The judges were informed about one trait of a fictitious target person. Their task was to estimate whether this person was more likely to have either a second trait that was evaluatively dissimilar but descriptively similar, or a second trait that was descriptively dissimilar but evaluatively similar. The different approaches led to convergent results with respect to the assessment of descriptive similarity. For 20 sets of four trait-descriptive terms, all estimates were influenced more by the descriptive meaning component than by the evaluative meaning component. Several implications of the findings are discussed.

Key words: Meaning components, description, evaluation, connotation, social desirability

Menschliche Beobachter und Beurteiler bilden seit jeher eine wichtige Datenquelle psychologischer Untersuchungen. Ihre Fähigkeit, hochgradig komplexe Information zu strukturieren und bedeutsame Aspekte individuellen Verhaltens quasi herauszufiltern, wird gegenwärtig von keinerlei Aufzeichnungsgeräten übertroffen. Entsprechend werden in vielen Bereichen psychologischer Forschung und Praxis Beurteiler herangezogen, um Daten zu gewinnen.

Dieses Vorgehen wurde jedoch schon frühzeitig problematisiert. So verwies etwa bereits Thorndike (1920) auf einen „konstanten Fehler“ in psychologischen Beurteilungen, welchen er als „Halo-Effekt“ bezeichnete. Er fand, daß solche Persönlichkeitseinschätzungen, welche positive Bewertungen implizieren, unangemessen hohe Interkorrelationen aufweisen. Weitere Hinweise auf die große Bedeutung von Bewertungsaspekten in der Personenbeschreibung gab der von Edwards (1953) berichtete Befund, daß die Zustimmungsfrequenz zu Fragebogenitems in einer Höhe von etwa .90 mit der sozialen Erwünschtheit des Iteminhalts korreliert. Weiterhin korrelieren die Ladungen der Items eines Fragebogens auf der ersten unrotierten Hauptkomponente, welche aus den Interkorrelationen eben dieser Items extrahiert wird, gewöhnlich hoch mit der sozialen Erwünschtheit des jeweiligen Iteminhalts (Edwards & Diers, 1962). Diese Befundlage veranlaßte einige Autoren (z. B. Edwards, 1977; Paulhus, 1981) dazu, die Faktorenwerte von Probanden bezüglich der ersten Hauptkomponente in Fragebogen als Maß sozial erwünschten Antwortens zu interpretieren. Diese Schlußfolgerung ist jedoch problematisch, wie an anderer Stelle gezeigt wurde (Borkenau & Amelang, 1985, 1986).

Ferner betonen Osgoods (1962; Osgood, Suci & Tannenbaum, 1957) Befunde zur Dreidimensionalität des affektiven Bedeutungsraumes die herausragende Stellung der evaluativen Dimension. In seinen Studien erwies sich nämlich die evaluative Komponente als besonders varianzstark und gut replizierbar. Auch konnte Osgood die transkulturelle Generalität des semantischen Raumes eindrucksvoll bestätigen. Allerdings wies bereits Osgood (1962) selbst auf denotative Einschränkungen hin, etwa daß die Korrelationen zwischen bestimmten Skalen in hohem Maße von den eingeschätzten Konzepten abhängen. In diesem Kontext ist die Anmerkung wichtig, daß sich Osgood auf *affektive* Bedeutungen bezog, welche insbesondere dann deutlich hervortreten, wenn sich die eingesetzten Skalen

nur sehr bedingt zur Erfassung der wesentlichen Charakteristika der einzuschätzenden Konzepte eignen. Dies ist etwa dann der Fall, wenn das Konzept *Vereinte Nationen* bezüglich der Polarität *warm-kalt* einzuschätzen ist. Osgoods Befunde schließen somit keineswegs aus, daß die Interpretation von Ereignissen sehr viel komplexer strukturiert ist, als es die Konzeption von der Dreidimensionalität des semantischen Raumes nahelegt. Die drei Dimensionen *Bewertung*, *Aktivität* und *Potenz* dürften eben nur die konnotativen Bedeutungen erfassen.

Einen originellen Ansatz zur Untersuchung deskriptiver und evaluativer Bedeutungskomponenten in Eigenschafts-beschreibenden Begriffen entwickelte Peabody (1967). Er bildete Trios (sets of three) und Quartette (sets of four) von Traitbegriffen in der Weise, daß Begriffe mit ähnlicher deskriptiver Bedeutung sehr unterschiedliche Bewertungen implizierten und, umgekehrt, evaluativ ähnliche Begriffe gegensätzliches kennzeichneten. Ein entsprechendes Quartett von Trait-Adjektiven ist das in Tabelle 1 dargestellte.

Tab. 1: Von Peabody (1967) vorgeschlagenes Konstruktionsprinzip für Adjektivquartette.

	Evaluative Dimension	
	positiv	negativ
Deskriptive Dimension	sparsam (1)	geizig (2)
	großzügig (3)	verschwenderisch (4)

Bei einem solchen Quartett kennzeichnen die in einer Spalte stehenden Begriffe ähnlich Bewertetes, jedoch deskriptiv Gegensätzliches. Hingegen bezeichnen die in einer Zeile stehenden Begriffe deskriptiv Ähnliches, implizieren jedoch eine gegensätzliche evaluative Nebenbedeutung.

Peabody (1967) stellte 15 entsprechende Quartette sowie 10 Trios zusammen und konnte zeigen, daß sich die implizite Persönlichkeitstheorie naiver Probanden stärker von den deskriptiven als von den evaluativen Bedeutungskomponenten leiten ließ. So wurde etwa eine hypothetische Person, über die die Information vorlag, sie sei *sparsam*, eher für *geizig* gehalten denn als *großzügig* eingestuft. Dieser Primat der deskriptiven gegenüber den evaluativen Bedeutungsaspekten zeigte sich durchgängig bei jedem einzelnen von insgesamt 70 vorgenommenen Vergleichen. Es handelt sich mithin um einen für psychologische Untersuchungen eindrucksvollen Befund. Allerdings produzierte Peabody im Rahmen der gleichen Untersuchungsreihe ein Artefakt, als er auf faktorenanalytischem Wege zu zeigen versuchte, daß evaluative Aspekte für die implizite Persönlichkeitstheorie unerheblich seien. Die entsprechende Methodenkritik findet sich bereits bei Rosenberg und Olshan (1970). Zudem konnte Felipe (1970) zeigen, daß der Primat deskriptiver Bedeutungsaspekte über die evaluativen nur dann auftritt, wenn die deskriptiven Bedeutungsähnlichkeiten sehr stark ausgeprägt sind. Andernfalls folgt die implizite Persönlichkeitstheorie der evaluativen Dimension und es läßt sich ein deutlicher Halo-Effekt nachweisen.

Gerade dieser Zusammenhang läßt sich jedoch zur Identifizierung solcher Traitbegriffe nutzen, welche bei konträrer impliziter Wertung dennoch deskriptiv Ähnliches kennzeichnen. Aus dem Umstand, daß im Kontext einer Traitinferenz ein zur vorgegebenen Eigenschaft evaluativ gegensätzlicher Begriff einem evaluativ ähnlichen vorgezogen wird, läßt sich nämlich rückschließen, daß der evaluativ gegensätzliche Begriff sich durch hohe deskriptive Ähnlichkeit zum vorgegebenen Begriff auszeichnet. Dieser Umstand läßt sich bei der Etablierung solcher Adjektivquartette nutzen, welche den von Peabody aufgestellten Kriterien genügen. Solche Quartette können für viele psychologische Fragestellungen ein nützliches Instrument sein, insofern sie Rückschlüsse dahingehend erlauben, ob Urteile über Personen primär auf die deskriptiven oder primär auf die evaluativen Bedeutungsaspekte der eingesetzten Skalen zurückgehen. So mag es etwa, um nur ein sehr willkürliches Beispiel herauszugreifen, für die Depressionsforschung wichtig sein, ob sich die Kognitionen Depressiver primär bezüglich der Bewertung von Ereignissen von den Kognitionen Nicht-Depressiver unterscheiden, oder ob die Hauptunterschiede sich bezüglich anderer Aspekte zeigen. Adjektivquartette der erwähnten Art können bei der Abklärung entsprechender Fragestellungen hilfreich sein.

Gegenstand der vorliegenden Arbeit ist die Darstellung der Schritte zur Etablierung eines entsprechenden Pools von 20 Adjektivquartetten für den deutschsprachigen Bereich. Berichte über einzelne Anwendungen bleiben zukünftigen Darstellungen vorbehalten.

Überblick

Die Erhebungen umfaßten eine vorgeschaltete Phase der Generierung geeignet erscheinender Adjektive und vier kleinere Studien. In der *ersten* Studie wurde die soziale Erwünschtheit der generierten Traitbegriffe durch eine Stichprobe naiver Beurteiler eingeschätzt. Da sich nur deutlich positiv bewertete und deutlich negativ bewertete Begriffe zur Weiterverwendung eigneten, wurden nach der ersten Studie solche Begriffe ausgeschieden, deren soziale Erwünschtheit als im Mittelbereich liegend eingeschätzt wurde. In der *zweiten* Studie wurde sodann eine neue Stichprobe von Beurteilern aufgefordert, solche Begriffe bezüglich ihrer deskriptiven Ähnlichkeit einzuschätzen, welche möglicherweise einem Quartett sensu Peabody zugeordnet werden konnten. Hierbei sollte von Aspekten der implizierten Wertung weitmöglichst abgesehen werden. In der *dritten* Studie wurde sodann versucht, die Befunde von Peabody (1967) zu replizieren. Die Probanden wurden also darüber informiert, daß eine fiktive Zielperson eine bestimmte Eigenschaft aufweise. Sie sollten nun angeben, ob diese Person eher die eine (deskriptiv ähnliche aber evaluativ gegensätzliche) oder eher die andere (evaluativ ähnliche aber deskriptiv gegensätzliche) von zwei zur Auswahl stehenden Eigenschaften aufweisen dürfte. Auf Basis der in den Studien 2 und 3 gewonnenen Daten wurde sodann eine weitere Reduktion der Adjektivquartette auf die endgültige Auswahl von 20 vorgenommen. Ausgeschlossen wurden dabei solche Quartette, bei denen sich unsere Vorannahmen bezüglich der deskriptiven Ähnlichkeiten empirisch nicht halten ließen. Da hierbei aber statistische Zufallsfluktuationen ins Spiel kamen, wurde schließlich in einer *vierten* Studie noch eine Kreuzvalidierung der

Befunde aus Studie 3 vorgenommen. Es wurde hierbei angestrebt, Peabodys Befunde an einer neuen Probandenstichprobe für diejenigen Eigenschaftsquartette zu bestätigen, welche den vorangegangenen Selektionsschritten zufolge als endgültige Auswahl angesehen werden konnten.

Gewinnung eines Ausgangspools von Adjektiven

Die Kriterien für die Aufnahme eines Eigenschaftsbegriffes in den Ausgangspool sind naturgemäß schwer rekonstruierbar. Zum einen wurden in diesem Kontext die von Peabody vorgeschlagenen Quartette übersetzt. Es wurden aber auch Eigenschaftstaxonomien konsultiert und Bekannte um Vorschläge gebeten. Auch die Intuition der Autoren verdient an dieser Stelle Erwähnung. Insgesamt wurden in dieser Weise 230 Adjektive gesammelt, welche zu maximal 41 Quartetten kombiniert werden konnten. Vielfach wurden in diesem Stadium auch mehrere parallele Begriffe aufgenommen, welche sowohl in deskriptiver als auch in evaluativer Hinsicht einander ähnlich waren. Dabei war von vornherein vorgesehen, nur einen dieser Begriffe beizubehalten. Die Entscheidung darüber, welcher dies sein würde, sollte aber unter Berücksichtigung der empirischen Befunde gefällt werden.

Studie 1

Methode

Probanden. Die soziale Erwünschtheit der 230 Adjektive wurde von 12 weiblichen und 14 männlichen Beurteilern eingestuft. Da anzunehmen war, daß sich bezüglich der Einschätzung der sozialen Erwünschtheit einzelner Adjektive bedeutsame Unterschiede zwischen Probandengruppen ergeben könnten, wurde auf Heterogenität der Beurteilergruppe bezüglich Alter, Geschlecht und Beruf geachtet. Das Durchschnittsalter betrug 34,7 Jahre bei einer Standardabweichung von 15,95 Jahren. Die Beurteilerstichprobe bestand zu etwa je einem Drittel aus Auszubildenden, Studenten und berufstätigen Erwachsenen. Dieser Umstand kommt auch in der recht hohen Altersstreuung zum Ausdruck.

Prozedur. Die 230 Adjektive des Ausgangspools wurden den 26 Beurteilern in jeweils unterschiedlicher Reihenfolge dargeboten. Sie erhielten dazu die folgende Instruktion:

„Im allgemeinen gebrauchen wir im Alltag eine Menge verschiedener Begriffe, um das Verhalten unserer Mitmenschen zu beschreiben. Auf den folgenden Seiten finden Sie eine Auswahl von Eigenschaftsworten, die man manchmal benutzt, um andere Personen zu beschreiben.

Eigenschaftsworte unterscheiden sich darin, daß sie mehr oder weniger wünschenswerte Eigenschaften von Personen kennzeichnen. Wenn Sie zum Beispiel jemanden für *dumm* halten, so werden Sie diese Eigenschaft wohl kaum für wünschenswert ansehen. Andererseits halten Sie sicher die Eigenschaft *intelligent* für wünschenswert.

Wir sind nun daran interessiert, inwieweit *für Sie* die auf den folgenden Seiten aufgeführten Eigenschaften *persönlich wünschenswert oder unerwünscht* sind. Hierzu wird jeweils immer eine Eigenschaft aufgeführt, und Sie sollen durch Ankreuzen einer Zahl angeben, ob für Sie diese Eigenschaft eher wünschenswert oder eher unerwünscht ist.“

Beispiel:

dumm: sehr unerwünscht -3 -2 -1 0 +1 +2 +3 sehr erwünscht

Die Instruktion hob mithin auf die persönliche Meinung der Beurteiler ab, statt nach der Bewertung durch „die Gesellschaft“ zu fragen. Aus diesem Grunde war auch auf die Heterogenität der Beurteilerstichprobe besonderes Gewicht gelegt worden. Die Beurteiler erhielten für ihre Mitarbeit eine kleine Bezahlung und schätzten die soziale Erwünschtheit der Traitbegriffe ohne Zeitbegrenzung ein.

Resultate

Zunächst wurde die Übereinstimmung der Beurteiler mittels Intraklassenkorrelation ermittelt. Die Reliabilität der Urteile einzelner Beurteiler belief sich demzufolge auf .69 (Koeffizient 2, 1 gemäß der Taxonomie derartiger Koeffizienten von Shrout & Fleiss, 1979). Die Reliabilität der über die 26 Rater gemittelten Urteile, welche im vorliegenden Kontext primär interessiert, belief sich entsprechend auf .98 (Koeffizient 2, 26). Die gemittelten Erwünschtheitsratings sind in Tabelle 2 für diejenigen 80 Begriffe angegeben, welche die endgültigen 20 Quartette konstituieren.

Tab. 2: Soziale Erwünschtheitsratings für die 80 in die Endauswahl aufgenommenen Begriffe.

Erwünschte Eigenschaften			Unerwünschte Eigenschaften		
Begriff	\bar{X}	s	Begriff	\bar{X}	s
beständig	1.81	1.06	unveränderlich	- 1.46	1.24
änderungsfähig	1.81	0.89	wechselhaft	- 1.35	1.33
standhaft	1.92	0.97	unflexibel	- 2.27	0.83
beweglich	2.00	0.85	unstetig	- 2.23	0.91
konsequent	1.92	1.01	starr	- 2.38	0.75
flexibel	2.27	0.73	unschlüssig	- 1.77	1.14
zielstrebig	1.65	1.20	gierig	- 2.77	0.44
genügsam	1.00	1.26	faul	- 2.38	0.85
selbstbewußt	2.23	0.71	eingebildet	- 2.81	0.40
bescheiden	1.01	1.23	schüchtern	- 1.08	1.09
gesellig	2.04	0.92	zudringlich	- 2.15	0.97
unaufdringlich	1.65	1.06	verschlossen	- 1.96	0.99

Erwünschte Eigenschaften			Unerwünschte Eigenschaften		
Begriff	\bar{X}	s	Begriff	\bar{X}	s
offen	2.19	0.98	aufdringlich	- 2.42	0.85
diskret	1.54	1.24	zugeknöpft	- 1.08	0.93
freimütig	1.35	1.17	taktlos	- 2.50	0.76
diplomatisch	1.62	1.45	unehrlich	- 2.92	0.26
kultiviert	1.77	1.31	gekünstelt	- 2.35	1.02
natürlich	2.54	0.86	naiv	- 1.92	1.13
locker	1.77	1.11	unzuverlässig	- 2.85	0.37
zuverlässig	2.81	0.49	pedantisch	- 1.92	1.13
aktiv	2.12	0.77	hektisch	- 2.15	0.93
entspannt	1.69	1.12	lahm	- 2.35	0.75
lebhaft	1.92	0.97	ruhelos	- 1.77	1.24
gemütlich	1.23	1.45	träge	- 2.31	0.79
munter	2.15	0.79	exaltiert	- 1.17	1.40
ruhig	1.15	1.37	schlaff	- 2.35	0.80
agil	1.27	1.28	überdreht	- 1.88	1.14
gelassen	1.65	1.09	apathisch	- 2.08	1.16
energievoll	2.00	0.89	explosiv	- 1.11	1.46
behutsam	1.58	1.17	energielos	- 2.08	1.26
ernsthaft	1.00	1.26	verbissen	- 2.11	1.14
humorvoll	2.46	0.70	kindisch	- 1.88	1.07
sachlich	1.50	1.17	humorlos	- 2.31	1.26
fröhlich	2.46	0.65	albern	- 1.77	1.31
sparsam	1.15	1.54	geizig	- 2.69	0.55
großzügig	1.31	1.09	verschwenderisch	- 2.19	1.02
gefestigt	1.62	1.47	eingefahren	- 1.35	1.23
anpassungsfähig	1.85	1.22	labil	- 1.85	1.44
kritisch	1.69	1.22	mäkelig	- 2.35	0.85
begeisterungsfähig	1.96	0.82	unkritisch	- 1.77	1.31

Anmerkung: \bar{X} informiert über das Durchschnittsrating der 26 Beurteiler und s über die Standardabweichung der einzelnen Beurteiler.

Auf der Basis der Erwünschtheitsratings wurden nun alle die Begriffe ausgeschlossen, die gemittelte Einschätzungen zwischen -1 und +1 aufwiesen, mithin also bezüglich ihrer sozialen Erwünschtheit eher als indifferent angesehen wurden. Außerdem wurden solche Begriffe eliminiert, bezüglich deren sozialer Erwünschtheit die Beurteilerübereinstimmung gering war. Weiterhin erfolgte jetzt eine Eliminierung bisher beibehaltener paralle-

ler Begriffe, so daß die Zahl der weiter untersuchten Begriffe genau viermal so hoch war wie die Zahl der Quartette. Insgesamt wurden 116 Begriffe beibehalten, welche 29 Adjektivquartette konstituieren sollten.

Studie 2

Die zweite Studie diente der Erhebung von Einschätzungen bezüglich der deskriptiven Ähnlichkeiten dieser 116 Begriffe. Wäre hierbei jeder Begriff mit allen 115 übrigen verglichen worden, so wären tausende von Einschätzungen erforderlich gewesen. Zudem waren primär die Vergleiche innerhalb der 29 Quartette wichtig. Entsprechend fanden die Vergleiche zwischen allen Begriffe ein- und desselben Quartetts statt. Bei sechs möglichen Vergleichen pro Quartett resultierten so 174 Ratings pro Beurteiler.

Methode

Probanden. Die Beurteilungen wurden von fünf Studentinnen und fünf Studenten der Universität Bielefeld vorgenommen. Diese erhielten für ihre Kooperation eine kleine Bezahlung und arbeiteten ohne vorgegebene Zeitbegrenzung.

Prozedur. Die Beurteiler wurden wie folgt instruiert:

„Nachfolgend finden Sie Begriffe, die Persönlichkeitseigenschaften bezeichnen. In dieser Untersuchung geht es darum, die Ähnlichkeit bzw. Gegensätzlichkeit der jeweiligen Wortbedeutungen zu bestimmen. Wichtig ist dabei, daß Sie von den wertenden Aspekten der Begriffe bitte absehen möchten.

Vergegenwärtigen Sie sich bitte die Adjektive *intelligent*, *gerissen* und *dumm*. Diese Begriffe implizieren eine Wertung in dem Sinne, daß *intelligent* positiv und die beiden anderen Begriffe eher negativ bewertet werden.

Sieht man aber von diesem wertenden Aspekt ab, so sind *intelligent* und *gerissen* in ihrer Bedeutung ähnlich, während *gerissen* und *dumm* trotz ihrer gleichermaßen negativen Tönung doch Gegensätzliches beschreiben.

Demzufolge sollten Sie hier die Adjektivpaare *intelligent* und *dumm*, ebenso wie *gerissen* und *dumm*, als Gegensätze bewerten. Demgegenüber wäre das Begriffspaar *intelligent* und *gerissen* als ähnlich in seiner Bedeutung zu bewerten. Kurz, versuchen Sie bei ihren Ähnlichkeitsurteilen von der bewertenden Seite der Begriffe so weit als irgend möglich abzusehen!“

Beispiel:

intelligent & gerissen: gegensätzlich -3 -2 -1 0 +1 +2 +3 identisch

Jeder der 10 Beurteiler bearbeitete sodann die 174 Vergleiche in jeweils einer anderen Reihenfolge.

Resultate

Die Beurteilerübereinstimmung wurde wiederum mittels Intraklassenkorrelation ermittelt. Für die Ratings der einzelnen Beurteiler belief sich die Reliabilität auf .43 (Koeffizient 2, 1 nach Shrout & Fleiss). Das gemittelte Rating aller 10 Beurteiler wies entsprechend eine Reliabilität von .88 auf (Koeffizient 2, 10).

Innerhalb jedes der 29 Quartette waren jeweils sechs Vergleiche vorgenommen worden, von denen je zwei vom gleichen Typ waren. Zwei Vergleiche bezogen sich jeweils auf deskriptiv ähnliche aber evaluativ gegensätzliche Begriffe (z. B. *beständig & unveränderlich* sowie *änderungsfähig & wechselhaft*). Zwei weitere Vergleiche bezogen sich auf evaluativ ähnliche aber deskriptiv konträre Begriffe (z. B. *beständig & änderungsfähig* sowie *unveränderlich & wechselhaft*). Die verbleibenden zwei Vergleiche schließlich bezogen sich auf solche Begriffspaare, welche sowohl in deskriptiver als auch in evaluativer Hinsicht Gegensätzliches bezeichnen (z. B. *beständig & wechselhaft* sowie *änderungsfähig & unveränderlich*). Als ein nahezu durchgängiger Befund ergab sich hierbei, daß die Ähnlichkeiten für den letztgenannten Typ von Vergleichen am niedrigsten eingestuft wurden. Offenbar gingen also instruktionswidrig Bewertungsaspekte in die Beurteilung der deskriptiven Ähnlichkeit ein.

Im vorliegenden Kontext ist allerdings der dritte Typ von Vergleichen weniger interessant. Hingegen lassen sich die beiden ersten Vergleichstypen für die Auswahl deskriptiv ähnlicher aber evaluativ gegensätzlicher Begriffe nutzen. Dabei lag dem eingeschlagenen Verfahren folgende Rationale zugrunde: Wenn entsprechend der Instruktion zu dieser Studie die Beurteiler die deskriptiv ähnlichen (aber evaluativ konträren) Begriffe als ähnlicher einstufen als die evaluativ ähnlichen (aber deskriptiv konträren), so kann dies als Hinweis darauf gewertet werden, daß die verwendeten evaluativ gegensätzlichen Begriffe tatsächlich deskriptiv Ähnliches beinhalten. Wenn z. B. das evaluativ konträre Begriffspaar *beständig & unveränderlich* als bedeutungsähnlicher eingestuft wird als das evaluativ ähnliche Begriffspaar *beständig & änderungsfähig*, so kann dies als Indiz dafür gewertet werden, daß die Begriffe *beständig & unveränderlich* sich in der Tat auf ähnliche Sachverhalte beziehen.

Bei dem erwähnten Begriffspaar bestätigten sich diese Zusammenhänge. Entsprechend der Nummerierung der Quadranten in Tabelle 1 erhielt zunächst der Begriff *beständig* die Nummer 1, der Begriff *unveränderlich* die Nummer 2, der Begriff *änderungsfähig* die Nummer 3 und der Begriff *wechselhaft* die Nummer 4. Für die sechs Begriffspaare ergaben sich folgende über die 10 Beurteiler gemittelten Ähnlichkeitsratings:

— Begriffe 1 & 2 (<i>beständig & unveränderlich</i>)	1.3
— Begriffe 3 & 4 (<i>änderungsfähig & wechselhaft</i>)	1.0
— Begriffe 1 & 3 (<i>beständig & änderungsfähig</i>)	— 1.6
— Begriffe 2 & 4 (<i>unveränderlich & wechselhaft</i>)	— 2.3
— Begriffe 1 & 4 (<i>beständig & wechselhaft</i>)	— 2.8
— Begriffe 2 & 3 (<i>unveränderlich & änderungsfähig</i>)	— 2.7

Hypothesenkonform zeigen sich bei diesem Quartett die höchsten Ähnlichkeitsschätzungen für die ersten beiden Vergleiche zwischen deskriptiv ähnlichen und evaluativ gegensätzlichen Begriffe. Die Schätzungen für die evaluativ ähnlichen aber deskriptiv konträren Paare (Zeilen 3 und 4) fallen demgegenüber niedriger aus, liegen aber noch über denen für sowohl deskriptiv als auch evaluativ gegensätzliche Adjektivpaare (Zeilen 5 und 6). Demgegenüber bestätigten sich bei anderen Quartetten unsere Vorhersagen nicht. So wurden etwa die Begriffe *kooperativ & eigenständig* als deskriptiv ähnlicher eingeschätzt als die Begriffe *kooperativ & unselbständig*. Entsprechend wurde denn auch das letztgenannte Begriffsquartett nicht in die Endauswahl aufgenommen.

In einem nächsten Schritt wurden aus den Ähnlichkeitsratings vier Differenzscores pro Begriffsquartett berechnet, nämlich (a) 1 & 2 — 1 & 3, (b) 3 & 4 — 1 & 3, (c) 1 & 2 — 2 & 4 sowie (d) 3 & 4 — 2 & 4, und jeweils einem der vier Adjektive als „Zentralbegriff“ zugeordnet (das Wort „Zentralbegriff“ wird in der Fußnote zu Tabelle 3 näher erläutert). In Vergleich (a) ist Begriff 1 (*beständig*) der Zentralbegriff und der Differenzscore subtrahiert das Ähnlichkeitsrating in Höhe von — 1.6 mit dem evaluativ ähnlichen aber deskriptiv gegensätzlichen Begriff *änderungsfähig* von dem Ähnlichkeitsrating in Höhe von + 1.3 mit dem evaluativ konträren aber deskriptiv ähnlichen Begriff *unveränderlich*. Entsprechend findet sich in der ersten Zeile der Spalte 1 der Tabelle 3 ein Differenzwert von 2.9. Für den Zentralbegriff *änderungsfähig* errechnet sich hingegen der Differenzscore aus dem Vergleich mit *wechselhaft* und *beständig* als 2.6. Dieser Wert findet sich neben dem Begriff *änderungsfähig* in der zweiten Zeile der Spalte 1 von Tabelle 3. In den nachfolgenden Zeilen von Spalte 1 finden sich die entsprechenden Differenzscores für alle diejenigen Quartette, welche zu den 20 endgültig ausgewählten gehören. Wie ersichtlich, haben 79 der 80 Differenzscores ein positives Vorzeichen, d. h. der deskriptiven Komponente kommt bei diesen Zentralbegriffen eine größere Bedeutung für das Ähnlichkeitsrating zu als der evaluativen. Selbstredend gilt dies zunächst nur für die hier untersuchten Vergleichsbegriffe mit stark ausgeprägten deskriptiven Ähnlichkeiten und Gegensätzlichkeiten.

Tab. 3: Ergebnisse der Vergleiche zur Prädominanz deskriptiver gegenüber evaluativen Bedeutungsaspekten.

Zentralbegriff	Differenzscore	Wahrscheinlichkeitsschätzung	
	in Studie 2	in Studie 3	in Studie 4
beständig	2.90	5.17	5.27
änderungsfähig	2.60	5.00	5.47
unveränderlich	3.60	5.85	6.60
wechselhaft	3.30	5.48	5.07
standhaft	2.30	5.10	4.93
beweglich	1.70	4.68	5.47
unflexibel	1.20	4.77	5.53
unstetig	0.60	4.88	5.27
konsequent	2.00	5.12	5.27
flexibel	0.40	4.08	4.40
starr	2.00	5.02	5.00
unschlüssig	0.40	4.18	4.47

Zentralbegriff	Differenzscore	Wahrscheinlichkeitsschätzung	
	in Studie 2	in Studie 3	in Studie 4
zielstrebig	1.20	5.22	5.53
genügsam	0.40	4.83	4.60
gierig	1.30	5.67	5.67
faul	0.50	5.08	5.20
selbstbewußt	1.40	4.77	4.67
bescheiden	2.20	5.23	5.47
eingebildet	2.30	5.77	5.20
schüchtern	3.10	5.75	5.73
gesellig	0.90	4.50	4.87
unaufdringlich	0.40	4.98	5.20
zudringlich	2.40	5.35	5.00
verschlossen	1.90	5.68	6.07
offen	1.80	4.47	4.60
diskret	1.90	5.18	5.73
aufdringlich	2.60	5.15	5.67
zugeknöpft	2.70	5.53	5.73
freimütig	1.50	4.55	5.07
diplomatisch	1.90	4.35	4.73
taktlos	0.40	5.00	5.47
unehrlich	0.80	4.43	4.33
kultiviert	0.40	4.80	4.80
natürlich	0.90	4.15	4.33
gekünstelt	1.30	4.05	4.40
naiv	1.80	5.15	4.87
locker	0.70	4.87	5.53
zuverlässig	1.60	4.58	4.60
unzuverlässig	2.20	5.35	5.67
pedantisch	3.10	5.63	5.60
aktiv	2.00	4.90	4.73
entspannt	2.30	3.98	4.20
hektisch	2.70	6.00	6.20
lahm	3.00	5.38	5.47
lebhaft	2.30	5.22	5.93
gemütlich	2.10	4.90	5.07
ruhelos	3.10	5.77	6.33
träge	2.90	5.63	6.13
munter	3.10	5.35	5.27
ruhig	2.10	4.65	5.07
exaltiert	2.30	5.60	4.87
schlaff	1.30	5.83	5.93
agil	2.60	4.95	5.33
gelassen	1.80	4.10	4.20
überdreht	3.20	5.57	5.73
apathisch	2.40	4.85	5.73

Zentralbegriff	Differenzscore	Wahrscheinlichkeitsschätzung	
	in Studie 2	in Studie 3	in Studie 4
energievoll	1.90	6.07	6.13
behutsam	0.30	4.48	4.13
explosiv	4.50	6.52	6.47
energielos	2.90	5.68	6.00
ernsthaft	2.60	5.00	4.87
humorvoll	1.90	4.55	4.27
verbissen	1.80	5.17	4.87
kindusch	1.10	4.98	4.80
sachlich	1.00	5.10	5.33
fröhlich	1.80	4.88	5.20
humoros	2.30	5.45	5.60
albern	3.10	5.40	5.40
sparsam	0.80	5.67	5.73
großzügig	2.90	5.65	6.13
geizig	1.10	6.35	6.73
verschwenderisch	3.20	5.95	6.40
gefestigt	0.90	4.62	5.20
anpassungsfähig	-0.20	4.38	4.00
eingefahren	1.30	5.05	5.40
labil	0.20	4.75	4.73
kritisch	0.60	5.00	4.80
begeisterungsfähig	0.40	5.05	4.73
mäkelig	1.10	5.32	4.67
unkritisch	0.90	5.45	4.53

Anmerkung: Unter *Zentralbegriff* wird im vorliegenden Kontext folgendes verstanden: In Studie 2 hatten die Probanden die rein deskriptive Ähnlichkeit zwischen einem Zentralbegriff einerseits und seinem deskriptiven Gegensatz (bei evaluativer Ähnlichkeit) bzw. seinem evaluativen Gegensatz (bei deskriptiver Ähnlichkeit) andererseits anzugeben. Das erste Rating wurde sodann vom zweiten subtrahiert. Zum Beispiel ist bei den Vergleichen *sparsam-großzügig* und *sparsam-geizig* das Wort *sparsam* der Zentralbegriff. Der entsprechende Differenzscore der beiden Ähnlichkeitsratings ist in der ersten Spalte neben dem Zentralbegriff *sparsam* aufgeführt und lautet 0.80. Bei den Studien 3 und 4 hingegen ist der Zentralbegriff die vorgegebene Eigenschaft der fiktiven Zielperson, also etwa *sparsam* für das im Methodenteil zu Studie 3 aufgeführte Beispiel.

Studie 3

Ebenso wie in Studie 2 wurden auch in die dritte Studie 29 Adjektivquartette mit insgesamt 116 Begriffen einbezogen. Die Studie diente der Replikation der Befunde von Peabody (1967) auf der Basis eines Satzes deutschsprachiger Adjektivquartette. Nur solche Adjektivquartette, für die sich die Befunde Peabodys replizieren ließen, sollten zudem in die Endauswahl eingehen.

Methode

Probanden. An dieser Studie nahmen 21 Studentinnen und 19 Studenten der Universität Bielefeld teil. Ihr Durchschnittsalter betrug 24.5 Jahre.

Prozedur. Die 40 Beurteiler wurden wie folgt instruiert: „Nachfolgend finden Sie Begriffe, die Eigenschaften von Personen bezeichnen. In dieser Untersuchung geht es darum, Wahrscheinlichkeiten einzuschätzen. Sie sollen die Wahrscheinlichkeiten angeben, mit der Personen mit vorgegebenen Eigenschaften auch andere, weitere Eigenschaften besitzen.

Hierzu wird jeweils eine Person mit einer charakteristischen Eigenschaft angegeben (z. B. eine *sparsame* Person) und Sie sollen jeweils die Wahrscheinlichkeit einschätzen, mit der diese Person eher die eine als die andere von zwei weiteren Eigenschaften besitzt. Also zum Beispiel:

eine *sparsame* Person ist

eher geizig ● ○ ○ ○ eher großzügig“

Wie aus dem Beispiel ersichtlich wird, hatten die Probanden stets zwischen dem deskriptiv ähnlichen (aber evaluativ gegensätzlichen) und dem evaluativ ähnlichen (aber deskriptiv gegensätzlichen) Begriff zu wählen. Da innerhalb jedes Quartetts jeder der vier Begriffe als vorgegebene Eigenschaft fungieren konnte, waren für die 29 Quartette insgesamt 116 Einschätzungen vorzunehmen. Die Probanden bearbeiteten diese in jeweils unterschiedlicher Reihenfolge.

Resultate

Wiederum wurde in einem ersten Schritt die Beurteilerübereinstimmung ermittelt. Die Reliabilität der Ratings eines einzelnen Beurteilers belief sich auf .45 (Koeffizient 2, 1 nach Shrout & Fleiss). Die Zuverlässigkeit des gemittelten Urteils der 40 Probanden, welche hier primär interessiert, belief sich auf .97 (Koeffizient 2, 40).

Die Angaben der Probanden wurden sodann in der Weise kodiert, daß Wahlen im Sinne deskriptiver Ähnlichkeit die höheren Scores erhielten und Wahlen im Sinne evaluativer Ähnlichkeit die niedrigeren Scores. Dabei fand eine von 1 bis 7 reichende Skala Verwendung. Wurde etwa bei der Beispielskala das große Quadrat markiert, so wurde diese Wahl als eine 7 kodiert, da diese Wahl ausschließlich die deskriptive Ähnlichkeit der Begriffe *sparsam* und *geizig* berücksichtigt. Entsprechend wurde etwa eine Markierung des zweitgrößten Kreises als eine 2 kodiert usw.. Folglich zeigen hohe Mittelwerte an, daß die Beurteiler ihre Entscheidungen vor allem unter Aspekten der deskriptiven Ähnlichkeit fällten. Aus hohen Mittelwerten kann also geschlossen werden, daß die von uns als deskriptiv ähnlich erachteten Begriffe von naiven Probanden ebenfalls als deskriptiv ähnlich aufgefaßt werden. Die über die 40 Beurteiler gemittelten Ratings sind in der zweiten Spalte der Ta-

belle 3 für diejenigen 20 Quartette aufgeführt, welche die Endauswahl bilden. Aus der Tabelle ist ersichtlich, daß 79 der 80 Mittelwerte über dem Skalenmittelpunkt 4.00 liegen. Der verbleibende Wert liegt mit 3.98 nur unwesentlich darunter. Der Gesamtmittelwert für alle 80 Vergleiche beläuft sich auf 5.11.

Auf der Basis der Ergebnisse der zweiten und dritten Studie wurde nunmehr die Endauswahl der 20 Adjektivquartette vorgenommen (vgl. Tabelle 2 und Tabelle 3). Es wurde hierbei angestrebt, daß (a) für alle vier Differenzscores, welche in Studie 2 ermittelt worden waren, ein positiver Wert vorliegen sollte und daß (b) alle vier in Studie 3 ermittelten Wahrscheinlichkeitsangaben pro Quartett über dem Skalenmittelpunkt 4.00 liegen sollten. Aus Tabelle 3 ist ersichtlich, daß die 20 Quartette der Endauswahl diese Kriterien mit zwei geringfügigen Ausnahmen erfüllen.

Möglicherweise sind jedoch die Werte in Spalte 1 und 2 der Tabelle 3 insofern systematisch verzerrt, als die Auswahl der 20 Adjektivquartette auf der Basis Meßfehler-belasteter Daten erfolgte. Zwar dürfte der Einfluß von Meßfehlern angesichts der sehr hohen Reliabilität der gemittelten Scores eher als gering einzuschätzen sein. Es wurde aber dennoch für notwendig erachtet, eine Kreuzvalidierung der Befunde aus Studie 3 vorzunehmen. Dies war der Zweck der vierten Studie.

Studie 4

Methode

Probanden. An dieser Studie nahmen 7 Studenten und 8 Studentinnen teil, deren Durchschnittsalter 21.9 Jahre betrug.

Prozedur. Material und Instruktion waren identisch mit Studie 3 und wurden dort bereits geschildert.

Resultate

Zunächst wurde wieder mittels Intraklassenkorrelation die Reliabilität der Urteile geschätzt. Die Reliabilität der Angaben einzelner Beurteiler belief sich auf .51 (Koeffizient 2, 1) und die Zuverlässigkeit der gemittelten Urteile aller 15 Beurteiler erreichte .94 (Koeffizient 2, 15).

Die gemittelten Einschätzungen werden in der letzten Spalte der Tabelle 3 berichtet. Der Gesamtmittelwert aller 80 Vergleiche beträgt 5.26 und liegt somit deutlich über dem Skalenmittelpunkt 4.00. Er liegt — entgegen der Erwartung — sogar etwas über dem Gesamtmittelwert in Studie 3. Weiterhin fällt auf, daß für keinen einzigen der 80 einzelnen Vergleiche ein Mittelwert auftritt, welcher unter dem Skalenmittelpunkt liegt. In einem Falle ergibt sich allerdings exakt der Wert 4.00.

Der Umstand, daß in der Kreuzvalidierung durchschnittlich höhere Angaben im Sinne deskriptiver Konsistenz gemacht wurden als in Studie 3, erscheint widersinnig. Die naheliegendste Erklärung besteht darin, daß die in Studie 3 und in Studie 4 eingesetzten Beurteilerstichproben nicht der gleichen Population angehörten. So fällt etwa das in der Kreuz-

validierungsstichprobe um 2.5 Jahre niedrigere Durchschnittsalter auf. Weiterhin gewannen wir den Eindruck, daß die Probanden der Studie 4 geringere Schwierigkeiten mit einzelnen Wortbedeutungen hatten als diejenigen der Studie 3. Möglicherweise waren die Probanden der Studie 4 verbal intelligenter als die der Studie 3. Dies wäre eine mögliche Erklärung der Befunde, wenn man die Annahme macht, daß sich Probanden bei mangelnder Klarheit über die deskriptiven Wortbedeutungen vor allem an der Konnotation der zu vergleichenden Begriffe orientierten.

Ergänzende Analysen

Ein Blick auf Tabelle 3 zeigt, daß sich das Überwiegen von Aspekten der deskriptiven Stimmigkeit gegenüber solchen der evaluativen Stimmigkeit für die einzelnen Zentralbegriffe recht unterschiedlich gestaltet. Der Primat von Aspekten der deskriptiven Stimmigkeit gegenüber solchen der evaluativen Stimmigkeit ist zwar für die verschiedenen Zentralbegriffe durchgängig gegeben; er ist jedoch anscheinend unterschiedlich stark ausgeprägt. Es stellt sich folglich die Frage, inwieweit diesbezüglich die Befunde der Studien 2 bis 4 zu vergleichbaren Resultaten führen.

Zwecks Abklärung dieser Frage berechneten wir die Korrelationen zwischen den drei Spalten der Tabelle 3 jeweils über die 80 Begriffe hinweg. Es ergaben sich dabei Korrelationen von .54 zwischen den Spalten 1 und 2, von .56 zwischen den Spalten 1 und 3, und von .82 zwischen den Spalten 2 und 3. Mithin zeigt sich ein erhebliches Ausmaß an Übereinstimmung bezüglich der verschiedenen Zugangsweisen, Begriffsquartette mit den gewünschten Charakteristika zu identifizieren. Dieser Umstand erleichtert es auch, aus den 20 hier dokumentierten Quartetten solche zu selektieren, welche die vorgegebenen Kriterien besonders gut erfüllen.

Weiterhin deutet sich in Tabelle 3 an, daß der Gesichtspunkt der deskriptiven Stimmigkeit bei den negativ bewerteten Begriffen besonders stark im Vordergrund steht. Dieser Befund war unter der von Goldberg (1982) formulierten Prämisse zu erwarten, daß negativ bewertete Traitbegriffe insbesondere extreme Merkmalsausprägungen kennzeichnen. (Entsprechend haben solche Begriffe, welche mit der Vorsilbe *über* beginnen, z. B. *überevorsichtig* und *überkritisch*, häufig eine negative Nebenbedeutung). Wenn aber beispielsweise die negativ bewerteten Begriffe *geizig* und *verschwenderisch* die Extreme der Dimension *Ausgabefreudigkeit* bezeichnen, während die positiv getönten Begriffe *sparsam* und *großzügig* die gemäßigeren Ausprägungen charakterisieren, so sollten die beiden negativ bewerteten Begriffe als deskriptiv unähnlicher eingestuft werden als die beiden positiv bewerteten. Entsprechend sollte den deskriptiven Aspekten bei negativ bewerteten Zentralbegriffen ein größeres Gewicht zukommen als bei positiven Zentralbegriffen. Dies wiederum sollte sich im Hinblick auf Tabelle 3 in stärker positiven Differenzscores (Studie 2) sowie in höheren Wahrscheinlichkeitsschätzungen (Studien 3 und 4) auswirken.

Zur Überprüfung dieser Hypothesen wurden 3 t-Tests für unabhängige Stichproben berechnet, in welche jeweils die 80 Zentralbegriffe als Beobachtungen eingingen. Es wurden hierbei die Differenzscores (Studie 2) bzw. die Wahrscheinlichkeitsschätzungen (Stu-

dien 3 und 4) für die 40 negativ bewerteten Zentralbegriffe mit den entsprechenden Scores für die 40 positiv bewerteten Zentralbegriffe verglichen. Alle drei Vergleiche erbrachten signifikante Mittelwertsdifferenzen in der erwarteten Richtung, $t(78) = 2.29$, $p < .05$ für Studie 2; $t(78) = 4.71$, $p < .001$ für Studie 3; $t(78) = 3.53$, $p < .001$ für Studie 4. Mithin zeigten sich systematische Tendenzen in dem Sinne, daß bei negativ bewerteten Zentralbegriffen die Ähnlichkeits- und Wahrscheinlichkeitsschätzungen stärker von der deskriptiven Bedeutungskomponente abhängen, als dies bei den positiv bewerteten Zentralbegriffen der Fall war. Eine Interpretation im Sinne einer größeren Extremität der negativ bewerteten Eigenschaften liegt nahe.

Diskussion

Die Befunde der hier berichteten Studien lassen sich wie folgt zusammenfassen: Es ist möglich, Quartette von Traitworten so zusammenzustellen, daß deskriptiven Aspekten für die Ähnlichkeitsrelationen zwischen den Worten eine größere Bedeutung zukommt als evaluativen Aspekten. Damit beinhalten unsere Ergebnisse zunächst einmal eine transkulturelle Replikation dieses zuerst von Peabody berichteten Befundes. Darüber hinaus ließ sich die postulierte Eigenschaft der Adjektivquartette in einer Kreuzvalidierung (Studie 4) bestätigen und schließlich führte ein anderer methodischer Zugang zu diesem Problem, nämlich der in Studie 2 gewählte, zu vergleichbaren Ergebnissen. Somit sprechen die Befunde deutlich gegen die Annahme einer absoluten Prädominanz evaluativer Gesichtspunkte, wie sie mitunter durch Untersuchungen mit dem semantischen Differential nahegelegt wird. Wird den Probanden durch die Art der vorgegebenen Entscheidungsalternativen vielmehr die Möglichkeit geboten, auch Bedeutungsähnlichkeiten nicht-evaluativer Art zu berücksichtigen, so machen sie von dieser Möglichkeit Gebrauch, und zwar mitunter in einem Ausmaß, welches den Einfluß evaluativer Bedeutungsinhalte überspielt. Dieses Überwiegen deskriptiver Bedeutungsaspekte scheint sich desto eher nachweisen zu lassen, je eindeutiger die Beziehungen zwischen den Begriffen in deskriptiver Hinsicht ausfallen. Dieser bereits in früheren Untersuchungen (Felipe, 1970) bestätigte Zusammenhang zeigte sich in der vorliegenden Studie darin, daß bei negativ bewerteten — und vermutlich deskriptiv extremeren — Zentralbegriffen die evaluative Komponente eine vergleichsweise geringere Rolle spielte als bei positiv bewerteten Zentralbegriffen, welche eher die gemäßigten Merkmalsausprägungen kennzeichnen dürften.

Die vorliegende Untersuchung wirft Fragen der Art auf, wie es zur Wahrnehmung deskriptiver Ähnlichkeiten und zu den Traitinferenzen kommt. Die hier dargestellten Studien sind in dieser Hinsicht wenig aussagekräftig. Befunde anderer Studien (Borkenau, 1986; Borkenau & Ostendorf, 1987) legen jedoch nahe, daß die Ähnlichkeit von Traitbegriffen mit dem Ausmaß einhergeht, in dem beide Begriffe sich auf überlappende Verhaltensbereiche beziehen. Deshalb geht möglicherweise der von Peabody (1967) berichtete und hier replizierte Befund auf den Umstand zurück, daß die Probanden bei einer Wahl im Sinne evaluativer Konsistenz implizit widersprüchliche Aussagen über das Verhalten der Zielperson machen würden. Davor aber scheuen sie offenbar dann zurück, wenn die-

se Widersprüchlichkeit offen auf der Hand liegt. Bei nachlassender deskriptiver Ähnlichkeit scheint demgegenüber die evaluative Komponente, welche im Sinne einer „guten Gestalt“ wirkt (vgl. Asch, 1946), die Oberhand zu gewinnen (Felipe, 1970).

Adjektivquartette der hier untersuchten Art sind zunächst einmal in allgemein- und sozialpsychologischer Hinsicht interessant, insofern sie Aufschlüsse über die Verarbeitung von Information über Personen liefern können. So zeigt etwa die vorliegende Studie, daß Halo-Effekte eine desto geringere Wirkung entfalten, je eindeutiger die Ähnlichkeitsbeziehungen hinsichtlich der deskriptiven Aspekte der Wortbedeutungen sind. Ein solcher Befund wird etwa von dem Wahrnehmungs- und Urteilsmodell von Upmeyer (1981) vorhergesagt.

Solche Adjektivquartette bieten möglicherweise aber auch differentialpsychologisch interessante Perspektiven und zwar dann, wenn sich Individuen identifizieren lassen, welche sich eher an deskriptiver Stimmigkeit orientieren und andere, welche sich eher an evaluativen Aspekten orientieren. Die Orientierung von Selbsteinschätzungen an ihrer positiven Bewertung wird gewöhnlich als sozial erwünschtes Antworten bezeichnet und ist traditionell Forschungsgegenstand der Psychodiagnostik. Zahlreiche Soziale-Erwünschtheits- bzw. Lügenskalen wurden bereits entwickelt, um diese Tendenz zur positiven Selbstdarstellung auf Kosten der deskriptiven Stimmigkeit zu erfassen. Das seit Crowne & Marlowe (1960) übliche Verfahren ist hierbei, deskriptive Stimmigkeit über Verhaltenswahrscheinlichkeiten zu bestimmen: Wer in einer Sozialen-Erwünschtheitsskala oft angibt, erwünschtes Verhalten zu zeigen, welches selten auftritt, dem wird eine sozial erwünschte Antworttendenz unterstellt. Wer angibt, besser zu handeln als die meisten seiner Mitmenschen, der ist folglich kein „besserer Mensch“, sondern ein Lügner. Diese problematische Rationale traditioneller Lügenskalen dürfte auch ihrer Akzeptanz in der Öffentlichkeit entgegenwirken.

Demgegenüber bieten Adjektivquartette der hier konstruierten Art die Möglichkeit, deskriptive Stimmigkeit gänzlich anders, nämlich stärker individuumszentriert (als an der sozialen Norm orientiert) zu erfassen. Wer etwa bei einer Frage *Würden Sie sich generell eher als sparsam denn als verschwenderisch bezeichnen?* zustimmend antwortet und das Item *„Würden Sie sich generell eher als großzügig denn als geizig bezeichnen?“* ebenfalls bejaht, wer mithin deskriptive Unstimmigkeiten zugunsten einer positiven Selbstdarstellung in Kauf nimmt, von dem kann man annehmen, daß er sich bei der Beantwortung von Fragebogen vor allem von der impliziten Evaluation leiten läßt. Seine Angaben widersprechen dabei nicht so sehr den allgemeinen Erfahrungen, welche der Autor traditioneller Lügenskalen mit der Schlechtigkeit der Menschen gemacht zu haben glaubt. Vielmehr wären die Angaben eines solchen Probanden *in sich* widersprüchlich. Dies gälte insbesondere dann, wenn gezeigt werden könnte, daß sich deskriptiv ähnliche Begriffe in der Tat auf weitgehend überlappende Verhaltensbereiche beziehen (Borkenau, 1986; Borkenau & Ostendorf, 1987).

Dieser Argumentation läßt sich entgegenhalten, daß das Verhalten von Individuen in hohem Maße situationsspezifisch sei und man etwa je nach Umständen sowohl *sparsam* als auch *großzügig* sein könne, etwa *sparsam* in Bezug auf allgemeine Ausgabenpolitik und *großzügig* als Gastgeber. Allein, läge ein solches Verhaltensmuster vor, so würde dar-

aus keinesfalls zwingend folgen, daß die Zielperson stets durch den positiv bewerteten Pol der beiden Skalen angemessen charakterisiert würde. Ebenso gut könnte sie ja bei der Frage *Würden Sie sich generell eher als sparsam denn als verschwenderisch bezeichnen?* an ihre Rolle als Gastgeber denken und die Frage verneinen. Analog könnte sie bei der Frage *Würden Sie sich generell eher als großzügig denn als geizig bezeichnen?* an ihre allgemeine Ausgabenpolitik denken und diese Frage ebenfalls verneinen. Allgemeiner ausgedrückt: Wenn einige Probanden sich systematisch an solche Situationen erinnern, in denen sich ihr Verhalten durch den positiv bewerteten Pol der Skala beschreiben läßt, so ist eben dies ein Element ihres sozial erwünschten Antwortverhaltens.

In seiner klassischen Arbeit versuchte Peabody (1967) zu zeigen, daß evaluative Bedeutungsaspekte von Trait-Adjektiven für einschlägige Trait-Inferenzen unerheblich seien. Er demonstrierte aber wohl nur die Möglichkeit, solche Wahlalternativen zu konstruieren, in denen der Aspekt der evaluativen Stimmigkeit von dem der deskriptiven Stimmigkeit überlagert wird. Keineswegs wurde damit die allgemeine Irrelevanz von Aspekten evaluativer Stimmigkeit bei Trait-Inferenzen nachgewiesen. Uns Autoren erscheint es nun als interessante Perspektive, Unterschiede in der Gewichtung von deskriptiver Stimmigkeit einerseits und positiver Evaluation andererseits als Persönlichkeitsvariable zu untersuchen. Die hier dargestellten Adjektivquartette können sich dabei als ausgesprochen hilfreich erweisen. Entsprechend verstehen wir die hier dokumentierten Konstruktionsschritte für 20 Adjektivquartette als Vorarbeiten zur Entwicklung einer neuartigen Sozialen-Erwünschtheitsskala. Einige Leser dieser Zeitschrift mögen solche Adjektivquartette freilich für gänzlich andere Zwecke verwenden wollen. Viele solcher Verwendungsweisen lassen sich denken.

Literatur

- Asch, S. E. (1946): Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258–290.
- Borkenau, P. (1986): Toward an understanding of trait interrelations: Acts as instances for several traits. *Journal of Personality and Social Psychology*, 51, 371–381.
- Borkenau, P. & Amelang, M. (1985): The control of social desirability in personality inventories: A study using the principal-factor deletion technique. *Journal of Research in Personality*, 19, 44–53.
- Borkenau, P. & Amelang, M. (1986): Zur faktorenanalytischen Kontrolle sozialer Erwünschtheitstendenzen. Eine Untersuchung anhand des Freiburger-Persönlichkeits-Inventars. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 7, 17–28.
- Borkenau, P. & Ostendorf, F. (1987): Retrospective estimates of act frequencies: How accurately do they reflect reality? *Journal of Personality and Social Psychology*, 52, 626–638.
- Crowne, D. P. & Marlowe, D. (1960): A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354.
- Edwards, A. L. (1953): The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37, 90–93.
- Edwards, A. L. (1977): Comments on Shweder's „Illusory correlation and the MMPI controversy“. *Journal of Consulting and Clinical Psychology*, 45, 925–929.
- Edwards, A. L. & Diers, C. J. (1962): Social desirability and the factorial interpretation of the MMPI. *Educational and Psychological Measurement*, 22, 501–509.

- Felipe, A. I. (1970): Evaluative versus descriptive consistency in trait inferences. *Journal of Personality and Social Psychology*, 16, 627—638.
- Goldberg, L. R. (1982): From Ace to Zombie: Some explorations in the language of personality. In: C. D. Spielberger & J. N. Butcher (Hrsg.), *Advances in personality assessment* (Bd. 1, S. 203—234). Hillsdale, N. J.: Erlbaum.
- Osgood, C. E. (1962): Studies on the generality of affective meaning systems. *American Psychologist*, 17, 10—28.
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957): *The measurement of meaning*. Urbana: University of Illinois Press.
- Paulhus, D. H. (1981): Control of social desirability in personality inventories: Principal factor deletion. *Journal of Research in Personality*, 15, 383—388.
- Peabody, D. (1967): Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology Monographs*, 7 (Whole No. 664).
- Rosenberg, S. & Olshan, K. (1970): Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, 16, 619—626.
- Shrout, P. E. & Fleiss, J. L. (1979): Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420—428.
- Thorndike, E. L. (1920): A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25—29.
- Upmeyer, A. (1981): Perceptual and judgmental processes in social contexts. In: L. Berkowitz (Hrsg.), *Advances in experimental social psychology* (Bd. 14, S. 257—308). New York: Academic Press.

Autorenhinweise

Die in diesem Artikel berichteten Studien wurden durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft gefördert (Az. Bo 774/2-1).

Diesen Artikel betreffende Korrespondenz sollte an folgende Adresse gerichtet werden: Dr. Peter Borkenau, Abteilung Psychologie der Universität Bielefeld, Postfach 8640, 4800 Bielefeld 1.