

Materialien aus der Bildungsforschung Nr. 4

Heinrich Meulemann

**WORTBEDEUTUNGSVERSTÄNDNIS UND
WORTBEDEUTUNGSEXPLIKATION**

**Eine empirische Analyse zweier Aspekte des
Sprachverhaltens und ihrer sozialen Determinanten
im Rahmen der Theorie der linguistischen Codes**

Teil I

**Max-Planck-Institut für Bildungsforschung
Berlin 1976**

F 761.1928-145

Materialien aus der Bildungsforschung

In dieser Reihe veröffentlicht das Max-Planck-Institut für Bildungsforschung Arbeitsmaterialien (Diskussionsgrundlagen und Dokumentation), die nicht den Charakter abgeschlossener Forschungsberichte tragen, aber dem jeweils interessierten Fachpublikum zugänglich gemacht werden sollen.

Bestellungen werden erbeten an die Verwaltung des Instituts, 1 Berlin 33, Lentzeallee 94, bei gleichzeitiger Überweisung von 33,- DM für Teil I und Teil II auf das Konto 09100588 der Sparkasse der Stadt Berlin West.

Nachdruck, auch auszugsweise, ist nur mit der Zustimmung des Instituts gestattet.

Vorbemerkung.

=====

Die vorliegende Arbeit besteht im wesentlichen in der Analyse eines soziolinguistischen Experiments, das zusammen mit drei weiteren soziolinguistischen Experimenten von Dr. U. Oevermann im Rahmen des von Oevermann, Krappmann und Kreppner bearbeiteten Projekts "Elternhaus und Schule" am Max-Planck-Institut für Bildungsforschung Berlin geplant und von dem Projektleiter mit Soziologie-Studenten der Universität Frankfurt durchgeführt wurde. Die Durchführung der Untersuchung sowie ihre Auswertung und Analyse sind das Ergebnis einer Teamarbeit, ohne die ein von einem einzelnen verfaßter Bericht, wie er hier vorliegt, natürlich nicht möglich wäre. Jürgen Krambeck hat zusammen mit dem Verfasser an der Durchführung und Analyse der Untersuchung gearbeitet. Herr Dr. Oevermann hat dem Verfasser die Möglichkeit gewährt, im Rahmen des Projekts "Elternhaus und Schule" als Doktorand mitzuarbeiten. Als Projektleiter war er an allen einzelnen Schritten, weit über die bei der Leitung von Examensarbeiten üblichen Hilfen hinaus, beteiligt. Insbesondere hat er die elektronischen Rechenarbeiten geleitet und die notwendigen Rechenprogramme verfaßt.

Zur Darstellung der Arbeit

=====

Da es für den Nachvollzug der empirischen Analysen behilflich ist, Text und Tabellen zugleich zu verfolgen, wurde die Arbeit in einen Text- und einen Tabellenband aufgeteilt. Im Textband wird auf alle Tabellen verwiesen, die an der betreffenden Seite für das Verständnis wichtig sind. Die Tabellen wurden nach der Abfolge der Analysen im Text angeordnet und durchnummeriert. Dem Tabellenband wurde ein Verzeichnis beigegeben, in dem alle Tabellen den Abschnitten des Texts zugeordnet sind, zu denen sie hauptsächlich - unabhängig von weiteren Verweisen - gehören. Der Tabellenband wird auf diese Weise unabhängig vom Textband lesbar. Um die durchgängige Lesbarkeit des Tabellenbandes nicht zu zerstören, wurde die Zuweisung weniger wichtiger Tabellen in einen Anhang vermieden. Stattdessen sind alle wesentlichen - rund ein Fünftel - im Verzeichnis des Tabellenbands durch ein " = " vor der Tabellen-Nummer gekennzeichnet.

Abbildungen stehen am Schluß des Tabellenbands; ein Verzeichnis findet sich am Anfang. Die Abbildungen enthalten grundsätzlich nur Aussagen, die auch in Tabellen dargestellt sind; zu jeder Abbildung ist die zugehörige Tabelle angegeben. Abbildungen wurden nicht nur für wichtige Sachverhalte, sondern auch für in Tabellenform schwer übersichtliche Zusammenhänge angefertigt.

I N H A L T S V E R Z E I C H N I S

Erstes Kapitel: <u>Begründung des Untersuchungsansatzes</u>		1 - 57
1.	Der theoretische und der praktische Bezug der Theorie der linguistischen Kodes in älteren und neueren Untersuchungen	1
2.	Empirische Ergebnisse zur Entwicklung des Wortschatzes - ein Bericht über ausgewählte Literatur zu den Problemen des Experiments	10
2.1.	Sprachentwicklung und Genese des Bedeutungssystems	10
2.1.1.	Kognitive Fähigkeiten, Bedeutungslernen und Sprachgebrauch	12
2.1.2.	Aufgaben und Mechanismen des Bedeutungslernens	16
2.1.3.	Die Entwicklung des Wortschatzes innerhalb der Sprachentwicklung	22
2.2.	Soziale Einflüsse auf die Wortschatzentwicklung	30
2.2.1.	Verschiedene Aspekte des Wortschatzes von Kindern in Abhängigkeit von dichotomer Sozialschicht	30
2.2.2.	Differenzierung sozialer Einflüsse auf den Wortschatz von Kindern	42
3.	Die Erklärung der schichtspezifischen Differenzen im Wortverständnis und im Wortgebrauch durch eine Theorie der sozialen Strategien des Sprachgebrauchs	50
Zweites Kapitel: <u>Ansatz, Methoden und Durchführung der Untersuchung</u>		58 - 74
1.	Design und Hypothesen des Experiments	58
2.	Stichprobe und Variable des Experiments	63
3.	Durchführung der Untersuchung	68
3.1.	Voruntersuchungen und Festlegung der Experimentalgruppen der zweiten Messung	68
3.2.	Parallelisierung der Versuchspersonen beider Messungen	69
3.3.	Technischer Ablauf beider Messungen	71
3.4.	Nacherhebung und Änderung der Parallelisierung der zweiten Messung	73

Drittes Kapitel: <u>Darstellung der Ergebnisse der ersten und zweiten Messung</u>	75 - 312
1. <u>Operationalisierung der Prüfvariablen der aktiven Bedeutungserklärung</u>	75 - 116
1.1. Konstruktion von Kategorien der Bedeutungserklärung	75
1.1.1. Vortests	75
1.1.2. Trennung von Bedeutungsverständnis und Bedeutungsexplikation	77
1.1.3. Auswertungsskalen für Bedeutungsverständnis und Bedeutungsexplikation	78
1.1.4. Reliabilität der Kategorisierung	87
1.1.5. Probleme der Kategorisierung	93
1.2. Bildung und Überprüfung von Maßen	95
1.2.1. Dimensionen der Maßkonstruktion	95
1.2.2. Aussagekraft der Maße	101
1.2.3. Überprüfung der Maße: Mittelwerte und Standardabweichungen	102
1.2.4. Überprüfung der Maße: Interkorrelationen	103
1.2.5. Schichtunterschiede der Interkorrelationen der Maße	109
1.2.6. Überprüfung der Maße: Faktorenanalyse	111
1.2.7. Zusammenfassung	115
2. <u>Darstellung der Untersuchungsgruppe der zweiten Messung</u>	117-- 141
2.1. Darstellung der Untersuchungsgruppe ohne Berücksichtigung des Erhebungszeitpunktes	117
2.1.1. Experimentalstichproben	117
2.1.2. Kontrolle des Schichteinflusses	121
2.1.3. Beziehungen der Kontrollvariablen zu anderen Variablen	122
2.1.4. Verteilung der Versuchspersonen auf eine differenzierte Einteilung und innerhalb der Dimensionen des Filterindex sozialer Schicht	124
2.1.5. Verteilung der Versuchspersonen nach den Arbeitsplatzkategorien für den Vater und für die Mutter	127
2.1.6. Verteilung der Versuchspersonen auf Schulen und Schulklassen	131
2.2. Mögliche Einflüsse der Nacherhebung	133

VII

3.	<u>Kurze Wiederholung der Hypothesen der ersten und zweiten Messung</u>	142
4.	<u>Ergebnisse der ersten Messung</u>	143 - 153
4.1.	Ergebnisse der ersten Messung für den gesamten Test	143
4.2.	Ergebnisse der ersten Messung für einzelne Worte und Wortgruppen	147
5.	<u>Ergebnisse der zweiten Messung: Bedeutungserklärung</u>	154 - 291
5.1.	Zu den in der zweiten Messung verwandten Profverfahren	154
5.2.	Einflüsse der dichotomen Schichteinteilung auf die Gesamtleistung der zweiten Messung nach dem Wilcoxon-Test	157
5.2.1.	Schichtunterschiede bei Kontrolle der passiven Wortschatzleistung	158
5.2.2.	Schichtunterschiede bei Kontrolle verschiedener Intelligenz- und Schulleistungen	159
5.3.	Einflüsse der dichotomen Schichteinteilung, des Geschlechts, der Kommunikationsform und der Experimentalgruppen auf die Gesamtleistung der zweiten Messung nach der Varianzanalyse	163
5.3.1.	Überblick über die Einzeleffekte und Interaktionen der Faktoren: Varianzanalysen mit allen Experimentalfaktoren	163
5.3.2.	Einflüsse der dichotomen Schichteinteilung	167
5.3.3.	Einflüsse des Faktors Geschlecht	174
5.3.4.	Einflüsse der Kommunikationsform	178
5.3.5.	Einflüsse der Experimentalgruppen	181
5.4.	Einflüsse der dichotomen Schichteinteilung auf Einzelleistungen und besondere Aspekte der zweiten Messung	187
5.4.1.	Schichtunterschiede für einzelne Worte	189
5.4.2.	Schichtunterschiede innerhalb der Wortgruppen extraverbaler Explizierbarkeit	193
5.4.3.	Schichtenspezifität und Reliabilität des Testmaterials	196
5.4.4.	Schicht- und Wortgruppenunterschiede bei den Differenzen zwischen Verständnis und Explikation	203
5.4.5.	Schichtunterschiede in der Verteilung der Antworten auf die Kategorien der Auswertung	208
5.4.6.	Schichtunterschiede der Anzahl falscher Antworten	215
5.4.7.	Bewertung der Einzelanalysen	218

5.5.	Zusammenfassung der Ergebnisse des dichotomen Schichtvergleichs	220
5.6.	Unabhängig von dichotomer Schicht erfaßte Einflüsse der gesellschaftlichen Lage der Eltern auf die Leistungen des Kindes	222
5.6.1.	Einflüsse einer weiter aufgegliederten Schichteinteilung	224
5.6.2.	Unterschiedliche Einflüsse der Einzelindikatoren sozialer Schicht	238
5.6.3.	Einflüsse der Arbeitsplatzsituation des Vaters auf die Wortschatzleistung der ersten und zweiten Messung	248
5.6.4.	Einflüsse der Arbeitsplatzsituation der Mutter auf die Experimentalleistung der zweiten Messung	257
5.6.5.	Der Einfluß des Buchbesitzes als eines Indikators der sozio-kulturellen Lebenswelt auf die Experimentalleistung der zweiten Messung	264
5.6.6.	Abschließende Beurteilung der Einflüsse unabhängig von sozialer Schicht erfaßter sozialer Faktoren	268
5.7.	Einflüsse der Kommunikations- und Kontrollweisen der Mutter auf die aktiven Worterklärungen des Kindes	271
5.8.	Zusammenhänge zwischen Intelligenzleistungen, Schulnoten und aktiver Worterklärung	276
5.8.1.	Zusammenhänge zwischen Intelligenzleistungen und aktiver Worterklärung	277
5.8.2.	Zusammenhänge zwischen Schulnoten und der aktiven Worterklärung	280
5.8.3.	Die relative Bedeutung der Intelligenztests und der Schulnoten für die aktive Worterklärung	285
5.8.4.	Die relative Bedeutung von Intelligenztests bzw. Schulnoten und Schichtindikatoren für die aktive Worterklärung	287
6.	<u>Ergebnisse der zweiten Messung: Sprachliche Mittel der Erklärung</u>	292 - 312
6.1.	Ziele der Analyse der sprachlichen Mittel der Erklärung	292
6.2.	Stellung der Analyse sprachlicher Mittel im Aufbau des Experiments	297
6.3.	Hypothesen und Auswertung der sprachlichen Mittel der Erklärung	302
6.4.	Ergebnisse des Schichtvergleichs der sprachlichen Mittel der Erklärung	306
6.5.	Zusammenfassung der Analyse der sprachlichen Mittel der Erklärung	311

Viertes Kapitel: Ergebnisse der dritten Messung

319 - 429

1.	<u>Operationalisierung der Prüfvariablen</u> <u>"kategoriale Differenzierung"</u>	
1.1.	Konstruktion der Testaufgaben	313
1.2.	Auswertung und Bildung von Maßen	318
2.	<u>Darstellung der Untersuchungsgruppe der</u> <u>dritten Messung</u>	329
2.1.	Kontrolle des Schichteinflusses	330
2.2.	Beziehungen der Kontrollvariablen WST sowie der Kontrollvariablen NONVER zu anderen Variablen	331
2.3.	Verteilung der Versuchspersonen auf eine diffe- renzierte Schichteinteilung und auf die Schulen und Schulklassen	334
3.	<u>Ergebnisse der dritten Messung</u>	335
3.1.	Kurze Wiederholung der Hypothese der dritten Messung und ihres Zusammenhangs mit der zweiten Messung	336
3.2.	Quantitative Analyse der Ergebnisse der dritten Messung	336
3.3.	Qualitative Analyse des "kategorialen Präzisionstests"	348
3.4.	Qualitative Analyse des "differenzierten Verständnistests"	349
3.4.1.	Ziele und Instrumente der qualitativen Analyse	349
3.4.2.	Ergebnisse der qualitativen Analyse	362
3.4.3.	Zusammenfassung der qualitativen Analyse	390
3.5.	Zusammenfassung und Bewertung der quantitativen und qualitativen Analysen sowie der beiden Testfälle	393
3.6.	Zusammenhänge zwischen der Leistung in der dritten Messung und anderen Faktoren als der dichotomen Sozialschicht	397
3.6.1.	Dimensionen der Sozialschicht und "kategoriale Differenzierung"	399
3.6.2.	Intelligenz, Schulleistung und "kategoriale Differenzierung"	405
3.6.3.	Sozialschicht oder Wortkenntnis: Ergebnisse der multiplen Regressionsanalyse	410
3.6.4.	Zusammenfassende Beurteilung des Einflusses von Schicht, Intelligenz und Wortschatz auf die "kategoriale Differenzierung"	415

3.7.	Einflüsse der Kommunikations- und Kontrollweisen der Mutter auf die "kategoriale Differenzierung" des Kindes	417
------	--	-----

Fünftes Kapitel: Diskussion und Bewertung der Ergebnisse

430 - 487

1.	Mögliche Ursachen für die den Hypothesen und den Ergebnissen ähnlicher Untersuchungen widersprechenden Ergebnisse in der Anläge des Wortschatzexperimentes	432
1.1.	Wahl der Schicht- und Altersgruppen	432
1.2.	Eliminierung möglicher Schichtunterschiede durch die Kontrolle des passiven Wortschatzes	439
1.3.	Mängel der Operationalisierung und Schwierigkeiten der Diskriminierung zwischen aktivem und passivem Wortschatz	447
2.	Kritische Diskussion des Evokationsmodells im Rahmen des Wortschatzexperimentes	454
3.	Konsequenzen der Ergebnisse des Wortschatzexperimentes für die Theorie der linguistischen Codes	461
3.1.	Die Abhängigkeit des Sprachgebrauchs von typischen Sozialbeziehungen und von typischen Situationen als Problem bei der Erhebung schichtspezifischer Sprachproben	462
3.2.	Nivellierung schichtspezifischer Arten des Sprachgebrauchs durch intentionale Elemente der familialen und schulischen Sozialisation	470
3.3.	Allgemeine Gesichtspunkte der Theorie der linguistischen Codes	473
4.	Praktische Folgerungen aus den Ergebnissen auf sozialpolitischem und pädagogischem Gebiet	483
	<u>Verzeichnis der benutzten und zitierten Literatur</u>	488

ANHANG: Zur zweiten Messung

A.1.	Testformen des Frankfurter Wortschatztests	495
2.	Für die Auswertung verbindliche Definitionen der Stichworte des WST	501
3.	Testanweisung des aktiven Wortschatztests	505
B.	Beschreibung des "komplexen Kategoriensystems"	508
C.	Bedeutungserklärungen der Vpn. zu ausgewählten Stimulusworten des Frankfurter WST mit den Bewertungen des Verständnisses und der Explikation	515

D.	Kategorien der Auswertung der sprachlichen Mittel der Erklärung:	
	1. Beschreibung der Kategorien	533
	2. Beschreibung der aus den Kategorien gebildeten Quotienten	542

ANHANG: Zur dritten Messung

E.	Testformular mit allen quantitativen und qualitativen Bewertungen	546
F.	Testformular mit Verteilung der Antworten der Vpn. beider Sozialschichten	559
G.	Aufgabenpaare des "kategorialen Präzisionstests" mit Verteilung der Antworten der Vpn. beider Sozialschichten	572

Erstes Kapitel: Begründung des Untersuchungsansatzes.
=====

In der folgenden Arbeit werden Hypothesen aus einem begrenzten Bereich einer soziologischen Theorie, der Theorie der linguistischen Kodes von Bernstein, überprüft. Bevor die Ergebnisse dargestellt werden können, soll (1) der breitere soziologische und sozialpolitische Bezugsrahmen, in dem die Theorie Bernstein verstanden sein will, angedeutet sowie (2) die spezifische Beschränkung der Untersuchung auf den Bereich des Wortschatzes innerhalb der Theorie Bernsteins begründet und in ihren Konsequenzen analysiert werden. Der ersten Aufgabe ist der folgende, der zweiten alle weiteren Abschnitte dieses einleitenden Kapitels gewidmet.

1. Der theoretische und der praktische Bezug der Theorie der linguistischen Kodes in älteren und neueren Untersuchungen.

Die Theorie der linguistischen Kodes stand seit ihren Anfängen zugleich in einem theoretisch-soziologischen wie in einem praktisch-pädagogischen und praktisch-sozialpolitischen Zusammenhang (Bernstein 1971: Introduction, Oevermann 1968, 1970). Unter einem theoretisch-soziologischen Gesichtspunkt werden schichtspezifische Arten des Sprachgebrauchs, "linguistische Kodes", als ein Erklärungsprinzip für schichtspezifische Sozialisationsprozesse angesehen; prinzipiell nicht-sprachliche Einstellungen zum Sprach- und Symbolgebrauch gelten als ein Vermittlungsmechanismus, durch den sich objektive Bedingungen der Sozialstruktur in individuelles Handeln und Meinen und auf Dauer auch in individuellen Lern- und Lebenserfolg umsetzen. Die wahrscheinlich aus den Bedingungen der Berufs- und Arbeitswelt herrührenden schichtspezifischen Formen familiärer Kommunikation und Interaktion können auf diese Art die seit langem bekannte Tatsache des geringeren Schulerfolgs von

Kindern aus unteren Schichten, selbst bei Kontrolle der Intelligenz erklären (Oevermann 1966, 1970 S. 13, für eine Zusammenfassung amerikanischer Untersuchungen, sowie Jencks et al. 1973, für eine ausführliche Bearbeitung dieser Untersuchungen). Innerhalb eines breiteren soziologischen Bezugsrahmens leistet die Sozialisationsforschung einen Beitrag zur Untersuchung der Mechanismen und Erscheinungsformen von sozialer Ungleichheit, von sozialer Kontrolle und von sozialem Wandel. Wenn die Steuerung der Interaktion durch kulturelle Symbole und damit die Notwendigkeit kultureller, im Gegensatz zu biologischer, Reproduktion Merkmale menschlicher Gesellschaften sind, so kann man den Gegenstand der Sozialisationsforschung - wie in einem engeren Sinn auch den Gegenstand der Soziologie institutionalisierter Bildung - in den Prozessen sehen, durch die die kulturelle Reproduktion zugleich eine soziale Reproduktion wird, eine Reproduktion der bestehenden Herrschaftsverhältnisse und ihrer Anerkennung bei den Beherrschten (Bourdieu/Passeron 1973).

Der praktische Bezug der Theorie der linguistischen Codes wird in der skizzierten breiteren soziologischen Betrachtung einsichtig. Er ergibt sich aus dem sozialpolitischen Postulat der Chancengleichheit, das in seiner strengeren Form nicht nur die gleichen Zugangs-, sondern auch die gleichen Erfolgchancen für die Kinder verschiedener Stände und Klassen im öffentlichen Bildungswesen fordert. Gleiche Erfolgchancen für alle Kinder bedeutet aber nicht nur eine gerechte, nicht offen (d.h. psychometrisch feststellbar) diskriminierende schulische Leistungsbeurteilung, sondern auch die Schaffung gleicher kognitiver und motivationaler Voraussetzungen in der vor-schulischen Sozialisation sowie die Beseitigung der latenten Diskriminierung durch die schulisch kanonisierte, im Doppelsinn "herrschende" Kultur. Positiv folgt aus der erweiterten Interpretation des Postulats der Chancengleichheit die Einrichtung vorschulischer Erziehungsprogramme mit Schwerpunkten in unterprivilegierten Regionen und die Institutionalisierung von Entscheidungsprozessen über Bildungsinhalte, die wissenschaftlich kontrolliert sind und interessierte soziale Gruppen zur folgen-

reichen Rechtfertigung ihrer Wünsche zwingen. Für beide - und nicht wie es scheinen könnte, allein für die erste - Forderungen ergeben sich Konsequenzen aus Forschungen zu schichtspezifischen Sozialisationsprozessen.

Daß die pädagogische Praxis im Vorschulbereich - und ebenso im Bereich der Grundschule - sich an der Sozialisationsforschung orientieren muß, liegt auf der Hand. Die psychologische Perspektive der Entwicklung individueller Fähigkeiten wird durch die von der Soziologie propagierte Sozialisationsforschung nicht durch die Einführung einer Variablen "soziale Schicht" erweitert - auf diese Weise werden psychische Prozesse nur mit theoretisch unverbundenen sozialen Rahmenbedingungen konfrontiert -, sondern durch die Untersuchung der Genese von Motivationen und Formen der Ichorganisation in der Struktur der familialen Interaktion, wie sie zum ersten Mal von Parsons und Bales (1953) theoretisch dargelegt wurde. Linguistische Codes sind in diesem Zusammenhang aus schichttypischen familialen Interaktionsmustern erwachsende Einstellungen zum Symbolgebrauch, die - situationsübergreifend, und insofern "formal" - die Explikation von Handlungsmotiven und die Darstellung von Individualität sowie die Wahrnehmung und Lösung von Problemen, kurz: Fähigkeiten des kommunikativen und instrumentellen Handelns, steuern. Mit einem derartigen Begriff der "linguistischen Codes" - und ähnlich mit dem vergleichbaren Begriff des "Habitus" (Bourdieu/Passeron 1973) - soll die soziale Genese individuell-konstanter Verhaltensstrategien empirisch zugänglich werden, ohne daß ihre relative Autonomie geleugnet und die Verhaltensorientierungen von Individuen mit den Normen und Werten des jeweils die Umgebung bildenden Sozialsystems identifiziert würden. Eine Bestätigung dieser Theorie würde die Institutionalisierung von vorschulischen Programmen nahelegen, die über ein "Training" von kognitiven und sprachlichen Leistungen hinausgehend versuchen, die sozialen Bedingungen für Strategien der bewußten Informationsverarbeitung und des autonomen Sprachgebrauchs herzustellen, etwa in Form von Rollenspielen und Verbalisierungsübungen, die weniger an Sprachformen als an Sprachfunktionen sich orientieren. Vielleicht liegt in der Konzentration auf kognitive

Leistungen eine Erklärung für die Kurzfristigkeit des Erfolgs bisheriger Vorschulprogramme und im Übergang auf sozial produzierte und reproduzierbare Einstellungen eine Chance, länger dauernde Erfolge zu garantieren.

Ebenso wie für die pädagogische Praxis im Vorschulbereich ergeben sich Konsequenzen der Sozialisationsforschung für die Entscheidung über Bildungsinhalte. Mit dem Schuleintritt trifft das Kind aus der Unterschicht nicht nur auf ihm bis dahin fremde Formen des Umgangs mit Symbolen, sondern auch schulische Inhalte, die ihm typischerweise vom Elternhaus her weniger vertraut sind als dem Mittelschichtkind; diese Distanz wird in der Schullaufbahn des Unterschichtkindes eher zu- als abnehmen. Entgegen einer verbreiteten Kritik (zuletzt in Niepold 1970) ist in der Theorie Bernsteins die Vermittlung von Inhalten durch linguistische Kodes ebenso berücksichtigt wie die Bedeutung der linguistischen Kodes für die Entwicklung "formaler" Fähigkeiten des sozialen Handelns; Bernstein betont, daß durch die linguistischen Kodes sowohl die Formen bestimmter Sozialbeziehungen symbolisch transformiert und bekräftigt, wie auch bestimmte Normen und Werte vermittelt werden (siehe als frühen Beleg: Bernstein 1958, sowie neuerdings: Bernstein 1969). Wenn Untersuchungen zum schichtspezifischen Sprachgebrauch besondere Qualitäten der Sprache der Unterschicht oder besondere Interessen der Unterschichtkinder aufzeigen, so hätte das auch für die wissenschaftliche Kontrolle curricularer Entscheidungen im Vor- und Grundschulbereich Konsequenzen.

Vor allem in zwei sozialen Formen kann die Theorie der linguistischen Kodes - wie jede Theorie schichtspezifischer Sozialisationsprozesse - auf die pädagogische Praxis Einfluß gewinnen: durch eine Verbreitung der Grundgedanken einer Theorie, die pädagogische Lehrmeinungen aufklären und pädagogisches Handeln subkutan verändern kann, und durch die von der Politik sanktionierte und von der Bildungsverwaltung kontrollierte Einführung wissenschaftlich legitimierter, spezifischer pädagogischer Neuerungen. Die Art ihrer praktischen Wirksamkeit hängt nicht zuletzt von der immanent-

theoretischen Entwicklung einer Theorie ab. Wenn im folgenden der theoretische Stellenwert und die praktische Bedeutung speziell der hier referierten Untersuchung diskutiert werden soll, so ist dazu ein - fast schon historischer¹⁾ - Überblick über die Untersuchungen der Theorie der linguistischen Kodes, ihre Aussagen und ihr Einfluß notwendig.

Die Anzahl der empirischen Untersuchungen, mit denen die Theorie der linguistischen Kodes ihre Popularität begründen konnte, ist erstaunlich gering. Lediglich drei Untersuchungen lagen bis etwa 1968 vor; zu diesem Zeitpunkt hatte die Theorie in der allgemeinen und pädagogischen Öffentlichkeit - als eine Theorie der Sprachbarrieren - breite Resonanz gefunden, die sich zum Teil aus der Hochschätzung der Sprache im traditionellen Bildungskanon und im Bildungsbürgertum erklären läßt und erst erschüttert wurde, als die Theorie der linguistischen Kodes - angesichts der vermuteten Motive ihrer Popularität, aber nicht ihres tatsächlichen Gehalts korrekt - wegen ihrer latenten Funktion der Integration der Arbeiterschaft in das spätkapitalistische Herrschaftssystem kritisiert wurde. Bernstein selber untersuchte Gruppendiskussionen von 15-18jährigen Jungen, die bezüglich der Verzögerungsphänomene (1962a) und der grammatischen Elemente (1962b) in zwei Aufsätzen ausgewertet wurden. Lawton wiederholte bei 12- bis 15jährigen Jungen die Gruppendiskussionsuntersuchung und erhob zusätzlich mündliche Sprachproben in gerichteten Interviews und schriftliche Sprachproben in Aufsätzen und Satzergänzungstests (Zusammenfassung in Lawton 1970, engl. 1968). Oevermann untersuchte 1965 in Deutschland Aufsätze einer 6. Realschulklasse mit einem gegenüber den englischen Untersuchungen stärker differenzierten linguistischen Instrumentarium (Oevermann 1970).

Alle drei genannten Untersuchungen konnten die Voraussagen Bernsteins recht gut bestätigen. Alle drei Untersuchungen aber hatten explorativen Charakter. Ungeachtet methodischer Einwände, die sich

¹⁾ Inzwischen liegt eine zusammenfassende Darstellung der Untersuchungen zur Theorie der linguistischen Kodes vor, die sich an ihrer Entstehungsgeschichte orientiert (Niepold 1970)

im einzelnen gegen jede dieser Untersuchungen vorbringen lassen, wie etwa ungenaue Operationalisierung der Schicht, fehlende Kontrolle des Einflusses der Schule oder des Geschlechts etc., konnten sie auch für die Theorie nur globale Bestätigungen erbringen; sie stellen eine Verbindung zwischen sozialer Schicht und breit streuenden, mehr oder minder systematisch ausgewählten sprachlichen Variablen in einer Vielzahl von Einzelhypothesen her und konnten weder Zusammenhänge zwischen positiven und negativen Ergebnissen noch differenzierte Hypothesen für einzelne Sprachfunktionen und sprachliche Bereiche testen. Trotz ihres für die Theorie nur explorativen Charakters haben diese Untersuchungen in der Öffentlichkeit breite Beachtung gefunden. Sie haben dem Theoriegebäude als Ganzem gleichsam eine pädagogische Legitimation gegeben und wahrscheinlich sogar einen großen Einfluß auf Orientierung und Handeln der Lehrer gehabt, bevor irgendwelche spezifischen Hypothesen bestätigt und in Handlungsanweisungen umgesetzt werden konnten.

Jedoch blieb die theoretische und empirische Weiterentwicklung der Theorie Bernsteins - zumindest in Deutschland - hinter der Resonanz zurück, die sie in der Praxis gefunden hatte. In Deutschland gibt es zwar eine Reihe von - an Bernstein und Oevermann mehr oder minder orientierten - soziolinguistischen Einzelforschungen. Keine dieser Untersuchungen versucht aber, den Zusammenhang zwischen den verschiedenen Dimensionen des Sprachgebrauchs, die in der Theorie der linguistischen Kodes angesprochen sind, empirisch zu sichern oder differenzierte Hypothesen für einzelne Sprachfunktionen oder Sprachbereiche zu testen. Die meisten dieser Untersuchungen beziehen sich eher auf die schichtspezifische Verteilung sprachlicher Merkmale, als auf - prinzipiell nichtsprachliche - Arten des Sprachgebrauchs. Entsprechend ergeben sich aus einigen dieser Untersuchungen für die Theorie Bernsteins vorwiegend positive (Schlee 1973), aus anderen vorwiegend negative (Sommer 1972, König 1972, Bühler 1972) und aus anderen wieder etwa gleich viel positive wie negative Einzelergebnisse (Jäger u.a. 1973, Wiederhold 1971), so daß sie insgesamt für die Theorie schichtspezifischen Sprachgebrauchs un-

schlüssig bleiben. Auf der anderen Seite wurden - ausdrücklich im Anschluß an die drei zuerst genannten Untersuchungen - in der Forschungsgruppe Bernsteins in London und der Forschungsgruppe Oevermanns in Berlin weitere soziolinguistische Untersuchungen durchgeführt, die sich meist auf die gleiche Stichprobe beziehen und eine Reihe von differenzierten Hypothesen über spezifische sprachliche Leistungen untersuchen. Aus den Untersuchungen der Londoner Forschungsgruppe wurden inzwischen verschiedene Berichte publiziert: über den Gebrauch von Fragen bei Müttern und Kindern (Robinson/Rackstraw 1972), über die sprachlich realisierten Formen und Inhalte von Erziehungstechniken bei Müttern und Kindern (Cook-Cumperz 1973, Turner 1973), über die Wahl verschiedener Wortklassen zur Beschreibung von Bildern und zur Lösung von Aufgaben (Brandis/Henderson 1970). Aus den Untersuchungen der Berliner Forschungsgruppe soll im folgenden über ein Experiment berichtet werden, das sich auf den Bereich des Wortschatzes beschränkt und versucht, spezifische Hypothesen über den aktiven und passiven Gebrauch des Wortschatzes, über die mündliche und schriftliche Form sprachlicher Darstellung und die "evokative Funktion" der Sprache zu prüfen.

Im Vergleich zu explorativen Untersuchungen, die eine Generalhypothese an einer Vielzahl von Variablen überprüfen, haben gezielte Experimente, die eine Kombination differenzierter Einzelergebnisse voraussagen, zwar eine höhere Aussagekraft, aber auch ein größeres Risiko. Lassen sich die Hypothesen nicht bestätigen, so ist nicht nur die besondere Ableitung der Hypothese aus der Theorie widerlegt, sondern auch die Suche nach Ad-Hoc-Modifikationen, die die ursprüngliche Ableitung korrigieren und damit die Theorie als Ganzes retten könnten, innerhalb der gegebenen Daten nur schwer möglich. Denn eine

nicht widerlegte Nullhypothese, um es in der Sprache der statistischen Hypothesenprüfung oder der Wissenschaftstheorie des Falsifikationismus auszudrücken (siehe z.B. Hays 1963, Kap. 9), läßt jede beliebige Alternativerklärung offen. Im Falle einer Nichtbestätigung der Voraussagen ist der Ertrag einer Untersuchung um so geringer, je spezifischer diese Voraussagen sind.

Aber auch die praktischen Konsequenzen eines Experiments mit genau spezifizierten Hypothesen sind andere als die einer explorativen Untersuchung. Wenn wir die Verbreitung eines orientierenden theoretischen Rahmens in der Schulpraxis und in der Öffentlichkeit und die Anleitung pädagogischer Maßnahmen durch sozialwissenschaftliche Einzelerkenntnisse als zwei Ebenen der Kommunikation zwischen Theorie und Praxis auseinanderhalten, so überwiegt bei den frühen Untersuchungen die erste, bei den Folgeuntersuchungen die zweite. In der Öffentlichkeit haben wissenschaftliche Erkenntnisse zudem einen entschieden größeren Multiplikator-Effekt als über die wissenschaftliche Anleitung pädagogischen Handelns. Neue Erkenntnisse können auf der ersten Ebene nahezu unbemerkt die Einstellungen der Handelnden neu orientieren, während sie auf der zweiten, weit hinausgehend über die Probleme einer bloßen "technischen Transformation", in die selbstkontrollierten Strategien der pädagogischen Handelnden eingehen müssen. Entsprechend verengt sich der Kreis der Adressaten: Während den frühen Untersuchungen auch von einer diffusen Öffentlichkeit Interesse entgegengebracht wurde, sehen die Folgeuntersuchungen sich nur noch der pädagogischen und soziologischen Fachwelt und den - personell weitgehend mit der Fachwelt identischen - Gremien der Politikberatung und Ausbildungsplanung als Adressaten gegenüber. Aber auch ganz abgesehen von ihrer eingeschränkten Resonanz in der Öffentlichkeit, ist es für Untersuchungen spezifischer Einzelhypothesen zwar möglich, aber keineswegs notwendig, daß ihnen pädagogische Maßnahmen Punkt für Punkt entsprechen. In vielen Fällen kann eine Hypothese, ohne unmittelbare praktische Konsequenzen zu enthalten, in einem breiteren theoretischen Rahmen praktische Bedeutung gewinnen. Am Ende dieses Berichts (Kap. V.4.) wollen wir diskutieren, ob aus ihren Ergebnissen pädagogisch praktische Konsequenzen unmittelbar oder erst über eine theoretische Vermittlung folgen.

Wir haben im vorangegangenen Abschnitt zu zeigen versucht, daß die praktischen Bezüge unserer Untersuchung nicht ohne Berücksichtigung der Entwicklung der Theorie der linguistischen Kodes analysiert werden kann. In den folgenden Abschnitten dieses Kapitels werden wir auf die theoretische Begründung unserer Untersuchungen im Rahmen der Theorie der linguistischen Kodes eingehen sowie theoretische Ansätze und empirische Ergebnisse für den gewählten Sprachbereich, den Wortschatz, referieren, um erst nach der Darstellung der Ergebnisse auf die möglichen praktischen Konsequenzen zurückzukommen.

2. Empirische Ergebnisse zur Entwicklung des Wortschatzes - ein Bericht über ausgewählte Literatur zu den Problemen des Experiments.

Das soziolinguistische Experiment, über das wir berichten, sollte Unterschiede im aktiven Wortgebrauch trotz kontrollierten passiven Wortverständnis bei 10jährigen Kindern verschiedener sozialer Milieus aufdecken und erklären. In diesem Literaturbericht konzentrieren wir uns daher auf die Analyse dieser beiden Funktionsweisen des Wortschatzes, Verstehen und Gebrauch. Sie fordert eine Darstellung der Entwicklung des Bedeutungssystems im Kontext der Entfaltung der Grammatik und des Aufbaus kognitiver Operationen, sowie ein Referat über empirisch festgestellte soziale Einflüsse auf die Entwicklung des Wortschatzes.

2.1. Sprachentwicklung und Genese des Bedeutungssystems.

Als eines der wichtigsten Ergebnisse der Forschungen zur Sprachentwicklung kann man die Diskrepanz zwischen der semantischen und grammatischen Entwicklung ansehen. Noch vor dem zweiten Lebensjahr beginnt ein nicht behindertes Kind Zwei-Wort-Sätze zu äußern und etwa um das vierte Lebensjahr verfügt es über alle wesentlichen Konstruktionen seiner Muttersprache: die Funktionen der Basisstruktur und einige Transformationen in spontan produzierten Sätzen (McNeill 1966, 1970: 16-36). Die Entwicklung des Lexikons jedoch geht, wenn nach dem Abschluß der holophrastischen Phase der Sprachentwicklung erstmals zwischen Lexikon und Grammatik getrennt werden kann, zunächst sehr langsam, nach Abschluß des Erwerbs der wesentlichen grammatischen Konstruktionsmittel zwischen dem 4. und 6. Lebensjahr merklich schneller und danach wieder langsamer, kontinuierlich und ohne sichtbaren Abschluß voran (McCarthy 1954:527, Lenneberg 1967:130). Die Schnelligkeit, mit der ein abstrakt-hierarchisches System von Operationen wie die Grammatik vom Kind gelernt wird, hat zu der Hypothese einer ange-

borenen humanspezifischen Anlage von "Spracherwerbsmechanismen" ("language acquisition device") geführt (McNeill 1966:39), die sich mit zahlreichen biologischen Argumenten wenn auch nicht im einzelnen belegen, so doch global unterstützen läßt (Lenneberg 1967). Die kontinuierliche quantitative Erweiterung des Wortschatzes dagegen deutet auf eine Verbindung mit der kognitiven Entwicklung: das Wort muß eine Kategorie, nicht nur Komplexe bezeichnen, um eine konstante für eine Vielzahl von Sprechern verständliche Bedeutung zu haben, und Hierarchien kognitiver Kategorien müssen aufgebaut sein, damit die Bedeutungen einzelner Worte in Beziehung gesetzt und der Wortschatz strukturiert werden kann; die kategoriale Erfassung der einzelnen Wortbedeutung und der Aufbau einer kategorialen Hierarchie, die die Beziehungen zwischen einer Mehrzahl von Wortbedeutungen darstellen, sind zwei Seiten des gleichen Prozesses (Wygotsky 1969).

Die Diskrepanz zwischen dem frühzeitigen Abschluß des Grammatikerwerbs und dem kontinuierlichen Wachsen des Wortschatzes bzw. der andauernden Strukturierung des Bedeutungssystems legt die Vermutung nahe, daß die Grammatik als ein "abstraktes Schema" ein Mittel darstellt, mit dem die Struktur der kognitiven Kategorisierungen aufgebaut und die Bedeutung eines Wortes analysiert und mit den Bedeutungen anderer Worte systematisch verbunden werden kann. In den folgenden Abschnitten wollen wir zunächst systematisch die Vorbedingungen für die Entfaltung des Bedeutungssystems in der Entwicklung der Grammatik und die Beziehungen zwischen Wortschatz und Bedeutungssystem analysieren (2.1.1.) und daraus Folgerungen für semantische Lern- und Lehrprozesse ziehen (2.1.2.), um dann einen Abriß der Entwicklung des Wortschatzes innerhalb der Sprachentwicklung zu geben (2.1.3.).

2.1.1. Kognitive Fähigkeiten, Bedeutungslernen und Sprachgebrauch.

Das Kind, das etwa mit dem vierten Lebensjahr die Grundoperationen der Grammatik beherrscht, verfügt über ein Medium, das durch Kategorialität und Hierarchie der Kategorien gekennzeichnet ist und in der kategorialen Hierarchie der universellen Basis-Struktur-Relationen von Subjekt-Prädikat, Subjekt-Objekt, Nominalphrase-Modifikation die symbolischen Operationen der Funktion, der Bewirkung und der Intersektion bereit hält¹⁾; es verfügt keineswegs über die kognitiven Operationen der Funktion, der Bewirkung oder der Intersektion, wie die Entwicklungspsychologie Piagets belegt (Bruner 1971:69-72). Zudem beherrschen die Kinder zwar vom Beginn der Sprachentwicklung an die Referenzfunktion: die ersten Dinge, die sie benennen, sind für sie von emotionalem Interesse (McCarthy 1954: 571, Nelson 1973: 33); aber erst mit etwa 10 Jahren beginnen sie die Bedeutung eines Wortes kategorial zu erfassen, und Beziehungen zu anderen Wortbedeutungen zu verstehen (Wygotsky 1969:115). Die erste Tatsache ist so selbstverständlich wie die zweite unbeachtet: Während das Kind Beziehungen zwischen Worten und Objekten schon früh herstellen kann, muß es die Bedeutung des Wortes lernen.

Wir können zwei Ebenen im Prozeß des Lernens von Wortbedeutungen unterscheiden. Einmal muß das Kind, mit Hilfe des Wortes als gleichsam eines Richtfadens, die für den Wortgebrauch kritischen

¹⁾-----
Damit ist nicht gesagt, daß das Kind diese Operationen grammatisch analysieren und damit in einem nicht-umgangssprachlichen Medium repräsentieren kann, sondern nur, daß es über diese Operationen verfügt, indem es Sätze seiner Muttersprache spontan produziert. Daß es in bezug auf gestellte Aufgaben sinnvolle Sätze mit Hilfe der genannten grammatischen Relation bildet, ist schon ein Produkt der Instrumentalisierung dieser Relationen (siehe dazu weiter unten).

Attribute aus der ungeordneten Ganzheit der Situation des Gebrauchs herauslösen und in einem Konzept wieder zusammenbringen; über die Vorstufen synkretistischer, komplexiver und "pseudobegrifflicher" oder "transduktiver" Konzepte entwickelt sich auf diese Weise ein Begriff als die kognitiv reifste Form der Wortbedeutung (Wygotsky 1969). Zweitens kann die Bedeutung eines Wortes an seinen Beziehungen zu anderen Worten abgelesen werden; die Summe dieser Bedeutungsbeziehungen (der Synonymie, Antynomie, Inkompatibilität u.a.) kann als die Bedeutung des Wortes angesehen werden (Lyons 1968:428). Die erste Ebene, die Verbindung eines Wortes mit einem Konzept und die verschiedenen Entwicklungsstufen von Konzepten, wurde traditionell von der Psychologie behandelt. Sie untersuchte, wie einzelne Bedeutungen von Worten, d.h. Konzepte, aus der Umwelt destilliert wurden. Die Psychologie ignorierte jedoch die Systematik des Lexikons, die das Lernen eines Wortes auch unabhängig von der den Lernerfolg kontrollierenden Objektumwelt ermöglicht. Die zweite Ebene, die Systematik der Lexikons, wurde traditionell von der Sprachwissenschaft behandelt. Sie versuchte vor allem die Bedeutung von "System des Lexikons" und die Strukturprinzipien eines Lexikons zu explizieren, die schon bei de Saussure in den Beziehungen zu allen anderen Termen der Sprache gesehen wurden (siehe Francescato 1973:74). Die Sprachwissenschaft konnte jedoch nie psychologische Hypothesen liefern, wie das System der Bedeutungen erlernt wird.

Wir können auf die Beziehung zwischen diesen beiden Ebenen nicht genauer eingehen, bevor wir nicht erläutert haben, inwiefern das abstrakte Schema der Grammatik im aktiven Sprachgebrauch zu einem Instrument für diese Lernprozesse werden kann; dafür muß wiederholt werden, daß die im Alter von vier Jahren in den Grundzügen entfaltete Grammatik ein System von Kategorien darstellt, deren Form der im gleichen Alter vorherrschenden komplexiven Kategorisierung der Wortbedeutungen überlegen ist und deren hierarchische Struktur im Gegensatz zur kategorialen Gliederung des Wortschatzes in diesem Alter fest organisiert ist.

Die sprachliche Darstellung von Erfahrungen erlaubt es dem Kind, widersprüchliche oder isolierte kognitive Kategorisierungen zusammenzubringen. Wir wollen zwei Beispiele anführen. Wenn erstens ein Kind bei Piagetschen Konservationsexperimenten zuerst sagt, die Ausgangsmenge sei größer als die transformierte und wenig später, die Ausgangsmenge sei kleiner, so kann ihm die Vereinigung der beiden Prädikationen des gleichen Subjekts in einem Satz den Widerspruch zum Bewußtsein bringen. Wenn zweitens ein Kind einen Gegenstand "Stuhl" und "Möbel" nennt, so kann es mit sprachlichen Mitteln (Sätzen: "gehört zu", Aufzählung weiterer Möbel) unterschiedlichen Allgemeinheitsebenen in Bezug setzen. Obwohl im ersten Fall die zweidimensionale simultane Transformation, in zweiten die Mengenintersektion als sprachfreie logische Operation zugrundeliegen, kann das Erlernen und Beherrschen dieser Operationen durch die Speicherung der kritischen Dimensionen in der Sprache erleichtert werden. Obwohl in beiden Fällen die Ablösung der Erfahrung von der Objektwelt durch ein symbolisches Medium überhaupt ermöglicht wird, so sind es doch die spezifischen Mittel des symbolischen Mediums Sprache, die Verbindungen verschiedener Prädikationen des gleichen Subjekts in einem Satzrahmen, die den Widerspruch bewußt machen oder die Verbindung herstellen (Wygotsky 1969:280, Bruner 1971:78, Oevermann 1968:335-339). Auf eine ähnliche Weise können die drei genannten universellen, und natürlich alle weiteren spezifischen Basis-Struktur-Relationen das Erlernen der Summe der Bedeutungsbeziehungen zu anderen Worten, die die Bedeutung eines Wortes darstellen, erleichtern. Denn die Formen der semantischen Relationen müssen nicht nur in einem Satz, sondern können auch schon in grammatischen Relationen ausgedrückt werden.

Die wesentliche Gemeinsamkeit der beiden skizzierten Lernprozesse beruht darin, daß die Beziehungen zwischen einzelnen Kategorisierungen und zwischen einzelnen Bedeutungen durch den Gebrauch der Sprache in grammatisch konstruierten Sätzen hergestellt wird. Der wesentliche Unterschied zwischen den beiden skizzierten Lernpro-

zessen liegt darin, daß die Grammatik im ersten Fall zur Organisation sprachlich explizierter Umweltkategorisierungen, im zweiten Fall zur rein innersprachlichen Integration von Bedeutungen und zum Aufbau eines Bedeutungssystems genutzt wird. Ob bestimmte Bedeutungen durch die Konfrontation des Wortes mit Umweltkategorisierungen oder durch die Summe innersprachlicher Verknüpfungen gelernt werden, ist nicht in allen Fällen durch die zu erlernende Bedeutung festgelegt: in einem großen Bereich des Wortschatzes können beide Prozesse füreinander eintreten. Auf der einen Seite ist die Entwicklung der verschiedenen Formen der Kategorisierung bis hin zum analytisch-synthetischen Begriff ein kognitiver Entwicklungsprozeß, der durch das innersprachliche Bedeutungslernen nicht ersetzt werden kann. Insofern Wortbedeutungen Begriffe sind, ist daher die Beherrschung der Klassenlogik für das Lernen von Bedeutungen grundlegend (Piaget 1969). Auf der anderen Seite werden spezifische Begriffe, in der Entwicklung nicht-behinderter Kinder, anhand der Lautgestalt von Worten (Wygotsky 1969:111) und der in grammatischen Relationen und Sätzen vorgezeichneten Bedeutungsbeziehungen gelernt. Insofern die Analyse- und Syntheseleistungen, die einen Begriff ausmachen, durch die Verbindung mit einem Wort auf Dauer als die Wortbedeutung festgehalten werden, ist die Sprache entscheidend für die Weiterführung der kognitiven Entwicklung; die im Vergleich verschiedener Begriffe und der Analyse ihrer Beziehungen, im Aufbau eines komplexen Begriffssystems besteht. Kurz: Kognitive Operationen bilden zwar ein Substrat für die begriffliche Präzisierung von Wortbedeutungen, aber die Worte speichern und die Konstruktionsmittel der Sprache verbinden und analysieren die Begriffe; Begriffssysteme können nur aufgebaut werden als Systeme von Wortbedeutungen.

Umgekehrt stellt der Wortschatz einer natürlichen Sprache ein Reservoir potentieller kognitiver Kategorisierungen dar. Vielleicht läßt es sich vollständig in der Form von Begriffen explizieren, wie Theorien der Semantik es annehmen (Katz/Fodor 1963); wahrscheinlich aber ist, daß Teile notwendig in der Form von Komplexen verbleiben,

wie Philosophen der Alltagssprache annehmen (Wittgensteins Theorie der "Familiennamen"). In jedem Fall hängt die Ausschöpfung des im Wortschatz enthaltenen Bedeutungssystems vom durch die objektiven Aufgaben - sachliche oder soziale - verlangten Sprachgebrauch ab. Wir können, schematisch zusammenfassend, drei Komplexe festhalten: Die Bildung von Begriffen setzt erstens sprachunabhängige Fähigkeiten der Analyse und Synthese voraus; das Wort und die Konstruktionsmittel der Grammatik sind jedoch zweitens notwendig für den Aufbau eines Begriffssystems, für das der Wortschatz einer Sprache ein natürliches Instrumentarium darstellt; wieweit drittens das im Wortschatz vorhandene Instrumentarium für den Aufbau eines Begriffssystems genutzt und der Wortschatz kategorial gegliedert wird, hängt vom Gebrauch des Wortschatzes in objektiv vorgegebenen, sozialstrukturierten Problemen ab.

2.1.2. Aufgaben und Mechanismen des Bedeutungslernens.

Wir wollen aus der Analyse der Zusammenhänge des Bedeutungslernens mit der kognitiven Entwicklung und mit dem Sprachgebrauch einige Folgerungen für den semantischen Lern- und Lehrprozeß ziehen. Eine erste Folgerung betrifft die im semantischen Lernen implizierte Aufgabe. Wenn - von Definitionen durch "Zeigehandlungen" einmal abgesehen - Wortbedeutungen durch Wortbedeutungen erklärt werden, dann ist das Wortbedeutungen verbindende Begriffssystem und das Sprechen über Wortbedeutungen selber Teil der Umgangssprache und die Umgangssprache ihre eigenen Metasprache. Für das Kind, das eine Sprache erwirbt, besteht die Aufgabe des Bedeutungslernens dann darin, unter den Sätzen, die es hört, diejenigen zu entdecken, die Definitionen enthalten; oder, "noch weiter gefaßt, das quasi-deduktive System, welches unbegrenzt viele Sätze der Sprache miteinander auf einer semantischen Grundlage verbindet, als Grundlage zu konstruieren" (Weinreich 1970:77). Eine solche Charakterisierung des semantischen Lernprozesses kann seine Langwierigkeit erklären (McNeill 1970:120). Auch wenn außersprachliche Umstände ihn unter-

stützen, setzt er eine implizite Kenntnis der metasprachlichen Funktionen der Umgangssprache und die Speicherung von Bedeutungen in einem Bedeutungssystem voraus. Mit dieser Folgerung haben wir die Aufgabe semantischen Lernens charakterisiert. In zwei weiteren Folgerungen wollen wir psychologische Mechanismen herausarbeiten, die die Lösung dieser Aufgabe ermöglichen; wir gehen auf die notwendige Kenntnis metasprachlicher Funktionen nicht näher ein und konzentrieren uns auf den Aufbau des Bedeutungssystems.

Eine zweite Folgerung betrifft die Leistung des Wortgebrauchs für den Aufbau eines Bedeutungssystems. Wenn Begriffe in Worten gespeichert und in Sätzen analysiert werden, dann ist der aktive Gebrauch des Wortschatzes auch entscheidend für die kategoriale Durchgliederung des passiven Repertories von Worten. Die Qualität des Wortverständnisses bedingt aber wiederum die Qualität des Wortgebrauchs. Auf diese Weise können wir den aktiven und passiven Wortschatz als Durchgangsstadien eines kreisförmigen Lernprozesses des Bedeutungssystems ansehen, in dem der entscheidende Anstoß immer wieder im aktiven Wortgebrauch liegt. Aktiver und passiver Wortschatz sind nicht nur Gebrauchs- oder Funktionsweisen des Wortschatzes, sondern zugleich Entwicklungsmechanismen. Ganz im Gegensatz zur Entwicklung der Grammatik, für die der Sprachgebrauch nur als eine auslösende, das endogene Entwicklungsprogramm hemmende oder fördernde Variable angesehen wird (Lenneberg 1967: 146), muß der Sprachgebrauch für das Lexikon als der Motor der Entwicklung betrachtet werden.

Eine dritte Folgerung betrifft Stellenwert und Erfolgsaussichten des Lehrens von Bedeutungen und schließt unmittelbar an die vorangegangene Folgerung an. Während das Kind grammatische Strukturen durch das "Erschließen latenter Strukturen" selbsttätig aufbaut und das intentionale Lehren grammatischer Formen durch Verbesserung ungrammatischer Sätze des Kindes so gut wie keinen Erfolg hat (Brown 1964:89 ff., 1969: 138 ff), kann der Lernprozeß von Wortbedeutungen und der mit ihnen verbundenen nichtlinguistischen Kategorisierungen durch Lehren erleichtert und beschleunigt werden;

für das erstmalige Erlernen von Kategorien durch Worte ist die Bezeichnung von Beispielen und die Kontrolle der Reaktionen des Kindes durch einen Tutor - was Brown "Original Word Game" (1958: 284-291) nennt - sogar der einzig denkbare Weg. Die Kategorisierung einer Objektklasse allein nach nicht-linguistischen Kriterien wäre für das Kind eine nahezu unmögliche Aufgabe, da die Umwelt auf viele Weisen diskriminiert werden kann. Die Benennung von Objekten durch den Tutor vereinfacht die Kategorisierung, indem sie selber für das Kind ein Merkmal der Objektklasse wird; das Kind nimmt sie als Schlüssel für die Suche nach den kritischen Attributen der Objektklasse. Die Vereinfachung, die die Benennung für den Kategorisierungsprozeß des Kindes leistet, ist für den Erwachsenen so selbstverständlich, daß er sich die Schwierigkeiten einer nicht-sprachlichen Umwelt-Kategorisierung, vor der das Kind ohne die Benennungen eines Tutors stünde, kaum mehr vorstellen kann. Die Effektivität dieses Lehrspiels beruht darauf, daß das Kind auf den Tutor antwortet und sein Verständnis durch die Praxis eigener Benennungen äußern kann. Der Tutor kann das Kind nicht nur durch wiederholte Nennung, sondern durch Ausschöpfung aller grammatischen Mittel der Sprache solange korrigieren, bis es das Konzept beherrscht, d.h. das Konzept internalisiert hat.¹⁾ Auf diese Weise können nicht nur Gegenstandsklassen, sondern auch - objektiv prüfbare ("heiß") wie sozial bedeutsame ("sicher")-Prädikate gelernt werden.

In dieser Konzeption wird angenommen, daß Kind und Erwachsener sich mit Hilfe eines Wortes verständigen, obwohl sie mit ihm Kategorisierungen unterschiedlicher kognitiver Qualität verbinden. Dies ist nur ein scheinbarer Widerspruch. Unterschiede der kategorialen Qualität, d.h. unterschiedliche Wortbedeutungen, können gerade dadurch aufgedeckt werden, daß das Kind durch Gebrauch des Wortes in

¹⁾-----
Alle Benennungen in diesem Spiel - auch wenn sie nur die Form eines Wortes haben - stellen auf der Ebene der Pragmatik natürlich Äußerungen dar.

einer neuen Situation dem Tutor ein Indiz für sein Wortverständnis und eine Korrekturmöglichkeit gibt. Durch den Gebrauch des gleichen Wortes für unterschiedliche, und im Sinne der üblichen falsche Referenten kann das Kind Stück für Stück die in der Sprache des Tutors herrschende Wortbedeutung lernen, und nur indem Kind und Tutor die gleichen Worte verwenden, können sie sich über die Unterschiede im Wortverständnis verständigen. Pointiert gesagt, fördert der Gebrauch des Wortes das Verständnis gerade dadurch, daß er über das Verständnis der Wortbedeutung hinausgeht. In der Konzeption des "Original Word Game" wird sichtbar, daß Kategorien durch den Gebrauch von Worten in Äußerungen erworben werden und der Bedeutungsgehalt der Worte sich aus den Kategorisierungsprozessen ergibt. Eine solche Konzeption behauptet keinen Kausaleinfluß von Worten auf Konzepte innerhalb des Sprecher-Hörers. Sie verdeutlicht, daß in den Worten einer Sprache objektiven Probleme angemessene und sozial verbindliche Kategorisierungen kondensiert sind, die dem Kind entfaltet werden müssen, damit es die Bedeutung der Worte versteht, und sie behauptet, daß dieser simultane Lernprozeß von Kategorien und Wortbedeutungen in der sprachlich geführten und praktisch kontrollierten Kommunikation zwischen einem Tutor und dem Kind stattfinden muß. Die Stärke dieser Konzeption liegt darin, daß der Erwerb von Kategorien und Wortbedeutungen ohne Rückgriff auf sprachdeterministische Argumente durch soziale Lehr- und Lernprozesse erklärt wird; ihre Schwäche liegt darin, daß die Systematik der Lehr- und Lernprozesse nicht berücksichtigt wird.

An zwei Einwänden läßt sich diese Schwäche erläutern. Einerseits kann zwischen den intentionalen Lehrprozessen eines Tutors und dem unbeabsichtigten Lehrerfolg, den jedes Spiel oder jedes Gespräch eines Erwachsenen mit einem Kind erbringen kann, nicht unterschieden werden, es sei denn durch die Tatsache der Intentionalität oder Nichtintentionalität. Andererseits wird nicht analysiert, wie die einzelnen gelernten Bedeutungen zueinander in Beziehung gesetzt werden können, sodaß der Lernprozeß einer jeden Bedeutung für sich durch praktische Testhandlungen kontrolliert

werden muß. Beide Einwände treffen dieselbe Schwäche; wenn wir die Intentionalität des Lehrprozesses charakterisieren, verstehen wir zugleich die Einheit der Lernprozesse. Eine Alternative, die beiden Einwänden gerecht wird, findet sich in Wygotskys Theorie alltäglicher und wissenschaftlicher Begriffsbildung (Wygotsky 1969: 167-290), die kurz mit der Theorie des "Original Word Game" verglichen werden soll.

Alltägliche Begriffe werden durch die Verbindung von Umweltkategorisierungen mit Worten gebildet. Sie können anschaulich demonstriert werden und ihr Verständnis kann durch praktische Differenzierungen in gegebenen Situationen geprüft werden. Sie sind konkret, aber zugleich voneinander isoliert. Nur auf alltägliche Begriffe bezieht sich das "Original Word Game". Wissenschaftliche Begriffe werden durch systematische Verknüpfungen untereinander gebildet. Sie werden durch die Explikation ihrer Systematik ohne anschauliche und situative Unterstützung gelernt. Wygotsky zeigt nun, daß das Lernen von alltäglichen Begriffen nur die Grundlage ist für ein Begriffssystem, das sich erst durch das Lehren wissenschaftlicher Begriffe entwickeln kann. Anhand der Systematik wissenschaftlicher Begriffe lernt das Kind die Strukturprinzipien der begrifflichen Verallgemeinerung, mit denen es auch die alltäglichen Begriffe ordnen kann; umgekehrt gewinnt es aus den alltäglichen Begriffen die Anschauung für wissenschaftliche Begriffe.¹⁾ So wie die beiden Einwände gegen die Theorie des "Original Word Game" bedingen sich zwei Neuerungen der Theorie Wygotskys: systematisch verbundene Begriffe können nur in abstrakten, d.h. nicht durch praktisches Handeln kontrollierten

1) -----
"Alltäglich" und "wissenschaftliche" dürfen nicht als substantielle, sondern müssen als funktionale Charakterisierung der Begriffe verstanden werden. Jeder Begriff kann "alltäglich" oder "wissenschaftlich" eingeführt werden, je nachdem, ob er isoliert oder in einer expliziten Systematik vorgestellt wird.

Lehrprozessen gelernt werden; umgekehrt kann die Besonderheit des Lehrens gegenüber lerneffizienten Alltagsinteraktionen durch seinen Gegenstand, die Systematik der Beziehungen zwischen Begriffen, charakterisiert werden.

Wir können nun die Mechanismen, durch die die eingangs dargestellte Aufgabe des semantischen Lernens bewältigt werden kann, zusammenfassen. In jedem Fall ist der Gebrauch bestimmter Worte für die Entwicklung ihres begrifflichen Verständnisses wichtig. Die Bedeutung einzelner Worte kann dabei in der Zusammenarbeit des Kindes mit einem Tutor gelernt werden, der - mehr oder minder gezielt - die Lernerfolge des Kindes kontrolliert. Der Aufbau eines Bedeutungssystems aber, der im semantischen Lernen ja gefordert wird, ist auf diese Weise nicht mehr möglich. Systematische Beziehungen kann das Kind dann zwischen den erlernten Begriffen konstruieren, wenn es die Strukturprinzipien der Verallgemeinerung gelernt hat, die ihm paradigmatisch in abstrakten Begriffssystematiken gelehrt werden müssen (Wygotsky 1969).

Die Theorie Wygotskys greift auf systematische und abstrakte, im sozial-evolutionären Sinne "schulische", Lehre zurück, um den Erwerb eines Bedeutungssystems beim Kind zu erklären. Sie ist unseres Wissens der einzige Versuch, die psychologischen Mechanismen beim Erwerb des Bedeutungssystems darzustellen - im Unterschied zur Explikation der Strukturprinzipien dieses Systems (wie sie etwa McNeill 1970: 116, Francescato 1973: 76-82, Lyons 1968: 443-469 u.a. geben). Wir wollen den Hinweis, daß ein solcher psychogenetischer Erklärungsversuch auf eine soziale Variable aufbaut, nicht weiter verfolgen. An dieser Stelle wollen wir nur noch einmal auf die Diskrepanz zwischen den linguistischen Explikationen des semantischen Bedeutungssystems und den psychologischen Erklärungen für den Erwerb eines solchen Systems hinweisen. Psychologische Erklärungen beziehen sich vorwiegend auf den Erwerb einzelner Wortbedeutungen und Konzepte. Wir haben in diesem Abschnitt argumentiert, daß derartige Erklärungen notwendig, aber nicht hinreichend für den Erwerb eines Bedeutungssystems sind.

Im folgenden Abschnitt wollen wir die Entwicklung des Wortschatzes innerhalb der Sprachentwicklung darstellen; als Beleg für die Entwicklung des Bedeutungssystems können wir dort nur Untersuchungen über den Zusammenhang des Bedeutungslernens mit dem Grammatikerwerb zitieren. Sie sollen jedoch denselben Gegenstand erfassen, für den Wygotsky eine psychogenetische Erklärung gibt: den Aufbau eines Bedeutungssystems hinter dem Erwerb einzelner Bedeutungen.

2.1.3. Die Entwicklung des Wortschatzes innerhalb der Sprachentwicklung.

An der bloßen Erweiterung des Wortschatzes im Laufe der Entwicklung läßt sich weder die Beziehung zwischen Wortschatz und Grammatik noch die interne Strukturierung des Wortschatzes zu einem System von Bedeutungen ablesen. Da zudem - wie wir in den beiden vorangegangenen Abschnitten erläutert haben - die Grammatik ein Mittel zur internen Strukturierung des Wortschatzes darstellt, werden wir die Darstellung der Wortschatzentwicklung nach den Daten des Grammatikerwerbs gliedern. Zuerst werden wir die "Lexiken" und "Grammatiken" darstellen, die sich aus den Äußerungen von Kindern vor dem Abschluß des Grammatikerwerbs gewinnen lassen; dann werden wir einige Untersuchungen zitieren, die die Bedeutung der Grammatik für den Aufbau eines Bedeutungssystems nach Abschluß des Grammatikerwerbs belegen.

Die ersten Äußerungen von Kindern in allen Sprachen treten zwischen dem 12. und 18. Lebensmonat auf und bestehen nur aus einem Wort (Lenneberg 1967: 280); sie werden "holophrastisch" genannt. Sie sind gekennzeichnet durch zweierlei: pragmatische Vieldeutigkeit und semantische Übergeneralisierung. Das gleiche Wort kann als Befehl, Frage und Aussage gebraucht, es kann aber auch in den unterschiedlichsten Situationen für verschiedene Referenten verwandt werden (Piaget 1969: 276). Hinter der semantischen Über-

generalisierung verbirgt sich Unsicherheit im durch das Wort vertretenen Konzept. Das Lernen des Wortschatzes und der Erwerb von Konzepten nämlich stellen zwei Entwicklungslinien dar. Auf der einen Seite löst das Kind aus seiner senso-motorischen Weltauffassung Konzepte heraus, auf der anderen Seite wird die Begriffsbildung durch die vorgegebenen sprachlichen Formen gesteuert. Im Zusammentreffen beider Entwicklungslinien bildet sich die Wortbedeutung, die Wygotsky als die nicht mehr auflösbare Grundeinheit des Denkens und Sprechens betrachtete. Auf dem Weg zu diesem Ziel werden verschiedene Vorstufen durchlaufen (Piaget 1969, Wygotsky 1969), in denen das Kind noch nicht die relevanten Merkmale finden und durchgängig anwenden kann. Wir haben versucht, diesen Prozeß in Abschnitt 2.1.2. systematisch darzustellen.

Die zweite Besonderheit der ersten Ein-Wort-Äußerungen, ihre pragmatische Vieldeutigkeit, deutet darauf, daß sie primitive undifferenzierte Sätze und damit zugleich den Ursprung der Grammatikentwicklung darstellen. In der Phase der Ein-Wort-Äußerungen kann das Kind bis zu 50 Worte aktiv produzieren, jedoch jeweils nur in Isolation. Semantisch decken die Ein-Wort-Äußerungen alle kommunikativen Funktionen von Sätzen ab. Der pragmatische Wert der Ein-Wort-Äußerungen kann vom Kind durch Intonation dargestellt werden; insofern sind phonologische das funktionale Äquivalent für grammatische Regeln (Lenneberg 1967: 283). Der pragmatische Wert der Ein-Wort-Äußerungen kann zudem durch die Situation bestimmt oder auch von den Zuhörern unterstellt werden. In jedem Fall haben die Ein-Wort-Sätze semantisch-pragmatische Bedeutungen, die sich in der späteren Grammatikentwicklung entfalten; die Semantik ist zugleich der Kern der Grammatik.

Auch wenn im Alter von zwei Jahren das Kind beginnt, zwei Worte zusammenzufügen, bleibt die anfängliche Einheit von Syntax und Semantik bestehen. Die Zwei-Wort-Sätze lassen sich in einer generativen Grammatik rekonstruieren, in der die Sätze aus einer funktionalen (pivot) und einer offenen (grammatisch undefinierten)

Klasse bestehen können (McNeill, 1966: 88). Beide Klassen sind, von der Erwachsenengrammatik gesehen, heterogen. Die pivot class enthält keineswegs nur Funktionsworte der späteren Grammatik; sie dient zur Ausdifferenzierung aller späteren Klassen von Worten.

Wenn wir als die entscheidenden Kennzeichen eines normalen Lexikons nehmen, daß es erstens einzelne Worte und Bedeutungen zuordnet und zweitens die Worte durch die Zugehörigkeit zu grammatisch nicht weiter differenzierbaren Klassen charakterisiert, so wird erst in der Phase der Zwei-Wort-Sätze ein Lexikon entwickelt, das wenigstens das erste Kriterium erfüllt. McNeill (1970: 115) gibt für den Aufbau eines Wort-Lexikons als untere zeitliche Grenze 18, als obere Grenze 30 Monate an. Ein solches Lexikon ist Vorbedingung für die Entwicklung von Transformationen, aber es ist noch so wenig differenziert wie die Kategorien einer Zwei-Wort-Satz-Grammatik. Vorformen eines solchen Wort-Lexikons lassen sich rekonstruieren, aber kaum belegen. In der Phase der Ein-Wort-Äußerungen müßte das Kind über ein "holophrastisches Lexikon" verfügen, das jeder Äußerung eines Wortes die Vielzahl der möglichen Interpretationen zuordnet. Auch in der Phase der Zwei-Wort-Sätze könnte das Kind fortfahren die Bedeutung jedes Zwei-Wort-Satzes in Form eines "Satz-Lexikons" zu speichern. Mit dem Anwachsen der möglichen Interpretationen von Zwei-Wort-Sätzen aber würde ein solches Vorhaben das Gedächtnis überfordern. Eine Lösung bietet dann ein Wort-Lexikon, in dem die Bedeutung jedes Wortes zusammen mit seiner Klassenzugehörigkeit in einer Zwei-Wort-Grammatik angegeben ist. Mit der weiteren Differenzierung der Kategorien der Grammatik differenzieren sich auch die Wortklassen des Lexikons, bis mit Abschluß des Grammatikerwerbs die Struktur eines normalen Lexikons erworben ist, d.h. eines Lexikons, das auch das zweite der oben genannten Kriterien erfüllt, und die weitere Entwicklung des Lexikons nur in einer quantitativen Erweiterung und semantischen, im Gegensatz zur grammatisch-kategorialen, Organisation des Wortschatzes bestehen kann.

Diese hypothetische Rekonstruktion (siehe McNeill 1966: 62-65, 1970: 115) verdeutlicht nicht nur die Verbindung zwischen semantischer und grammatischer Entwicklung vor Abschluß des Grammatikererwerbs, sie zeigt zudem, daß die Antriebskraft für grammatische Differenzierungen in der Vielzahl der bei den gegebenen grammatischen Mitteln möglichen semantischen Interpretationen liegt, die nicht mehr gespeichert werden können und für die neue Mittel des Ausdrucks gesucht werden.¹⁾ Solange die grammatischen Kategorien der Basis sich noch ausdifferenzieren, können semantische und grammatische Entwicklungen nicht getrennt werden. Insofern die Vielzahl semantischer Interpretationen, die gespeichert werden müssen, zur Differenzierung grammatischer Kategorien zwingt, treibt die Semantik den Aufbau der Syntax voran.

Sobald aber mit etwa vier Jahren der Grammatikerwerb abgeschlossen ist, können Syntax und Semantik getrennt werden und gewinnt die Grammatik Bedeutung für die weitere Entwicklung des Wortschatzes. Belege für den Abschluß des Grammatikererwerbs finden sich erstens in der Konstanz des Verhältnisses der "parts of speech", d.h. der wichtigsten grammatischen Klassen, bei weiter wachsendem Umfang sprachlicher Äußerungen (Deutsch u.a. 1964, Templin 1957); der Erwerb neuer syntaktischer Funktionen müßte sich in einer quantitativen Umgruppierung der Wortklassen niederschlagen. Damit ist aber noch nichts über die Akkuratheit des Regelgebrauchs gesagt. Rekonstruktionen der kindlichen Äußerungen in einer generativen Grammatik geben

1) Die Erklärung der Entwicklung aus der Notwendigkeit der Vereinfachung sagt nichts über die spezifische Art und Weise, mit der die Vereinfachung geleistet wird, und die Fähigkeit des Kindes, solche Mittel zu finden. Insofern schließen die Aussagen, daß Notwendigkeiten kognitiver Ökonomie den Antrieb für sprachliche Differenzierungen darstellen und daß die Wahl spezifischer grammatischer Mittel nur durch die angeborene Fähigkeit zur Sprache erklärt werden kann, einander nicht aus.

deshalb eine zweite Art von Belegen. P. Menyuk (1963) konnte an Sprachproben von Mittelschichtkindern nachweisen, daß alle Basis-Strukturen von 4jährigen Kindern beherrscht werden, und Regeln, die in diesem Alter noch nicht erworben waren, auch von 9jährigen noch nicht beherrscht wurden. Lenneberg (1967, S. 133) fand, daß von 500 MS-Kindern und US-Kindern mit 36 Monaten fast alle spontan komplexe Sätze produzierten.

Die Grammatik ist ein "abstraktes Schema", d.h. ein rein formales Organisationsprinzip, mit dem Realitätsbereiche in Beziehung gesetzt, in zeitliche, funktionale und hierarchische Ordnung gebracht werden können. Der Abschluß des Grammatikerwerbs markiert den Beginn der zweiten Phase der Sprachentwicklung, in der Grammatik und Lexikon sowohl als unabhängige wie auch miteinander interagierende Entwicklungslinien verfolgt werden können. Die weitere Grammatikentwicklung bringt nur noch den quantitativen Ausbau der prinzipiell beherrschten Basis-Strukturen zu größerer Komplexität.¹⁾ Die weitere Entwicklung des Wortschatzes besteht in einer Erweiterung des Umfangs und einer internen Strukturierung der Wortbedeutungen zu einem Bedeutungssystem.

¹⁾ -----
Untersuchungen über die Entwicklung der Syntax nach dem 5. Lebensjahr behandeln vor allem: die grammatischen Konsequenzen performativer Verben: ask/tell, promise (C. Chomsky 1969), die Entdeckung in der Oberflächenstruktur nicht aufgezeigter Tiefenstrukturbeziehungen ("Easy to See") (C. Chomsky 1969), die Entdeckung systematischer Ausnahmen von Regeln (Pronominalisierung) (C. Chomsky 1969) oder die Entdeckung von Ambiguitäten (Shultz/Pilon 1973). Derartige Konstruktionen und Regeln werden oft auch von Erwachsenen nicht spontan beherrscht, aber durchaus verstanden; sie können mit dem starken Anwachsen des passiven Vokabulars bei nur geringer Erweiterung des aktiven Vokabulars verglichen werden (C. Chomsky 1969:2-5); sie können als die Bereiche der Grammatik angesehen werden, deren aktive Verfügung zwischen Individuen bei gleicher sprachlicher Kompetenz variiert. Daß sie erst nach dem 5. Lebensjahr erworben werden, spricht nicht gegen die Tatsache, daß die Regeln der Basisstruktur und die wichtigsten Transformationen in diesem Alter beherrscht werden.

Der Umfang des Vokabulars wächst monoton während der ganzen Entwicklung, zuerst recht langsam, dann in der Vorschulzeit, d.h. nach Abschluß des Grammatikerwerbs ziemlich schnell, und dann wieder langsam (McCarthy, 1954). Die interne Strukturierung des Wortschatzes kann durch die verschiedenen semantischen Beziehungen der Synonymie, Hyponymie und Antonymie (siehe dazu Lyons 1968: 446-469) oder durch ein System semantischer Merkmale (Katz/Fodor 1963) formal dargestellt werden. Eine ausgeführte Theorie der Bedeutungsstruktur gibt es jedoch für keine natürliche Sprache, so daß es schwer ist, Belege für die Entwicklung des Bedeutungssystems zu finden. Wir haben im letzten Abschnitt Wygotskys Theorie der alltäglichen und wissenschaftlichen Begriffe als einen möglichen Entwicklungsmechanismus für den Aufbau eines Bedeutungssystems diskutiert. Wir haben außerdem dargelegt, daß die Grammatik ein Mittel zum Aufbau der Bedeutungsstruktur sein kann; für diese Annahme wollen wir abschließend Belege aus zwei Bereichen der empirischen Psycholinguistik bringen: aus Untersuchungen von Wortassoziationen und der Bedeutungskonstruktion sinnloser Worte.

Zwischen dem 6. und 8. Lebensjahr wechseln die Kinder von primär syntagmatischen zu vorwiegend paradigmatischen Wortassoziationen über (Entwisle u.a. 1963). Sie verbinden mit einem Reizwort nicht mehr ein Reaktionswort, das mit ihm im alltäglichen Sprachgebrauch stark verbunden und daher aus einer anderen Wortklasse ist, sondern ein Wort aus der gleichen Wortklasse. Der Wechsel von syntagmatischen zu paradigmatischen Wortassoziationen ist unabhängig von der gemessenen Intelligenz und von sozialer Schicht; die Zunahme paradigmatischer Assoziation zwischen dem 6. und 8. Lebensjahr ist in einer extrem kommunikationsfeindlichen ländlichen Subkultur nur im Trend verschoben; schließlich ist die Rangfolge verschiedener Wortklassen nach der Häufigkeit paradigmatischer Assoziationen in drei Altersstufen für Gruppen verschiedener Intelligenz und sozialer Herkunft dieselbe (Entwisle 1966). Der Gebrauch der Syntax als eines Strukturprinzips des Wortschatzes, der uns selbstverständlich erscheint, wird also erst zwischen dem 6. und 8. Lebensjahr fest erlernt. - Die Entwicklung paradigma-

tischer Assoziationen und die des korrekten semantischen Verständnisses einer grammatischen Kategorie verlaufen außerdem parallel: Wer aus einer Stichprobe von 6- bis 8-jährigen Kindern und Erwachsenen zu Alltagswörtern mehr paradigmatische Assoziationen gab, konnte auch eher einem Nonsense-Wort, von dem allein die grammatische Klasse bekannt war, einen Sinn beilegen, der in der gleichen Wortklasse ausgedrückt war (Brown/Berko 1960). - Die grammatische Kategorie kann gleichsam als Filter bei der Bedeutungsbestimmung eines unbekanntes Wortes dienen: Einem "count noun" ordneten 3- und 4-jährige Kinder am ehesten ein Objekt, einem Verb eine Handlung und einem "mass noun" eine Substanz zu (Brown 1957). Das Kind gewinnt also aus der grammatischen Kategorie eines Wortes Aufschluß über die Bedeutung. Obwohl der erwachsene Sprecher weiß, daß jede grammatische Kategorie von Worten zahlreiche Beispiele enthält, die der allgemeinen Bedeutung dieser Wortklasse widersprechen, wird er wahrscheinlich ebenso wie das Kind einem unbekanntes Wort zuerst eine Bedeutung unterstellen, die mit der allgemeinen Bedeutung der betreffenden Wortkategorie vereinbar ist.

Nicht nur die Stelle eines Wortes im Satz entscheidet über die Bedeutung mit, sondern der sprachliche Kontext insgesamt. Für das Bedeutungslernen ist nicht bloße Kenntnis grammatischer Kategorien, sondern das Erfassen der gesamten Satzstruktur erforderlich. Werner/Kaplan (1963) berichten von einem Experiment, in dem die Bedeutung eines Nonsense-Wortes aus sechs sinnvollen Sätzen, in denen es auftrat, konstruiert werden mußte. Sie analysieren zwei komplementäre Entwicklungsprozesse: Einerseits werden die Bedeutungselemente eines Wortes aus den Kontexten genommen, andererseits wird die Gesamtbedeutung von jedem spezifischen Kontext unabhängig. Beides sind semantische Prozesse, aber beide wären nicht möglich ohne implizite grammatische Analyse der Sätze. - In einem ähnlichen Experiment konnte die Wirksamkeit von semantischen Selektionsregeln bei der Verbindung von Worten nachgewiesen werden. Ein Nonsense-Wort wurde von den Vpn. zuerst in einer bestimmten Be-

deutung gelernt und dann in einem Satzrahmen mit einem weiteren Nomen vorgelesen. Die Vpn., die eine Bedeutung für das Nonsense-Wort gelernt hatten, die mit den Selektionsregeln des Satzverbs übereinstimmten, konnten zu ihm das weitere Nomen signifikant häufiger assoziieren, als die Vpn., die das Nonsense-Wort in einer Bedeutung gelernt hatten, die nicht mit den Selektionsregeln des Satzverbs übereinstimmte (Arnold/Bower/Bobrow 1972). Wieweit innerhalb eines Satzes verschiedene Bedeutungen miteinander verbunden werden können, hängt wiederum von der Struktur des ganzen Satzes ab.

Wir wollen die Entwicklung des Wortschatzes nach dem Abschluß des Grammatikerwerbs im nächsten Abschnitt weiterverfolgen, der soziale Bedingungsfaktoren in die Betrachtung einbezieht.

2.2. Soziale Einflüsse auf die Wortschatz-Entwicklung

2.2.1. Verschiedene Aspekte des Wortschatzes von Kindern in Abhängigkeit von dichotomer Sozialschicht.

Die Sprachentwicklung bis zum Abschluß des Grammatikerwerbs scheint sich nach einem endogenen Programm zu richten, das gegen äußere Einflüsse relativ immun ist (Lenneberg 1967). Sie kann als eine von der linguistischen Theorie her notwendige, unumkehrbare Abfolge von Entwicklungsstadien beschrieben werden, in der auf die Entwicklung grammatischer Kategorien die von Basisregeln und schließlich von Transformationen folgt (McNeill 1966). Die Sprachentwicklung nach dem Grammatikerwerb kann nur noch als quantitative Erweiterung eines erworbenen Grundschemas angesehen werden und sich nicht mehr an durch die linguistische Theorie vorgegebenen Entwicklungsmarken orientieren. Sie muß daher grundsätzlich unter dem Gesichtspunkt möglicher Umwelteinflüsse gesehen werden. Auch wenn sich empirisch keine Differenzen zwischen sozialen Gruppen, sondern nur konforme Entwicklungen ergeben, so können soziale wie genetische Erklärungen herangezogen werden. Spezifisch für den Wortschatz gilt ein zweites, empirisches Argument. Das Vokabular ist einerseits weitgehend unabhängig von anderen Sprachmaßen, andererseits korreliert es hoch mit gemessener Intelligenz und - mittelbar - sozialer Schicht (Deutsch u.a. 1964, Hurlburt 1954). Dies zwingt, die Wortschatzentwicklung nach Abschluß des Grammatikerwerbs aus der immanent linguistischen Perspektive zu lösen und mit der intellektuellen und sozialen Entwicklung in Zusammenhang zu stellen. Wir wollen prüfen, ob empirisch festgestellte Schichtunterschiede verschiedener Aspekte des Wortschatzes auf schichtspezifische Strategien des Sprachgebrauchs als ein Erklärungsprinzip bezogen werden können.

Schichtdifferenzen im quantitativen Umfang des Wortschatzes ließen sich durch standardisierte Wortschatztests und experimentell stimulierte Sprachproben belegen. Beide Methoden haben ihre Problematik. Während die standardisierten Tests subkulturelle Differenzen in Bedeutung und Häufigkeit der Worte vernachlässigen, weiß man bei experimentellen Sprachproben nichts über Schichtunterschiede im Reizwert der Experimentalsituation und der Fähigkeit der Kinder, deren Künstlichkeit zu überwinden. In Wortschatztesten erhält die Mittelschicht durchgängig höhere Mittelwerte (Siller 1957, John 1963). Deutsch (1967:360) berichtet Korrelationen zwischen Schicht und verschiedenen Wortschatztesten von .22 und .33 für 6jährige und .49 für 11jährige; die Schichtabhängigkeit der Wortkenntnis nimmt also mit dem Alter zu. In experimentellen Sprachproben ergaben sich Korrelationen zwischen Wortschatz (gemessen durch type-token-ratio) und Schicht von ca. $r = .30$ für beide Altersstufen (Deutsch u.a. 1964:43). Eine Faktorenanalyse über die Korrelationen von Sprachmaßen (Gesamtzahl der Sätze, der Worte; Anzahl der verschiedenen Wortklassen, der untergeordneten Sätze etc.) und demographischen Variablen (IQ, SES, Alter) ergab einen ersten Faktor, der nur mit sprachlichen Maßen geladen war; die die quantitative wie grammatische Sprachkenntnis darstellten, und einen zweiten Faktor, der nur mit Vokabelmaßen (type-token-ratio) und Intelligenz wie Schicht geladen war (Deutsch u.a. 1964).¹⁾ Wortschatztestwerte messen daher ebenso sehr sozial akkumulierte Erfahrung wie intellektuelle und sprachliche Fähigkeiten; Schichtunterschiede spiegeln in erster Linie Unterschiede des allgemeinen Anregungspotentials wider und können nur mittelbar als Indizien schichtspezifischer Strategien des Gebrauchs von Sprache gesehen werden.

¹⁾-----
Ein Nachteil dieser Faktorenanalyse ist allerdings, daß nicht zwischen Korrelationen, die sprachlogisch notwendig, und solchen, die empirisch aussagekräftig sind, getrennt wurde.

Schichtdifferenzen in der Struktur des Wortschatzes lassen sich anhand von Standardwortlisten sowie anhand der Häufigkeitsprofile grammatischer Wortklassen feststellen. Aus dem Ergebnis, daß bei 7- und 11jährigen die Korrelationen von Schicht mit der (durch die type-token-ratio gemessenen) Diversifikation des Wortschatzes durchgängig höher sind als mit allgemeiner sprachlicher Produktivität, schließt Deutsch (Deutsch u.a. 1964:111), daß die Schichten sich weniger im Umfang als in der Struktur des sprachlichen Ausdrucks unterscheiden. In schriftlichen Sprachproben gebrauchen die Mittelschichtkinder häufiger Worte, die in einer Standard-Liste der "100 gebräuchlichsten Wörter" enthalten sind, als die Unterschichtkinder, obwohl beide Gruppen auf einem passiven Wortverständnis parallelisiert waren; die Differenz verkleinert sich allerdings bei den 15- gegenüber den 13jährigen (Lawton 1970: 153). Thomas (1962) fand in einer Untersuchung der mündlichen Sprache von Unterschichtkindern, daß sie zwar seltener die häufigsten Worte der Standardlisten gebrauchten, jedoch viele Worte verwenden, die auf diesen Listen überhaupt nicht genannt sind. Die Prozentzahlen der ersten 1 000 Worte von 5 Standardlisten, die nicht von Unterschichtkindern gebraucht wurden, gingen von 18,1 für die Rineland- bis 46,9 für die Thorndike-Lorge-Liste. Umgekehrt tauchten 157 von insgesamt 1 284 Worten der Unterschichtkinder in keiner der 5 Listen auf. Nur ein Viertel dieser Worte fand sich in den üblichen Schullesebüchern, während die Unterschichtkinder nur 50% der Lesebuch-Worte gebrauchten.

Ein Vergleich der prozentualen Häufigkeit der Wortklassen bei 6jährigen Kindern aus Unter- und Mittelschicht (Thomas 1962, Templin 1957) ergab ein Überwiegen von Nomina und Konjunktionen in der Unterschicht, von Pronomina und Adverbien in der Mittelschicht; zugleich aber eine wohl sprachlogisch bedingte große Ähnlichkeit beider Häufigkeitsverteilungen trotz dieser Differenzen. Die gleiche Übereinstimmung und ähnliche Differenzen in der relativen Häufigkeit der Wortklassen konnte Oevermann (1970) für deutsche

Sprachproben nachweisen. Innerhalb der funktionalen Wortklassen (Präpositionen, Konjunktionen) gebraucht die Mittelschicht mehr Konjunktionen, die Unterschicht mehr Präpositionen; innerhalb der lexikalischen Wortgruppen (Substantive, Adjektive und Adverbien, Verben) die Mittelschicht mehr Adjektive und Adverbien, die Unterschicht mehr Substantive. Der Anteil der Verben und der lexikalischen wie funktionalen Wortgruppen insgesamt ist in beiden Schichten jedoch fast gleich. Ein Vergleich der prozentualen Anteile der einzelnen Wortklassen an der Gesamtzahl mit der Zahl der verschiedenen gebrauchten Wörter zeigt ferner, daß die Mittelschichtkinder Adjektive und Adverbien, die Unterschichtkinder Substantive am wenigsten wiederholen. Die erhobenen Strukturmerkmale des Wortschatzes in beiden Schichten kennzeichnen unter theoretischen Gesichtspunkten beliebige Untermengen des Vokabulars der Hochsprache; sie können erst dann Belege für schichtspezifische Strategien des Sprachgebrauchs darstellen, wenn für jede Schicht die Prinzipien der Auswahl aus dem Vokabular der Hochsprache herausgearbeitet werden.

Als dritter Aspekt der Wortschatzentwicklung nach Umfang und Struktur soll das qualitative Verständnis des einzelnen Wortes betrachtet werden. Hierher gehören Variable wie das Abstraktionsniveau und die logische Form von Worterklärungen, die man als die mit einem Wort verbundene kognitive Kategorisierungsweise bezeichnen kann. Hierher gehört auch die unterschiedliche Beherrschung des Wortschatzes in Rezeption und Gebrauch, mündlicher und schriftlicher Äußerung. Wenn sich die bisher referierten Variablen auf zählbare Aspekte objektiv vorliegender Sprachproben bezogen, so erfassen die folgenden Beurteilungen der Qualität einzelner sprachlicher Verhaltensweisen. Diese Variablen zielen nicht auf das Ergebnis, sondern den Prozeß sprachlicher Produktion und können daher durch soziale Strategien des Sprachgebrauchs, wie sie in der Theorie Bernsteins (1964, 1965) konzipiert sind, beeinflußt werden. Wir werden bei der Darstellung dieses dritten Aspekts der Wortschatzentwicklung, der der restliche Teil des Abschnitt 2.2.1. gewidmet ist,

zuerst allgemeine Entwicklungsfolgen sowie Schichtunterschiede im qualitativen Verständnis von Worten feststellen und dann fragen, in welcher Weise diese Schichtunterschiede auf soziale Strategien des Sprachgebrauchs bezogen werden können.

Der aktive Wortgebrauch erfordert eine bewußte und flüssige Vergegenwärtigung (recall) der Bedeutung, der passive lediglich die Wiedererkennung (recognition). Hurlburt (1954) gab 15- und 17jährigen Schülern einen aktiven und einen passiven Vokabeltest mit den gleichen "items", außerdem einen Intelligenztest. Im Durchschnitt konnten die Schüler nur 45% der Worte, die sie passiv wiedererkannten, aktiv gebrauchen, die intelligentere Gruppe jedoch 52% und die weniger intelligente nur 40%. Die Korrelationen zwischen aktiver und passiver Leistung waren mit ca. .60 nur mäßig hoch und für die weniger wie mehr intelligenten Gruppen gleich. Die Leistungen auf beiden Wortschatztesten korrelierten etwa gleich gut mit Intelligenz. Als Gruppe gesehen nutzen die intelligenteren Schüler ihr passives Vokabular stärker im aktiven Gebrauch aus; für jeden Schüler einzeln jedoch, gleich ob er zur mehr oder weniger intelligenten Gruppe gehört, hat die Leistung auf dem einen Test die gleiche, relativ geringe Voraussagekraft für den anderen Test.

Der viel größere Umfang des passiven Vokabulars und die relative Unabhängigkeit beider Vokabular-Arten läßt sich verstehen, wenn man das Alter der Schüler berücksichtigt. Mit 15 bis 17 Jahren treten keine neuen Alltagserfahrungen mehr auf, die zu einer Erweiterung des aktiven Vokabulars zwingen, während durch Lektüre und Lernen ständig neues und von der erfahrenen Umwelt immer weiter entferntes Wissen vermittelt wird, das nur noch den passiven Wortschatz bereichert. Zu Beginn der sprachlichen Entwicklung treten jedoch neue Dinge, Personen und Situationen sowie neue Worte zugleich auf; die Worte werden dem Kind in der Interaktion mit Erwachsenen bekannt und es lernt durch ihre Hilfe und Korrektur, die Worte richtig zu gebrauchen. Man kann daher ein unterschiedliches Verhältnis von aktivem und passivem Vokabular vermuten je nach den

Entwicklungsstufen des Kindes und den mit ihnen verbundenen Arten des Lernens. Solange das Kind noch in alltäglichen sozialen Interaktionen neue Situationen und Regeln kennenlernt, lernt es mit neuen Worten zugleich auch neue Formen des aktiven Umgangs mit diesem Wort. Sobald es aber für sich alleine aufbereitete Erfahrung aus Büchern und in der Schule übernimmt, erweitert es lediglich sein passives Vokabular. Die Differenz zwischen beiden Wortschaft-Arten wäre dann das Ergebnis einer Entwicklung: sie müßte sich ständig vergrößern. Endpunkt der Ausdehnung des passiven Vokabulars ist der gesamte Wortschatz einer natürlichen Sprache; auch wenn sein Umfang - Fremdworte? technische Vokabularien? etc. - strittig für die Linguistik ist, für die Sprecher der Sprache ist er eine objektive Größe. Nur wenige Sprecher aber kennen mehr als 10% der Worte ihrer Muttersprache (Ullmann 1973:297).

Die in einem Wort latenten Kategorisierungen können auf zwei Weisen manifest gemacht werden: durch eine Aufforderung zur aktiven Wortdefinition und durch den Test des passiven Verständnisses, das sich in der Wahl zwischen semantisch richtigen, aber kategorial unterschiedlichen Definitionen zeigt. Feifel/Lorge (1950) protokollierten die aktiven Definitionen, die je 100 Kinder aus den Altersgruppen von 6 bis 14 Jahren auf die 45 Stimulus Worte des Stanford-Binet Vokabel Tests gaben, und klassifizierten sie in 5 Kategorien: 1. Synonym, 2. Gebrauchsweise und Beschreibung, d.h. nicht-definitorische Angabe singulärer Objektmerkmale, 3. Umschreibung, d.h. nicht-kategoriale Erklärung des Wortsinns, 4. Beispiele, 5. Fehler. Die Anzahl der Synonyme stieg zwischen den 6- und 14jährigen sehr stark, die der Umschreibungen leicht an; die Anzahl von Gebrauchsweisen und Umschreibungen und die von Beispielen nahm leicht ab. Die jüngeren Kinder verstehen die Worte konkret und orientieren sich an partikularen Aspekten; erst die älteren verallgemeinern und entdecken sozusagen den Klassencharakter einer Wortbedeutung. L. P. Kruglov (1953) und Russell/Saadeh (1962) konnten dieses Ergebnis mit der zweiten der genannten Methoden bestätigen. In beiden Experimenten wurde ein multiple-choice Test,

in dem die 5 zur Wahl stehenden Möglichkeiten der Definition des Stimulus-Wortes den 5 Kategorien von Feifel/Lorge entsprachen, Altersgruppen zwischen 9 und 15 Jahren gegeben. Es ergab sich ein Entwicklungstrend von konkreten zu abstrakten Kategorisierungen. Ihm entspricht außerdem eine Differenz zwischen normalen und schizophrenen Erwachsenen.¹⁾ Beide der betrachteten Testarten verlangen den Abruf einer Kategorisierung, der Definitionstest aber zusätzlich eine angemessene sprachliche Explikation.²⁾ Daher kann er über das wirklich verstandene konzeptuelle Niveau täuschen, das sich isoliert nur in der Auswahl aus vorgegebenen Kategorisierungen zeigt. Die Bestätigung des bei aktiven Definitionen beobachteten Trends von konkreten zu abstrakten Erklärungen auch in multiple-choice Tests zeigt, daß er tatsächlich eine Entwicklung der Konzepte darstellt. Obwohl die Kinder Definitionen auf einer höheren konzeptuellen Ebene hätten wählen können, entschieden sie sich für die ihrem Entwicklungsstand gemäßen, aber weniger kategorialen Umschreibungen.

1)-----
In einem multiple-choice Test der gleichen Art wählten Schizophrene weniger Klassen- und mehr Beispiels- und Funktionsdefinitionen; zugleich korrelierte die Wahl der "unreifen" Definitionsarten bei Schizophrenen wie Normalen signifikant negativ mit einem Index "intellektueller Verarmung". (Chodorkoff/Mussen 1952). Rabin u.a. (1955) gaben normalen und schizophrenen Erwachsenen den verbalen Subtest eines Intelligenztests und einen Bild-Vokabel-Test, in dem die Bedeutung durch Zeigen auf das richtige von vier Bildern erklärt werden mußte; schließlich mußten die Subjekte die nichtverbal getroffene Wahl verbal begründen. Auf allen drei Maßen waren die Normalen signifikant besser, die geringsten Unterschiede ergaben sich jedoch auf dem nicht-verbalen Bild-Vokabel-Test. Außerdem wurden die Erklärungen für abstrakte und konkrete Worte verglichen: Die Normalen gaben für Abstrakta mehr funktionale und deskriptive, für Konkreta mehr konkretistische Erklärungen und Synonyma oder Definitionen. Die Schizophrenen gaben generell mehr irrelevante Antworten und unterschieden kaum in der Art der Erklärungen für konkrete und abstrakte Worte.

2) Dies zeigt sich auffällig in der Differenz des Prozentanteils der Fehler, der bei aktiven Definitionen (Feifel/Lorge 1950) zwischen 86,0 bei 6- und 61,3 bei 14jährigen liegt; bei allen multiple-choice Tests höchstens bei 13% selbst bei den jüngsten Gruppen.

Welchen Einfluß übt nun die Schichtzugehörigkeit auf die beschriebene Entwicklungsfolge im Wortverständnis aus? Siller (1957) untersuchte schichtspezifische Differenzen der Kategorisierung auf der auch die Entwicklung kennzeichnenden Dimension Abstrakt-Konkret. Er gab Kindern aus der Unter- und Mittelschicht einen multiple-choice Test semantisch gleichwertiger, aber logisch unterschiedlicher Definitionen; die Alternativen waren ähnlich wie bei Feifel/Lorge klassifiziert. Einerseits zeigten sich signifikante Unterschiede zwischen beiden Gruppen: Die Mittelschicht wählte mehr Klassen-Definitionen und weniger Beispiele, außerdem seltener die einzige falsche der fünf möglichen Antworten. Andererseits war in beiden Schichten die Häufigkeitsrangfolge der Alternativen gleich; beide wählten am meisten Beschreibungen, gefolgt von Klassen-Definitionen, Funktionsangaben, Beispielen und Fehlern. In Sillers Experiment waren jedoch die Schichtgruppen auf keiner Variablen kontrolliert; die Intelligenzunterschiede waren hochsignifikant. Sie könnten die unterschiedliche Wahl der qualitativen Kategorien bewirkt haben, ohne daß intelligenzunabhängige soziale Einflüsse zur Erklärung notwendig wären.

Die große Ähnlichkeit der referierten Untersuchungen des Wortverständnisses von Feifel/Lorge (1950), Kruglov (1953), Russell/Saadeh (1962) einerseits und Siller (1957) andererseits erlaubt zumindest einen groben Vergleich zwischen der Entwicklung kognitiver Kategorisierungen und dem Einfluß der sozialen Schicht auf den Gebrauch kognitiver Kategorisierungen:

1. Die Kategorisierungen, deren Prozentanteil sich im Laufe der Entwicklung am stärksten verändert, sind auch am empfindlichsten für soziale Einflüsse. Während in der Ontogenese die kategorialen und synonymen Erklärungen sehr stark zunehmen und die Beispiel-Erklärungen, wenn auch in geringerem Maße, abnehmen, liegt die Überlegenheit der Mittelschichtkinder genau darin, daß sie weniger Beispiele und mehr Synonyme als die Unterschichtkinder wählen.

2. Während die Umschichtungen innerhalb der möglichen Kategorisierungen im Laufe der Entwicklung beträchtlich sind, sind die Differenzen zwischen den Schichten relativ gering. Nur im Vergleich von Entwicklungsstufen, nicht im Vergleich von Schichtgruppen treten Änderungen der Häufigkeitsrangfolgen der Kategorien auf. Die Stufen der Konzeptbildung scheinen daher von der Entwicklung, nicht von sozialen Einflüssen abhängig zu sein. Im durch die Entwicklung gesetzten Rahmen können soziale Einflüsse jedoch systematisch zusammenhängende Veränderungen bewirken. Dies zeigt sich bei Siller (1957) in der Systematik, mit der die Unterschichtkinder, verglichen mit Mittelschichtkindern, anstelle von Klassendefinitionen Beispiele wählen, ohne daß die gemeinsame Rangfolge der Kategorien zerstört wird.¹⁾

Inwiefern können wir Schichtunterschiede im Wortverständnis als Ergebnisse schichtspezifischer Strategien des Sprachgebrauchs verstehen? Zwei Bedingungen lassen sich angeben: Zunächst darf es sich bei den festgestellten Schichtdifferenzen nicht einfach um sprachimmanente Merkmale, sondern es muß sich um Merkmale des Gebrauchs von Sprache für außersprachliche Aufgaben handeln. Ferner müssen die festgestellten Schichtdifferenzen des Sprachgebrauchs unabhängig von allgemeinen Schichtdifferenzen in sprachlichen und intellektuellen

¹⁾-----
Wir wollen die Belege für den Vergleich kurz nachtragen: Der Prozentanteil der Synonyme oder Klassifizierungen nimmt zwischen 8 und 13 Jahren bei Feifel/Lorge (aktive Definitionen) von 6% auf 22,1% und bei Kruglov (multiple-choice) von 20% auf 53% aller Antworten zu. Beide Male zeigt diese Kategorie gegenüber den anderen die stärksten Veränderungen. Der Prozentanteil der Beispiel-Kategorie nimmt zwar zwischen 8 und 10 Jahren leicht zu, fällt dann aber von 23% auf 9% für 13jährige. Der Prozentanteil der im multiple-choice Test bei Siller gewählten Klassendefinitionen liegt für die Mittelschicht bei 33,6%, die Unterschicht bei 27,9%; der der Beispiele für die MS bei 4,0%, die US bei 9,2%; der der Fehler für die MS bei 1,2%, für die US bei 3,0%. Auf den anderen Kategorien ergeben sich keine Unterschiede; die Unterschiede bei Beispiel und Klassendefinitionen sind genau komplementär. - Der Mangel dieser Vergleiche ist, daß nur für die Altersgruppe der 11jährigen, die schon 6 Jahre zur Schule gegangen sind, ein Schichtvergleich vorliegt. Wenn Schichtdifferenzen sich primär in zeitlich verschobenen Entwicklungslinien ausdrücken, so werden sie in diesem Alter schon wieder stark vermindert sein.

Fähigkeiten sein, damit eine Erklärung der Unterschiede durch eine schichtspezifische Verteilung psychischer Fähigkeiten ausgeschlossen ist. Wir wollen einige Untersuchungen zitieren, die Schichtunterschiede unter diesen beiden einschränkenden Bedingungen belegen können.

John (1963) gab Mittel- und Unterschichtkindern einen Test, in dem die Fähigkeit zu konkreter und zu abstrakt-zusammenfassender Bezeichnung von Objekten verglichen wurde. Die Kinder mußten zuerst die auf einem Bild gezeigten Objekte einzeln mit dem zugehörigen Wort aufzählen, dann dem ganzen Bild einen zusammenfassenden Namen geben. Die 5jährigen Kinder aus beiden Schichten lösten beide Aufgaben gleich gut. Bei den 11jährigen jedoch konnten die Mittelschichtkinder für mehr Bilder als die Unterschichtkinder einen integrierenden Namen finden, während die Kinder beider Schichten sich bei der Aufzählung der dargestellten Einzelobjekte nicht unterschieden. Solange lediglich eindeutigen Objekten sprachliche Etikette zugeordnet werden müssen, zeigen Kinder beider Schichten die gleichen Fähigkeiten. Sobald jedoch Sprache als Mittel dient, die gegebenen Stimuli zu überschreiten und sie in neuen Begriffen zusammenzubringen, sind die Mittelschichtkinder im Vorteil. Ferner gab John den Kindern einen Objekt-Sortier-Test, in dem die Gegenstände kategorial und funktional klassifiziert werden konnten, und fragte sie anschließend nach den angewandten Zuordnungsregeln. Die Unterschichtkinder klassifizierten häufiger funktional und konnten ihre Regeln weniger gut verbal erläutern als die Mittelschichtkinder. In einer ähnlichen Untersuchung verglichen Sigel u.a. (1966) das Kategorisierungsverhalten von Mittel- und Unterschichtkindern. Die Kinder mußten Objekte auf drei Repräsentationsebenen sortieren: als konkrete Gegenstände, schwarz-weiß photographiert und farbphotographiert. Die Unterschichtkinder klassifizierten seltener nach analytischen als nach funktionalen Gesichtspunkten. Sie konnten außerdem - im Gegensatz zu den Mittelschichtkindern - ihre Klassifikationsweise nicht auf allen Repräsentationsebenen durchhalten. Die Unterschichtkinder

verfügten also nicht mit der gleichen Sicherheit wie die Mittelschichtkinder über von der Darstellungsform unabhängige Klassifikationsregeln. Da sprachliche Begriffe gerade ein Mittel sind, Klassifikationen unabhängig von der Anschauung zu speichern, deuten diese Ergebnisse auch auf eine sprachliche Benachteiligung. Beide Untersuchungen zeigen, daß die Benachteiligung der Unterschichtkinder besonders im Gebrauch der Sprache liegt: Sie zeichnen in der Sprache eher die Gegebenheiten der Objektwelt nach, als daß sie Sprache als Medium zur Analyse und Synthese der Objektwelt nutzen. Sie verstehen es schlecht, sprachliche Formen als Mittel der kognitiven Kategorisierung (siehe 2.1.1.) zu nutzen.

Daß Schichtdifferenzen im Kategorisierungsverhalten nicht durch Mittelschichtfreundlichkeit der Aufgaben künstlich entstanden und fernerhin unabhängig von gemessener Intelligenz sind, belegen Untersuchungen von Findlay/McGuire (1957) und Siller (1957). Findlay/McGuire gingen von der Kritik an den Intelligenztests aus: Wenn diese einen kulturellen "bias" enthalten, die die Unterschichtkinder gemessen an ihren wahren Fähigkeiten zu schlecht beurteilt, so müßten Unterschichtkinder auf subkulturell neutralen Aufgaben besser abschneiden als Mittelschichtkinder gleicher Intelligenz. Gruppen von 9-, 12-, und 15jährigen Kindern beider Schichten mit jeweils gleichem Intelligenzmittelwert mußten Klötze nach abstrakten, sprachfreien Prinzipien klassifizieren. Ganz im Gegensatz zur Hypothese schnitten die Mittelschichtkinder signifikant besser ab. Als eine Erklärung könnte man wiederum vermuten, daß die Mittelschichtkinder implizite sprachliche Begriffe besser als Lösungshilfe einsetzen können. Diese Annahme wird gestützt durch die Besonderheit der von Findlay/McGuire gestellten Aufgabe. Eine korrekte Lösung verlangte die Kombination von Werten in zwei Dimensionen, außerdem wechselten die Dimensionen von Subtest zu Subtest. Gerade zur Analyse und Synthese von Dimensionen der Objektumwelt können sprachlich fixierte Begriffe - wie wir an den Ergebnissen Johns (1963) und Sigels u.a. (1966) sehen konnten -

ein Mittel sein. Siller (1957) verglich in dem schon zitierten Experiment die Leistungen von 11jährigen Unter- und Mittelschichtkindern in nichtverbalen Klassifikations- und Analogisierungsaufgaben und in verbalen Klassifikations- und Analogisierungsaufgaben. Wenn die Kinder nach den Ergebnissen der nicht-verbalen Tests parallelisiert waren, zeigte die Mittelschicht bessere Leistungen in den verbalen Tests. Umgekehrt aber ergab eine Parallelisierung nach verbalen Leistungen nicht den erwarteten Vorsprung der Unterschicht in den nicht-verbalen Werten. Diese Ergebnisse zeigen eine Überlegenheit der Mittelschichtkinder vor Unterschichtkindern auf verbalen Intelligenzleistungen, obwohl beide Gruppen gleichwertig waren auf nicht-verbalen Intelligenzleistungen. Sie können jedoch nicht belegen, daß die Unterschichtkinder gleichsam von einem höheren nicht-verbalen Intelligenzniveau ausgehen müssen, um die gleiche verbale Intelligenz zu erreichen.

Auch bei gleichen sprachlichen Fähigkeiten wurden jedoch Unterschiede zwischen Kindern verschiedener sozialer Milieus im Wortverständnis festgestellt. Carson/Rabin (1960) haben die aktiven Worterklärungen von Kindern aus drei sozialen Milieus untersucht: Neger aus dem Süden und Norden, und Weiße aus dem Norden der Vereinigten Staaten. Die Kinder waren zuvor nach ihrer Leistung in einem sprachfreien Wortverständnisstest, der die Zuordnung eines Wortes zu dem passenden von vier Bildern verlangte, parallelisiert worden. Trotz der gleichen Ausgangsbedingungen im Wortverständnis waren die Worterklärungen der Negerkinder schlechter als die der weißen, und die der Negerkinder aus dem Süden schlechter als die von Negerkindern aus dem Norden. Die jeweils schlechtere Gruppe gebrauchte signifikant weniger Synonyme, Kategorisierungen und treffende Beschreibungen und gab häufiger falsche oder keine Antworten. Sowohl die Rassezugehörigkeit wie das kulturelle Milieu hatten einen signifikanten Einfluß auf den Sprachgebrauch unabhängig vom Sprachverständnis. Obwohl Kinder aus verschiedenen

Milieus über gleich viele Kategorisierungen verfügen, können sie sie nicht in gleicher Weise explizieren. Verständnis ist zwar Bedingung einer guten kategorialen Explikation, die Explikation aber nicht durch die Qualität des Verständnisses allein bestimmt. Diese Tatsache spricht für die Abhängigkeit der Bedeutungsexplikation nicht nur von sozialer Schicht, sondern von sozialen Strategien des Sprachgebrauchs. Die zweite Messung des hier referierten Experiments ist eine Replikation des Experiments von Carson/Rabin.

2.2.2. Differenzierung sozialer Einflüsse auf den Wortschatz von Kindern

In der Regel wurden soziale Einflüsse auf den kindlichen Wortschatz nur in der Form der globalen Schicht-Dichotomie untersucht. Unabhängig aber von einer weiteren Differenzierung der Dimension soziale Schicht, müssen in genetischer Perspektive die spezifischen sozialen Bedingungen genannt werden, die in beiden Schichtgruppen jeweils den typischen Leistungsstand bedingen. Wir werden in Abschnitt 2.2.2. einige Untersuchungen zitieren, die auf die Bedeutung spezifischer sozialer Faktoren für die Entwicklung des Wortschatzes und allgemeiner der Intelligenz hinweisen. Wir wollen in Abschnitt 3. die Theorie schichtspezifischer Strategien des Sprachgebrauchs einführen als eine mögliche Erklärung für die Schichtunterschiede in der sprachlichen und intellektuellen Entwicklung.

Spezifische soziale Bedingungen für die Genese allgemeiner Leistungsunterschiede zwischen Kindern sozialer Schichten sind in erster Linie die Formen der Interaktion zwischen Eltern und Kind, in zweiter Linie auch ökologische Bedingungen. Diese spezifischen Bedingungen haben eine größere Voraussagekraft für Intelligenz und Schulleistung als die Dimensionen der sozialen Schicht. Bloom (1964:78 und 124) referiert Untersuchungen von Dave und Wolf, in denen verschiedene Variable der Eltern-Kind-Interaktion erhoben wurden: die Leistungserwartungen und -aspirationen

der Eltern, die Bewertung verschiedener Aspekte des Sprachgebrauchs durch die Eltern, sowie sozialökologische und soziokulturelle Bedingungen allgemeinen Lernens. Eine multiple Korrelation dieser Variablen mit der allgemeinen Intelligenz lag bei .76, mit der Leistungsmotivation bei .80. Bloom vergleicht diese multiplen Korrelationen mit den durchschnittlichen einfachen Korrelationen einzelner Schichtindikatoren wie Ausbildung, Einkommen und Beruf, die bei .40 liegen.

Der Einfluß sozialer Bedingungen verstärkt sich im Laufe der Entwicklung. Deutsch (1967:361) hat ein kumulatives Defizit in der Entwicklung sozial, rassisch und milieumäßig benachteiligter Kinder festgestellt; der Vorsprung der bevorteilten Kinder nimmt vom 5. zum 11. Lebensjahr zu. Die Korrelationen fast aller ermittelten Sprach- und Intelligenzwerte mit Schicht liegen im fünften Schuljahr erheblich höher als im ersten. Varianzanalysen zeigen einen signifikanten Interaktionseffekt von Schulklasse (Alter) und einem Index sozialer Deprivation (Deutsch 1967: Tabellen 16.7 und 16.13). Der Deprivationsindex faßt 6 Variable zusammen: den Zustand der Wohngegend, das Aspirationsniveau der Eltern für ihr Kind, die Zahl der Kinder unter 18 in der Familie, die Häufigkeit von Tischgesprächen mit Kindern, vom Kind vorausgesagte kulturelle Aktivitäten der Familie für das nächste Wochenende, Kindergartenbesuch des Kindes; der Deprivationsindex umfaßt also, wie die Erhebungen Daves und Wolfs, Formen der Interaktion zwischen Eltern und Kindern sowie allgemeine sozialökologische und soziokulturelle Bedingungen des Familienlebens. Die kumulative Wirkung der durch die sozialstatistischen Variablen Einkommen, Beruf und Ausbildung gegebenen Benachteiligung könnte auf die mit ihnen verbundene Familien- und Milieuverhältnisse zurückgeführt werden.

Deutsch selbst hat diese Hypothese ausführlich untersucht. Zuerst verglich er die Wirkung der drei sozialen Faktoren: Rasse, Schicht und Deprivation. Sowohl auf Wortschatz- wie Intelligenzleistungen zeigt die mit dem beschriebenen Index gemessene Deprivation Effekte

unabhängig von Schicht und Rasse. Umgekehrt müßten Kompensationen der Deprivation Wortschatz und Intelligenz steigern können unabhängig von Verbesserungen der objektiven Klassenlage. Als zweites untersucht Deutsch typische Einflußmuster dieser Faktoren auf Wortschatz und nicht-verbale Intelligenz in der Ontogenese. Interaktionseffekte von sozialen Variablen und Alter, die der statistische Ausdruck für das "kumulative Defizit" sind, ergaben sich bei den Wortschatz-Werten für Rasse, Status und Deprivation; bei den Intelligenzwerten jedoch nur für Rasse und Deprivation. Deutsch schließt daraus, daß die Schichteinflüsse auf die nicht-verbale Intelligenz im wesentlichen schon vor dem 6. Lebensjahr stattgefunden haben, auf den Wortschatz jedoch erst später einsetzen und länger andauern. Beide Fähigkeiten sind zwar abhängig von Sozialfaktoren, die kumulative Wirkung von Schichteinflüssen jedoch ist nicht dieselbe. Die Intelligenzentwicklung richtet sich stärker nach einem immanenten Programm, das sich qualitativ in den Stadien Piagets, quantitativ mit den Zahlen von Bloom (1964) darstellen läßt: Die Intelligenzwerte von 4jährigen determinieren etwa 50%, die von 8jährigen etwa 80% der Varianz der Intelligenzwerte von 17jährigen. Schichteinflüsse wirken daher am stärksten in der Vorschulzeit und können später kaum mehr zusätzliche Differenzen bewirken. Der Wortschatz ist hingegen primär durch die Lernanregung der Umwelt bestimmt; Schichteinflüsse können sich daher während der ganzen Entwicklung geltend machen und kumulieren.¹⁾

¹⁾-----
Die unterschiedlichen Entwicklungsmuster für Wortschatz und Intelligenz ändern nichts an ihrer hohen Interkorrelation, die in der gleichen Untersuchung bei $r = .67$ lag. Beide müssen also synchron, wie dargestellt, in einen Zusammenhang gebracht werden, auch wenn sie sich in den spezifischen Entwicklungsstadien unterscheiden, in denen sie von sozialen Faktoren am meisten abhängig sind (Deutsch 1967:343).

Als drittes schließlich versucht Deutsch, aus einer Korrelationsmatrix einen Entwicklungsablauf von objektiven Sozialdaten (Status, Beruf, Ausbildung) und Bedingungen der Umgebung (Deprivations-Index) zu Fähigkeiten (Wortschatz, Intelligenz) und Leistungen (Lesen) zu rekonstruieren. Mit Hilfe des Schemas dieser vier kausal wie genetisch aufeinanderfolgenden Komplexe lassen sich diejenigen Größen des Deprivations-Index, die sowohl von objektiven sozialstrukturellen Bedingungen abhängig sind wie auch individuelle Leistungen determinieren, von solchen trennen, die lediglich mit objektiven Bedingungen verknüpft sind, ohne Einfluß auf individuelle Fähigkeiten zu haben. Da der Deprivations-Index hier außerdem um einige Variablen erweitert wurde, ergeben sich zwei interessante Vergleiche (Deutsch 1967: Tab. 17.1). Beide betreffen die Interaktion des Kindes mit Erwachsenen. Erstens liegt die Bedeutung der Eltern-Kind-Interaktion für die Leistungen des Kindes nicht im bloßen Zusammensein, sondern darin, daß es zu einer Unterhaltung genutzt wird. Die Variablen "Gegenwart von Vater bzw. Mutter bei Frühstück bzw. Mittagessen" korrelierten weder mit Sozialstatus der Familie noch mit (Lese-)Leistungen des Kindes, erwiesen sich also in keiner der beiden Richtungen der Kausalkette relevant. Die Variable "Unterhaltung mit dem Kind beim Essen" jedoch hatte in beiden Richtungen signifikante Korrelationen. Zweitens haben Tätigkeiten des Kindes mit erwachsenen Verwandten nicht an sich Bedeutung für die Entwicklung des Kindes, sondern nur, wenn es sich um gemeinsame Aktivitäten außerhalb der Familie handelt. Die Variable "Anzahl der vom Kind für das nächste Wochenende antizipierten Aktivitäten mit Verwandten" zeigte keine, die Variable "Anzahl der antizipierten kulturellen Aktivitäten" beide der genannten Korrelationen. Die Eltern-Kind-Interaktion hat für die sprachliche und intellektuelle Entwicklung des Kindes vor allem dann Bedeutung, wenn sie zum einen sprachlich vermittelt, zum anderen auf Interessengebiete jenseits der Familie gerichtet ist. John/Goldstein haben die Bedeutung der Eltern-Kind-Interaktion für das Lernen sprachlicher Bedeutungen untersucht; F. L. Strodtbeck hat darauf hingewiesen, daß sprachliche Präzision nicht an

sich vom Kind als Wert erkannt und gelernt werden kann, sondern nur, wenn sie in einem Problemkontext sinnvoll und handlungsrelevant wird.

John/Goldstein (1964) gaben 4jährigen Kindern aus der schwarzen Unterschicht einen Bild-Vokabel-Test. Die Kinder konnten vor allem Verben, Worten, die sich auf das Landleben, und solchen, die sich auf in der Unterschicht seltene Gegenstände beziehen, nicht die richtigen Bilder zuordnen. Das Scheitern an den beiden letzten Wortklassen kann durch die offensichtlich andere Erfahrungswelt der Kinder erklärt werden, das Scheitern an den Verben jedoch nicht. Die Kinder versagen bei der richtigen Bezeichnung der Verben, nicht weil die entsprechenden Handlungen ihrer Erfahrungswelt fremd sind, sondern weil sie die verschiedenen beobachteten Handlungsformen nicht sprachlich diskriminieren können. Je weniger eine Bedeutung nämlich durch eindeutige Referenten bezeichnenbar ist, desto mehr muß sie dem Kind durch einen ständig korrigierenden Dialog mit Erwachsenen verdeutlicht werden; dies gilt besonders für Verben, da sie Abstrakta, Handlungen und Beziehungen, bezeichnen. Mangelnde sprachliche Interaktion mit Erwachsenen kann also die Unsicherheit der Unterschichtkinder gerade im Gebrauch von Verben erklären. Sie sind beim Lernen von Bedeutungen auf das rezeptive Zuhören und stabile Wort-Referenz-Beziehungen angewiesen, während die Mittelschichtkinder in aktiver verbaler Interaktion mit Erwachsenen sich in sprachlichen Bedeutungen einüben können.

John/Goldstein zeigen am Beispiel der Verben den allgemeinen Mechanismus des Bedeutungslernens. Das Kind formt entsprechend dem Sprachgebrauch der Erwachsenen eine Hypothese über die Beziehung der Äußerung zu einer Referentenklasse, die seinen eigenen Wortgebrauch leitet. Sein Wortgebrauch wird dann von den Erwachsenen solange korrigiert, bis er mit dem sozial üblichen übereinstimmt (siehe 2.1.1.). Wortbedeutungen stellen jedoch nicht nur eine soziale Norm, sondern zugleich ein Mittel der Umweltbewältigung dar. Soll das Kind beides lernen, so darf seine Interaktion mit Er-

wachsenen nicht lediglich von Sprache begleitet sein. Sprache muß vielmehr für die in Interaktionen zu lösenden Aufgaben relevant werden. Eine bloße Sprachanreicherung ist daher sinnlos, wenn das Kind nicht die Gelegenheit und die Macht hat, sich mit ihrer Hilfe in seiner Gruppe - der Familie - durchzusetzen. Dieser Prozeß kann von beiden Ausgangspunkten her betrachtet werden: Auf der einen Seite muß die objektive Familienstruktur differenziert genug sein, um die Artikulation individueller Standpunkte und Rollen überhaupt erst zu stimulieren; auf der anderen Seite darf eine verfeinerte Sprache nicht gleichsam in der Luft hängen, sondern muß für das Kind einsichtig zur Lösung der Probleme in der Familie beitragen. Strodtbeck (1967) nennt vor allem zwei Problemkontexte, in denen das Kind zu differenziertem Sprachgebrauch motiviert und zugleich durch Erfolge in ihm bekräftigt wird. Erstens wird, bei egalitärer Machtverteilung zwischen den Eltern, das Kind zu wechselnden und begrenzten Koalitionen mit einem der beiden Partner angeregt, in denen es seine jeweilige Stellung zu beiden genau darstellen muß; zugleich kann es auf diese Weise seine eigenen Wünsche gezielt vorbringen und optimal durchsetzen. Eine balancierte Interaktion in der Eltern-Kind-Triade ist selbst der Problemgegenstand, in dem es Sprache einsetzen muß und durch ihren Gebrauch gratifiziert wird. Zweitens wird es durch gemeinsame Aktivitäten, die auf Ziele außerhalb der Familie gerichtet sind, aber interne Beratschlagung verlangen, in differenziertem Sprachgebrauch gefördert. Beide strukturellen Bedingungen - Machtgleichgewicht und Orientierung außerhalb der Familie - begünstigen eher das Mittelschichtkind.

Diese Hypothesen sind von Strodtbeck literarisch dargestellt, aber nicht überprüft worden. Einen bestätigenden Hinweis liefern Wolf und Dave (referiert bei Cazden 1966:196). Eine Bewertung der objektiven Gelegenheiten, die in der Familie zum Gebrauch eines differenzierten Vokabulars bestanden, korrelierte hoch mit der Intelligenz und der Schulleistung des Kindes; eine Beurteilung der Sprache der Mutter durch den Interviewer jedoch nicht. Qualität

der Sprachmodelle und Effektivität des Sprachgebrauchs müssen daher getrennt analysiert werden. Hess/Shipman (1965, 1967) untersuchten die mütterlichen Lehrstile bei einer von Mutter und Kind gemeinsam zu lösenden Aufgabe, deren Struktur jedoch nur der Mutter bekannt war. Die Unterschichtmütter waren unfähig, den Zusammenhang ihrer verbalen Anweisungen mit der gestellten Aufgabe den Kindern deutlich zu machen; sie konnten keine problemgerecht spezifizierten Anweisungen geben. Vor allem aber konnten sie die möglichen Alternativen dem Kind nicht vor der Entscheidung auseinanderlegen, während sie die vom Kind gewählten Schritte dennoch belohnten oder bestrafte. Für das Kind war es daher unmöglich, verallgemeinerte Beziehungen zwischen seinen Handlungen und der Reaktion der Mutter wie zwischen dem jeweiligen Problemstand und den notwendigen Folgeschritten herzustellen, die es ja erst durch die Reaktion der Mutter lernen sollte. Die Unterschichtmutter ist nicht in der Lage, Sprache in Kooperation mit dem Kind zur Problemlösung zu nutzen.

Schichtspezifische Unterschiede der Lehrstile der Mütter sind auch in bezug auf die Erklärung von Wortbedeutungen untersucht worden. Henderson (1970) ließ Mütter von 7jährigen Kindern aus der Mittel- und Unterschicht aus vier alternativen Umschreibungen eines Wortes diejenigen auswählen, die sie ihrem Kind als die beste und zweitbeste Erklärung geben würden. Zu jedem Wort stand jeweils eine allgemeine Definition, ein Antonym, ein spezifisches Beispiel und ein unspezifisches Beispiel zur Wahl. Die Mütter konnten also ihrem Kind das gleiche Wort mit qualitativ unterschiedlichen kognitiven Kategorisierungen erläutern. Die Mittelschichtmütter gaben nicht nur eher allgemeine Definitionen, sie wählten auch von den Beispielen häufiger die spezifischen als die unpräzisen - im Gegensatz zu den Unterschichtmüttern, die beide Beispielarten etwa gleich häufig wählten. Außerdem diskriminierten die Mittelschichtmütter eher in ihren Erklärungen zwischen den Worten; sie wählten unterschiedliche Kombinationen von bester und zweitbesten Erklärungsart, während die Unterschichtmütter für alle Wörter die

gleichen Erklärungsarten bevorzugten. Schließlich gaben die Unterschichtmütter eher als erstes ein Beispiel und dann erst eine Definitionserklärung, während die Mittelschichtmütter als beste Erklärung eine Definition und erst als Erläuterung ein Beispiel wählten. Die Mittelschichtmütter zeigten also einen deduktiven, die Unterschichtmütter einen induktiven Lehrstil.

Diese Untersuchung ist aus zwei Gründen wichtig: Sie kann erstens, als eine Untersuchung des Wortverständnisses der Mutter mit den in 2.2.1. referierten Untersuchungen (Kruglov 1953, Russell/Saadeh 1962) in Zusammenhang gebracht werden, in denen das Kind zwischen semantisch gleichen Umschreibungen logisch unterschiedlicher Qualität zu wählen hat. Die größere Fähigkeit der Mittelschichtkinder, abstrakte und präzise Konzepte mit einem Wort zu verbinden, läßt sich dann aus den abstrakteren und präziseren Erläuterungen verstehen, die die Mittelschichtmütter ihren Kindern geben. Zweitens wird deutlich, daß die Mütter und Kinder aus der Unterschicht sich nicht durch die Kenntnis der Worte von Müttern und Kindern aus der Mittelschicht unterscheiden, sondern durch ihre Art, mit den Worten kognitive Kategorien zu verbinden, durch eine Variable also, die als ein Indikator für Strategien des Sprachgebrauchs angesehen werden kann. Mütter und Kinder der Unterschicht wählten, obwohl ihnen auch präzise Definitionen vorgegeben waren, die unpräzisen Umschreibungen. Die Mütter der Mittelschicht gebrauchen - entgegen gängigen Vermutungen - nicht alleine abstrakte Erklärungen, sondern auch spezifische Beispiele häufiger als Mütter der Unterschicht. Die Art des Sprachgebrauchs in der Mittelschicht gibt den Kindern also nicht nur ein logisch, sondern auch sachlich präziseres Wortverständnis.

3. Die Erklärung der schichtspezifischen Differenzen im Wortverständnis und Wortgebrauch durch die Theorie der sozialen Strategien des Sprachgebrauchs

Wir können unterschiedliche Leistungen und Fähigkeiten von Kindern aus verschiedenen sozialen Milieus aus ihrer Sozialisationsgeschichte erklärt werden? Das ist das Ausgangsproblem der Theorie Bernsteins (1964, 1965). Sie will Unterschiede des Denkens und Handelns zwischen Individuen aus Unterschieden ihrer sozialstrukturellen Herkunft erklären. Weil in der soziologischen Theorie soziale Strukturen als Bedeutungskomplexe (Rose 1962) und soziales Handeln als sinnhaft (Mead, Weber) analysiert wurden, kann das wichtigste symbolische Medium, die Sprache, als Vermittlungsprinzip zwischen sozialer Struktur und sozialem Handeln gewählt werden. Weil zugleich Unterschiede des Handelns aus Unterschieden der Struktur erklärt werden sollen, muß ein theoretischer Aspekt der Sprache gewählt werden, unter dem die gemeinsame Hochsprache sich unterschiedlich betrachten läßt. Einen solchen Aspekt stellen die "linguistischen Kodes" oder präziser die Arten oder Strategien des Sprachgebrauchs dar.¹⁾

7) -----
Die Wahl des Vermittlungsprinzips Sprache ist also durch theoretische Überlegungen motiviert, keineswegs durch die Identifikation mit sozial gängigen Bewertungen eines bestimmten Kulturguts. Ebenso ist die Beschreibung der beiden "linguistischen Kodes" an ihrer Abhängigkeit von Sozialstrukturen und ihrer Notwendigkeit für Handlungschancen, nicht an schichtspezifischen Vorurteilen orientiert. Schließlich ist die Wahl bestimmter manifester sprachlicher Mittel als Indikatoren für "linguistische Kodes" durch die Brauchbarkeit dieser Mittel im sozialen Handeln und nicht durch eine normative Stilistik gerechtfertigt.

Da als Vermittlungsprinzip zwischen sozialer Struktur und sozialem Handeln ihre Sinnhaftigkeit, als ein allgemeines Merkmal, gewählt wird, können mit Hilfe der "linguistischen Kodes" eine Reihe von sozialstrukturell bedingten Unterschieden individueller Motive und individueller Fähigkeiten auf die gleiche Weise erklärt werden: Unterschiede der Leistungsmotivation, der Kontrolle von Impulsen und Affekten, des Planungshorizonts, der kognitiven Stile. Die Theorie der "linguistischen Kodes" ist also hervorragend geeignet eine Vielzahl deskriptiv erfaßter schichtspezifischer Sozialisationsprozesse erklärend zusammenzufassen.

Um das theoretische Konstrukt "Strategien des Sprachgebrauchs" zu bestimmen, geht Bernstein von der Trennung von "Language" und "Speech" aus, die die vom Gesichtspunkt der Grammatik systematischen und kontingenten Elemente der Sprache voneinander ablöst und in der modernen Linguistik als die Trennung von "Kompetenz" und "Performanz" wiedererscheint. Die Strategien des Sprachgebrauchs stellen weder die manifeste Sprache noch die zugrundeliegende Grammatik dar, sondern die Prinzipien der Auswahl innerhalb der Möglichkeiten der Sprachstruktur. Sie werden als eine abhängige Variable nicht nur psychischer Bedingungen (Konzentration, Gedächtnis, Planung) wie in der Linguistik, sondern sozialer Bedingungen betrachtet. Mindestens drei Arten sozialer Bedingungen sind denkbar, die Optionen innerhalb der Struktur der Gemeinsprache lenken können. Zum ersten verpflichtet die Gruppenzugehörigkeit, in Berufen, Parteien oder regionalen Gemeinschaften, zu einem bestimmten Jargon, einem Rotwelsch oder einem Dialekt. Soziologisch handelt es sich um Stände, die sich, prinzipiell unabhängig von der Dimension ungleicher objektiver Lebenschancen, nach funktionalen oder regionalen Kriterien bilden, und um ständische Sprachen, die wie jeder andere Teil der ständischen Kultur normativ kontrolliert werden. Die Zugehörigkeit ist relativ dauerhaft, eindeutig und kann daher vom Individuum nur schwer nach Belieben manipuliert werden. Zweitens verpflichten mehr oder minder standardisierte Situationen zur Wahl sprachlicher Formen, Tonlagen und Inhalte. Soziologisch handelt es sich um Situationsrollen und entsprechende sprachliche Register. Die Zugehörigkeit ist äußerst flüchtig, oft nur durch objektive Gegebenheiten, wie z.B. Räume, gesichert und daher sehr leicht vom Individuum kontrollierbar. Drittens können die besonderen Qualitäten der Sozialbeziehungen eine bestimmte Art des Sprachgebrauchs fördern; sie sind unabhängig von sektoralen Normen des Handelns und situationsübergreifend, weil sie sich aus den typischen Handlungs- und Interaktionsformen im Arbeits- und Familienmilieu sozialer Schichten ergeben, die von den Handelnden internalisiert sind. Sie sind daher den Individuen ebenso dauerhaft wie selbstverständlich und ohne Sanktion eigen.

Auch wenn Bernstein nun zuweilen Beispiele ständischer und situativer Sprachen (Militär, Begrüßung auf Cocktailparty) gibt, so setzt er für die soziologische Erklärung des Sprachgebrauchs doch ausschließlich die Ebene der "Qualität von Sozialbeziehungen" an, und versucht in Parallele zur Klassifikation der "Qualität von Sozialbeziehungen" die "Linguistischen Kodes" zu charakterisieren. Idealtypisch stehen personenorientierte, rollenspezifische und konfliktoffene Sozialbeziehungen als Bedingungen des elaborierten Kodes statusorientierten, rollendiffusen und konsensbezogenen Sozialbeziehungen als Bedingungen des restringierten Kodes gegenüber. Wenn unter universalistischem Sprachmodell ein allen Mitgliedern der Gesellschaft zugängliches und unter partikularistischer Bedeutung eine nur im pragmatischen Kontext erklärbar zu verstehen ist, so zeichnet sich der restringierte Kode durch universalistische Sprachmodelle und partikularistische Bedeutungen, der elaborierte durch partikularistische Sprachmodelle und universalistische Bedeutungen aus. Beide "Linguistischen Kodes" werden zwar in den jeweils typischen Formen sozialer Beziehungen laufend bekräftigt, sie stellen aber darüberhinaus in der frühen Sozialisation internalisierte Einstellungskomplexe dar, die das Handeln auch in den jeweils untypischen Formen sozialer Beziehungen bestimmen können.

Wir können die Charakterisierung von Strategien des Sprachgebrauchs verallgemeinern auf den Gebrauch von Symbolen überhaupt: Auch noch die Entscheidung, ob eine Ausdrucksintention sprachlich oder in anderen symbolischen Medien, Mimik oder Gestik, ausgedrückt oder ob sie überhaupt inexplizit gelassen wird, unterliegt einem sozialen "Kode". Allein aus kontingenten Gründen - weil wir uns, ebenso wie Bernstein, mit sprachlichem Material beschäftigen - sprechen wir von "Linguistischen" Kodes oder Strategien des Sprachgebrauchs. Auf der Ebene der Sprache können soziale Strategien Entscheidungen bezüglich der Grammatik wie des Wortschatzes bestimmen. In unserem

Experiment beschränken wir uns auf den Wortschatz. Wir wollen prüfen, ob erstens schichtspezifische Strategien bei Grundschulkindern den aktiven Gebrauch des Wortschatzes beeinflussen und ob zweitens mit schichtspezifischen Strategien des aktiven Gebrauchs des Wortschatzes soziologisch erklärbare Konsequenzen für die kognitive Entwicklung verbunden sind. Wir wollen diese beiden Fragen im folgenden erläutern und mit der ersten beginnen.

Im Wortschatz einer Sprache stehen dem Sprecher wie dem Hörer eine Vielzahl möglicher Kategorisierungen zur Verfügung, mit denen sie Objekte - gleich ob sachliche, subjektive oder soziale - erfassen können. Das zu kategorisierende Objekt hat an sich keine eindeutige Klassenzugehörigkeit. Ein Löffel kann als "Löffel", "Besteck", "Silber", "Kunstgewerbe" oder "Gebrauchsgegenstand" kategorisiert werden. Eine Gefühlsregung ist kein Ding jenseits der Sprache; was sie ist, hängt davon ab, ob sie als "Liebe", "Zuneigung" oder "Sympathie" bezeichnet wird. Die Leistung des Sprechers besteht darin, seinen persönlichen Intentionen und der sozialen Situation angemessene Kategorisierungen, das treffende Wort, zu finden; die Leistung des Hörers darin, die wichtigsten mit dem Wort verbundenen Kategorien wiederzufinden und sie in der Situation wiederanzuwenden. In vielen Fällen ist die Verbindung von Situation und Sprachgebrauch sozial eingespielt, so daß sie auf psychologischer Ebene habitualisiert ist und kaum als Wahl bewußt wird. In anderen Fällen ist sie offen und stellt eine produktive Leistung zur Meisterung der Situation dar; auch in der offenen Situation müssen jedoch Sprecher und Hörer auf ihre jeweils nächstliegenden und verständlichsten Kategorisierungen zurückgreifen und ihren jeweils typischen, sozial eingespielten Arten des Sprechens und Verstehens folgen. Die Auswahl wie das Verständnis von Worten kann, mit einem Wort, gleichermaßen in Abhängigkeit von sozialen Strategien des Sprachgebrauchs gesehen werden.

Wenn wir nun die Abhängigkeit des aktiven Wortgebrauchs von sozialen Strategien des Sprachgebrauchs überprüfen wollen, so muß bei der Überprüfung die Wortkenntnis kontrolliert werden. Wenn Kinder beider Schichten - als Test des Wortgebrauchs - die Bedeutung verschiedener Worte in einer relativ unnatürlichen Testsituation¹⁾ erklären müssen, so werden Kinder der Unterschicht der Strategie reduzierter, Kinder der Mittelschicht aber der Strategie expliziter Verbalisierung folgen. Wir können voraussagen, daß trotz gleichwertigem allgemeinen Wortverständnis die Worterklärungen der Mittelschichtkinder eher eine genaue Explikation der Kategorien der Wortbedeutung enthalten werden. Die objektiv - d.h. an den sprachlichen Symbolen nachprüfbar - besseren Erklärungen der Mittelschichtkinder werden aber beim Zuhörer genauere Kategorisierungen des gemeinten Sachverhalts hervorrufen und ihn zu einer anspruchsvolleren Reaktion veranlassen, die für das Mittelschichtkind wiederum größere Lernchancen bietet; dieser Prozeß verläuft unabhängig vom tatsächlichen Verständnis der Kinder und ist allein durch den objektiven Gehalt sprachlicher Symbole vermittelt.²⁾ Wenn wir über die Folgen der Reaktionen des Zuhörers für die Entwicklung des Wortschatzes der Kinder spekulieren, gelangen wir schon zur zweiten Ausgangsfrage unseres Experiments, der Konsequenz sozialer Strategien des Sprachgebrauchs für die kognitive Entwicklung.

¹⁾ -----
Zur Frage, ob die natürliche Aktivierung des Wortschatzes in Alltagssituationen in einem Test von Wortbedeutungserklärungen abgebildet werden kann und ob eine solche Konkretisierung nicht schon schichtspezifisch diskriminiert, siehe Kapitel II.1.

²⁾ Wir haben versucht neben dem objektiven Gehalt der Erklärungen auch ihre sprachliche Form auszuwerten. Wenn die Reaktion des Zuhörers nicht auf den durch die sprachliche Form evozierten Kategorien, sondern auf einer Bewertung der sprachlichen Form selber beruht, bleiben die Konsequenzen für den Sender doch in der gleichen Weise unabhängig von seinem tatsächlichen Verständnis (siehe III.6.).

Kognitive Kategorisierungen können vom Hörer nur mit der Präzision empfangen werden, mit der der Sprecher sie sprachlich enkodiert; die reduzierte Explikation des Unterschichtsprechers evoziert, unabhängig von dem, was der Sprecher tatsächlich sagen will, die kognitiven Kategorisierungen des Zuhörers. Entsprechend wird der Hörer eine Antwort mit geringerem Anspruch und geringerer Informationseingabe vorbringen.¹⁾ Der Sprecher wiederum enthält nun auch tatsächlich weniger sprachliche und kognitive Differenzierungen und wird gleichsam die nächste Runde der Kommunikation mit tatsächlichen geringeren Kenntnissen beginnen. In einem zirkulären sozialen Kommunikationsprozeß kann sich eine den Mittelschichtkindern vergleichbare Ausgangskennntnis der Unterschichtkinder durch die von den Unterschichtkindern internalisierte Strategie reduzierter sprachlicher Explikation in eingeschränkte Reaktionen der Zuhörer und einen tatsächlich geringeren Lernertrag der Unterschichtkinder umsetzen. Fortgesetzt über die ganze primäre Sozialisation kann dieser Zirkel eine zunehmende Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung verursachen.²⁾ Anders als in der Theorie Whorfs (1963), die eine intrapsychische Kausalbeziehung zwischen Sprache und Denken postuliert, hat die Sprache in dieser Konzeption Einfluß auf die kognitive Entwicklung nur über die sozialen Folgen internalisierter Strategien des aktiven Sprachgebrauchs.

1) -----
Die Reaktion des Zuhörers auf die durch die Sprache des Sprechers evozierten Kategorien wird in den meisten Fällen habituell festgelegt. Daß die Reaktion des Zuhörers auch bewußt kalkuliert werden kann, bezeichnet die Stelle pädagogischer Eingriffe, die ebenso fördernd wie gerade durch eine "Kindgemäßheit" retardierend sein können. Umgekehrt könnte man die habituellen Reaktionen des Zuhörers als eine implizite Pädagogik bezeichnen.

2) Die Wirkung dieses Prozesses wird noch verstärkt, wenn man zwei für sozialisatorische Kommunikationen typische Bedingungen hinzunimmt: das Kind als Sprecher kann weniger als ein Erwachsener die außersprachlichen Bedingungen manipulieren, die seine Absicht trotz allen Ausdrucksmängeln verdeutlichen könnten; und es ist auf das Entgegenkommen des erwachsenen Zuhörers angewiesen, der, was das Kind inexplizit gelassen hat, explizieren oder das Kind durch neue Informationen herausfordern kann.

Um nun die Schichtabhängigkeit der kognitiven Entwicklung allein auf den langfristigen Einfluß schichtspezifischer Arten des aktiven Sprachgebrauchs zurückzuführen, müssen wir zugleich die Unabhängigkeit des passiven Sprachverständnisses von der Schichtzugehörigkeit nachweisen. Dazu müssen wir wiederum die Wortkenntnis kontrollieren. Wenn Kinder beider Schichten den gleichen sprachlichen Formen ausgesetzt sind, werden sie aus ihnen bei gleicher Wortkenntnis die gleichen kognitiven Kategorisierungen gewinnen; die Entwicklung unterschiedlicher kognitiver Kategorisierungen kann nur noch von allgemeinen Fähigkeiten der Kinder abhängen, und zwar in beiden Schichten in gleicher Weise.

Wenn beide Hypothesen zugleich sich bestätigen lassen, die Überlegenheit der Mittelschichtkinder in der aktiven Erklärung und die Gleichheit von Kindern beider Schichten im passiven Verständnis, jeweils bei Kontrolle der allgemeinen Wortkenntnis, dann können wir eine soziologische Erklärung für die Tatsache der im Vergleich zu den Mittelschichtkindern geringeren Intelligenz der Unterschichtkinder (Oevermann 1966, 1970, Bosse u.a. 1972) und für den geringeren passiven Wortschatz der Unterschichtkinder (Carson/Rabin 1960, Sommer 1972) - zugleich unsere Hypothese für die erste Messung - durch die langfristigen Wirkungen schichtspezifischer Arten des Sprachgebrauchs geben, ohne auf Erklärungen durch schichtspezifische Verteilungen von Fähigkeiten zurückgreifen zu müssen. Verstehen und Gebrauch des Wortschatzes werden in der hier ausgeführten soziologischen Erklärung angesehen als Stadien in einer Mikro-Entwicklung, des Lernens einzelner Bedeutungen, die immer wieder durchlaufen werden, bis das gesamte System der Bedeutungen aufgebaut ist. Wir können diese soziologische Erklärung in unsere entwicklungspsychologische Darstellung (2.1.1.) an der Stelle einfügen, an der wir die Notwendigkeit des Sprachgebrauchs für den Aufbau des Bedeutungssystems herausgearbeitet haben.

Abschließend soll die hier ausgeführte Begründung der beiden Hypothesen gegen ähnliche Begründungen abgegrenzt werden, die durch unsere Annahmen zwar nicht ausgeschlossen, für das angestrebte Ziel einer soziologischen Erklärung von Unterschieden in der kognitiven Entwicklung zwischen Kindern beider Schichten aber unzureichend sind. - Zum ersten kann die Überlegenheit der Mittelschichtkinder in der aktiven Erklärung nicht aus der besseren Anregung durch mittelschichttypische Sprachmodelle erklärt werden. Unterschiede der Umwelt-Anregung können Unterschiede des passiven Wortschatzes, nicht aber Unterschiede des aktiven Wortschatzes bei gleichem passiven Wortschatz erklären.

Zweitens spielen für die Voraussage des besseren aktiven und passiven Wortschatzes bei den Mittelschichtkindern die Strategien des Sprachgebrauchs auf seiten der Interaktionspartner des Kindes nur eine untergeordnete Rolle. Eine solche Erklärung liegt zwar durchaus im Rahmen der Theorie der linguistischen Kodes und kann die geringere Informationseingabe der Unterschichtkinder zusätzlich erklären; auch könnte sie die trotz gleichen passiven Wortverständnisses geringere aktive Erklärungsleistungen der Unterschichtkinder voraussagen, da unterschichttypische Strategien des Sprachgebrauchs für Kinder wie Eltern gelten. Sie könnte jedoch nicht erklären, wie langfristig eine Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung allein aus ihrem Sprachgebrauch, unabhängig von faktischen Fähigkeiten, sich ergeben kann. Das ist nur möglich, wenn man sich auf die unabhängig von der passiven Wortkenntnis wirksamen Strategien des Sprachgebrauchs auf seiten der Kinder konzentriert und versucht, die Folgen einer reduzierten Explikation im sozialen Kommunikationsprozeß zu extrapolieren.

Zweites Kapitel: Ansatz, Methoden und Durchführung der
=====
Untersuchung.
=====

1. Design und Hypothesen des Experiments.

Wir haben im ersten Kapitel ausgeführt, daß die Theorie der linguistischen Codes eine Erklärung für die empirisch nachgewiesene (evtl. sogar zunehmende: Whiteman/Brown/Deutsch 1967) Diskrepanz zwischen Unter- und Mittelschichtkindern in der kognitiven Entwicklung geben kann; sie zeigt, wie schichtspezifische Arten des Sprachgebrauchs beim Sprecher über die Evokation kognitiver Kategorien beim Hörer den Lernertrag des Sprechers auf Dauer beeinflussen. Im Experiment sollen zwei wichtige Etappen dieses sozialen Kommunikationsprozesses geprüft werden, mit Hypothesen, die wir im vorangehenden Abschnitt theoretisch begründet haben. Mittelschichtkinder werden Wortbedeutungen besser aktiv explizieren als Unterschichtkinder, obwohl beide über ein gleich großes passives Repertoire verfügen. Und: Kinder beider Schichten werden aus den gleichen sprachlichen Formen die gleichen kognitiven Kategorisierungen herausholen. Entsprechend den beiden Annahmen gliedert sich das Experiment in zwei Teile.

Im ersten Teil soll geprüft werden, wieweit Kinder, die über ein gleich großes Repertoire an Konzepten verfügen, dennoch schichtspezifisch sich in der sprachlichen Explikation unterscheiden. Es wurde, leicht verändert, das Experiment von Carson/Rabin (1960) wiederholt. 10jährige Kinder aus unterschiedlichen sozialen Milieus erhielten einen standardisierten multiple-choice Wortschatztest, der das konzeptuelle Verständnis der Worte messen sollte. Dieses wäre zwar am besten durch die sprachfreie Zuordnung eines Wortes zu Bildern gemessen worden. Da es aber keinen deutschsprachigen Bild-Vokabel-Test gibt und eine Neukonstruktion umfangreiche Vor-

arbeiten erfordert, wurde ein rein sprachlicher Test übernommen, bei dem sowohl das Stimulus-Wort wie die Erklärungsalternativen sprachlich formuliert sind. Obwohl er also in gewissem Maße schon die sprachliche Enkodierung mit erfaßt, verstehen wir diesen Test als Test vor-sprachlicher Kategorisierungen. - Nach der Leistung in diesem Test wurden Kinder gleichen Geschlechts aus der Unter- und Mittelschicht parallelisiert. Sie erhielten dann eine Parallelform des gleichen Tests, dessen 60 Stimulus-Worte sie nun erklären mußten, jeweils 30 Worte schriftlich und mündlich. Die Erklärungen der Wortbedeutungen sollen messen, wie groß der aktive Wortschatz der Kinder ist, der Wortschatz also, den sie in beliebigen Sprechsituationen gebrauchen können. Die Bedeutungserklärung der Worte eines standardisierten Tests stellt nun gegenüber dem alltäglichen Gebrauch des Wortschatzes, den sie erfassen soll, eine zweifache Einschränkung dar, die wir in einem kurzen Exkurs diskutieren wollen.

Erstens wählt ein Sprecher seinen aktiven Wortgebrauch gewöhnlich nach den Notwendigkeiten der gegebenen Situation, ohne notwendig den allgemeinen Sinn und die spezifische Situationsangemessenheit seiner Worte explizieren zu können. Die Bedeutungserklärung verlangt dagegen die Herausarbeitung des allgemeinen Sinns unabhängig von irgendeiner Situation und stellt eine spezifische Situation auf höherer Ebene dar. Die bewußte Konzentration auf Sprache in einer für alle zu erklärenden Worte gleichen Situation soll über den habitualisierten Wortgebrauch in wechselnden Alltagssituationen Auskunft geben. Möglicherweise aber verfügen Kinder der Mittelschicht eher über die in der Testsituation geforderte bewußte Einstellung zur Sprache und schneiden bei der aktiven Worterklärung besser ab, ohne daß sie in alltäglichen Situationen tatsächlich einen größeren Wortschatz als die Kinder der Unterschicht aktivieren. Zur Rechtfertigung der Operationalisierung läßt sich zunächst anführen, daß es kaum praktikable Alternativen gibt; Beobachtungen des natürlichen Sprachgebrauchs scheitern am Problem der Vergleichbarkeit von Situationen und ausgedrückten Bedeutungen zumindest praktisch, wenn nicht sogar prinzipiell. Fernerhin müßten die Strategien des

Sprachgebrauchs, die als situationsüberdauernde Einstellungen zu Sprache konzipiert sind, sich auf die Erklärungen von Worten in Testsituationen und die Aktivierung des Wortschatzes im Alltag in der gleichen Weise auswirken; der mutmaßliche Vorteil von Kindern der Mittelschicht bei der Worterklärung, der keinen faktisch größeren aktiven Wortschatz vertritt, müßte sich ebenso wie der vermutete Überschuß der Worterklärung über den passiven Wortschatz durch die für die Mittelschicht typische Strategie expliziter Verbalisierung ergeben. Schließlich kann man die Worterklärung, gleichgültig wie weit sie den Wortgebrauch ungetreu abbilden mag, als eine Leistung an sich auffassen, die zusätzlich zum impliziten Wortverständnis, das im alltäglichen Gebrauch aktiviert wird, eine bewußte Darstellung verlangt und von sozialen Strategien des Sprachgebrauchs abhängt. Die Bewertung einer solchen Leistung müßte sich an ihrer Notwendigkeit für soziales Handeln orientieren und kann nicht allein schon deshalb negativ ausfallen, weil sie vermutlich von Mittelschichtkindern besser beherrscht und ausgebeutet wird. In jedem Fall ist ihre Verwandtschaft mit tatsächlich geforderten Schulleistungen deutlich (siehe auch Kapitel III.5.8.2.).

Zweitens werden in der Testsituation nur eine beschränkte Anzahl von Worten abgefragt, die nicht unter soziologischen Gesichtspunkten ausgewählt wurden. Wir haben die Möglichkeit der Schichtdiskriminierung durch den Test selber in Kapitel III.5.4.3., überprüft und konnten keine Begünstigung der Mittelschicht feststellen. - Wir kehren nach diesem Exkurs zur Darstellung des Experiments zurück.

Im zweiten Teil des Experiments soll überprüft werden, wie der aktive Sprachgebrauch in der sozialen Kommunikation die kognitive Entwicklung der Unterschichtkinder beeinflusst. Da die vom Sprecher der Unterschicht gemäß seiner Strategie des Sprachgebrauchs gewählte weniger explizite sprachliche Form unabhängig von seiner Ausdrucksintention die kognitive Kategorisierung des Hörers evoziert, könnten die entsprechend reduzierten Reaktionen des Hörers auf die Dauer zu einem weniger gegliederten System

sprachlicher und kognitiver Kategorien und einem geringeren passiven Wortschatz beim Unterschichtkind führen. Allgemein gesprochen, führt eine Faktorenkette vom Sprachgebrauch des Sprechers, zu evozierten Kategorien und Reaktionen des Zuhörers und wieder zum passiven Wortschatz bzw. den kognitiven Kategorien des ursprünglichen Sprechers. Diese Vermittlung von Sprache und Denken durch soziale Kommunikation sollte durch eine experimentelle Zerlegung der Gesprächssituation überprüft werden, die wir kurz beschreiben wollen.

Die aktiven Erklärungen der Kinder wurden zu einem neuen passiven Wortschatztest der multiple-choice-Form zusammengestellt, der als Stimulus-Wort und Antwort-Alternativen Erklärungen der Kinder zum Stimulus-Wort des Ursprungs-Tests enthielt. Diese Alternativen sind nicht falsch oder richtig, sondern stellen mehr oder minder präzise Kategorisierungen oder verschiedenartige Verständnisweisen des Stimulus-Wortes dar. Wir betrachten die aktiven Erklärungen als Äußerungen in einer Gesprächssituation, deren Empfängerseite im neukonstruierten passiven Test imitiert ist. Die als neuer Stimulus verwandte aktive Erklärung bildet die vom Sender übermittelte sprachlich manifeste Botschaft und die im neuen multiple-choice-Test gewählte Antwort die beim Empfänger evozierte kognitive Kategorisierung.

Dieser Test wurde einer zweiten Gruppe von Kindern gegeben, die ebenfalls nach ihrer Leistung im standardisierten passiven Wortschatztest parallelisiert worden waren. Die Hypothese ist, daß die auf denselben Wort-Stimulus zurückgehenden Bedeutungsexplikationen der Kinder verschiedener sozialer Schichten jetzt bei den Kindern beider Schichten die gleichen Kategorisierungen hervorrufen. Da mit dem ersten passiven Wortschatztest die allgemeine Wortkenntnis bei Kindern der beiden Schichten kontrolliert ist, dürfte die Wahl der Kinder nur noch von der Qualität des neuen Ausgangs-Worts und den zur Wahl stehenden Alternativen abhängen. Das neue Stimulus-Wort müßte bei Kindern beider Schichten die gleichen kognitiven Kategorisierungen evozieren, die zur Wahl der gleichen Response-Alternativen führen.

Wenn diese Hypothese bestätigt werden kann, so eröffnen sich Rückschlüsse auf die im passiven Wortschatztest des ersten Teils gemessenen Differenzen. Wenn Kinder beider Schichten aufgrund derselben Art des Sprachgebrauchs ihre Ausdrucksintentionen gleich gut explizieren und die Zuhörer gleich ausführliche und präzise sprachliche Reaktionen geben, dann werden kurzfristig die Eingabe für Kinder beider Schichten in jeder Gesprächssituation und langfristig die Wirkung des Sprachgebrauchs für die sprachliche und kognitive Entwicklung von Kindern beider Schichten sich nicht unterscheiden. Unterschiede des passiven Wortschatztests können sich dann aus den durch soziale Strategien des Sprachgebrauchs bedingten geringeren Eingaben für die Unterschichtkinder ergeben haben. In dieser Ableitungskette von den Strategien des Sprachgebrauchs beim Sprecher über die evozierten Kategorien beim Hörer und die sprachlichen Reaktionen des Hörers zu den evozierten Kategorien des Sprechers sind nur die beiden letzten Glieder im neu konstruierten passiven Wortschatztest operationalisiert. Für unterschiedliche Strategien des Sprachgebrauchs beim Sprecher wären Schichtunterschiede in der aktiven Worterklärung, die in der zweiten Messung an einer anderen Gruppe untersucht wurden, ein Indiz.

Wir wollen die Messungen und Hypothesen der beiden Teile des Experiments kurz zusammenfassen. Das Experiment besteht aus insgesamt drei Messungen: die passiven und aktiven standardisierten Tests des ersten Teils und den neukonstruierten passiven Teil des zweiten Teils. Für diese Messungen ergeben sich folgende Hypothesen. (1) Die Mittelschichtkinder erreichen im passiven Wortschatztest bessere Leistungen als die Unterschichtkinder, (2) auch unabhängig von gemessener Intelligenz. (3) Auch bei Kontrolle der Leistung im standardisierten passiven Wortschatztest geben die Mittelschichtkinder im aktiven Test bessere Antworten als die Unterschichtkinder. Sollte diese Hypothese bestätigt werden, so werden wir folgern, daß soziale Strategien des Sprachgebrauchs die Auswahl von Worten aus dem passiven Repertoire steuern. (4) Die beiden Schichtgruppen, wiederum bei Kontrolle der Leistung im standardisierten passiven Wortschatztest, unterscheiden sich nicht mehr in den Antworten im dritten Test, vielmehr hängt die Antwort von der sprachlichen Qualität der Stimuli ab. Sollte auch diese Hypothese bestätigt werden, so werden

wir folgern, daß auch das passive Repertoire von Worten und allgemeiner die kognitive Entwicklung von sozialen Strategien des Sprachgebrauchs beeinflußt werden. Wenn die Hypothesen (3) und (4), der aktiven Worterklärung und des passiven Wortverständnisses, zugleich bestätigt werden, dann wäre eine soziologische Erklärung für die vielfach belegte Benachteiligung der Unterschicht gegenüber den Mittelschichtkindern in der kognitiven Entwicklung gewonnen: sie ließe sich, unabhängig von Unterschieden psychischer Fähigkeiten, als Ergebnis unterschiedlicher Strategien des Sprachgebrauchs verstehen.

2. Stichprobe und Variable des Experiments.

Für die Wahl des Alters der Stichprobe waren zwei Gründe maßgebend: Die Kinder sollten auf der einen Seite alt genug sein, um die spezifischen Strategien des Sprachgebrauchs ihres sozialen Herkunftsmilieus wirksam internalisiert zu haben; sie sollten auf der anderen Seite noch nicht durch das Schulsystem schichtspezifisch so ausgelesen sein, daß die Einflüsse von Schicht und Ausbildung miteinander verquickt wären. Beide Bedingungen sind für das 4. Schuljahr der Grundschule erfüllt.

Für die Wahl des Orts der Stichprobe konnten Gesichtspunkte der Repräsentativität nicht maßgebend sein. In der Stichprobe sollten vielmehr Milieus der Unter- und der Mittelschicht etwa gleich stark vertreten und möglichst viele allgemeine wirtschaftliche, ökologische und kulturelle Einflüsse ausgeglichen sein. Aus diesem Grund sollte die Schülerpopulation eines Orts vollständig für den 4. Grundschuljahrgang untersucht werden. Es wurde die Stadt Oberursel gewählt, die als Villenvorort Frankfurts dient, aber auch aus größeren Siedlungen von Arbeitern der örtlichen Maschinenbau-Industrie besteht, so daß ein Gleichgewicht der Schichtgruppen gesichert erschien. Dennoch ergab sich ein zu niedriger Prozentsatz von Unter-

schichtkindern, so daß die zentrale Schule einer städtischen Vorort-Siedlung, wiederum vollständig für den 4. Grundschuljahrgang, nacherhoben werden mußte, die vorwiegend von kinderreichen Familien der Unterschicht bewohnt wird (Frankfurt-Goldstein). Die Stichprobe kann als repräsentativ für großstädtische Ballungszentren in Deutschland gelten.

Es wurden 9 Klassen mit 274 Kindern in Oberursel, 2 Klassen mit 65 Kindern in Goldstein getestet. Die insgesamt 339 Kinder erhielten den Frankfurter Wortschatztest und einen Intelligenztest, von ihnen wurden 83 Kinder in der zweiten und 49 Kinder in der dritten Messung erneut getestet. Das Experiment ist eine unter mehreren soziolinguistischen Untersuchungen, die auf einer Auswahl aus der gleichen Ausgangsstichprobe von 339 Kindern beruhen.

Über diese Ausgangsstichprobe haben wir Informationen aus dem Wortschatz- und dem Intelligenztest, aus den Klassenbüchern (Schulnoten) und aus einem Elternfragebogen. Der Fragebogen enthielt Fragen nach der Ausbildung und dem Beruf der Eltern und Großeltern, nach der Anzahl und der Ausbildung der Geschwister, nach dem Familiennetteinkommen und dem Buchbesitz der Familie, sowie Einstellungsfragen zu Kommunikations- und Kontrollweisen der Mutter. Wir werden die mit diesen Informationen gebildeten Variablen im Zusammenhang mit den Einzelanalysen zur zweiten (III.5.6.ff.) und dritten (IV.3.6.ff.) Messung unseres Experiments beschreiben; wir werden die abhängigen Variablen, die Experimentalleistungen, bei der Darstellung der Ergebnisse der zweiten (III.1.) und dritten (IV.1.) Messung ausführlich beschreiben; wir wollen an dieser Stelle nur die für beide Experimentalmessungen identische unabhängige Variable "soziale Schicht" und die für beide Messungen identischen Kontrollvariablen "Wortschatz" und "nichtverbale Intelligenz" beschreiben.

Darstellung der unabhängigen Variablen. Die Festlegung der Schichtzugehörigkeit beruht in der Regel auf einer gewichteten Kombination von Werten in den drei Dimensionen Einkommen, Beruf und Ausbildung. Es wird ein summarisch-deskriptiver Index gebildet, der eher statistische Gruppen als faktische Sozialstrukturen erfaßt und keine das Handeln unmittelbar erklärende Variable darstellt. Dies gilt auch für die hier vorgenommene Schichteinteilung. Jedoch wurde zusätzlich noch die Ausbildung der Mutter, als wichtiger Einflußfaktor für das Sprachverhalten des Kindes, hinzugenommen und die Dimensionen miteinander nicht nach einem statistischen, sondern einem Verfahren verbunden, das die verschiedenen Bereiche flexibel nach ihrer jeweiligen Bedeutung zusammenbrachte.

Der Fragebogen gab Auskunft über Ausbildung und gegenwärtigen Beruf des Vaters, Ausbildung der Mutter und Familien-Nettoeinkommen. In diesen 4 Dimensionen wurden a priori Schichtgrenzen für 5 Schichten festgelegt und jedes Individuum in allen 4 Dimensionen klassifiziert. Bei der Kombination dieser Schichteinteilungen dienten die Dimensionen in der Reihenfolge ihrer Wichtigkeit gleichsam als Filter. Der Beruf des Vaters bildete den Anker der Schichtzuweisung, mit dem die anderen Dimensionen nach der Reihe verglichen wurden: zuerst die Ausbildung des Vaters, dann die Ausbildung der Mutter und zuletzt das Familien-Nettoeinkommen. Die durch die nachfolgenden Dimensionen festgelegte Schichtzuweisung eines Individuums konnte die der Ankerdimension nur noch nach oben, nicht mehr nach unten verändern. Deshalb wurden die Kriterien in der Anker-Dimension ziemlich streng angesetzt. Als Obergrenze der "Unteren Unterschicht" galten angelernte Arbeiter; der "Oberen Unterschicht" Facharbeiter; der "Unteren Mittelschicht" kleine Selbständige, selbständige Landwirte und einfache nicht-manuelle Angestellte bzw. Beamte; der "Mittleren Mittelschicht" mittlere Selbständige, mittlere und gehobene Angestellte bzw. Beamte und freie Berufe; zur "Oberen Mittelschicht" gehörten alle Kinder, die in einer der vier Dimensionen über der "Mittleren Mittelschicht" lagen. Im Zweifelsfall konnte anhand der im Fragebogen verlangten

Arbeitsplatzbeschreibung über die Schichtzuweisung entschieden werden (Bosse u.a. 1971:61 und Anhang Tab. 2.2.4,1).

Durch die sukzessive Überprüfung aller 4 Dimensionen wurde erreicht, daß Informationen aus allen status-relevanten Lebensbereichen genutzt wurden. Man kann annehmen, daß z.B. eine hohe Ausbildung der Mutter bei geringem Berufs-Status des Vaters mit einer typischen Mittelschichtsozialisation ähnlich zusammenhängt wie ein hohes Familien-Nettoeinkommen bei geringer Ausbildung des Vaters. Die Einteilung in 5 Schichten (KFIL3) wurde für die Verrechnung in Korrelationen innerhalb der Gesamtstichprobe (N=359) z-transformiert (IFILT), ebenso die Rangfolge der Berufe und der Ausbildungsabschlüsse von Vater und Mutter.

Darstellung der Kontrollvariablen. Die Kinder der Unter- und Mittelschicht wurden für die Testvergleiche der zweiten und dritten Messung auf zwei Variablen parallelisiert: den Werten des Frankfurter Wortschatztests und der nichtverbalen Subtests des P-S-B. Die Kontrolle des passiven Wortschatzes folgte direkt aus der theoretischen Absicht des Experiments, das aktive Worterklärungen und Qualität mit Worten verbundener Kategorisierungen unabhängig vom passiven Wortschatz untersuchen sollte. Die Kontrolle der nicht-verbalen Intelligenz wurde aus folgenden Gründen hinzugezogen. Erstens messen die nicht-verbalen Subtests des P-S-B eine allgemeine Gewandtheit bei der Anwendung intellektueller Operationen, die für die Analyse unterschiedlicher und die Wahl optimaler kognitiver Konzepte wahrscheinlich wichtiger ist als verbale Kenntnisse. Die nicht-verbale Intelligenz stellt eher zugrundeliegende Fähigkeiten dar als Lernergebnisse - im Gegensatz zur verbalen. Zweitens gaben als zusätzliche Kontrolle zur Wortschatzleistung die nicht-verbalen Werte mehr neue Information als die mit dem Wortschatz eng zusammenhängende verbale Intelligenz.

Der Frankfurter Wortschatztest ist ein multiple-choice Test. Von den fünf Alternativen gibt nur eine die Bedeutung des Stimuluswortes wieder. Dennoch lassen die anderen verschiedene Interpretationen der Mißverständnisse zu, die zu ihrer Wahl geführt haben.

Bei einem Teil beruhen sie auf Lesefehlern, bei einem anderen auf Verwechslung mit naheliegenden, aber klar unterschiedenen Bedeutungen, bei wieder anderen auf völliger Unkenntnis oder rein phonetischer Assoziation. Bei Kindern mit gleich geringer Kenntnis kann daher eine größere Geschicklichkeit im Raten zu besseren Ergebnissen führen. Zu jedem Wort gibt es fünf Alternativerklärungen, der Zufall allein müßte also zu ca. 20% richtigen Lösungen führen. Die falschen Wahlen mindern den Testwert bei der Auswertung nicht. Die Alternativen sind nach statistischen Ergebnissen in Vortests so zusammengestellt, daß (1) weder von allen noch von keinem Schüler die richtige Antwort gewählt wurde, die Aufgabe also weder zu schwer noch zu leicht ist, (2) eine meist gewählte, d.h. eine eindeutig richtige Antwort zu jeder Aufgabe vorhanden war, (3) der Prozentsatz der richtigen Wahlen mit dem Alter zunimmt, die Worte also zum entwicklungsrelevanten Teil des Wortschatzes gehören. Die benutzte Form "W-S-T 5-7" gilt für die Altersgruppe der 10-13jährigen, ist also für das 4. Schuljahr eher zu schwer. Dies spiegelt sich auch in der linkssteilen Form der Verteilung, die in den oberen Bereichen stärker diskriminiert. Ausgezählt wurden die Rohwerte, sie wurden auf c-Werte transformiert. Die Parallelisierung geschah nach den c-Werten.

Die nicht-verbale Intelligenz wurde mit 6 von insgesamt 10 Subtests des P-S-B von Horn gemessen. Den Kindern wurde der gesamte Test gegeben, aber in diesem Zusammenhang nur der nicht-verbale Teil berücksichtigt. Der P-S-B hat Normen für die Gesamtleistung und für jeden Subtest, nicht aber für Kombinationen von Subtests. Der nicht-verbale Wert wurde deshalb von den durchschnittlichen c-Werten der Subtests gebildet, obwohl korrekterweise aus den addierten Rohwerten der Subtests in der Eichstichprobe c-Werte hätten gebildet werden müssen. Der Arbeitsaufwand stand jedoch in keinem Verhältnis zum Gewinn, der Vermeidung geringer Verzerrungen. Der P-S-B gliedert sich nach Thurstones "primary abilities"; die nicht-verbale Subtests messen "logisches Schließen" (3,4), "Raumvorstellung" (7), "Analytische Fähigkeit" (kognitive Stile) (8),

"Rechenfähigkeit" (9), "Konzentration" (10). Die Parallelisierung geschah ebenfalls nach den c-Werten. - Die restlichen 4 Subtests des P-S-B messen verbale Fähigkeiten: "Wortkenntnis" (1/2), "Worteinfall" (5), "Erkennen verstümmelt geschriebener Wörter" (6).

3. Durchführung der Untersuchung

3.1. Voruntersuchungen und Festlegung der Experimentalgruppen der zweiten Messung.

Lediglich für den Test der aktiven Erklärungen waren einige Voruntersuchungen notwendig. Es mußte entschieden werden, ob die Erklärungen (a) mündlich oder schriftlich, (b) jeweils von Untergruppen oder von der ganzen Experimentalgruppe gegeben werden sollten. Für die Hauptuntersuchung wurde nach einigen Erprobungen festgelegt, daß sowohl mündliche wie schriftliche Erklärungen der Worte verlangt werden sollten, und zwar nicht von verschiedenen Gruppen, sondern von allen Kindern. Da der W-S-T 5-7 in zwei Parallelförmigkeiten erhältlich ist, wurde den Kindern für die aktive Erklärung jeweils die Form gegeben, die sie beim ersten Test nicht bekommen hatten. Da die Abfolge der Worte in der Testform nach Schwierigkeit gestaffelt ist, wurden alle Worte mit ungerader Testnummer zur mündlichen, alle mit gerader zur schriftlichen Erklärung gegeben, oder umgekehrt. Durch die Kombinationen der "geraden" und "ungeraden" Wortgruppen der beiden

Testformen A und B mit den beiden Kommunikationsformen mündlich und schriftlich ergaben sich also 4 Experimentalgruppen:

1. Form A, ungerade Worte, mündlich
Form A, gerade Worte, schriftlich
2. Form A, gerade Worte, mündlich
Form A, ungerade Worte, schriftlich
3. Form B, ungerade Worte, mündlich
Form B, gerade Worte, schriftlich
4. Form B, gerade Worte, mündlich
Form B, ungerade Worte, schriftlich

In jeder der 4 Experimentalgruppen wurden Jungen und Mädchen aus der Unter- und Mittelschicht getestet. Wir haben in der zweiten Messung einen Experimentalplan, in dem alle Kombinationen der Faktoren Schicht, Geschlecht, Experimentalgruppe und Kommunikationsform (mündlich-schriftlich) vorhanden sind, und in dem die Versuchspersonen auf dem Faktor Kommunikationsform wiederholt gemessen werden. Der vollständige Experimentalplan der zweiten Messung ist in Abb. 16 dargestellt; wir werden auf die Probleme des Experimentalplans und der angemessenen Prüfverfahren in den Abschnitten III.5.1. und III.5.3.1. ausführlich eingehen.

3.2. Parallelisierung der Versuchspersonen beider Messungen.

Bei der Parallelisierung der Kinder beider sozialer Schichten mußten drei Entscheidungen getroffen werden. Erstens: Aus welchen Bereichen des Schichtkontinuums sollten die Unterschichtgruppe bzw. die Mittelschichtgruppe ausgewählt werden. Zweitens: Welche weiteren Merkmale der Vpn. sollten bei der Paarbildung konstant gehalten werden. Drittens: Welche Grenzen sollte man für die "Gleichheit" der Vpn. auf den Kontrollvariablen festlegen?

Zur ersten Entscheidung: Wie beschrieben, wurden die Vpn. mit Hilfe eines Informations aus vier Bereichen filternden Verfahrens in fünf Schichten eingeteilt. Von diesen fünf Schichtgruppen wurden für die Parallelisierung die Untere und Obere Unterschicht sowie die Mittlere und Obere Mittelschicht zusammengelegt; die Untere Mittelschicht wurde nicht berücksichtigt. Auf diese Weise wurde die Mittelgruppe ausgelassen, ohne daß die Parallelgruppen künstlich zu Extremgruppen gemacht wurden. Die Kinder, deren Eltern sich geweigert hatten, den Fragebogen auszufüllen, wurden nicht hinzugezogen. Für beide Messungen wurde die Unterschichtgruppe aus 118, die Mittelschichtgruppe aus 140 Kindern zusammengestellt; die nicht berücksichtigte Untere Mittelschicht enthielt lediglich 34 Kinder (siehe für die zweite Messung Tab. 24, für die dritte Messung Tab. 169).

Zur zweiten Entscheidung: Es war notwendig, Paare des gleichen Geschlechts zu bilden, da in fast allen Sprachuntersuchungen sich Geschlechtsdifferenzen, meist zugunsten der Mädchen, ergeben haben (McCarthy 1954). Weiterhin hätten wir die Paare noch aus Kindern der gleichen Schulklasse oder der gleichen Schule bilden können. Dies erwies sich jedoch bei der Größe der Ausgangsstichprobe als unmöglich.

Zur dritten Entscheidung: Da innerhalb der Geschlechter nach zwei Variablen zugleich "gematcht" wurde, konnten nicht immer zwei mal zwei annähernd gleiche Werte gefunden werden. Bei den Wortschatzwerten wurden Differenzen von höchstens 0,2 c-Werten zugelassen; bei den nicht-verbalen Intelligenzwerten wurde eine Obergrenze von 0,4 c-Werten Differenz angestrebt, aber in manchen Fällen auch größere Differenzen zugelassen. Die Mittelwerte und Streuungen der Gruppen auf den Kontroll-Variablen zeigten keine signifikanten Unterschiede (siehe Tab. 16 für die zweite, Tab. 161 für die dritte Messung).

Die parallelisierten Paare hätten korrekterweise nach dem Zufallsprinzip auf die beiden Messungen verteilt werden müssen. Da die Kinder in Frankfurt-Goldstein aber wegen des bevorstehenden Schulwechsels stark durch Schultests belastet waren und der Rektor soweit wie möglich weitere Einschränkungen des Unterrichts vermeiden wollte, wurden die zeitraubenden Einzeltests der zweiten Messung allein in Oberursel, die den Unterricht nur einmal unterbrechenden Gruppentests der dritten Messung mit Schwerpunkt in Goldstein durchgeführt. Für die zweite Messung sind also nur Kinder aus Oberursel genommen worden, während von den 42 in der dritten Messung getesteten Kindern 12, d.h. 29% aus Goldstein stammen. Dieser Prozentanteil liegt etwas über dem Anteil der Kinder aus Frankfurt-Goldstein an der Gesamtstichprobe von 19%.

3.3. Technischer Ablauf beider Messungen.

Beide Experimentalmessungen wurden während der Unterrichtszeit in der Schule durchgeführt. Für die zweite Messung wurden die Worte einzeln auf Zettel notiert und immer in der gleichen, durch die Testform vorgegebenen Reihenfolge von den Kindern erklärt. Da die Kinder das Wort schriftlich vor sich hatten, ist sowohl bei den mündlichen wie den schriftlichen Erklärungen eine Leseleistung vorausgesetzt. Die Abfolge von "gerader" und "ungerader" Wortgruppe einer Testform wurde variiert, immer aber kamen die mündlichen vor den schriftlichen Erklärungen.

Die mündlichen Erklärungen wurden individuell getestet; die Kinder blieben mit dem Tester alleine in einem Raum. Sie sprachen in das an einen Kopfhörer montierte Mikrofon einer Gegensprechanlage. Durch den (als solchen nicht gebrauchten) Kopfhörer waren die Kinder von der Umwelt relativ isoliert. Sie sprachen für sich, konnten sich auf die vor ihnen liegenden Zettel mit den Worten konzentrieren und nach eigener Entscheidung zum nächsten Wort weiter-

blättern. Die Tester gaben lediglich kurz die Testanweisung und zogen sich dann aus dem Blickfeld der Kinder zurück. Sie waren ausdrücklich angewiesen, keine Hilfe zu geben. Die schriftlichen Erklärungen konnten sehr einfach mit mehreren Kindern in einem Raum getestet werden. Da die Kinder die Aufgabe schon kannten, war lediglich ein kurzer Hinweis auf die Änderung und eine geringe Überwachung gegen Abschreiben notwendig. Jedes Kind verließ den Raum für sich, wenn es glaubte, fertig zu sein. Die Kinder wurden also bei mündlichen wie schriftlichen Erklärungen so wenig wie möglich gedrängt.

Die Testanweisung (siehe Anhang A.3) wies die Kinder an, den Sinn des Wortes zu erklären, so als wollten sie das Wort einem, der es nicht kennt, verständlich machen. Sie gaben einige Erklärungsbeispiele, die in der Form weit genug variierten (Funktionsangabe, Beispiel, Umschreibung), um das Kind nicht schon von Anfang auf eine bestimmte Definitionsart festzulegen. Den Kindern wurde keine Zeitbegrenzung gesetzt. Sie brauchten im Durchschnitt etwa eine halbe Stunde für die mündlichen und schriftlichen Erklärungen der insgesamt 60 Worte. Nur ein äußerst verschüchtertes Kind benötigte mehr als eine Stunde.

Zur Auswertung wurden die aktiven Worterklärungen der mündlichen bzw. schriftlichen Messung in der Original-Orthographie bzw. -aussprache transkribiert und für jeweils ein Stichwort, einmal für die mündliche und einmal für die schriftliche Messung, ohne Angabe der Vpn.-Nrn. und in wechselnder Reihenfolge untereinander geschrieben.

Der neukonstruierte Test der dritten Messung wurde Gruppen von etwa 10-20 Kindern unter Aufsicht eines Testleiters und ohne Mitwirkung des Lehrers gegeben. Ihnen wurde erklärt, daß die Alternativen zu dem Neuen Reizwort alle "irgendwie dasselbe" bedeuten, sie jedoch diejenige auswählen sollten, die sie für die beste hielten. Mit

der gesamten Gruppe wurden vier Beispiele durchgegangen. Jede Wahl, die die Kinder dabei trafen, wurde als richtig bestätigt, und der Zusammenhang mit dem Reizwort der gesamten Gruppe erklärt. Dieser Test ist kein Leistungstest. Daher wurde zwar zu Beginn eine Zeitgrenze von 20 Minuten angegeben, die Kinder durften aber die Zeit solange überschreiten, bis sie fertig zu sein glaubten. Wenn ein Kind eine Aufgabe nicht gelöst hat, kann man daher Unkenntnis des gesamten angesprochenen Bedeutungsfelds annehmen.

3.4. Nacherhebung und Änderung der Parallelisierung der zweiten Messung.

Durch ein Mißverständnis unter den Testern ergab sich bei der zweiten Messung ein Fehler: nicht alle "gematchten" Paare erhielten die gleichen Teile der Testform zur mündlichen und schriftlichen Erklärung, so daß die Partner in verschiedene Versuchsgruppen fielen. Um die Trennung zwischen mündlichen und schriftlichen Erklärungen und damit die ursprünglichen Testgruppen beizubehalten, mußten einige Kinder nacherhoben und das "Match-Design" geändert werden. Da jedoch die Auswahl der Worte jetzt geringer war, gingen folgende Möglichkeiten der Parallelisierung verloren: (1) Bei allen Vpn. der zweiten Messung konnte nur noch die Leistung im standardisierten Wortschatz-, nicht mehr im Intelligenztest kontrolliert werden.¹⁾ (2) Mittelwertdifferenzen zwischen den Geschlechtsgruppen konnten nicht mehr ganz ausgeschaltet werden. Sie sind zwar für die Gesamtgruppe mit 0.2, für die gesamte Unterschichtgruppe mit 0.2, und für die gesamte Mittelschichtgruppe mit 0.1 ziemlich klein; doch reichen sie innerhalb der einzelnen Test-

¹⁾-----
Für die dritte Messung bleiben jedoch beide Maße Kontroll-Variable. Man kann annehmen, daß die logische Analyse eher als die aktive Produktion von Erklärungen die Beherrschung allgemeiner intellektueller Operationen wie dem Trennen und Überschneiden von Mengen, der Bildung von Beispielen etc. abhängig ist. Die nicht-verbale Intelligenz blieb daher als Kontroll-Variable in der entscheidenden Messung erhalten (siehe IV.2.).

gruppen bis 1.0 c-Werten. (3) Die c-Wert-Differenz zwischen den Partnern überstieg oft die gesetzte Grenze von 0.2. (4) 9 der 19 nacherhobenen Kinder mußten die gleiche Testform wie im ersten Test erhalten. Indes fand die erste Wortschatzmessung im November 1969, die Nacherhebung im August 1970 statt, so daß über die Zwischenzeit von insgesamt 9 Monaten wohl kaum mehr Erinnerungen vom ersten Test die Leistung beeinflussten. - Trotz allem blieb beim neuen Parallelisieren die Vergleichsmöglichkeit der Schichteinflüsse gesichert. Die schichtspezifischen Mittelwertdifferenzen der Testgruppen auf der Kontrollvariable Wortschatz lagen bei höchstens 0.13 c-Werten, der Gesamtgruppe bei 0.02 c-Werten (siehe Tab. 16 und 17). Da die Haupterhebung unmittelbar gegen Ende des Schuljahres stattgefunden hatte, die Kinder aber im nächsten Schuljahr auf die verschiedenen weiterführenden Schulen verstreut waren, mußte die Nacherhebung im Elternhaus stattfinden, zu Beginn und Ende der Sommerferien. Wir haben für alle gebildeten Maße die Signifikanz des Einflusses der Nacherhebung überprüft (siehe III.2.2.).

Der standardisierte Wortschatz- und der Intelligenztest wurden im November 1969 durchgeführt, die zweite und dritte Wortschatzmessung im Juli 1970, die Nacherhebung der zweiten Messung im August 1970. Die Bewertung der aktiven Worterklärungen sowie die Auswertung des neugebildeten passiven Tests (dritte Messung) wurde im Winter 1970/71 durchgeführt; die Auswertung der sprachlichen Mittel der Worterklärungen im Winter 1971/72.

Drittes Kapitel: Darstellung der Ergebnisse der ersten und
zweiten Messung.

1. Operationalisierung der Prüfvariablen der aktiven Bedeutungserklärung.

1.1. Konstruktion von Kategorien der Bedeutungserklärung.

1.1.1. Vortests.

In einem Vortest (Dezember 1969) haben wir jeweils 30 Worte der "geraden" und "ungeraden" Hälfte jeder Parallelfarm des Frankfurter Wortschatztests von jeweils 10 (insgesamt 40) Kindern schriftlich erklären lassen, so daß für jedes der insgesamt 120 Stichworte bis zu 10 Erklärungen vorlagen. Insgesamt vier Auswerter versuchten für diese Erklärungen qualitative Beschreibungen zu finden, aus denen sich, verallgemeinernd, Kategorien für die spätere Auswertung entwickeln lassen sollten. Sie sollten bei der Auswertung der aktiven Erklärungen der zweiten Messung und bei der Konstruktion des Tests der dritten Messung verwandt werden.

Aus diesem Grunde sollten die Kategorien nicht nur die Ränge einer Bewertungs-Skala umschreiben, sondern gleichsam die Methodik der Erklärung unabhängig von der bestimmten erfragten Bedeutung qualitativ möglichst genau erfassen. Bewertung und qualitative Kategorisierung sollten nicht übernommen, sondern durch eine induktive Inventarisierung des vorliegenden Materials gewonnen werden.

Das Ergebnis dieser Voruntersuchungen war das "komplexe Kategoriensystem"; es bestand aus 26 beschreibenden Kategorien, die in 8 bewertende Güteklassen und 2 nicht-bewertende "Methodiken" zusammengefaßt waren (siehe Anhang B). Mit der mehrdimensionalen Zusam-

menfassung einer Vielzahl von Kategorien war angestrebt, die Zuweisung einer einzelnen Kategorie in eine Folge von Einzelentscheidungen aufzulösen und zu einem intersubjektiv kontrollierbaren, statt individuell intuitiven Prozeß zu machen. Aus drei zusammenhängenden Gründen haben wir das "komplexe Kategoriensystem" schließlich nicht verwandt und den Versuch, mit ihm den Test der dritten Messung zu konstruieren, aufgegeben.

Erstens zeigte sich eine starke Stimulusabhängigkeit der Erklärungen; auf denselben Stimulus fanden wir ähnliche, auf verschiedene Stimuli jedoch sehr unterschiedliche Erklärungen, so daß eine durchgängige Kategorisierung der Erklärungen erschwert wurde. Deshalb ließ sich - zweitens - die Verbindung von qualitativer Beschreibung und Bewertung, die im Kategoriensystem a priori festgelegt war, bei der Auswertung nicht durchhalten; Erklärungen verschiedener Stimulusworte, die in die gleiche Kategorie fielen, schienen dennoch nicht gleich gut zu sein. Drittens war deshalb die Reliabilität der 7 Bewertungs-Niveaus gering; die Reliabilität der 26 beschreibenden Kategorien ließ sich überhaupt nicht einschätzen. Nachträglich gesehen, hätte das Problem mangelnder Reliabilität durch eine längere Einübung der Bewerter gelöst werden können; die Stimulusabhängigkeit der Erklärungsarten spricht nach wie vor gegen den Versuch, die Methodik der Erklärung zu kategorisieren, und für eine weniger explizite Skalierung, die den Bewertern Freiheit läßt, jede Kategorie an die zu einem Stimulus vorliegenden Erklärungen anzupassen.

Als zweites versuchten wir deshalb das Bewertungssystem von Carson/Rabin (1960), das aus 7 sehr weit gefaßten Kategorien besteht, auf unser Material anzuwenden. Carson/Rabin bewerteten die Worterklärungen von 10jährigen Kindern auf folgender Skala:

- 1 Kategorisierung und Synonym
- 2 Wesentliche Beschreibung
- 3 Wesentliche Funktion
- 4 Beispiele
- 5 Vage Beschreibungen oder vage Funktion
- 6 Fehler
- 7 Keine Antwort

Drei Bewerter versuchten, diese Kategorien auf einen Teil unseres Materials anzuwenden. Dabei erwiesen sie sich jedoch als nicht ausreichend und zu ungenau bestimmt.

Nach verschiedenen Versuchen der Erweiterung verwandten wir schließlich ein Kategoriensystem, das - wie bei Carson/Rabin - eine eindimensionale Bewertungsskala darstellt, für jedes Bewertungsniveau aber die gleich bewerteten Arten der Erklärung aufzählt und genau definiert.

Wir haben also versucht, die Vorteile der Skala von Carson/Rabin, leichte Anwendbarkeit, mit der Ausführlichkeit unseres "komplexen Kategoriensystems" zu verbinden. Außerdem haben wir versucht, Streitigkeiten, die bei der Bewertung der Erklärungen auftauchten und einen systematischen Grund hatten, durch eine zweifache Bewertung jeder Antwort nach "Bedeutungsverständnis" und "Bedeutungsexplikation" zu lösen.

1.1.2. Trennung von Bedeutungsverständnis und Bedeutungsexplikation.

Bei der Diskussion der Bewertungen ergaben sich häufig Differenzen zwischen den Bewertern, die sich weder aus einer unterschiedlichen Auffassung der Kategorien noch einem unterschiedlichen Bedeutungsverständnis des Ausgangswortes oder der Erklärung ergaben und in immer gleicher Weise geklärt werden konnten. Während ein Bewerter eine Erklärung gut bewertete, weil alles wichtige aus ihr hervorging, kritisierte ein anderer, daß die kritischen Bedeutungselemente nicht expliziert seien. Die Differenz lag im Bezugspunkt der Bewertung: der erste Bewerter bezog sich auf das Verständnis des Erklärenden, das von jemandem, der die Bedeutung schon kennt, erschlossen werden kann; der andere Bewerter auf den Wert der Erklärung für einen Fragesteller, der das Wort noch nicht kennt. Wenn man für einen Augenblick die strukturalistische Semantiktheorie (Katz/Fodor 1963) als Modell nimmt, so schreibt der erste Bewerter die in den semantischen Redundanzregeln ausgedrückten Inklusionen der Erklärung gut, während der zweite die allgemeinen, aber redundanten Kategorien expliziert sehen will. Beide Bewerter können

Stimulus, Erklärung und Bewertungskategorien gleich verstanden haben und trotzdem zu einer unterschiedlichen Bewertung gelangen. Eine Entscheidung zwischen den unterschiedlichen Bezugspunkten der Bewertung ist innerhalb der Experimentalsituation nicht möglich, sie hängt von der Qualität der Alltagssituationen ab, in denen Sprache gebracht wird.

Die Trennung beider Dimensionen erlaubt zudem eine Wiederholung des Tests der Hypothesen. Wenn wir - für das ganze Experiment - angenommen hatten, daß die Unterschichtkinder trotz gleich gutem Wortverständnis (d.h. trotz gleicher Leistung in der ersten Messung) den Mittelschichtkindern in der aktiven Erklärung unterlegen seien (d.h. in der zweiten Messung), so könnten wir jetzt - innerhalb der zweiten Messung - annehmen, daß Kinder beider Schichten gleich gutes Wortverständnis, aber die Mittelschichtkinder eine bessere Wortexplikation zeigen; oder daß - anders gesagt - die Differenzen zwischen Verständnis und Explikation bei den Unterschichtkindern größer sind. In der Mittelschicht - so läßt sich nach der Theorie des linguistischen Kodes vermuten - werden Werte der individuellen Selbstdarstellung betont; die Kinder lernen eine allgemeine Einstellung, fremde Meinungen und Absichten beim Gesprächspartner zu vermuten, und explizieren ihr Verständnis vollständig. In der Unterschicht - so wird weiter angenommen - verbietet die Betonung des Werts der Solidarität eine Erziehung zu individueller Selbstdarstellung; die Kinder lernen eine allgemeine Einstellung, auf das Wissen und Verständnis des Gesprächspartners ungefragt zu vertrauen, und lassen eine Kluft zwischen Verstandenem und Explikation.

1.1.3. Auswertungsskalen für Bedeutungsverständnis und Bedeutungsexplikation.

Als Richtlinie für die Auswertung nach dem Bedeutungsverständnis nahmen wir die Rückerschließbarkeit des Stimulusworts aus der Erklärung (Wieviel Verständnis des Stimulus geht aus der gegebenen Erklärung hervor?), als Richtlinie für die Auswertung nach der

Bedeutungsexplikation die Rückübersetzbarkeit der Erklärung in das Stimuluswort (Wieweit ist das ursprüngliche Stimuluswort seinerseits eine angemessene Erklärung der gegebenen Erklärung?). Die Trennung von Verständnis und Explikation soll nun an einigen Beispielen erläutert werden; weitere Bewertungsbeispiele mit allen Erklärungen für einzelne Stichworte finden sich in Anhang C.

Sehr oft wird eine unterschiedliche Bewertung nach Verständnis und Explikation bei prägnanten Beispielen als Erklärungen notwendig: "drollig" = "wenn kleine Hunde sich auf die Hinterpfoten stellen"; "brutal" = "wenn ein großer Junge auf einem Pony reitet und es immerzu tritt und mit der Peitsche schlägt". Derartige Erklärungen können nicht ohne ein präzises Verständnis der Bedeutung gegeben worden sein und erhielten die beste Verständnisbewertung; aus ihnen lassen sich jedoch keine Kategorien entnehmen, die die Bedeutung des Stimulusworts explizieren, deshalb wurden sie unter Explikation nur als "Beispiel" (6) bewertet. Ein gutes Verständnis ist aus der Erklärung erschließbar, aber sie kann nicht mehr in das Stimuluswort übersetzt werden. Ohne die Trennung beider Bezugspunkte ist der Streit, ob diese Erklärungen gut (voll verstanden) oder schlecht (nur Beispiel) bewertet werden sollen, nicht zu schlichten. Ein zweiter typischer Fall, der die Trennung notwendig macht, ist eine allgemeine Kategorie als Erklärung, die treffend genug ist, um nur aus genauer Wortkenntnis genannt worden zu sein, deren spezifische Differenz jedoch nicht expliziert ist: "Armbrust" = "eine Waffe aus dem Mittelalter"; "Puma" = "ein wildes Tier"; "Gier" = "wenn jemand Lust nach etwas bekommt". Alle Erklärungen lassen ein genaues Verständnis erschließen, explizieren aber würden sie es erst, wenn sie etwa "ein in der Ritterzeit verwendetes Gewehr", "eine Raubkatze" und "maßloses Verlangen" lauteten. Ein dritter Fall ist eine treffende Umschreibung ohne einen Kategorisierungsversuch: "Spesen" = "das bekommt man, wenn man eine Sitzung gehalten hat"; "heimzahlen" = "wenn einer einem was getan hat und der will es ihm wiedergeben". Diese Erklärungen verraten ein genaues Verständnis des Sachverhaltes, können aber nicht mehr in das ursprüngliche Stimuluskonzept zurückübersetzt werden, wie etwa "Aufwandsentschädigung" oder "sich rächen".

An den genannten Beispielen lassen sich auch die Probleme der Auswertung auf die beiden Skalen ermessen. Sie sollen zusammen mit den gewählten Lösungen aufgezählt werden. Zunächst erwies es sich als schwierig, die analytische Trennung zwischen Verständnis und Explikation bei der Auswertung durchgängig zu befolgen. Denn die sprachlich nicht realisierte Leistung "Verständnis" kann ja nur aus den expliziten sprachlichen Äußerungen erschlossen werden. Zweitens erwies es sich als schwierig, für die Auswerter verbindlich zu regeln, wann der Prozeß der Bedeutungserschließung beendet werden kann, d.h. die Verständnisbewertung zu skalieren. Denn jede Differenzierung von Stufen des Verständnisses lief tendenziell auf eine Entweder-Oder-Entscheidung "Verstanden - Nicht verstanden" hinaus. Für diese beiden Schwierigkeiten haben wir folgende Lösung gewählt. Da wir die Differenzen zwischen Verständnis und Explikation messen wollten, haben wir für "Verständnis" eine Skala gleicher Länge wie für "Explikation" beibehalten. Zur Erleichterung der Auswertung und aufgrund der sprachlogisch sinnvollen Annahme, daß die Explikation nie besser als das Verständnis sein kann, haben wir dann die Regel festgelegt, eine Erklärung auf der Verständnisskala immer mit einem gleich guten oder besseren Punkt als auf der Explikationsskala zu bewerten. Anders gesagt: Wir haben eine Kommensurabilität beider Skalen künstlich hergestellt, um die Differenz zwischen ihnen messen zu können. Die Gleichwertigkeit beider Skalen ist jedoch nur äußerlich, tatsächlich ist die Verständnisskala nur ein Anhang der Explikationsskala.

Eine dritte Schwierigkeit ist die Beurteilung der Antwortverweigerungen. Einerseits verraten sie eine größere verbale Selbstkontrolle der Vp., andererseits sollte eine falsche Erklärung besser als eine verweigerte bewertet werden. Wir haben

schließlich falsche und verweigernde Antworten auf beiden Skalen mit dem schlechtesten Skalenpunkt bewertet, für die Bildung von Maßen aber Antwortverweigerungen und falsche Antworten getrennt auf Datenkarten gebracht (siehe auch 1.2. und 5.6.1.).

Eine vierte Schwierigkeit bei der Auswertung war die Behandlung phonetischer Mißverständnisse. Wir wählten folgende Lösung: Wenn alle Kinder das Stimuluswort gleich mißverstanden, wurde dies Mißverständnis als das zu bewertende Stimuluswort angesehen; alle Kinder lasen z.B. "Gabe" statt "Garbe". Wenn einzelne Kinder etwas falsch verstanden, wurde, falls das Mißverständnis ein schwierigeres Wort als das Ausgangswort ergab, die Erklärung des schwierigeren Worts bewertet, falls das Mißverständnis ein leichteres Wort ergab, die Antwort als falsch angesehen; manche Kinder erschwerten sich die Aufgabe, z.B. indem sie "Verlag" statt "Vertrag" oder "Einwand" statt "Einband" lasen.

Die beiden Bewertungsskalen sind, so wie sie den Auswertern zur Hand waren, im folgenden dargestellt. Die Skala "Explikation" besteht aus ausführlichen Beschreibungen der logischen Form von Erklärungen, die z.T. aus dem "komplexen Kategorienschema" übernommen und zusammengefaßt wurden. Diese Skala versucht, die Detailliertheit des "komplexen Kategorienschemas" mit der Flexibilität des Bewertungsschemas von Carson/Rabin zusammenzubringen. Die Skala "Verständnis" konnte natürlich nicht mehr auf klaren Definitionen identifizierbarer sprachlicher oder logischer Formen beruhen, es besteht daher aus Umschreibungen von Gütestufen des Verständnisses und aus Kategorien, die bestimmte, häufig auftauchende und relativ genau rekonstruierbare Formen des Verständnisses definieren und aus dem "komplexen Kategorienschema" übernommen sind ("pragmatisch-singuläres Synonym", "Umdeutung durch falsches Verständnis von Prä- oder Suffixen"). Beide Skalen sind Rangskalen von Güteklassen. Dennoch haben wir sie - bei der Auswertung wie bei der Bildung von Maßen - als Intervallskalen behandelt. Bei der Auswertung haben wir versucht, die Abstände zwischen den Punkten intuitiv möglichst gleich zu halten, ohne daß dies sich belegen läßt. Am nächsten kommen die Skalen einer schulischen Benotung, die intersubjektiv kontrolliert ist.

Kategorien für die Analyse der Bedeutungsexplikation der Antworten im aktiven Wortschatztest.

(Kriterium: Rückübersetzbarkeit der Antwort in das Reizwort)

1. Synonym oder zutreffende Kategorisierung

- a) Synonym im Sinne der umgangssprachlichen Bedeutungs-
gleichheit
- b) Kategorisierung - Angabe eines Oberbegriffs und der
spezifischen Differenz, so daß der Bedeutungsumfang des
Reizwortes genau umschrieben ist
- c) "komplexe" Synonyme - z.B. erschöpfende Aufzählungen
(Lebewesen: Tier, Mensch oder Geschöpf / Tier, Mensch)
oder präzise Synonymisierungen durch einen Wortkomplex

2. Sehr prägnante Umschreibung der Bedeutung des Reizwortes,
auch wenn keine Kategorisierung angestrebt worden ist. Z.B.
auch sehr prägnante exemplarische Umschreibungen

3. Beschreibungen der wesentlichen Bedeutung

- a) durch Umschreibung (typisch: "wenn man ...")
- b) durch relativ nahen Oberbegriff ohne prägnante Spezifi-
zierung (z.B. Komponist, Musiker)
- c) nicht ganz zutreffendes, aber doch nahes Synonym (z.B.
vermählen: heiraten / häufig, wenn nicht dieselbe
Wortklasse verwendet wurde)

4. Beschreibung nicht der wesentlichen Bedeutung, aber der we-
sentlichen Funktion, Bedingung, Begleitumstände, Ursache usf.
des mit dem Reizwort gemeinten Sachverhalts oder Referenten.
Es muß sich um eine klare Funktionsbeziehung zwischen Reiz-
wort und Antwort handeln.

5. Partielle oder zu allgemeine Explikation, Beschreibung oder
Kategorisierung

- a) partielle, d.h. nicht die zentrale Bedeutung eines Reiz-
wortes treffende Antwort (Beispiele: zärtlich: vorsichtig/
Export: Bierart). Entweder handelt es sich um eine Neben-
bedeutung oder die Antwort liegt im Fokus eines verwandten,
aber perspektivisch und kategorisch anderen Bedeutungsfeldes
- b) zu weiter Oberbegriff (Bsp: Puma: Tier / Komponist: Beruf)

Bis hierher handelt es sich bei den Antworten um Versuche, die Bedeutung des Reizwortes kategorial-allgemein zu explizieren.

6. Richtige beispielhafte Verwendung des Reizwortes und Umschreibung eines konkreten Beispiels (Sachverhalt/Situation) für die Verwendungsmöglichkeit des Reizwortes

- a) Das Reizwort wird in einem Phrasen-Rahmen verwendet, der Aufschluß über die richtige Bedeutung des Reizwortes gibt (Bsp.: "Der Fall ist logisch")
- b) Partikulares Mitglied der Menge von Referenten, auf die sich das Reizwort als Klassenbegriff bezieht (Bsp.: Blume : Tulpe)
- c) Beschreibung einer konkreten Situation, auf die die Bedeutung des Reizwortes ungefähr zutrifft

7. Beschreibung einer peripheren Bedeutungskomponente oder -funktion des Reizwortes

- a) periphere Funktion des mit dem Reizwort Gemeinten
- b) peripheres Beispiel für die Anwendung des Reizwortes
- c) Synonym oder Kategorie für ein peripheres Bedeutungsmerkmal des Reizwortes (Bsp.: röcheln : schlapp)

(Peripher heißt hier, daß die Beschreibung oder das Beispiel zentral in einem mit der Bedeutung des Reizwortes kaum zusammenhängenden Bedeutungsfeld liegt, aber mit unwesentlichen Merkmalen der Bedeutung des Reizwortes plausibel in Verbindung gebracht werden kann. Eine Rückübersetzbarkeit der Antwort in das Reizwort ist im Unterschied zu 5 in keinem Fall möglich)

8. Vage Assoziation

- a) Eine lockere, aber verschlungene und kategorial-allgemein nicht ausgewiesene Assoziationskette läßt sich zwischen Reizwort und Antwort herstellen.
- b) Beispielhafte Verwendung des Reizwortes in einem Phrasen-Rahmen, der keinen indirekten Aufschluß über die Bedeutung des Reizwortes gibt - ebenso: Informationsloses Wechseln der Wortklasse des Reizwortes

9. Eindeutig falsche Antwort oder keine Antwort

Kategorien für die Analyse des Bedeutungsverständnisses nach den Antworten im aktiven Wortschatztest

1. Verständnis der prägnanten Bedeutung des Reizwortes.

Die Antwort muß erkennen lassen, daß nur ein klares Verständnis der mit der Reizwortbedeutung gegebenen allgemeinen Kategorisierung von Sachverhalten zu ihr führen konnte.

2. Verständnis der wesentlichen Bedeutung des Reizwortes.

Wesentliche Aspekte des Bedeutungsfeldes, das mit dem Reizwort gegeben ist, sind verstanden worden, ein genaues Verständnis der engen kategorialen Bedeutung ist jedoch nicht voll zu erkennen.

3. Partielle oder pragmatisch singuläre Bedeutung des Reizwortes ist verstanden worden (also Bedeutungen, die sich fast schon zu einem Homonym des Wortes verselbständigt haben).

Verständnis nur eines partiellen Bedeutungsmerkmals

Bis hierher ist die Bedeutung des Reizwortes kategorial erkennbar richtig lokalisiert worden.

4. Hierarchisierungsfehler im Verständnis: Die Bedeutung des Reizwortes ist im Schema der entsprechenden Begriffshierarchie nicht richtig lokalisiert worden: sie wird entweder zu eng oder zu weit verstanden.

5. Verwechslung mit einem anderen Mitglied (lexikalisch) derselben Oberkategorie: Die richtige und die verstandene Bedeutung lassen sich unter denselben Oberbegriff subsumieren (Bsp.: Komponist: Dirigent)

Die Kategorien 4 und 5 beziehen sich auf ein Verständnis, das kategorial nicht ganz richtig ist, sich jedoch mühelos vom Beurteiler als Kategorienfehler rekonstruieren läßt.

6. Umdeutung der Wortbedeutung durch falsches Verständnis von Präfixen oder Suffixen (ersinnen: sich erinnern)

7. Konkretisierendes Mißverständnis:

Interpretation eines Bedeutungspartikels ohne Verständnis des Gesamtsinns oder konkretistisches falsches Verständnis durch einfache Kombination der Einzelbedeutung der Wortlemente.

8. Vage Assoziation

9. Wiederholung, Tautologie oder keine Antwort.

Beispiele zu den verschiedenen Kategorien der Bedeutungsexplikation

1. a) drollig = niedlich, vergeuden = verschwenden
b) Haustier = Tier, das bei den Menschen lebt und zahm ist
2. vergeuden = wenn man viel Geld für unnützes Zeug ausgibt
3. a) vergeuden = wenn man viel Geld hat und es schnell ausgibt
b) drollig = süß, Gerümpel = Sachen auf dem Boden
4. a) Asthma = wenn man nicht genug Luft bekommt
b) Gerümpel = wenn etwas ganz durcheinander ist
5. a) vergeuden = trödeln, drollig = wenn jemand klein und dick ist
b) Gerümpel = altes Eisen
6. a) Schätzung = man schätzt wie groß das Haus ist
b) Gerümpel = Haufen Papier
7. a) Traktor = man fährt mit
b) röcheln = wenn man so ein Geräusch macht
8. a) heimzahlen = bereuen, verzweifelt = beleidigt
b) Schätzung = wenn man schätzt

Beispiele zu den verschiedenen Kategorien des Bedeutungsverständnisses (Kategorien der Bedeutungsexplikation in Klammern)

1. Traktor = mit dem Traktor fährt der Bauer auf das Feld (6)
= Wagen der auf den Feldern arbeitet, damit es schneller geht (3)
2. Traktor = Fahrzeug (5)
= Maschine der Bauern (5)
Ampulle = was der Arzt gut gebrauchen kann (7)
3. Aufschneider = wenn man einen reinlegen will (5)
logisch = das ist doch logisch (6)
4. Kongress = Abstimmung (5)
= wo Leute zusammen sprechen (7)
5. Periode = wenn man in die Wechseljahre kommt (7)
liederlich = wenn man in der Schule schlecht ist (8)
verwenden = behalten (7)

6. vergeuden = wenn man einem etwas gibt und nach einer
Weile möchte man es wiederhaben (8)
unbeholffen = man kriegt nicht geholffen (8)
7. Aufschneider = Schneider der was zuschneidet (9)
Wolkenkratzer ≠ Kratzer am Himmel (9)
8. Quittung ≠ wenn man was ausrechnet (8)
Mailand = ist ein Land (8)

1.1.4. Reliabilität der Kategorisierung.

Mündliche wie schriftliche Erklärungen der Kinder aus der Haupterhebung wurden in der Original-Orthographie bzw. -Aussprache für jedes Wort auf ein Blatt mit der Maschine geschrieben. Die Abfolge der Erklärungen der verschiedenen Kinder wechselte dabei von Stichwort zu Stichwort, Vp.-Nrn. waren nicht angegeben. Es bereitete keinerlei Schwierigkeiten, mündliche und schriftliche Erklärungen mit den gleichen Skalen auszuwerten; im Schriftlichen tauchten gelegentlich Äußerungen eines Kindes auf, die zwei Erklärungen zum gleichen Stimuluswort enthielten; in diesem Fall berücksichtigten wir die bessere der beiden Erklärungen. Im Mündlichen wie Schriftlichen zielte die Bewertung auf die Güte der Erklärungen und des Verständnisses unabhängig von der sprachlich-grammatischen Form wie von konventionellen Maßstäben: Rechtschreibung, Dialekt, Stil. Wir haben versucht, diese Merkmale gesondert auszuwerten (siehe 6.).

Zwei Bewerter kategorisierten alle vorliegenden Erklärungen, ein dritter etwa ein Drittel der vorliegenden Erklärungen. Alle drei Bewerter verglichen alle Kategorisierungen und legten alle endgültigen Bewertungen fest. Da Unstimmigkeiten zwischen den Bewertern sich nicht nur aus dem Kategoriensystem, sondern auch aus einem unterschiedlichen Verständnis der Stimulusworte ergeben konnte, wurden zuvor von allen drei Bewertern verbindliche Definitionen der 120 Stimulusworte festgelegt (siehe Anhang A.2). Die Reliabilität der Bewertungsskalen können wir auf zwei Weisen überprüfen: Durch die Korrelation der Bewertungen zwischen den Bewertern und durch den Vergleich der Häufigkeiten der Werte der endgültigen Bewertungen zwischen der mündlichen und schriftlichen Messung.

Die Inter-Rater-Reliabilitäten sind für die Wörter aller Testformen innerhalb der Verständnis- und Explikationsbewertung der mündlichen und schriftlichen Messung in Tab. 1 dargestellt; sie sind nicht zufriedenstellend. Die drei Bewerter haben jedoch in langen

Diskussionen, in denen jeder sein Urteil für die anderen einsichtig begründen mußte, das unter Berücksichtigung aller aufgetauchten Gesichtspunkte beste Urteil zu fällen versucht. Unzuverlässigkeiten der Beurteilung konnten sich ja nicht nur aus der Vagheit der Kategorien, sondern auch aus Unterschieden des Verständnisses der Erklärungen zwischen den Bewertern ergeben haben. Unter diesem Gesichtspunkt ist die geringe Reliabilität nicht mehr so überraschend, die gemeinsame Beurteilung erscheint um so notwendiger. Offensichtlich sind mehrere Sprecher notwendig, schon um die Eindeutigkeit des zu beurteilenden Objekts, einer sprachlichen Erklärung, herzustellen. Deshalb ist die Übereinstimmung zwischen den ursprünglichen isolierten Beurteilungen der drei Bewerter für die Einschätzung der beiden Bewertungsskalen weniger wichtig, als die Konsistenz der endgültigen Bewertungen.

Der Vergleich der prozentualen Häufigkeiten der endgültigen Werte der beiden Skalen zwischen verschiedenen Untergruppen, die keinen Experimentaleffekt darstellen, kann nach der Logik der "split-half-reliability" einen Eindruck von der Reliabilität der endgültigen Bewertung geben. Als Dimensionen des Experimentalaufbaus, die keinen ursprünglich zu prüfenden Effekt darstellen, bieten sich die Kommunikationsformen und die Experimentalgruppen an: Wir können die mündliche und schriftliche Messung innerhalb der Gesamt-, der beiden Schicht- und der beiden Geschlechtsgruppen vergleichen und wir können die vier Experimentalgruppen untereinander vergleichen. Dabei werden wir die prozentualen Häufigkeiten der 9 Explikations- und der 9 Verständniskategorien, sowie der Antwortverweigerungen, bezogen auf alle möglichen Antworten darstellen (Abb. 1-5); außerdem werden wir, nur für die Explikationsskala, anhand von Rangkorrelationen der Häufigkeiten der 9 Kategorien die Ähnlichkeit der Profile in den genannten Gruppen einschätzen (Tab. 2). Die Ausgangsdaten für Profile und Rangkorrelationen sind in den Tab. 3, 4 und 5 wiedergegeben.

Der Vergleich der mündlichen und schriftlichen Profile innerhalb der Gesamtgruppe (Abb. 1) ergibt für die Verständnisbewertung eine nahezu vollständige Übereinstimmung - aber auch eine starke Tendenz zur U-Verteilung, die auf eine zugrundeliegende Entweder-Oder-Entscheidung verweist -; für die Explikationsbewertung eine sehr ausgeprägte Kurve, deren Verlauf in der mündlichen wie schriftlichen Messung exakt gleich ist. Die Rangfolgen der prozentualen Häufigkeiten der 9 Explikationskategorien korrelieren .77. Betrachtet man die beiden Schichtgruppen getrennt (Abb. 2 und 3), so ergibt sich genau das gleiche Bild wie in der Gesamtgruppe für beide Schichtgruppen. Die Rangfolgen der prozentualen Häufigkeiten der 9 Explikationskategorien korrelieren zwischen mündlicher und schriftlicher Messung in jeder Schichtgruppe höher als in der Gesamtgruppe (Tab. 2). Betrachtet man die beiden Geschlechtsgruppen getrennt (keine gesonderte Abb.), so ergibt sich das genau gleiche Bild in jeder Geschlechtsgruppe. Die Rangfolgen der prozentualen Häufigkeiten der 9 Explikationskategorien korrelieren zwischen mündlicher und schriftlicher Messung in beiden Geschlechtsgruppen höher als in der Gesamtgruppe (Tab. 2), jedoch eher niedriger als in den Schichtgruppen. In allen genannten Gruppen finden wir also eine starke Ähnlichkeit der Profile der Verständnis- wie der Explikationsbewertung in der mündlichen und schriftlichen Messung.

Bis jetzt haben wir die Profile der Antwortbewertungen der mündlichen und schriftlichen Messung so verglichen, so daß die verglichenen Vpn. in beiden Messungen die gleichen Worte erklärt haben. Im folgenden sollen die Profile innerhalb der Wort- und innerhalb der Experimentalgruppen zwischen der mündlichen und schriftlichen Messung verglichen werden; dabei werden entweder die gleichen Worte zwischen verschiedenen Vpn. oder die gleichen Vpn. zwischen verschiedenen Worten verglichen. Damit bietet sich die Möglichkeit, die Reliabilität der Explikationsskala, auf die wir uns im folgenden beschränken, schärfer als bisher zu prüfen. Die Explikationsskala muß stabil mit den zu erklärenden Worten verbunden sein, um die unterschiedlichen Leistungen der Vpn. zur

Erklärung messen zu können. Dann aber müssen die Profile verschiedener Vpn. mit den gleichen Worten einander stärker gleichen, als die Profile gleicher Vpn. mit verschiedenen Worten. Der Vergleich von (2) und (3) in Tab. 2 bestätigt genau diese Vermutung.¹⁾

Dieselben Vergleiche lassen sich auch an Abb. 4 verfolgen, die die weitgehende Ähnlichkeit der mündlichen und schriftlichen Explikationsprofile innerhalb der vier Wortgruppen zeigt. Weicht das Profil einer Wortgruppe von dem Explikationsprofil für die Gesamtgruppe (Abb. 1) ab - wie für die Kategorie 1 in "A ungerade", 6 in "A gerade" oder 5 in "B ungerade" -, so gelten die Abweichungen für die mündliche und schriftliche Messung gleichermaßen. Noch besser läßt sich die Abhängigkeit der Profile von den Worten aus dem Vergleich der nach der Testform zusammengefaßten Experimentalgruppen ansehen (Tab. 2, Korrelationen mit "b" in (5), Abb. 5). Die Profile der Experimentalgruppen mit der gleichen Testform ähneln einander stark, die Profile der Experimentalgruppen mit verschiedener Testform unterscheiden sich deutlich. Offensichtlich konnten die Kategorien bei den Worten der Testform B besser differenziert werden, während bei den Worten der Testform A die Kategorien 3 ("Wesentliche Bedeutung") und 6 ("Beispiel") besonders häufig gewählt wurden. Die Übereinstimmung der Verteilung innerhalb der beiden Experimentalgruppen mit den gleichen Testformen zeigt jedoch, daß die genannten Unterschiede nicht auf die Auswertung, sondern auf die Testform zurückgeführt werden müssen. Wir finden hier ein Indiz für

1)-----
Die geringeren Rangkorrelationen bei den Wortgruppen "A ungerade" und "A gerade" wie bei den entsprechenden Experimentalgruppen 1 und 2 ergeben sich aus zweierlei: 1. Die besonders schlechte Position der Experimentalgruppe 1 auf der Kontrollvariablen und ihrer besonders schlechten Leistung in der 2. Messung, 2. Die im Vergleich zur Testform B größere Schwierigkeit der Testform A, besonders des Teils "A gerade" (siehe Tab. 63, Abschnitt 5.3.5.). Durch die großen Unterschiede zwischen den Experimentalgruppen 1 und 2 auf der Kontrollvariablen ergeben sich die geringeren Profilähnlichkeiten bei den Wortgruppen A im Vergleich zu den Wortgruppen B, deren entsprechende Experimentalgruppen 3 und 4 auf der Kontrollvariablen gleichwertige Leistungen zeigen (Tab. 2 (2)). Durch die größere Schwierigkeit der Wortgruppe "A gerade" können die geringeren Profilähnlichkeiten in den Experimentalgruppen 1 und 2, neben den Unterschieden auf der Kontrollvariablen, bedingt sein (Tab. 2 (3)). Die beschriebenen Zusammenhänge sind natürlich keine kausalen, sie ergeben sich aus unterschiedlichen Darstellungen des gleichen Sachverhalts.

Unterschiede zwischen den beiden Testformen des Frankfurter Wortschatztests, das durch die Ergebnisse der ersten Messung (siehe 4.2.) und der zweiten Messung (5.3.5.; 5.4.3.) unterstützt werden wird.

Noch weiter kann man die Bedingungen der Reliabilitätsprüfung verschärfen, wenn man die Wortgruppen und die Experimentalgruppen innerhalb der mündlichen und schriftlichen bzw. innerhalb der gesamten Messung vergleicht; dabei werden Worte und Vpn.variiert (Tab. 2, (4) und die Korrelationen ohne "b" in (5)). Die Profile der Explikationskategorien zwischen verschiedenen Wortgruppen und verschiedenen Vpn. (4) sind überraschenderweise stabiler als zwischen verschiedenen Wortgruppen und gleichen Vpn.(3). Ein Grund dafür ist sicherlich, daß in (4) die Kommunikationsform konstant gehalten wurde, während die Vpn. und die Worte variierten. Dennoch erreichen die Korrelationen nicht die Höhe von (2), wo die Stimulusworte konstant gehalten wurden, während die Kommunikationsform und die Vpn.variierten; und zeigen die Vergleiche in (5), wo die Kommunikationsform nicht mehr berücksichtigt wurde¹⁾ und sowohl die Vpn. wie die Worte variieren, ähnlich hohe Korrelationen. Die Unterschiede der Höhe der Korrelationen zwischen (4) und (5) einerseits und (3) andererseits belegen daher, daß die Explikationsprofile verschiedener Wortgruppen nicht mehr, sondern weniger Ähnlichkeit erhalten, wenn man sie zwischen den gleichen, statt zwischen verschiedenen Vpn. vergleicht. Die Explikationsskala ist also nicht nur - wie zuvor gezeigt - in der Lage, Besonderheiten der Worte zu erfassen, sondern auch die Leistungen der gleichen Vpn. zwischen verschiedenen Aufgaben zu differenzieren.

Wir können folgende Argumente zur Reliabilität unserer Kategorisierung resümmieren. Erstens sind die Inter-Rater-Reliabilitäten weder für die Verständnis- noch für die Explikationsskala zufried-

¹⁾-----
In (5) ist allerdings die Anzahl der Antworten, die in die Prozentverteilung der verglichenen Profile eingehen, doppelt so hoch wie in (2), (3) und (4), was zu den vergleichsweise höheren Korrelationen beiträgt.

denstellend. Wir haben versucht, die endgültige Bewertung durch alle drei Auswerter um so gewissenhafter festzulegen. Ob dies uns gelungen ist, kann durch die Profilvergleiche der endgültigen Bewertung zwischen Gruppen und Messungen unseres Experiments überprüft werden, die keine ursprünglich zu testenden Effekte darstellen. Zweitens stimmen die Profile der Verständnisbewertung in der mündlichen und schriftlichen Messung der Gesamtgruppe, der Schicht- und der Geschlechtsgruppen weitgehend überein. Allerdings war durch die Tendenz der Verständnisprofile zur U-Form auch kaum eine Variationsmöglichkeit im mittleren Bereich der Kategorien gegeben. Drittens zeigen die Profile der Explikationsbewertung in der mündlichen und schriftlichen Messung der Gesamtgruppe, der Schicht- und der Geschlechtsgruppen eine sehr ausgeprägte und dennoch in allen Gruppen gleiche Form. Viertens sind die Profile der Explikationsskala stabil mit den zu erklärenden Wortgruppen, nicht mit den erklärenden Vpn. verbunden; sie sind für verschiedene Wortgruppen und gleiche Vpn. weniger konstant als für verschiedene Wortgruppen und verschiedene Vpn. Die Explikationsskala erfaßt also Besonderheiten der Worte und kann zwischen verschiedenen Leistungen der gleichen Vpn. differenzieren. - Trotz der geringen Inter-Rater-Reliabilität glauben wir aufgrund dieser Argumente, daß die Explikationsskala ein verlässliches Instrument gewesen ist.

Da unsere Untersuchung Schichtunterschiede der Bedeutungserklärung prüfen soll, ist es wichtig, daß keine schichtspezifischen Vorurteile in die Bewertung der Bedeutungserklärungen eingegangen sind. Die Profile der Verständnis- und Explikationsskala für die mündliche und schriftliche Messung sind in der Unterschicht- wie in der Mittelschichtgruppe einander sehr ähnlich. Die Ähnlichkeit der Profile beider Schichten innerhalb jeder Kommunikationsform ist sehr groß und größer als die Ähnlichkeit der Profile beider Geschlechter - als einer vergleichbaren Zweiteilung unseres Experimentalsamples. So wenig wie von einer Schichtenspezifität der Auswertung kann außerdem - wie in 5.4.3. gezeigt werden wird - von einer Schichtenspezifität des Testmaterials geredet werden.

1.1.5. Probleme der Kategorisierung.

Abschließend soll noch auf einige Probleme unseres Kategoriensystems und auf Probleme von Auswertungen sprachlichen Materials im Allgemeinen hingewiesen werden. Unser Kategoriensystem enthielt zwei miteinander verbundene Probleme: Die Trennung der Dimensionen des Verständnisses und der Explikation und die Differenzierung qualitativer Stufen in jeder Dimension.

Die Trennung der Dimensionen des Verständnisses und der Explikation scheint uns von den praktischen Motiven der Auswertung und den theoretischen Zielen des Experiments nach wie vor gerechtfertigt. Die Lösung jedoch, die wir gewählt hatten, beide Dimensionen zu skalieren und "Verständnis" auf einer ebenso langen Skala wie "Explikation" hat sich nicht bewährt: Die Verständnisskala zeigt eine U-Form, in 5 ihrer Kategorien (4-8) fallen (im mündlichen wie im schriftlichen) weniger als 10% der Antworten, während in eine Kategorie (1) über 23% der Antworten gehören (siehe Tab. 4). Der Grund für dieses Scheitern scheint zu sein, daß wir die Trennung beider Dimensionen, die sich bei einigen Erklärungen plausibel durchführen läßt, für alle Erklärungen erzwungen haben. Die Trennbarkeit beider Dimensionen hängt stark von der Bedeutung der Wörter ab.¹⁾ Nachträglich wäre es wohl besser gewesen, die Erklärungen auf einer Dimension zu skalieren und bestimmte typische Weisen der Differenz zwischen Verständnis und Explikation (wie sie weiter oben dargestellt sind) als eigene Untergruppe oder eigene Kategorie auszuwerten; aus der zweidimensionalen Skalierung könnte man Kombinationen von Verständnis- und Explikationsbewertungen, die genereller oder auf bestimmte Stimulusworte besonders häufig auftauchen, für diesen Zweck herausgreifen. Dies führt allerdings zu einer weiteren qualitativen Differenzierung der Explikationsskala.

¹⁾-----
Einen zusätzlichen Beleg für diese Vermutung können wir aus dem Ergebnis nehmen, daß die mittleren Differenzen (zwischen Verständnis und Explikation) sich zwischen den Wortgruppen stark unterscheiden (siehe 5.4.4.): Die Redundanzregeln, mit deren Hilfe sich aus einer gegebenen Erklärung das zugrundeliegende Verständnis erschließen läßt, sind offensichtlich bei den verschiedenen Worten unterschiedlich klar definiert und anwendbar.

Die Explikationsskala scheint jedoch schon in ihrer jetzigen Form in zuviele Stufen zu zerfallen. Wie anhand des Häufigkeitsprofils erkennbar, könnten z.B. die Kategorien 7 und 8 zusammengefaßt und die Kategorie 2 aufgegeben werden. Daß eine zusätzliche Differenzierung der Explikationsskala nach "Verständnis-Merkmalen", ihre Kategorien trennschärfer und ihre Anwendung leichter gemacht hätte, ist dennoch möglich. Da die Explikationsskala eine Verbindung von Beschreibung und Bewertung ist, könnte eine genauere Beschreibung der auf jeder Bewertungsstufe zusammengefaßten qualitativen Kategorien die Zuordnung vereinfachen; wegen der Vielzahl der Bewertungsstufen aber könnte ihre Rangfolge dann unplausibel erscheinen. Kurz: Das allgemeine Problem aller Kategoriensysteme, beginnend schon mit dem "komplexen" und dem Carson/Rabin-schen Schema, ist die Verbindung von differenzierter Bewertung und differenzierter Beschreibung. Nachträglich wäre es wohl besser gewesen, weniger als 9 bewertende Stufen mit viel mehr qualitativen Kategorien zu verbinden, so daß die Zuordnung zu Rangstufen erleichtert wird, ohne daß sich intuitiv unsinnige Rangfolgen ergeben. Wenn unsere Explikationsskala eine Überdifferenzierung der Bewertung mit einer Unterdifferenzierung der Beschreibung verbindet, so hätte die gegenteilige Verbindung vielleicht eine verlässlichere Skala ergeben.

Jenseits dieser sich aus der Besonderheit unseres Kategoriensystems ergebenden Probleme deutet die Schwierigkeit, für die Wortbedeutungserklärungen eine befriedigende Reliabilität der Bewertung zu erlangen, u.E. auf allgemeine Probleme bei der Auswertung sprachlichen Materials. Problematisch ist nicht nur die Tauglichkeit der Kategorien der Bewertung, sondern auch die Bedeutung der zu bewertenden Objekte. Was ein Kind mit einer Erklärung gemeint hat, muß festgelegt werden, noch vor einer Entscheidung, wie diese Erklärung bewertet werden soll. Aus diesem Grunde scheint uns eine Diskussion zwischen verschiedenen Auswertern über die verschiedenen Bedeutungsaspekte einer vorliegenden Worterklärung, die zu einem Konsens über das zu bewertende Objekt und die korrekte Bewertung führt, für die "Reliabilität" der Bewertung wichtiger, als eine Übereinstimmung isoliert getroffener Urteile, durch die üblicherweise "Reliabilität" gemessen wird.

1.2. Bildung und Überprüfung von Maßen

1.2.1. Dimensionen der Maßkonstruktion.

Ausgangsmaterial für die Konstruktion der Maße ist die Bewertung aller Antworten auf den beiden Skalen des Bedeutungsverständnisses und der Bedeutungsexplikation sowie die Bewertung aller Antwortverweigerungen mit dem schlechtesten Wert beider Skalen, die jedoch unabhängig von gegebenen, aber falschen Erklärungen verkodet wurde. Außerdem wurden alle, auch verweigerte, Antworten in eine Rangfolge gebracht, die jedoch nur innerhalb der Stufen der Explikationsskala feinere Abstufungen herstellte, ohne die Abfolge dieser Stufen selber jemals zu durchbrechen. Alle Maße wurden für die mündliche und schriftliche Messung gesondert und in der genau gleichen Weise errechnet, so daß wir auf dieses Einteilungsprinzip nicht mehr weiter eingehen werden; jedes Maß ist durch den letzten Buchstaben ("M" oder "S") als mündliches oder schriftliches gekennzeichnet. Alle Maße können ferner, danach unterschieden werden, ob sie sich auf die Verständnis- oder Explikationsskala beziehen; jedes Maß ist durch die beiden Buchstaben VE oder EX an der 4. und 5. Stelle der insgesamt 6 Stellen für die abgekürzten Namen der Maße entsprechend gekennzeichnet.

Der Konstruktion der Maße liegen zwei voneinander unabhängige Dimensionen zugrunde. Da alle Maße Quotienten darstellen, können sie mit dem Zähler und dem Nenner identifiziert werden: Die Basis, d.h. die Gesamtheit der Erklärungen, auf die sich die Maße beziehen (der Nenner), und die Gewichtung der Bewertung, d.h. die Entscheidung, ob eine bestimmte Bewertung als Einzelfall einer Klasse von Skalenpunkten gezählt oder mit der ursprünglichen Punktzahl verrechnet wird (der Zähler). Diese Dimensionen sollen im folgenden kurz erläutert werden; in Tab. 6 sind alle Maße so dargestellt, daß die Prinzipien ihrer zweidimensionalen Konstruktion und ihre Querbezüge deutlich werden.

Drei verschiedene Basen für die Errechnung der Maße lassen sich unterscheiden: Die Maße können sich beziehen (1) auf die Gesamtheit aller möglichen Antworten, d.h. die richtigen Antworten, die falschen Antworten und die Antwortverweigerungen, so daß für jede Vp. die gleiche Anzahl von Aufgaben berücksichtigt wird. Die Maße können sich beziehen (2) auf die Gesamtheit aller von einer Vp. gegebenen Antworten, d.h. die richtigen Antworten und die falschen Antworten oder (3) die richtigen Antworten ohne die falschen Antworten, so daß für jede Vp. in diesen beiden Fällen eine von ihr bestimmte Anzahl von Aufgaben berücksichtigt wird. Wir haben die drei Gruppen von Maßen, die sich auf dieser Dimension unterscheiden lassen, Maße der A-, B- oder C-Kategorisierung genannt; sie sind meist durch ein A, B oder C an der dritten Stelle der Abkürzung gekennzeichnet.

Die Maße der A-Kategorisierung beurteilen jede Vp. automatisch in dem durch die Testanordnung gesetzten Rahmen. Die Maße der B- und C-Kategorisierung beurteilen jede Vp. in einem von ihr selbst gesetzten Rahmen: dem ihrer gegebenen Antworten, ihrer "Gesprächigkeit" gleichsam (B), und dem ihrer gegebenen nicht falschen Antworten, gleichsam ihres "Könnens" (C). Wichtig ist, daß alle Maße einen wie immer gefaßten Rahmen geben. Auch wenn bei vielen Maßen der A-Kategorisierung keine Division in die Definition aufgenommen wurde, stellen sie Quotienten dar: Entscheidend ist nicht die Durchführung der Rechenoperation, sondern die Relativierung der vorhandenen auf eine mögliche Leistung in der Anlage der Messung. Man kann sich leicht davon überzeugen, daß eine Division aller Maße der A-Kategorisierung, die keinen Nenner haben, durch 30, an der Reihenfolge und den Abständen der Vpn. auf diesen Maßen nichts ändern würde - ebensowenig wie eine Multiplikation der Mittelwerte mit 30 eine solche Änderung zur Folge hätte. Aufgrund dieser Überlegung wird auch verständlich, daß die Basen der B- und C-Kategorisierung ihrerseits Maße der A-Kategorisierung darstellen:

GATEX = die Zahl der gegebenen Antworten, und RATEX = die Zahl der gegebenen nicht falschen, in diesem Sinn "richtigen" Antworten müssen natürlich als Teil aller möglichen Antworten interpretiert werden, auch wenn sie nicht explizit durch 30 dividiert worden sind.¹⁾

Die Gewichtung der Bewertung - die zweite Dimension der Maßkonstruktion - sollte Ungenauigkeiten der in 9 Stufen differenzierten Skalen durch sukzessiv gröbere Zusammenfassungen kontrollierbar machen. Entsprechend läßt sich diese Dimension als ein Kontinuum verstehen. Je größer die Klasse von Bewertungsstufen, auf denen eine Erklärung nur als ein Fall dieser Klasse gezählt wird, desto geringer ist das Gewicht der einzelnen Bewertung; andererseits kann man die Verrechnung der Bewertungen mit ihrem tatsächlichen Wert auch als Zusammenfassung aller Klassen von Bewertungen nur einer Stufe verstehen. Die getroffene Einteilung in "nicht gewichtete", "grob gewichtete" und "gewichtete Bewertungen" ist willkürlich. Innerhalb der A-Kategorisierung lassen sich die Maße, beginnend mit den am wenigsten gewichtenden, folgendermaßen aufzählen: GATEX, RATEX, die Einteilung der 9 Kategorien in 2 Stufen (S12, S22), die Einteilung der 9 Kategorien in 3 Stufen (S13, S23, S33), die Mittelwerte, die Rangsummen (RAGSU). Die gleiche Reihenfolge wäre prinzipiell auch für die B- und C-Kategorisierung möglich, wenn man alle analogen Maße konstruiert. Sie sind jedoch sowohl rechnerisch redundant (d.h. durch 1 oder 2 andere Maße vollständig determiniert) wie ohne besonderen Informationsgehalt. Lediglich für die B-Kategorisierung sind 3 derartige Maße gebildet worden,

¹⁾-----
RATEX könnte ebensogut auch auf GATEX, statt auf 30 bezogen werden; da wir aber diesen Quotienten ausdrücklich gebildet haben (PRIAT) und da zudem der Unterschied zwischen beiden Maßen RATEX und PRIAT gering ist (siehe Tab. 8), haben wir in Tab. 6 auch RATEX auf die Gesamtzahl möglicher Antworten bezogen.

da sie Informationen anderer Maße in besonderer Weise zusammenfassen; PVEAT = der Anteil der am besten "verstandenen" aktiven Worterklärungen an den gegebenen Erklärungen, PEXAT = der Anteil der am besten "explizierten" aktiven Worterklärungen an den gegebenen Worterklärungen, PRIAT = der Anteil der "richtigen" aktiven Worterklärungen an den gegebenen Erklärungen. Da die Mittelwerte, von den Rangsummen abgesehen, am feinsten gewichtet, wählten wir sie für die durchgehende Konstruktion auf allen drei Kategorisierungsebenen aus. Sie bilden daher gleichsam das Rückgrat der Systematik unserer Maße. Alle Mittelwerte können zwischen 1 und 9 (A- und B-Kategorisierung) bzw. 1 und 8 (C-Kategorisierung) schwanken und sind negative Maße.

Zwei Maße fallen etwas aus der vorgelegten Systematik: die Rangsummen und die Differenzen. Die Rangsummen sind negative Maße und nicht ausdrücklich auf eine Basis bezogen; sie sind als das am feinsten gewichtende Maß der A-Kategorisierung dargestellt, weil sie die nicht gegebenen und falschen Antworten berücksichtigen und weil sie die gegebenen Antworten noch weiter als die 9 Stufen der Explikationsbewertung differenzieren. Auf der anderen Seite gehen die nicht gegebenen Antworten in die Rangsummen mit einem Rangplatz ein, der von der gegebenen Anzahl der Antworten nicht einer Vp. zu allen Worten, sondern aller Vpn. zu einem Wort abhängt. Insofern gehören sie gleichsam zu einer gruppen-, nicht individuenbezogenen B-Kategorisierung. Sie bringen als einziges Maß Informationen über die Erklärungsschwierigkeit der Worte in der ganzen Gruppe in die Bewertung einer einzelnen Vp.: eine falsche Erklärung, wo alle Vp. nichts erklären, wiegt wenig; eine falsche oder auch schlechte Erklärung, wo die meisten Vp. gute Erklärungen geben, wiegt sehr schwer. Da der individuellen Vp. aber in jedem Fall, ob sie eine Erklärung gibt oder nicht, ein Rangplatz zugewiesen wird, kommt das Maß RAGSU der A-Kategorisierung unserer Systematik am nächsten.

Die Differenzmaße wurden aufgrund der Vermutung gebildet, daß die Unterschichtkinder zwar eine schlechtere Explikation liefern, doch über ein ebenso gutes Verständnis verfügen würden wie die Mittelschichtkinder, daß daher die durchschnittlichen Differenzen der Unterschichtkinder größer seien als der Mittelschichtkinder (siehe 1.1. 2.). Da wir die Differenzmaße in zwei Abschnitten ausführlicher behandeln werden (5.4.4. und 5.6.1.), wollen wir auf sie etwas genauer eingehen. Rein von der Maßkonstruktion her können die Differenzen, da "Verständnis" für jede einzelne Erklärung besser angesetzt wurde als "Explikation", um so größer werden, je schlechter die Explikationsbewertung einer Erklärung ist. Von der tatsächlichen Verteilung der Bewertungen auf die Skalenpunkte her zeigen sich bei "Explikation" Schwankungen um etwa den gleichen Prozentsatz, während "Verständnis" zu einer U-Form tendiert (Tab. 3 und 4, Abb. 1). Auf die Häufigkeit bestimmter Differenzwerte kann jedoch aus diesen Verteilungen nicht geschlossen werden, da jede Explikationsbewertung mit verschiedenen Verständnisbewertungen, solange sie nur niedriger (besser) sind, kombiniert werden kann.

Die Differenzmaße wurden nicht mehr für einzelne Erklärungen interpretiert, sondern nur für eine Summe von Erklärungen, sei es über eine Gruppe von Worten für eine einzelne Vp., sei es über eine Gruppe von Vpn. für einzelne oder mehrere Worte, als entsprechender Durchschnittswert errechnet. Da für eine Reihe von Erklärungen der durchschnittliche Differenzwert auf der einen Seite um so leichter hoch sein kann, je höher der durchschnittliche Explikationswert ist, auf der anderen Seite dieser Zusammenhang - wie gezeigt - nicht notwendig besteht, so müssen wir unterschiedliche Differenzen zwischen Gruppen von Vpn. oder Worten immer im Zusammenhang mit den entsprechenden Unterschieden der Explikation analysieren. Wenn bei einem Gruppenvergleich die Rangfolge der Explikations- und Differenzmittelwerte nicht gleich ist, so können wir beide unabhängig interpretieren. Wenn bei einem Gruppenvergleich

die Rangfolge der Explikationswerte und Differenzen gleich sind, wird eine Interpretation schwierig, da offensichtlich primär die Logik der Maßkonstruktion für die Unterschiede der Differenzen verantwortlich ist.

Für die Differenzen hat die Trennung der Kategorisierungsarten nicht die gleiche Bedeutung wie für die Verständnis- und Explikationsdimension allein. Während bei ihnen die "k.A." mit dem sinnvoll an die Skala anschließenden Wert "9" berücksichtigt werden, gehen sie in die Differenzen mit "0" ein und drücken den Gesamtwert durch eine unverhältnismäßige Verkleinerung des Nenners herab. Deshalb stützen wir uns vor allem auf die Maße DIBVE, wo die Null-differenz der als "9" bewerteten falschen Erklärungen ebenso informationshaltig sind wie bei den mit "1" bewerteten Antworten, und DICVE, wo überhaupt keine mit "9" bewerteten Antworten berücksichtigt werden.

Zwei weitere Maße fallen nur scheinbar aus der Systematik: GATEX und RATEX. Da falsche und nicht gegebene Erklärungen auf beiden Skalen mit dem schlechtesten Wert "9" bewertet wurden, wären die nicht gebildeten Maße GATVE und RATVE mit diesen beiden Maßen identisch. Anders als die Differenzmaße stehen also GATEX und RATEX nicht außerhalb von Verständnis und Explikation, sondern können auf beide zugleich bezogen werden; die Zuordnung zu den Explikationsmaßen ist willkürlich. Dieselbe Überlegung gilt für das Maß PRIAT. Nur weil sie Verständnis wie Explikation umfassen, können GATEX und RATEX ja auch als Basis für die Verständnis- wie Explikationsmaße der B- und C-Kategorisierung dienen.

1.2.2. Aussagekraft der Maße.

In die Konstruktion der Maße gingen zwei Dimensionen ein: die Berücksichtigung oder Nichtberücksichtigung der individuellen Antwortbereitschaft bzw. Antwortfähigkeit und die unterschiedliche Gewichtung der ursprünglichen 9er Skalen. Der Informationsgehalt der Maße läßt sich anhand der ersten und - ergänzend - auch der zweiten Dimension darstellen. Die Maße zerfallen in drei Gruppen: (1) Maße, die nur die Häufigkeit der Antworten unabhängig von der Güte der Antworten bewerten, (2) Maße, die nur die Güte der Antworten berücksichtigen, unabhängig von Unterschieden zwischen den Kindern in der Häufigkeit der Antworten, (3) Maße, die sowohl die Häufigkeit wie die Güte der Antworten berücksichtigen. In die Gruppe (2) fallen alle Maße der B- und C-Kategorisierung, in die Gruppe (1) und (3) alle Maße der A-Kategorisierung; hier besteht eine klare Grenze. In die Maße der Gruppe (1) fallen die nicht bzw. weniger gewichtenden Maße der A-Kategorisierung; hier besteht keine klare Grenze, sondern ein kontinuierlicher Übergang. Die Systematik der in den Maßen enthaltenen Aussagen ist in Tab. 7 zusammengefaßt; sie kann leicht auf Tab. 6 zurückgeführt werden, wenn man die Maße der B- und C-Kategorisierung zusammenfaßt und die Maße der A-Kategorisierung zwischen "nicht gewichteten" und "grob gewichteten Bewertungen" teilt. PRIAT wurde in Tab. 7 als Maß der Antwortgüte dargestellt, obwohl es aus zwei Maßen der Antworthäufigkeit gebildet ist; wie alle Maße der Antwortgüte bewertet es die Leistung der Vp. in einem von ihr selbst gesetzten Rahmen, beide Größen sind jedoch einander soweit wie möglich angenähert.

Zehn der insgesamt 25 gebildeten Maße sind insofern redundant, als sie durch ein oder zwei der übrigen Maße vollständig determiniert sind; sie sind in Tab. 6 und 7 in Klammern gesetzt und zerfallen in drei Gruppen. Innerhalb der Zweier-Einteilung und der Dreier-Einteilung der Skalen ist ein beliebiges Maß redundant; wir haben die Maße S22 und S23 als redundant dargestellt. Weiterhin sind alle Differenzen redundant, weil sie sich aus den Unterschieden

zwischen den zugehörigen Explikations- und Verständnismittelwerten ergeben. Schließlich sind alle Maße der B-Kategorisierung, die keine Mittelwerte darstellen, redundant, weil sie als Quotienten von zwei Maßen der A-Kategorisierung gebildet sind. Auch bei den letzten beiden Gruppen redundanter Maße ist es natürlich prinzipiell beliebig, welches der drei jeweils untereinander verbundenen Maße als redundant bezeichnet wird; die in Tab. 6 und 7 getroffene Wahl orientiert sich an der Bedeutung der Maße für die Darstellung der Ergebnisse. Konstruktionsweise und Aussagekraft der Maße sollen im folgenden an den Mittelwerten und Interkorrelationen aller und einer Faktorenanalyse der nicht redundanten Maße überprüft werden.

1.2.3. Überprüfung der Maße: Mittelwerte und Standardabweichungen.

Sie sind für die mündliche und schriftliche Messung in Tab. 8 zusammengestellt. Zunächst kann an der weitgehenden Übereinstimmung beider Messungen für alle Maße in der Gesamt- wie in der Mittel- und Unterschichtgruppe erneut die Reliabilität unserer endgültigen Kategorisierung abgeschätzt werden (siehe 1.1.4.). Im folgenden sollen jedoch am Vergleich der Maße innerhalb einer Gruppe und Messung die Dimensionen der Maßkonstruktion überprüft werden.

Im Durchschnitt hat jede Vp. etwa 14 Stichworte zu erklären versucht und 12-13 nicht völlig falsch erklärt (GATEX und RATEX).¹⁾ Setzt man beide Zahlen zur Gesamtzahl der möglichen Antworten in Relation (30), so kann man die Unterschiede zwischen den drei Kategorisierungsarten ermessen: Zwischen der B- und C-Kategorisierung besteht ein geringer Unterschied, während beide sich scharf von der A-Kategorisierung absetzen; das Verhältnis der Basen der B- und C-Kategorisierung stellt PRIAT dar. Die Mittelwerte der Mittelwerte bestätigen dies Verhältnis der drei Kategorisierungs-

↑)-----

Bei diesen Maßen ergibt sich der einzige auffällige Schichtunterschied, der jedoch nicht signifikant ist (siehe 5.3.2.).

arten: sie liegen für die B- und C-Kategorisierung eng zusammen, aber weit entfernt von der A-Kategorisierung. Die Differenzen bezeichnen genau den Abstand zwischen den entsprechenden Verständnis- und Explikationsmittelwerten. Die zwischen den Kategorisierungsarten unterschiedlichen Standardabweichungen der Mittelwerte verlangen eine Erklärung; sie sind durchweg in der B-Kategorisierung an größten, in der C-Kategorisierung am kleinsten. Die Erklärung liegt in der Lage der entsprechenden Mittelwerte, die für die B-Kategorisierung die größten Abstände erlauben innerhalb des Rahmens von 9 möglichen Werten, für die C-Kategorisierung wegen der extremen guten Lage der Mittelwerte jedoch nur geringe Abstände innerhalb eines Rahmens von 8 möglichen Werten.

Von den Maßen der Zweier- und Dreier-Einteilung ist nur jeweils eins dargestellt; die dargestellten ergänzen sich aber selbstverständlich mit den nicht dargestellten entsprechenden Werten zu 30. Die Tendenz zur U-Verteilung in "Verständnis" kann auch hier erkannt werden: Obwohl in den positiven Verständnismaßen der Zweier- und Dreier-Einteilung (S12, S13) weniger Kategorien zusammengefaßt sind als in den entsprechenden Maßen der Explikation (siehe Tab. 6), sind sie viel stärker besetzt. Die Maße der Zweier- und Dreier-Einteilung können, wie die Mittelwerte, mit den prozentualen Häufigkeiten der einzelnen Kategorien, aus denen sie sich zusammensetzen, verglichen werden (Tab. 3 und 4).

1.2.4. Überprüfung der Maße: Interkorrelationen.

Sie sind für die mündlichen und schriftlichen Maße getrennt dargestellt, für die Gesamtgruppe in Tab. 9, für die Unterschichtgruppe in Tab. 10, für die Mittelschichtgruppe in Tab. 11; eine Zusammenfassung dieser Korrelationen nach den verschiedenen Maßdimensionen - mit und ohne Berücksichtigung der Kommunikationsform - findet sich in Tab. 13. Anhand dieser Korrelationen wollen wir das Verhältnis der drei Kategorisierungsarten zueinander klären und einzelne Maße den Kategorisierungsebenen bzw. Informationselementen

zuordnen; die Beziehung zwischen den Verständnis- und Explikationsmaßen genauer beschreiben, was eine ausführliche Betrachtung der Differenzmaße einschließt; und schließlich die Interkorrelationen analoger Maße der mündlichen und schriftlichen Messung darstellen. Anders als bei den Mittelwerten spiegelt der Vergleich zwischen den Kommunikationsformen und Schichtgruppen hier nicht unmittelbar die Experimentalergebnisse und kann zur Prüfung der Reliabilität der Maße dienen.

Die Gesamtzahl der Antworten und die Gesamtzahl der nicht falschen Antworten (GATEX und RATEX) korrelieren natürlich in allen Messungen und Gruppen sehr hoch. Deshalb müssen die entsprechenden Maße der B- und C-Kategorisierung innerhalb jeder Messung auch sehr eng zusammenhängen: MIBEX mit MICEX, MIBVE mit MICVE und DIBVE mit DICVE korrelieren in der mündlichen wie schriftlichen Messung in allen Gruppen um .90. Die Korrelationen der Mittelwerte des Verständnisses bzw. der Explikation der A-Kategorisierung mit denen der B- und C-Kategorisierung sind jedoch entschieden niedriger: sie liegen zwischen .18 und .77 und schwanken um .40; auffällig niedrig sind sie für Verständnis und Explikation in der mündlichen Messung der Unterschichtgruppen, wo offensichtlich die Antwortbereitschaft mit der Erklärungsfähigkeit am wenigsten zu tun hat. Insgesamt ergibt sich aus den Korrelationen das gleiche Bild über das Verhältnis der drei Kategorisierungsarten wie aus den Mittelwertvergleichen.

Für die Zuordnung einzelner Maße zu den Kategorisierungsarten müssen wir ihre Korrelationen mit den Mittelwerten als Maßstab nehmen. GATEX und RATEX gehören - wie aus den Korrelationen mit den Mittelwerten ersichtlich - zu den Maßen der A-Kategorisierung und können ebenso als Maßstab für die Zuordnung weiterer Maße zu dieser Kategorisierung genommen werden. - Die Maße der Zweier- und Dreier-Einteilung korrelieren in allen Messungen und Gruppen hoch miteinander und innerhalb der gleichen Skala am höchsten: S12VE und S13VE bzw. S12EX und S13EX unterscheiden sich ja, genau wie GATEX und RATEX, nur durch die Hinzunahme einer weiteren Kategorie

auf der jeweiligen Skala (siehe Tab. 6). Die Maße der Zweier- und Dreier-Kategorisierung korrelieren in allen Messungen und Gruppen höher mit den Verständnis- und Explikationsmittelwerten der A- als der B- und C-Kategorisierung - ein deutlicher Hinweis für ihre Zuordnung zu den Maßen, die Antwortgüte und Häufigkeit verarbeiten und für die Irrelevanz der rechnerischen Relativierung eines Maßes auf die Gesamtzahl möglicher Antworten, wenn es von der Anlage der Messungen schon darauf bezogen ist. Die Korrelationen der Maße der Zweier- und Dreier-Einteilung mit den Mittelwerten der B- und C-Kategorisierung für Verständnis und Explikation sind jedoch für die Gesamtgruppe im Mündlichen deutlich niedriger als im Schriftlichen; dies zeigt sich verstärkt wiederum in der Unterschichtgruppe, während in der Mittelschichtgruppe die Korrelationen in beiden Kommunikationsformen gleichmäßig hoch sind. Mit PVEAT korrelieren die Mittelwerte der B- und C-Kategorisierung für Verständnis, mit PEXAT für Explikation sehr hoch (um .90); dies gilt in allen Messungen und Gruppen, so daß die Beziehungsstruktur der die Antwortgüte verarbeitende Maße als durchweg konstant angesehen werden kann. Auch PRIAT korreliert höher mit den Maßen der B- und C- als mit denen der A-Kategorisierung und gehört daher zu den Maßen, die nur die Antwortgüte berücksichtigen.

Die Rangsummen konnten schon von der Konstruktion nicht umstandslos einer Kategorisierung zugeordnet werden. Sie korrelieren mit den Mittelwerten für Verständnis und Explikation der A-Kategorisierung geringer als die Maße der Zweier- und Dreier-Einteilung, die die Mittelwerte in etwa gleichem Maße vergrößern wie die Rangsummen sie verfeinern. In den Fällen, wo die Rangsummen mit GATEX und RATEX nur mäßig hoch korrelieren (schriftliche Messungen aller Gruppen), sind die Unterschiede zwischen ihren Korrelationen mit den Verständnis- und Explikationsmittelwerten der A- und denen der B- und C-Kategorisierung gering; in den Fällen sehr hoher Korrelation mit GATEX und RATEX jedoch (mündliche Messung aller Gruppen)

sind diese Unterschiede sehr groß. Wenn die Vpn. durch viele Erklärungen einen guten Gesamtrangplatz erhalten haben, so garantiert dies noch nicht einen guten Rang der gegebenen Erklärungen (im Mündlichen); wenn der Gesamtrangplatz der Vpn. sich weniger aus der Häufigkeit von Erklärungen ergeben hat, so hängt er stärker mit der Güte der gegebenen Erklärungen zusammen (im Schriftlichen). Kurz: In der mündlichen Messung sind die Rangsummen eher ein Maß der A-Kategorisierung, in der schriftlichen Messung sind sie, im Vergleich zu anderen Maßen, gegenüber den Kategorisierungsebenen ziemlich neutral, obwohl sie auch hier noch am ehesten der A-Kategorisierung zuzuordnen sind. Einen Eindruck von der Differenzierungskraft der Rangsummen geben ihre Korrelationen mit PEXAT: sie sind für die Gesamtgruppe mittelmäßig hoch, aber im schriftlichen deutlich höher, für die Unterschichtgruppe im schriftlichen gering und im mündlichen null, für die Mittelschichtgruppe in beiden Messungen gleichmäßig sehr hoch.

Bis jetzt haben wir die Zuordnung von Maßen zu den Kategorisierungsebenen dargestellt, ohne zwischen den beiden Skalen zu unterscheiden. Im folgenden wollen wir die Beziehungen zwischen Maßen der Explikation und des Verständnisses untersuchen. - Die Korrelationen zwischen den Mittelwerten für das Verständnis und der Explikation sind in allen Gruppen und Messungen sehr hoch und nehmen mit den Kategorisierungsarten kontinuierlich ab, so daß sie in der A- bei .95, der B- bei .85 und der C-Kategorisierung bei .70 liegen.¹⁾ Die Korrelationen zwischen PVEAT und PEXAT liegen in der Gesamtgruppe bei .70, also so hoch wie die Korrelationen zwischen Verständnis und Explikation für die C-Kategorisierung, in den beiden Schichtgruppen zwischen .49 und .62. Die Interkorrelationen zwischen den Verständnis- und Explikationsmaßen der Zweier- bzw. der Dreier-Einteilung betragen in allen Messungen und Gruppen unge-

1) -----
Eine wahrscheinlich bedeutungslose Ausnahme ist die mündliche Messung der Mittelschicht, wo die Korrelation zwischen MICVE und MIBEX .52 ist.

fähr .80. Da Antwortverweigerungen in beiden Skalen ja gleich bewertet wurden, müssen wir die Maße der B- und C-Kategorisierung als besten Maßstab für die Interkorrelation von Verständnis und Explikation ansehen.

Die Interkorrelationen zwischen Verständnis und Explikation sind bedeutsam für die Interpretation der Differenzen, da diese mit der Höhe der Explikation größer, mit der Höhe des Verständnisses kleiner werden können. Durch die Maße der Explikation und des Verständnisses zusammen sind die Differenzen der gleichen Kategorisierungsebene voll determiniert. Anders gesagt: Die Partialkorrelation des Verständnisses mit den Differenzen, unabhängig von der Explikation, müssen für alle Kategorisierungsarten -1.00 ; die Partialkorrelationen der Explikation mit den Differenzen, unabhängig vom Verständnis, für alle Kategorisierungsarten $+1.00$ betragen. Dieser Zusammenhang läßt sich anhand der in der Tab. 12 angegebenen Interkorrelationen und der Formel für Partialkorrelationen verifizieren. - Die Gegenläufigkeit der Einflüsse des Verständnisses und Explikation auf die Differenzen braucht sich nicht auf allen drei Kategorisierungsebenen in entgegengesetzten Vorzeichen der Korrelationen auszudrücken. In der A-Kategorisierung zählt jede nicht gegebene Antwort als Nulldifferenz und geht dennoch in die Basis, auf die die Summe der Differenzen bezogen wird, ein; da im Durchschnitt mehr als die Hälfte der Worte nicht erklärt wurden, werden die Differenzen mit beiden Maßen positiv, mit der Explikation aber stärker positiv, korrelieren: Je besser (d.h. auf den Skalen je niedriger) Verständnis und Explikation bewertet werden, desto größer ist die Differenz; dies ist eine Artefakt der zahlreichen verweigerten (auf den Skalen mit "9" bewerteten) Erklärungen, für die keine Differenz berechnet werden konnte. Innerhalb der B- und C-Kategorisierung jedoch müßte sich der eigentliche Zusammenhang zeigen: Je besser (d.h. auf der Skala niedriger) das Verständnis und je schlechter (d.h. auf der Skala höher) die Explikation bewertet wird, desto größer ist die Differenz. Auch

diese Zusammenhänge können in Tab. 12 verifiziert werden. - Für die Zusammenhänge beider Dimensionen gilt aber im Vergleich der drei Kategorisierungsebenen: Je weniger Verständnis und Explikation interkorrelieren, desto größer wird der Unterschied zwischen ihren Korrelationen mit der Differenz. Auch dies läßt sich in Tab. 12 verifizieren.

Mit GATEX und RATEX müßten von der Maßkonstruktion her die Differenzmaße der A-Kategorisierung hoch, die Differenzmaße der B- und C-Kategorisierung so gut wie nicht korrelieren. Für die Differenz der A-Kategorisierung kann dies in allen Gruppen und Messungen bestätigt werden. Für die Differenzen der B- und C-Kategorisierung jedoch ergibt sich eine Überraschung (Tab. 12): in der Unterschichtgruppe korrelieren sie positiv mit der Zahl der gegebenen wie der Zahl der nicht falschen Antworten, in der Mittelschichtgruppe aber negativ.

Bis jetzt haben wir die Beziehungsstruktur der Maße der drei Kategorisierungsebenen und der beiden Skalen in der Gesamt-, Unter- und Mittelschichtgruppe dargestellt. In den drei Gruppen wie in den beiden Kommunikationsformen blieb die Beziehungsstruktur der Maße konstant - mit einigen Ausnahmen in der mündlichen Messung der Unterschicht und bei den Differenzmaßen, auf die wir in 1.2.5. eingehen werden. Wenn wir also die Konstanz der Beziehungsstruktur der Maße in verschiedenen Gruppen und Messungen als Indiz nehmen, so ist die Reliabilität der Maße befriedigend. Eine zweite Möglichkeit der Reliabilitätsprüfung unserer Maße bieten die Interkorrelationen der analogen Maße¹⁾ der mündlichen und schriftlichen Messung (Tab. 14). Sie stellen eine Art "split-half-reliability"

1)-----

Zwei Maße bezeichnen wir als "analog", wenn alle Dimensionen der Maßkonstruktion (Verständnis-Explikation; A-, B-, C-Kategorisierung) mit Ausnahme der Kommunikationsform gleich sind. Zwei Maße bezeichnen wir als "nicht analog", wenn außer der Kommunikationsform noch die Skala oder die Kategorisierung oder beide Dimensionen verschieden sind.

innerhalb der gesamten Messung für jede Vp. dar. Nimmt man die kaum "gewichtenden", d.h. kaum Bewertungen einschließenden, Maße GATEX und RATEX als Vergleichspunkt, so sind die Interkorrelationen der am stärksten gewichtenden, d.h. nur Bewertungen enthaltenden, Maße wie MIBEX und MICEX in allen Gruppen sehr hoch. Die Interkorrelationen der Mittelwerte nehmen von der A- zur C-Kategorisierung kontinuierlich ab; das ist durch den Fortfall des Einflusses der Anwohnhäufigkeit erklärlich, die zwischen den Kommunikationsformen konstant bleibt (siehe GATEX und RATEX). Die Höhe der Interkorrelationen für PEXAT und PVEAT entspricht etwa der für MIBEX. Die Interkorrelationen der Differenzmaße sind recht niedrig; sie vereinen gleichsam die Instabilität der Verständnis- und Explikationsmaße. Auffällig sind die im Vergleich zur Mittelschicht hohen Interkorrelationen der Differenzen in der Unterschicht. Fast alle der 20 in Tab. 14 dargestellten Paare analoger Maße zeigen in der Mittelschichtgruppe höhere Interkorrelationen als in der Unterschichtgruppe, mit den Ausnahmen GATEX und RATEX, DIAVE und DICVE. Für N=83 sind die Interkorrelationen etwas niedriger als für N=65 und N=66.

1.2.5. Schichtunterschiede der Interkorrelationen der Maße.

Nur an wenigen Stellen finden sich Unterschiede zwischen den beiden Schichtgruppen in der Beziehungsstruktur der Maße; diese Tatsache haben wir als Indiz für die Reliabilität unserer Maße genommen. Unabhängig von der formalen Prüfung der Maße wollen wir diese Unterschiede im folgenden inhaltlich interpretieren; es handelt sich um zwei Ergebnisse.

Zunächst haben wir anhand der Korrelationen der Mittelwerte der A-Kategorisierung und der Maße der Zweier- und Dreier-Kategorisierung für Verständnis und Explikation mit den Mittelwerten der B- und C-Kategorisierung und an der Korrelation der Rangsummen mit PEXAT festgestellt, daß in der Unterschicht "Güte" und "Güte und

Häufigkeit" der Antworten in der mündlichen Messung kaum, in der schriftlichen Messung jedoch recht gut zusammenhängen, während in der Mittelschicht die beiden Dimensionen der Erklärungsleistung sowohl in der mündlichen wie in der schriftlichen Messung zusammenhängen; in beiden Schichtgruppen wurden in der mündlichen und schriftlichen Messung dieselben Vpn. getestet. Für die Wirkung der Kommunikationsform haben wir die Hypothese, daß die schriftliche Kommunikationsform von Kindern beider Schichten bessere Leistungen erzwingt, für die Kinder der Unterschicht jedoch in stärkerem Maße (siehe 5.3.4.). Unabhängig von absoluten Unterschieden der Erklärungsleistung finden wir hier Unterschiede in der Struktur der Erklärungsleistung zwischen Kindern beider Schichten in Abhängigkeit von der Kommunikationsform, die diesen Hypothesen entspricht. Die mündliche Messung übt für die Kinder der Unterschicht kaum einen leistungsstrukturierenden Zwang aus: Kinder, die viele Antworten bringen, geben vergleichsweise schlechte Erklärungen, und Kinder, die wenige Antworten bringen, geben vergleichsweise gute Erklärungen. Dieselben Kinder aber geben in der schriftlichen Messung durchweg in Häufigkeit und Güte vergleichbare Erklärungen. Für die Unterschichtkinder verlangt die schriftliche Messung - unabhängig vom absoluten Leistungsniveau - eine Konzentration der Kräfte, eine zusätzliche Anpassung oder eine geänderte Auffassung der Aufgabe, die für die Mittelschichtkinder nicht spürbar ist. Anders gesagt: der Wechsel der Kommunikationsformen bewirkt für die Unterschichtkinder zugleich einen Wechsel in Strategien des Sprachgebrauchs, während er die Strategien des Sprachgebrauchs bei den Mittelschichtkindern unberührt läßt.¹⁾

Als zweites zeigte sich, daß die Differenzen der B- und C-Kategorisierung in der Unterschichtgruppe positiv, in der Mittelschichtgruppe negativ mit GATEX und RATEX korrelieren. Anwohnhäufigkeit

1)-----
"Strategien des Sprachgebrauchs" sind dabei in einem "elementaristischen" Sinne zu verstehen, der in Oevermann 1970 und in Kapitel V.3.1. dieser Arbeit erläutert wird.

hängt also bei den Unterschichtkindern mit einer, gemessen am Wortverständnis, schlechten Erklärung zusammen, bei den Mittelschichtkindern jedoch mit einer dem Verständnis vergleichbaren Erklärung. Bei der Konstruktion der Differenzmaße hatten wir angenommen, daß die Unterschichtkinder ihr Wortverständnis schlechter explizieren können als die Mittelschichtkinder und daher größere Differenzen aufweisen. Wir finden hier wiederum einen Unterschied in der Struktur der Worterklärung zwischen beiden Schichten, der der ursprünglichen Hypothese über die absoluten Unterschiede der Worterklärung zwischen beiden Schichten entspricht.¹⁾ Hatten wir zuerst in den Schichten unterschiedliche Zusammenhänge zwischen "Güte" und "Güte und Häufigkeit" der Erklärungen in Abhängigkeit von der Kommunikationsform festgestellt, so finden wir als zweites in den Schichten unterschiedliche Zusammenhänge zwischen Verständnis und Explikation in Abhängigkeit von der Antwoorthäufigkeit. In beiden Fällen sind die Zusammenhänge in der Mittelschicht konsistent, in der Unterschicht inkonsistent. Wenn die schichtspezifischen Strukturzusammenhänge der Maße als ein Indiz für schichtspezifische Strategien des Sprachgebrauchs genommen werden, so unterscheidet sich die Unterschicht- von der Mittelschichtgruppe der zweiten Messung durch eine größere Abhängigkeit der Erklärungen von der Kommunikationsform und der Antwoorthäufigkeit.

1.2.6. Überprüfung der Maße: Faktorenanalyse.

Die Faktorenanalyse mit allen nicht redundanten mündlichen und schriftlichen Maßen (Tab. 15) destilliert die wichtigsten Dimensionen aus der Matrix der besprochenen Interkorrelationen. Wir wollen als Ausgangspunkt der Faktorenanalyse die Ergebnisse der Korrelationsanalyse kurz zusammenfassen und beschränken uns dabei auf die Gesamtgruppe, auf die sich die Faktorenanalyse ebenfalls bezieht (siehe Tab. 13). Die Interkorrelationen der Verständnis-

¹⁾-----
Allerdings wurde unsere Hypothese über die absoluten Schichtunterschiede der Differenzmaße nicht bestätigt: Wir finden eher in der Mittelschicht die größeren Differenzen (siehe 5.4.4.).

mit den Explikationsmaßen sind etwa doppelt so stark wie die der A- mit der B- und C-Kategorisierung; die Interkorrelationen der mündlichen mit den schriftlichen Maßen liegen ungefähr dazwischen. Zwischen den Maßen, die nur die Anwothhäufigkeit, und denen, die nur die Antwortgüte verarbeiten, besteht so gut wie kein Zusammenhang; die Maße, die Anwothhäufigkeit und Antwortgüte verarbeiten, korrelieren mit denen, die Anwothhäufigkeit verarbeiten, sehr hoch, mit denen, die Antwortgüte verarbeiten, jedoch nur mittelmäßig. Danach muß in der Faktorenanalyse folgende Gruppierung der Maße erwartet werden: Die Maße für Anwothhäufigkeit und für Antwortgüte oder: die Maße der A-Kategorisierung müßten, gleich ob auf das Verständnis oder Explikation bezogen, eine Einheit bilden, von der sich B- und C-Kategorisierung, wiederum unabhängig von Verständnis oder Explikation, deutlich abheben. Eine ähnliche Gruppierung ist auch von der Maßkonstruktion her (Tab. 7) zu erwarten.

Die Faktorenanalyse bestätigt diese Gruppierung der Maße. Nur vier Faktoren zeigten einen Eigenwert über 1, d.h. umfaßten mehr Varianz als eine einzelne Variable (Tab. 15). Die unrotierte Faktorenmatrix weist den typischen ersten, allgemeinen Faktor auf, einen zweiten Faktor, der die A- von der B- und C-Kategorisierung trennt, einen dritten Faktor, der die Kommunikationsformen, und einen vierten der wiederum, allerdings schwächer als der zweite und mit Vertauschung der Vorzeichen zwischen den Kommunikationsformen, die A- von der B- und C-Kategorisierung trennt. Die Rotation der Faktoren verdeutlicht sehr schön, daß sie von der Kommunikationsform und den Kategorisierungsarten bestimmt sind. Die mündlichen Maße laden hoch auf den beiden ersten, die schriftlichen auf den beiden weiteren Faktoren. Innerhalb jeder Kommunikationsform zeigt sich klar die Trennung zwischen Maßen der A- und der B- und C-Kategorisierung in der Höhe wie den Vorzeichen der Ladungen. Die schriftlichen Maße der A-Kategorisierung laden geringer auf dem Faktor der mündlichen Maße der A-Kategorisierung und umgekehrt; ebenso laden die schriftlichen Maße der B- und C-Kategorisierung auf dem Fak-

tor der mündlichen Maße der B- und C-Kategorisierung geringer und umgekehrt. Innerhalb jeder Kommunikationsform finden wir also ein abgeschwächtes Spiegelbild der Faktorenstruktur der anderen Kommunikationsform. In jedem Faktor sind die Verständnis- mit den zugehörigen Explikationsmaßen zusammengefaßt. Insgesamt ist die Symmetrie der Faktorenstruktur sehr eindrucksvoll.

Die Unterscheidung zwischen Maßen, die nur die Anwohthäufigkeit, und Maßen, die Anwohthäufigkeit und -güte verarbeiten, ist - wie von den Interkorrelationen und dem kontinuierlichen Übergang zwischen beiden Extremen zu erwarten war - in der Faktorstruktur nicht erhalten geblieben. Weder in Faktor I noch in Faktor III ist die in Tab. 6 und 7 prognostizierte Reihenfolge der Gewichtung wiederzuerkennen. Die Trennung zwischen Maßen der A- und B- bzw. C-Kategorisierung ist in der schriftlichen Messung (Faktor III und IV) weit weniger deutlich als in der mündlichen: dies rührt von den höheren Korrelationen der Maße der A- mit den Maßen der B- u. C-Kategorisierung in der schriftlichen Messung her (siehe Tab. 9). Es läßt sich auch an den unterschiedlichen Varianzanteilen der Faktoren verfolgen: Zwar erklären in beiden Messungen die Faktoren der A-Kategorisierung (I und III) die meiste Varianz; Faktor IV ist jedoch bezüglich des Varianzanteils Faktor III viel eher ebenbürtig als Faktor II dem Faktor I. Die Rangsummen schließlich wechseln zwischen den Kategorisierungsarten. Während sie in der mündlichen Messung eindeutig der A-Kategorisierung zugerechnet werden müssen, liegt ihr Schwerpunkt in der schriftlichen Messung zwar bei der B- und C-Kategorisierung, auf den Faktoren der A-Kategorisierung zusammen (I und III) laden sie jedoch noch stärker und in der entgegengesetzten Richtung. Zudem hat RAGSUS die niedrigste Kommunalität aller Maße. Schwerpunkt wie Verstreutheit der Ladung der schriftlichen Rangsummen waren von den Korrelationen zu erwarten.

Die Trennung zwischen der A- und B- und C-Kategorisierung bzw. zwischen den Maßen der Antwortgüte und Maßen der Antworthäufigkeit oder Antworthäufigkeit und -güte ist also die entscheidende Dimension für das Verständnis der Maße innerhalb jeder Kommunikationsform. Muß indes die Trennung der Maße nach den Kommunikationsformen, in denen doch jeweils exakt dieselben Maße auftauchen, und dieselben Vpn. gemessen werden, als Indiz für die Stabilität oder Instabilität unserer Maße gewertet werden? Wäre eine Struktur, in der sich Maße der mündlichen und schriftlichen Messung, so wie hier Maße der Explikation und des Verständnisses, in von der Logik der Maßkonstruktion bestimmten Gruppen zusammenfinden ein besseres Zeichen für die Reliabilität der Maße? Hatten wir nicht bis jetzt hohe Korrelationen zwischen mündlichen und schriftlichen Maßen als ein Maß der Reliabilität genommen (Tab. 14)?

Zur Beantwortung dieser Fragen müssen wir uns die möglichen Beziehungen zwischen den Maßen der mündlichen und schriftlichen Messungen verdeutlichen. Die in Tab. 14 aufgeführten Korrelationen beziehen sich nur auf die analogen Maße der mündlichen und schriftlichen Messung. In die Faktorenanalyse gehen dagegen die Korrelationen aller 15 mündlichen mit allen 15 schriftlichen Maßen ein, insgesamt 435 Korrelationen. Sieht man von den jeweils 105 Interkorrelationen der Maße innerhalb der mündlichen bzw. schriftlichen Messungen ab, so bleiben 225 Korrelationen zwischen einem mündlichen und einem schriftlichen Maß, von denen die 15 analogen für die Einschätzung der Stabilität der Maße zwischen den beiden Messungen eingesetzt wurden und die 210 nichtanalogen zwischen (bei Bereinigung des unterschiedlichen Vorzeichen-Charakters der Maße) $-.01$ und $+.66$ mit einem Schwerpunkt um $.20$ schwanken. Die Korrelationen der nicht-analogen Maße müssen viel geringer sein, da sie verschiedenen Kommunikationsformen und verschiedene Meßdimensionen zugleich verbinden. Die Korrelationen aller Maße zwischen mündlicher und schriftlicher Messung sind durch die wenigen Korrelationen der analogen Maße viel zu hoch eingeschätzt.

In einer Faktorenanalyse aller mündlichen und schriftlichen Maße ist daher die Extraktion eines Faktors der Kommunikationsform in der unrotierten bzw. die Gruppierung der Faktoren nach der Kommunikationsform in der rotierten Faktorenmatrix sehr wahrscheinlich, die Alternative zwischen den Kommunikationsformen nicht differenzierender Faktoren unwahrscheinlich. Bei der geringen Abhängigkeit beider Messungen voneinander verlieren vielmehr die Tatsachen an Selbstverständlichkeit, daß sich mündliche und schriftliche Faktoren auf fast die gleiche Weise differenzieren und daß in der einen Messung das abgeschwächte Spiegelbild der Struktur der anderen Messung wiedererscheint. Anders gesagt: daß aus den jeweils 105 Interkorrelationen der mündlichen bzw. schriftlichen Maße die gleiche Struktur und aus den jeweils 105 Interkorrelationen der mündlichen mit nicht-analogen schriftlichen bzw. der schriftlichen mit nicht-analogen mündlichen Maßen in gleicher Weise ein schwaches Abbild der Interkorrelationen der analogen Maße der gleichen Kommunikationsform gewonnen werden kann. Diese beiden Tatsachen sprechen aber für die Stabilität der Beziehungsstruktur und der Aussagekraft unseres Maßes. Sie sind auch ein letztes Argument für die Reliabilität unseres Maßes.

1.2.7. Zusammenfassung.

Die Maße der zweiten Messung geben Informationen über die Häufigkeit, die Güte, und die Häufigkeit und Güte der Erklärungen. Wegen relativ hoher Korrelationen brauchen wir zwischen Verständnis und Explikation und zwischen B- und C-Kategorisierung nur dann zu unterscheiden, wenn es die jeweilige Analyse erfordert. Als für die drei Informationen repräsentative Maße können wir GATEX, MIBEX und MIAEX ansehen.

Die Frage der Reliabilität unserer Maße muß von der Reliabilität der ihnen zugrundeliegenden Bewertung, die wir im vorangegangenen Abschnitt behandelt haben, getrennt werden. Für die Reliabilität

der Maße läßt sich zuerst die Höhe der Interkorrelationen der analogen Maße zwischen mündlicher und schriftlicher Messung angeben: Die Interkorrelationen der bewertenden Maße (MIBEX) sind fast so hoch wie die nur ein Minimum an Bewertung einschließenden Maße der Häufigkeit (GATEX). Zweitens entspricht die tatsächliche Beziehungsstruktur der Maße - wie sich aus den Interkorrelationen und der Faktorenanalyse entnehmen läßt - ziemlich genau den bei der Konstruktion verwandten Prinzipien. Drittens ergeben sich keine gravierenden Unterschiede der Maße zwischen der Unterschicht und der Mittelschicht - wie sich aus den schichtspezifischen Korrelationen der analogen Maße zwischen mündlicher und schriftlicher Messung und den schichtspezifischen Beziehungsstruktur der Gesamtheit der Maße entnehmen läßt.

2. Darstellung der Untersuchungsgruppe der zweiten Messung

Wie schon in Kapitel II. dargestellt, wurde durch ein Mißverständnis unter den Testern eine Nacherhebung erforderlich. Bei der folgenden ausführlichen Beschreibung der Versuchsanordnung werden wir zuerst die Untersuchungsgruppen des Experiments ohne Berücksichtigung möglicher Differenzen zwischen erst- und nacherhobenen Vpn. auf Unterschiede in relevanten Variablen hin überprüfen und mit der zugrundeliegenden Gesamtstichprobe vergleichen, um dann den möglichen Einflüssen der Nacherhebung auf die abhängigen Variablen gesondert nachzugehen.

2.1. Darstellung der Untersuchungsgruppen ohne Berücksichtigung des Erhebungszeitpunktes.

2.1.1. Experimentalstichproben.

Die Stichprobe des Experiments stellt eine durch die Schichtparallelisierung bedingte Auswahl von 83 Individuen aus der Gesamtstichprobe von 339 Kindern dar. Von den insgesamt 83 Vpn., die an der zweiten Messung des Wortschatzexperimentes beteiligt waren, konnten wiederum 66 in das endgültige "Matching-design" aufgenommen werden. Sie bilden die Experimentalstichprobe I, in der die beiden Schichten und die beiden Geschlechter ausgewogen sind und die aus Paaren des gleichen Geschlechts aus beiden Schichten besteht (Tab. 16). An dieser Gruppe wurden mit den verschiedenen Testverfahren alle Hypothesen überprüft. Die 83 insgesamt getesteten Vpn. bilden die Experimentalstichprobe II, an der zusätzlich einige Hypothesen überprüft wurden. Da die Experimentalstichprobe II nicht mehr aus ausgewogenen Gruppen parallelisierter Paare und nicht mehr aus gleich großen Gruppen mit der gleichen Testform besteht, konnten mit ihr keine Wilcoxon-Tests mehr, aber

noch einige Varianz- und Korrelationsanalysen durchgeführt werden. Die zusätzlichen Vpn. der Experimentalstichprobe II sind in Tab. 17 mit den Werten auf der Kontrollvariablen aufgeführt. Vp. 1300 wurde bei der ersten Erhebung schon einmal als Vp. 113 der Gruppe 4 getestet. In vielen Analysen wurde die zweite Testmessung (Vp. 1300) deshalb nicht berücksichtigt.

Zwei Irrtümer konnten erst nach Abschluß der Nacherhebung korrigiert werden. Erstens bekam Vp. 110, die zur Gruppe 1 (A gerade mündlich/ A ungerade schriftlich) gehört, im Mündlichen nicht die richtige Testform "A gerade", sondern "B ungerade". Da keine Vpn. mehr zur Verfügung stand, die Form "A gerade" mündlich beantwortet hatte, wurden die schriftlichen Antworten "A gerade" von Vp. 400 als mündlicher Testteil zu den schriftlichen Antworten von Vp. 110 hinzugenommen. Der Wechsel der Kommunikationsform stellt für die Bewertung der Explikations- und Verständnislänge, die ja die von der Darstellungsform relativ unabhängige logische Qualität der Antwort erfassen und von dem Stimulus-Material (vorgegebene Worte) abhängen, eine geringere Verzerrung dar, als sie der Wechsel der Wortgruppen bedeutet hätte. Für die Bewertung der rein linguistischen Form der Antwort jedoch (siehe 6.), wurden die mündlichen Antworten der Vp. 400 in der Form "A ungerade" hinzugezogen, da hier - umgekehrt - die Kommunikationsform stärker als das Stimulus-Material die Antwort bestimmt. Der Wert der beiden Vpn. auf der Kontrollvariablen wurde für die Übersicht der Versuchsgruppen gemittelt (siehe Tab. 16). In die Varianzanalysen der Kontrollvariablen mit den Untersuchungsgruppen als Effekten (Tab. 18-20) gingen nur die Werte von Vp. 110 ein; es wurde zur Kontrolle jedoch eine Korrelation ohne die "künstliche" Vp. 110/400 gerechnet (N=65) (siehe Tab. 14). Die Auswirkung dieses Irrtums konnte so für die jeweilige Analyse möglichst gering gehalten werden.

Zweitens ging Vp. 237 in die entsprechenden Varianzanalysen mit den Intelligenz- und Wortschatzwerten der Vp. 201 ein; da die Vpn. Geschwister waren, unterschieden sie sich nicht in den Sozialdaten und den Antworten der Eltern auf die Sprach- und Kommunikationsfragen. Die Differenz auf der Kontrollvariablen liegt bei 0.6 (237:3.9, 201:4.5). In der Übersicht (Tab. 16) ist der zutreffende Wert der Vp. 237 berücksichtigt. Innerhalb der Untersuchungsgruppe 1 sind die Differenzen der Mittelwerte für die Schichtgruppen, wenn (irrtümlich) der Wortschatztestwert von Vp. 201 eingeht - 0.028; wenn (richtig) der Wortschatztestwert von - 0.113 zugunsten der Mittelschicht. Die Auswirkungen des Irrtums sind also quantitativ nicht groß; sie haben Testbedingungen in den statistischen Analysen (mit Vp. 237) im Vergleich zu den Angaben in der Experimentalgruppenübersicht (mit Vp. 201) eher zu Ungunsten der Hypothesen verschärft.

Trotz der beschriebenen Mängel ergeben sich in der Experimentalstichprobe I keine Differenzen für die Schichtgruppen-Mittelwerte; die Geschlechtsgruppen unterscheiden sich lediglich in der 4. Gruppe nennenswert (1.050 zugunsten der Mädchen) (Tab. 16). Ebenso ergab die Varianzanalyse mit den Wortschatzwerten als Kriterium und den Experimentalgruppen als Faktoren keine signifikanten Unterschiede (Tab. 17). Die Experimentalstichprobe II zeigt natürlich stärkere Unterschiede zwischen den Zellenmittelwerten der Kontrollvariablen.¹⁾ In ihr werden wir die einzelnen Experimentalgruppen - auch wegen der ungleichen Häufigkeit - nicht mehr berücksichtigen. Für die Schicht- und Geschlechtsgruppen insgesamt zeigen sich jedoch auch in der Experimentalstichprobe II keine Unterschiede der Kontrollvariablen.

1)-----
Vp. 110 wurde in ihr zur Gruppe 2 gerechnet, obwohl sie ebenso der Gruppe 3 hätte zugerechnet werden können; Vp. 400 wurde der ursprünglichen, korrekten Gruppe 1 zugerechnet.

Nach der Vorstellung der zwei Experimentalstichproben soll Abschnitt 2.1. zwei weitere, eng miteinander verknüpfte Aufgaben erfüllen: die Überprüfung des Experimentalplans der zweiten Messung und die Beschreibung der Verteilung weiterer Variablen in den Experimentalgruppen, die für die Interpretation des Schichteinflusses wichtig sein oder als eigene Einflußfaktoren auf die Experimentalleistung überprüft werden können. Die Überprüfung des Experimentalplans bezieht sich vorwiegend auf die Stichprobe I. Hier soll zuerst überprüft werden, ob bei der Bildung von Experimentalgruppen der Schichteinfluß auf die Experimentalleistung wichtiger Variablen kontrolliert worden ist: auf den Wortschatztest (Tab. 18), aber auch auf weitere verbale und nichtverbale (Tab. 19, 20) Intelligenzwerte und Schulnoten. Als Prüfverfahren werden wir für die Stichprobe I Varianzanalysen mit den Schicht- und Geschlechtsgruppen des Experiments und für die Stichproben I und II Korrelationsanalysen mit den z-transformierten Werten des Filterindex verwenden, der ursprünglich ja 5 Stufen umfaßte (siehe II.2.). Zweitens soll untersucht werden, wieweit durch die Auswahl auf dem Wortschatztest gleicher Paare aus beiden Schichten sich die Korrelationen des Wortschatztests mit den übrigen Tests in den Experimentalgruppen I und II gegenüber der Gesamtstichprobe geändert haben.

Die Beschreibung weiterer, nicht in den Experimentalplan eingegangener Variablen wird sich zunächst auf die feinere Aufgliederung und die Dimensionen des Filterindex der sozialen Schicht konzentrieren. Ferner soll die Verteilung der Vpn. nach den Arbeitsplatzkategorien des Vaters und der Mutter und auf die verschiedenen Schulen und Schulklassen dargestellt werden. Da diese Faktoren innerhalb des Experimentalplans nicht ausgewogen sind, kann ihr Einfluß in der Stichprobe II ebensogut wie in der Stichprobe I untersucht werden. Deshalb werden wir die Verteilung beider Stichproben darstellen.

2.1.2. Kontrolle des Schichteinflusses.

Die Varianzanalyse der Kontrollvariablen WST mit den Schicht- und Geschlechtsgruppen des Experiments (Tab. 18) zeigt, daß für keinen der untersuchten Effekte sich auch nur ein F-Wert über 1 ergibt. Die Korrelationen zwischen dem z-transformierten Filterindex (IFILT) und dem Wortschatztest sind von .30 ($p < .01$) in der ursprünglichen auf .02 bzw. .04 in der Experimentalstichprobe gesunken (Tab. 21). Die völlige Ausschaltung des Schichteinflusses auf den WST durch die Parallelisierung ist das wichtigste Ergebnis des Abschnitts 2.1.

Außer dem Wortschatztest haben wir noch 6 weitere Variable verbaler Leistungen überprüft: die verbale Gesamtintelligenz des P-S-B, die drei verbalen Subtests des P-S-B und die Schulnoten für Deutsch mündlich und Deutsch schriftlich. Bei den Varianzanalysen mit den Experimentalgruppen (Tab. 19) ergaben sich hier zwei signifikante Schichteffekte für Test 5 (Worteinfall) und Deutsch schriftlich zugunsten der Mittelschicht. Bei den Korrelationen (Tab. 21) finden sich signifikante Korrelationen mit dem Filterindex in der Experimentalgruppe wiederum für Test 5 und Deutsch schriftlich und zusätzlich für Deutsch mündlich. Während die Korrelationen des Filterindex mit der nichtverbalen Intelligenz und Test 1/2 (Wortkenntnis) und Test 6 ("verstümmelte Wörter") durch die Auswahl der Experimentalgruppe verschwunden sind, bleiben die Korrelationen des Filterindex mit den Deutsch-Noten im Experimentalsample bestehen, die Korrelation mit Test 5 steigt sogar noch an. Diese Ergebnisse müssen wohl zufällig sein, von den Korrelationen des Wortschatztests mit den genannten Test- und Schulleistungen her (Tab. 21) sind sie nicht verständlich. Wenn die bessere Leistung der Mittelschichtkinder auf Test 5 und den Deutsch-Noten überhaupt einen Einfluß auf die Leistung in der zweiten Messung des Experiments hat, so zugunsten unserer Hypothesen; um so erstaunlicher ist, daß sich keine Schichtunterschiede in der aktiven Worterklärung gefunden haben (siehe 5.5.). Der signifikante Experimentalgruppeneffekt bei Test 6 ist zu vereinzelt, um mehr als

zufällige Bedeutung zu haben. Insgesamt ist durch die Vpn.-Auswahl der Schichteinfluß auf die Wortschatzleistung vollständig und auf die nicht kontrollierten verbalen Testleistungen weitgehend ausgeschaltet worden.

Der Schichteinfluß auf die übrigen Testleistungen und Schulnoten konnte durch die Kontrolle des Wortschatztests desto weniger eliminiert werden, je unabhängiger diese Variablen von der Wortschatzleistung sind. Trotzdem ergaben sich signifikante Schichteffekte nur für Test 9 (Rechenfähigkeit) und der Schulnote Rechnen, bei der Note zugunsten der Mittelschicht, dem Subtest der Unterschicht (Tab. 20). Dieses Ergebnis wird durch die Korrelationen bestätigt (Tab. 21), zusätzlich findet sich eine signifikante positive Korrelation der Note in Heimatkunde mit der Schichtzugehörigkeit. Alle diese Korrelationen bestanden in der Gesamtgruppe jedoch in ähnlicher Stärke, während die Korrelationen aller übrigen Noten mit der Schichtzugehörigkeit in der Experimentalgruppe nicht mehr auftauchen. Die signifikanten Geschlechtseffekte zugunsten der Mädchen bei den Schulnoten für Fleiß, Aufmerksamkeit, Religion (Tab. 20) und Deutsch mündlich (Tab. 18) zeigen, daß die allgemein bessere Benotung der Mädchen (siehe Bosse u.a. 1971) auch in unserer Experimentalstichprobe wiedererscheint.

2.1.3. Beziehung der Kontrollvariablen WST zu anderen Variablen.

In der Gesamtgruppe fanden sich hohe Korrelationen zwischen dem Wortschatztest und allen Intelligenztests (außer Test 10), allen Schulnoten und allen Schichtindikatoren (Tab. 21). In den Experimentalstichproben I und II finden sich signifikante Korrelationen nur noch zwischen den verbalen Intelligenztests und dem Wortschatztest; die Korrelationen des Wortschatztests mit der Gesamtintelligenz bleiben zwar signifikant, sinken aber stark. Die Korrelationen des Wortschatztests mit den Schulnoten bleiben bestehen, für die Deutsch-Noten steigen sie sogar noch. Die Korrelationen des Wortschatztests mit den Schichtindikatoren sind völlig eliminiert. Die schon beschriebene Ausschaltung des Einflusses des

Schicht-Filterindex (KFIL3) gilt also für alle Dimensionen in der gleichen Weise. Durch die Parallelisierung sind - wie beabsichtigt - die Korrelationen des Wortschatztests mit den nichtverbalen Intelligenzwerten reduziert worden. Die Interkorrelationen der vom Wortschatztest abhängigen (bzw. unabhängig gewordenen) Variablen müssen sich in der Experimentalstichprobe entsprechend geändert haben.

Die Interkorrelationen der Intelligenzgesamtmaße (Tab. 22, Feld AA) haben sich in der Experimentalstichprobe I verändert. Die Korrelation der verbalen Intelligenz mit der Gesamtintelligenz sinkt und die mit der nichtverbalen Intelligenz verschwindet fast vollständig. Für die verbale Intelligenz gilt also dasselbe wie für den Wortschatz: eine Auflösung des Zusammenhangs zwischen verbalen Testleistungen und der nichtverbalen Intelligenz. Die Experimentalgruppe stellt eine Auswahl dar, in der der Zusammenhalt der verbalen mit den nichtverbalen Intelligenzmaßen gelockert ist, während letztere unter sich in gleich starker Beziehung wie in der Gesamtstichprobe bleiben. Die Vpn. sind in verbalen und nichtverbalen Bereichen nicht konsistent ausgewählt worden. Die ursprüngliche Absicht, auch die nichtverbale Intelligenz zu parallelisieren, findet nachträglich Rechtfertigung und der durch den Irrtum entstandene Verlust wird ersichtlich.¹⁾

Die Korrelationen der verschiedenen Tests und Schulnoten mit der Kontrollvariablen (Tab. 21) sowie deren Interkorrelationen (Tab. 22) sind wichtig für die Interpretation der Korrelationen und der multiplen Regressionen der Tests und Schulnoten mit den Experimentalleistungen (Abschnitt 5.8.). Wenn Tests oder Noten mit den

1)-----
In der dritten Messung, in der die nichtverbale Intelligenz zusätzlich kontrolliert wurde, fällt die Korrelation des Wortschatztests mit der nichtverbalen Intelligenz in der Experimentalgegenüber der Gesamtstichprobe zwar ebenfalls ab. Die Interkorrelationen zwischen Wortschatztest und nichtverbaler Intelligenz sind aber höher als in der zweiten Messung (siehe IV.2.).

Experimentalleistungen korrelieren, kann an den Korrelationen der Tests oder Noten mit der Kontrollvariablen die Eigenständigkeit eines solchen Zusammenhangs eingeschätzt werden. An den Interkorrelationen der Tests und Noten, sowie der Schichtindikatoren, läßt sich das für die multiple Regressionsanalyse typische Problem der Multiikollinearität der Prädiktoren beurteilen. Auffällig ist, daß die Interkorrelationen der Schulnoten durchweg über den Interkorrelationen der Intelligenzsubtests liegen.

2.1.4. Verteilung der Vpn. auf eine differenzierte Einteilung und innerhalb der Dimensionen des Filterindex.

Die Verteilung der Vpn. auf 5 Schichten gibt Aufschluß über den Schwerpunkt der sozialen Herkunft der Vpn., damit über den sozialen Abstand zwischen den dichotomen Vergleichsgruppen, und über die durch die Parallelisierung bedingte Auswahl der relativ schlechteren Vpn. im oberen und der relativ besseren im unteren Schichtbereich. Wir betrachten zunächst die Anzahlen der Vpn. in den Schichtgruppen und dann die Gruppenmittelwerte auf der Kontrollvariablen.

Daß die "Untere Mittelschicht" nicht in das Experimentalsample aufgenommen wurde (siehe II.2.), wird in Tab. 24 erkenntlich. Aus den übrigen vier Schichtgruppen standen nur diejenigen Vpn. für die Auswahl zur Verfügung, deren Eltern den Fragebogen beantwortet hatten (N=292). Innerhalb der beiden unteren Straten wurde jeweils ungefähr ein Viertel der Vpn. in die Experimentalstichproben I und II aufgenommen, innerhalb der beiden oberen Straten jedoch relativ mehr Vpn. aus der "Oberen Mittelschicht". Absolut sind die beiden mittleren Straten am stärksten, die beiden extremen geringer vertreten.

Bezüglich der Gruppenmittelwerte auf der Kontrollvariablen zeigt sich in beiden Experimentalstichproben, daß die durch die Parallelisierung notwendige positive Auswahl in den unteren Straten allein für die "Obere Unterschicht" gilt, während die ebenso notwendige

negative Auswahl in den oberen Straten in beiden Schichten nahezu gleichmäßig ist. Dies gilt im wesentlichen auch in den beiden Geschlechtsgruppen. Bei den Jungen konzentriert sich die negative Auswahl in den oberen Straten auf die "Mittlere Mittelschicht", bei den Mädchen auf die "Obere Mittelschicht"; in den unteren Straten müssen die unterschiedlichen Mittelwerte der Geschlechtsgruppen wegen der geringen Anzahlen innerhalb der "Unteren Unterschicht" nicht mehr beachtet werden. Für die gesamte "Untere Unterschicht" bleibt aber die Frage, warum entgegen der Tendenz der Parallelisierung negativ ausgelesen wurde. Zur Erklärung muß die Varianz der Wortschatzwerte für die 5 Schichten in der Gesamtstichprobe herangezogen werden (siehe die Standardabweichung in Tab. 37). Sie ist in den beiden unteren Straten wesentlich kleiner als in den beiden oberen¹⁾; zudem sind die Mittelwertunterschiede zwischen den beiden unteren Straten größer als zwischen den beiden oberen und steigen die Mittelwerte bis zur Mittleren Mittelschicht an, um dann wieder leicht abzufallen (Tab. 24). Unter diesen Bedingungen mußten wohl für die untere dichotome Schicht die besseren Vpn. aus der Oberen Unterschicht, die weniger guten aus der Unteren Unterschicht gewählt werden, während für die obere dichotome Schicht die Wahl zwischen Mittlerer und Oberer Mittelschicht relativ frei blieb.

Die Tatsache, daß durch die Parallelisierung in der Unteren Unterschicht nicht positiv ausgelesen wurde, wird wichtig für die Interpretation von Schichtunterschieden nach KFIL3 bei einigen Maßen der zweiten Messung (siehe 5.6.1.). Um sie nicht fälschlich Schichtunterschieden der Kontrollvariable innerhalb der beiden dichotomen Schichten statt Unterschieden auf der Experimentalleistung zuzuschreiben, haben wir Varianzanalysen der Kontrollvariablen innerhalb

1)-----
Der Quotient der größten durch die kleinste Varianz beträgt 1.998. Es handelt sich um 5 Varianzen, die größte Gruppe hat ein N von 155. Für 5 und 60 Freiheitsgrade beträgt der F(max)-Wert auf dem 5%-Niveau 2.04, auf dem 1%-Niveau 2.4. Es handelt sich also nicht um homogenen Varianzen.

der Stichproben I und II mit KFIL3 als Faktor gerechnet (Tab. 25). Es ergaben sich jedoch keine signifikanten F-Werte.

Für den Filterindex wurden vier Dimensionen der sozialen Schicht verwertet: der Beruf des Vaters (Tab. 26), die Ausbildung des Vaters (Tab. 27), die Ausbildung der Mutter (Tab. 27) und das Familiennettoeinkommen (Tab. 28). Zum Verständnis der späteren Korrelationsanalysen (siehe 5.6.) sind die innerhalb der Gesamtstichprobe z-transformierten Werte angegeben. Da die Untere Mittelschicht nicht in die Experimentalstichprobe einbezogen wurde, muß die Verteilung der Experimentalstichprobe für alle diese Dimensionen im mittleren Bereich etwas verzerrt sein. Nur besondere Abweichungen sollen kurz erwähnt werden. Von den Berufen der Väter sind in beiden Experimentalstichproben kleine Selbständige (Gruppe 5) unter- und mittlere Selbständige (Gruppe 7) überrepräsentiert. Von den Bildungsabschlüssen sind in der Experimentalstichprobe II sowohl für die Väter wie die Mütter Hochschulabschlüsse (Gruppe 9) überrepräsentiert. Beim Familiennettoeinkommen zeigt sich eine weitgehende Übereinstimmung zwischen den Stichproben. Das Gewicht der einzelnen Dimensionen für die Schichtzuweisung geht aus ihren Korrelationen mit dem Filterindex hervor (Tab. 21). In allen Stichproben ist der Beruf des Vaters die wichtigste Dimension. In beiden Experimentalstichproben hat sich die Rangfolge der Bedeutung der Dimensionen getreu erhalten. Selbst die Struktur der Interkorrelationen zwischen den Dimensionen (Tab. 22, Feld DD), die sowieso nicht sehr differenziert ist, wurde in der Experimentalstichprobe I vollständig bewahrt. Dies verdient festgehalten zu werden, weil bei der Parallelisierung nach dem schichtabhängigen Merkmal Wortschatz aus dem oberen Bereich des Schichtkontinuums relativ schlechtere, aus dem unteren relativ bessere ausgewählt werden mußten. Die Eliminierung des Zusammenhangs zwischen dem Wortschatztest und dem gesamten Filterindex wie seiner Dimensionen blieb also ohne Folge für die Beziehung der Dimensionen untereinander.

2.1.5. Verteilung der Vpn. nach den Arbeitsplatzkategorien für die Mutter und für den Vater.

Im Fragebogen wurden beide Eltern um eine möglichst detaillierte Beschreibung ihrer Tätigkeit am Arbeitsplatz gebeten. Die Antworten auf diese beiden offenen Fragen wurden zunächst in 31 qualitativen Kategorien verkodet, die auf drei Ebenen variierten: Art der Tätigkeit, Qualifikation der Tätigkeit, Wirtschaftssektor, dem der Arbeitsplatz zugehört. Die größte Einteilung bezog sich auf die "Art der Tätigkeit": manuelle oder nicht-manuelle Tätigkeit. Die entscheidende Dimension der Qualifikation reichte von Hilfsarbeiten bis zu hochqualifizierten Tätigkeiten. Unter "Wirtschaftssektor" wurde zwischen Industrie, Dienstleistungsgewerbe, Handel, Handwerk, Handelsgewerbe, Ausbildungssektor, Führungstätigkeit, Lebensfürsorge usw. unterschieden; diese Dimension geht in Beschreibungen der "Substanz der Arbeit" (Kohn 1969) über.

Für die Kategorisierung des Arbeitsplatzes des Vaters wurden diese ursprünglichen 31 Kategorien nach soziologischen und Verteilungsgesichtspunkten zu 14 Kategorien zusammengefaßt (siehe Tab. 29). Die auch die Schichtdichotomie bestimmende Trennung zwischen manueller und nicht manueller Arbeit liegt bei den 14 Kategorien zwischen den Kategorien 6 und 7; hinter der Einteilung der Arbeitsplätze muß daher zugleich ein Schichtunterschied stehen¹⁾: In unserer Experimentalstichprobe I (N=63) gehören alle Väter von Vpn. in den Kategorien 1-5 zu dichotomen Unterschicht, von den 6 Vätern der Gruppe 6²⁾ gehören jeweils die Hälfte zur Unter- bzw. zur Mittelschicht und alle Väter von Vpn. in den Kategorien 7 bis 14 gehören zur dichotomen Mittelschicht.³⁾ Mit dem Arbeitsplatz des

1) -----
In unserer Experimentalgruppe wird diese Tendenz im Vergleich zur Gesamtgruppe noch verstärkt (siehe Bosse u.a. 1971:264); das erklärt sich aus dem Fortfallen der am breitesten über die Arbeitsplatzkategorien des Vaters streuenden UMS.

2) Entscheidend war innerhalb dieser Gruppe der Berufsstand: die 3 Väter dieser Gruppe, die zur Unterschicht gezählt wurden, waren Arbeiter, die 3 Väter, die zur Mittelschicht gezählt wurden Angestellte und Beamte.

3) Ebenso gehören in der Experimentalstichprobe II alle Väter der Kategorien 1-5 zur Unter-, alle Väter der Kategorien 7-14 zur Mittelschicht.

Vaters zusammenhängende Unterschiede zwischen den Vpn. in der aktiven Worterklärung, müssen deshalb zuerst innerhalb der dichotomen Schichten analysiert werden.

Gewisse Zusammenhänge bestehen ferner zwischen der Schicht, dem Beruf und dem Arbeitsplatz des Vaters. In der Experimentalstichprobe I sind von den 33 Vätern der Unterschicht 9 Facharbeiter, die zugleich eine "manuell spezialisierte Tätigkeit in der Industrieproduktion" (Kategorie 08 der 31 Kategorien) ausüben. Innerhalb der 33 Väter aus der Mittelschicht ist die Verbindung von bestimmten Arbeitsplatz-Kategorien und dem Beruf weniger fest; dennoch sind 10 Väter mittlere, gehobene oder leitende Angestellte, die qualifizierte oder hochqualifizierte Tätigkeit in Bürokratien mit flexiblen oder fixen Karrieremustern ausüben (Kategorie 19, 20, 22 und 23 der 31 Kategorien).

Betrachtet man die Werte der Vpn. der beiden Experimentalstichproben auf der Kontrollvariablen (siehe auch Abb. 12) in den nach dem Arbeitsplatz des Vaters gebildeten Gruppen, so zeigen sich innerhalb der 5 der Unterschicht zuzurechnenden Gruppen kaum Unterschiede. In der beiden Schichten zugehörenden Gruppe 6 zeigen die Vpn. leicht bessere Werte, ohne daß dabei innerhalb der Gruppe zwischen beiden Schichten ein Unterschied bestünde. Innerhalb der der Mittelschicht zuzurechnenden Gruppen zeigen sich jedoch größere Unterschiede. Während die Vpn. in beiden Experimentalstichproben, deren Väter hochqualifizierte Tätigkeiten in der Technologie oder Bürokratien mit fixem Karrieremuster ausführen, deutlich bessere Leistungen auf der Kontrollvariablen zeigen, fallen die Vpn., deren Väter Arbeitsplätze der Kategorien 12, 13, 14 innehaben, stark ab. Im Vergleich zur Gesamtstichprobe mußten in den Experimentalsamples innerhalb der Arbeitsplätze der Mittelschicht aus Gründen der Parallelisierung insgesamt relativ schlechtere Vpn. ausgewählt werden. Die Gegenüberstellung der Gruppen 10 und 11 mit den Gruppen 12, 13, 14 und weiterhin auch 7, 8, 9 spiegelt jedoch nicht die Situation der Gesamtstichprobe wider, sondern ist erst durch die Auswahl der Experimentalsample entstanden.

Für die Kategorisierung des Arbeitsplatzes der Mutter wurden dieselben, ursprünglich 31, qualitativen Kategorien zugrundegelegt, mit der Ausnahme, daß "Verkäuferin" wegen der Häufigkeit als besondere, 32., Kategorie gebildet wurde. Die Zusammenfassung der Kategorien richtet sich jedoch nach anderen soziologischen und anderen Verteilungsgesichtspunkten als bei der Kategorisierung des Arbeitsplatzes des Vaters. An erster Stelle hat die berufstätige Frau in vielen Fällen den Status des nicht lebenslangen "Zweitverdieners", der aus materiellen oder psychologischen Motiven, die Berufstätigkeit aufnimmt, unterbricht oder wiederaufnimmt. In der Gesamt- wie in den beiden Experimentalstichproben sind die meisten Mütter früher berufstätig gewesen, gegenwärtig aber nicht mehr berufstätig (Tab. 30); ebenfalls in allen Stichproben ist etwa ein Zehntel aller Mütter nie berufstätig gewesen. Nur etwa 10% (Kategorien 3 und 4 von Tab. 30) sind auf der anderen Seite früher wie heute berufstätig gewesen, haben also einen den Vätern vergleichbaren Erwerbsstatus, was eine Vergleichbarkeit der "Substanz der Arbeit" und eine Vergleichbarkeit der familiendynamischen Bedeutung der Erwerbstätigkeit noch keineswegs einschließt. In der Klassifizierung des Arbeitsplatzes der Mutter (Tab. 31) wurde daher der Erwerbsstatus in der Dichotomie "Gegenwärtig berufstätig - gegenwärtig nicht berufstätig" berücksichtigt, und die Gruppe der Mütter, die keine Angabe oder die Angabe "nie berufstätig" lieferten, gesondert aufgeführt.

An zweiter Stelle sind die vorwiegenden Berufstätigkeiten der Mütter nicht mit denen der Väter vergleichbar. Während, in unserer Experimentalstichprobe, innerhalb des manuellen Bereichs die Väter, wie gezeigt, vorwiegend als Facharbeiter in spezialisierter Industriearbeit tätig sind, leisten die Frauen vorwiegend Hilfs- und Routinearbeiten in der Industrie oder im Dienstleistungsgewerbe. Umgekehrt sind, innerhalb des nichtmanuellen Bereichs, die Mütter unserer Stichproben seltener in qualifizierten und hochqualifizierten Tätigkeiten in Bürokratien mit fixen Karrieremustern, der Technik und der Kultur zu finden; Ausnahmen bilden hier nur

"qualifizierte Tätigkeiten in Bürokratien mit flexiblen Karriere-
mustern" (Kategorie 8 in Tab. 31, entspricht Kategorie 9 in Tab.
29) und "qualifizierte technische Kontrolltätigkeiten oder Tätig-
keiten in der Lebensfürsorge" (Kategorie 9 in Tab. 31). Als Kon-
sequenz der beiden Tendenzen ergibt sich, daß die zusammenfassende
Klassifikation der Arbeitsplätze für die Mütter im "unteren" Be-
reich differenzierter, im "oberen" Bereich weniger differenziert
sein mußte als die entsprechende Klassifikation für die Väter;
insgesamt ergeben sich für die Mütter weniger Kategorien, nur zwei
sind mit den Kategorien für die Väter identisch.

Betrachten wir die Verteilung der Vpn. nach der Berufstätigkeit
und der Art des Arbeitsplatzes der Mutter: Wenn die Mütter ohne
Angabe zur Berufstätigkeit (Kategorie 1 und 6 aus Tab. 30) nicht
eingeschlossen werden, so ist in unseren beiden Experimentalstich-
proben (Tab. 31) nur ein Fünftel der Mütter gegenwärtig berufstä-
tig, unabhängig von Angaben über Dauer und Art der früheren Berufs-
tätigkeit (Kategorien 3, 4, 5, 7 aus Tab. 30), während vier Fünftel
aus der bezogen auf den Erwerbstatus homogenen Gruppe der gegenwär-
tig nicht, aber früher berufstätigen Mütter stammen (Kategorie 2
aus Tab. 30). Auffällig ist, daß in unseren Stichproben die gegen-
wärtig berufstätigen Mütter entweder einfach manuelle Hilfs- und
Routinetätigkeiten (Kategorie 1 und 2, Tab. 31) oder qualifizierte
und hochqualifizierte Tätigkeiten (Kategorien 8, 9 und 10) ausfül-
len und die mittleren Kategorien so gut wie nicht besetzt sind; für
die Häufigkeit früherer Berufstätigkeit gilt ähnliches. Offensicht-
lich handelt es sich um jeweils schichttypische Arbeitsplätze mit
schichttypischen Gründen der Berufstätigkeit: materiellen Zwängen
einerseits und individuellen Motiven der Lebensbereicherung anderer-
seits.

Auch die Kategorien des Arbeitsplatzes der Mutter hängen in unseren
Experimentalstichproben mit der Sozialschicht der Familie zusammen,
sowie mit der Ausbildung der Mutter. Der Zusammenhang mit der
Sozialschicht der Familie ist sehr stark, zeigt jedoch nicht die-
selbe Eindeutigkeit wie bei den Arbeitsplätzen der Väter. Während

von den 27 Müttern der Experimentalstichprobe I, die Arbeiten der Kategorien 1 bis 3 ausüben oder ausübten, 21 zu Familien der dichotomen Unterschicht gehören, kommen 24 von den 28 Müttern der übrigen Arbeitsplatzkategorien aus der dichotomen Mittelschicht. Ähnlich eng ist der Zusammenhang zwischen Ausbildung und Arbeitsplatz der Mutter. Von den 24 Müttern, die Arbeiten der Kategorien 1 und 2 ausüben, haben 18 nur einen Volksschulabschluß ohne Lehre, während die Mütter aller übrigen Kategorien wenigstens eine Lehre abgeschlossen haben. Der Einfluß der beruflichen Arbeit der Mutter auf die Leistungen des Kindes muß also stark mit dem der Schichtzugehörigkeit der Familie und der Ausbildung der Mutter identifiziert werden; auf der anderen Seite bietet die Kategorisierung der Arbeitsplätze eine Möglichkeit differenzierter Betrachtung der beiden ersten Einflußfaktoren.

Betrachtet man die Werte der Vpn. der beiden Experimentalstichproben auf der Kontrollvariablen in den nach dem Arbeitsplatz der Mutter gebildeten Gruppen, so zeigen sich keine großen Unterschiede; lediglich die Kinder von Verkäuferinnen (Kategorie 5), die in unser Experimentalsample gekommen sind, zeigen besonders schlechte, die eine Vp., deren Mutter früher eine qualifizierte Tätigkeit in der Bürokratie mit fixem Karrieremuster ausgeübt hat, besonders gute Leistungen im standardisierten Wortschatztest. Zwischen den Kindern gegenwärtig berufstätiger, und gegenwärtig nicht berufstätiger Mütter jedoch besteht eine beachtliche Differenz auf der Kontrollvariablen, insgesamt wie in allen einzelnen Arbeitsplatzkategorien (mit einer Ausnahme).

2.1.6. Verteilung der Vpn. auf Schulen und Schulklassen.

Die Schulen und Schulklassen repräsentieren die soziale Zusammensetzung ihres Einzugsgebietes und besondere Didaktiken. Mit der Verteilung der Vpn. auf Schulen und Schulklassen (Tab. 31a) soll geprüft werden, wieweit diese beiden Einflüsse bei der Bildung von Experimentalgruppen nach Schicht und Geschlecht in unterschiedlicher Weise, vermittelt über die Anzahl und Leistungshöhe der ausgewählten Vpn., durchgesetzt haben. Für "Geschlecht" lassen sich

weder in der Anzahl der Vpn. noch in den Mittelwerten der Kontrollvariablen Unterschiede zwischen den Schulen feststellen. Für "Schicht" läßt sich der Einfluß des Schuleinzugsgebiets noch an dem großen Anteil von Unterschichtkindern aus der Schule Nord und dem entsprechenden großen von Mittelschichtkindern aus der Schule Mitte ablesen. Die für diese beiden Schulen spezifischen Schichtdifferenzen auf der Kontrollvariablen sind ebenfalls durch die Parallelisierung bedingt. Für die Schulen insgesamt unterscheiden sich die Wortschatztest-z-Werte kaum. -

Die Untersuchungsgruppe ist aus zwei Gründen so ausführlich charakterisiert worden. Einmal sollten Ähnlichkeiten und Unterschiede zur Gesamtstichprobe belegt, dann Differenzen zwischen den Schichtgruppen für eine Interpretation der Ergebnisse möglichst genau herausgearbeitet werden. Die Beschreibung beschränkte sich auf die sozialen Ausgangsdaten, die ja von möglichen Einflüssen der experimentellen Erhebung unabhängig sind. Umgekehrt muß die folgende Prüfung möglicher Einflüsse des Erhebungszeitpunktes sich nicht auf die Ausgangsdaten, sondern die Ergebnisse des Experiments beziehen; insofern fällt sie schon unter die "Darstellung der Ergebnisse".

2.2. Einflüsse der Nacherhebung.

Durch die Nacherhebung ließen sich unterschiedliche Einflüsse der Testsituation in der Schule und im Elternhaus nicht vermeiden. Aber die Möglichkeit solcher Einflüsse kann durch eine zwischen den Experimentalgruppen unterschiedliche Verteilung der nacherhobenen Vpn. und durch unterschiedliche Werte der in der Schule und der im Elternhaus getesteten Vpn. auf der Kontrollvariablen unterschiedlich vorgefiltert worden sein. Bevor wir deshalb die inhaltliche Wirkung des Erhebungszeitpunkts und -orts prüfen können, müssen wir die Verteilung der nacherhobenen Vpn. auf die Experimentalgruppen darstellen.

Für alle Experimentalgruppen insgesamt (Tab. 32) sind die nacherhobenen und endgültig in den Experimentalplan eingegangenen ("gematchten") Vpn. gleichmäßig auf die Schicht- und Geschlechtsgruppen verteilt. Das gleiche gilt für die 3. und 4. Experimentalgruppe, während die 2. überhaupt keine nacherhobenen Vpn. enthält. Lediglich in der 1. Experimentalgruppe sind die beiden nacherhobenen Vpn. (Vp. 720 und 301, siehe Tab. 16) Jungen aus der Mittelschicht. Die inhaltlichen Wirkungen der Nacherhebung können sich also nur in dieser Gruppe unterschiedlich ausgewirkt haben. In jedem Fall müssen wir neben der generellen Prüfung des Effekts des Erhebungszeitpunkts die möglichen unterschiedlichen Leistungen der Vpn. auch innerhalb der 4 Experimentalgruppen verfolgen.

Die mündlichen Tests wurden im Elternhaus wie in der Schule individuell für jede Vp. mit der den Sprecher akustisch isolierenden Gegensprechanlage, d.h. unter erzwungener Konzentration der Vpn. durchgeführt. Die schriftlichen Tests wurden auch in der Schule nicht als Gruppentests, sondern individuell durchgeführt. Die Anwesenheit von Eltern, Geschwistern oder peers konnte in der Nacherhebung vermieden werden. Von der Technik der Tests also weitgehend gleich, unterschieden sich die beiden Erhebungen vermutlich vor allem durch die sozialen Situationen im Elternhaus bzw. in der Schule. Art und Richtung dieses Unterschieds sind jedoch schwer angebbbar. Die Ungewohnheit eines Tests kann durch die Vertrautheit

des Elternhauses relativiert worden sein, ebenso aber diese Vertrautheit zerstört haben. Der Leistungsanspruch der Kinder kann durch das Elternhaus gesenkt, ebenso aber angespannt worden sein, usf. Da wir keinen theoretisch faßbaren Einfluß eindeutig mit der Nacherhebung verbinden können, läßt sich nur sagen, daß sie für jedes Maß zusätzliche Fehlervarianz auf Kosten der durch die Faktoren des Experimentalplans produzierten Varianz produziert, die Maße in der gesamten Gruppe weniger verläßlich gemacht hat.

Im folgenden werden wir zuerst mit Hilfe der Varianzanalyse Maße mit signifikanten Effekten des Erhebungszeitpunkts suchen, d.h. Maße, bei denen die Nacherhebung die Reliabilität der gesamten Messung nennenswert beeinträchtigt hat; zweitens die Unterschiede zwischen den Erhebungen in den verschiedenen Zellen des Experiments verfolgen; und drittens diese Unterschiede mit Unterschieden auf der Kontrollvariablen vergleichen, so daß die unterschiedlichen Ergebnisse eventuell auf unterschiedliche Ausgangsbedingungen außerhalb der Testsituation zurückgeführt werden können. Auf diese Weise werden wir zuerst die Maße der Bedeutungserklärung, dann die Maße der sprachlichen Mittel darstellen.

In Tab. 33 sind die Ergebnisse einer Einwegvarianzanalyse mit dem Erhebungszeitpunkt als Faktor zusammengefaßt. Bei den Maßen der Bedeutungserklärung zeigen sich signifikante Effekte des Erhebungszeitpunkts unter dem 10%-Niveau bei S33VE im Mündlichen sowie bei S33VE, S13EX, S33EX, RATEX im Schriftlichen, und ein unter dem 5%-Niveau signifikanter Effekt bei MIAEX im Schriftlichen. Wenn der Erhebungszeitpunkt also auf die Bedeutungserklärung einen Einfluß hatte, dann vorwiegend auf die schriftliche Messung und ausschließlich auf die Maße der A-Kategorisierung, die Anzahl und Qualität der Antworten messen; bei den Maßen der B- und C-Kategorisierung ergeben sich in den meisten Fällen nicht einmal F-Werte über 1. Auf den Maßen der A-Kategorisierung sind die im Elternhaus getesteten Kinder den in der Schule getesteten in der mündlichen wie der schriftlichen Messung überlegen. Auf den Maßen der B- und C-Kategorisierungen ist, wenn die Unterschiede auch minimal sind, die mündliche Leistung der im Elternhaus, die schriftliche Leistung der

in der Schule getesteten Kinder besser. Wir können also zusammenfassen, daß durch die Nacherhebung die absolute Anzahl der schriftlichen Antworten gesteigert, alle übrigen Leistungen der aktiven Worterklärung nicht berührt wurden.

Daß die Erhebung im Elternhaus nur die absolute Anzahl der Antworten, nicht aber ihre Qualität gesteigert hat, ist nachträglich gut verständlich. Offensichtlich steigert das Elternhaus die Motivation der Kinder und kann die Antwortbereitschaft viel eher als die tatsächliche Wortkenntnis durch Situationsunterschiede beeinflusst werden. Warum aber gilt dies nur für die schriftliche Messung, obwohl man doch eine leichtere Beeinflußbarkeit der mündlichen Leistungen erwarten würde? Hier scheint doch die Testtechnik eine zusätzliche Wirkung gehabt zu haben. Während die mündliche Testbedingung durch die Gegensprechanlage Einflüsse der Situation überhaupt nicht sich durchsetzen ließ, konnten Einflüsse der Elternhaus-Situation im Schriftlichen zur Geltung kommen.

Lassen sich ähnliche Effekte des Erhebungszeitpunkts auch innerhalb der Schicht-, der Experimental- und noch weiter aufgeächerten Gruppen des Experimentalplans aufdecken? Betrachtet man zunächst die beiden Schichtgruppen einzeln (Tab. 33a), so finden wir bei den Maßen der A-Kategorisierung (MIAEX und RATEX) in jeder Schicht die gleichen Unterschiede wie in der Gesamtgruppe, bei den Maßen der B- und C-Kategorisierung (MIBEX als Beispiel) jedoch bessere Leistungen der Unterschicht bei der Erhebung in der Schule, bessere Leistungen der Mittelschicht bei der Erhebung im Elternhaus. Die besseren Leistungen der Unterschicht in der Schule sind besonders ausgeprägt in der mündlichen, die besseren Leistungen der Mittelschicht im Elternhaus besonders ausgeprägt in der schriftlichen Messung, so daß die zuvor für die Gesamtgruppe dargestellten Ergebnisse der Maße der B- und C-Kategorisierung - im Mündlichen bessere Leistungen in der Schule, im schriftlichen Bessere Leistungen im Elternhaus - auf die Kinder der Unter- bzw. der Mittelschicht

entscheidend zurückgeführt werden müssen. Wir finden also bei den Maßen der B- und C-Kategorisierung einfache Interaktionen von Erhebungszeitpunkt und Schicht sowie von Kommunikationsform und Schicht sowie eine doppelte Interaktion der Effekte Erhebungszeitpunkt, Kommunikationsform und Schicht. Offensichtlich werden die Unterschichtkinder durch die schulische Situation besonders herausgefordert. Daß die Herausforderung der Unterschichtkinder stärker in der mündlichen Messung wirksam ist, hätten wir nicht vorausgesagt; vielmehr hätten wir angenommen, daß durch das Schreiben als dem schulnahen und konzentrationsfördernden Modus die Kinder der Unterschicht stärker angespornt würden. Wir haben jedoch schon zuvor gesehen, daß - durch die Gegensprechanlage - gerade die mündliche Messung die Leistungen erzwingt, die wir der schriftlichen Messung zugeschrieben haben; in eine solche Interpretation paßt, daß die Unterschichtkinder durch die mündliche Messung in der Schule besonders herausgefordert werden.

Betrachtet man die 4 Experimentalgruppen einzeln (Tab. 34), zeigen die Maße mit signifikantem Effekt der Nacherhebung (Zweier- und Dreier-Kategorisierung und vor allem MIAEX und RATEX) folgende Unterschiede zwischen den beiden Erhebungen: in der Gruppe 1 stark zuungunsten, der Gruppe 4 stark und der Gruppe 3 leicht zugunsten der nacherhöbten Kinder.¹⁾ Vergleicht man diese Unterschiede mit den Unterschieden der beiden Erhebungsgruppen auf der Kontrollvariablen (Tab. 36), so zeigt sich, daß in der Gruppe 1 die in der Schule, in der Gruppe 4 die im Elternhaus getesteten Kinder überlegen sind. Die beiden Experimentalgruppen mit den größten Differenzen auf der Testleistung zeigen auch die entsprechenden Differenzen auf der Kontrollvariablen, während die geringen Unterschiede

1) -----
Die Gruppe 2 enthält keine nacherhöbten Vpn. Sie ist in die Einwegvarianzanalyse eingegangen, fällt aber aus dem gruppeninternen Vergleich der Mittelwerte heraus. Ihre Mittelwerte fallen etwa zwischen die der Gruppe 1 und der gesamten Gruppen (Tab. 61) und entsprechen ungefähr dem Mittel aller Gruppen. Die Gruppe 2 kann daher zu den Unterschieden zwischen den Erhebungen nicht beigetragen haben.

auf der Testleistung zwischen beiden Erhebungen in der Gruppe 3 keine Entsprechung auf der Kontrollvariablen finden; in der Gesamtgruppe entsprechen die Unterschiede der Testleistung Unterschieden auf der Kontrollvariablen. Ein weiteres Indiz für die Abhängigkeit der Testleistungen von der Kontrollvariablen ist die Gleichheit der Differenzen zwischen den Erhebungsgruppen in der mündlichen und schriftlichen Messung in der Gesamt- wie den einzelnen Experimentalgruppen.

Eine noch weitere Differenzierung des Vergleichs der beiden Erhebungen läuft natürlich Gefahr, wegen der geringen Zahlen in den Gruppen zufällige Ergebnisse herauszustellen. Dennoch läßt sich aus dem Vergleich der beiden Erhebungen innerhalb der Experimental-, Schicht- und Geschlechtsgruppen (Tab. 35) zumindest kein Beweis gegen die Abhängigkeit der Leistungen in beiden Erhebungen von der Kontrollvariablen erbringen. Insbesondere entsprechen in der 1. Gruppe, in der die beiden nacherhobenen Vpn. Jungen der Mittelschicht sind (siehe Tab. 32), die Unterschiede zu den in der Schule erhobenen Jungen der Mittelschicht ziemlich genau Differenzen auf der Kontrollvariablen. Ebenfalls wiederholen sich die Parallelen der Differenzen in den Experimentalleistungen und der Kontrollvariablen in den Gruppen 1 und 4; allein die Gruppe 3, in der sich bei der bloßen Aufteilung nach Experimentalbedingungen (Tab. 34) diese Parallele nicht ergab, muß hier genauer analysiert werden (Tab. 35). Für die Unterschicht- wie Mittelschicht-Jungen der Gruppe 3 lassen sich die Unterschiede der Erhebung jetzt auch auf Unterschiede der Kontrollvariablen zurückführen. Für die Mädchen beider Schichtgruppen ergeben sich unterschiedliche Differenzen in der mündlichen und schriftlichen Messung und daher keine Möglichkeit,

sie auf Unterschiede der Kontrollvariablen zu beziehen.¹⁾

Die Maße der sprachlichen Mittel zerfallen in 3 Gruppen: Maße der sprachlichen Oberflächenstruktur (Satzteile, Pronominalgebrauch, Satzverknüpfung), Maße des Erklärungsstils und Fehlermaße. Von jeder Gruppe dieser Maße, die in Abschnitt 6.3. und Anhang D genau dargestellt sind, wurden nur die wichtigsten Maße überprüft. Alle überprüften Maße sind Quotienten, in denen das jeweilige sprachliche Merkmal auf die übergeordnete Gesamtheit (Zahl der Worte, Verbphrasen oder Antworten) relativiert wurde.

Bei den Maßen der sprachlichen Oberflächenstruktur ist auffällig, daß die Vpn. bei der Erhebung im Elternhaus im Mündlichen ($p < .09$) wie Schriftlichen ($p < .03$) signifikant weniger Worte für eine Erklärung brauchten (Tab. 33), obwohl auch die Gesamtzahl der Antworten (GATEX) im Elternhaus, ebenfalls in mündlicher und schriftlicher Messung, höher ist. Weiterhin gebrauchen die Vpn. im Elternhaus weniger Verbphrasen pro Antwort, allerdings nur im Schriftlichen signifikant ($p < .01$). Bei den übrigen Maßen der sprachlichen Oberflächenstruktur finden sich nur minimale Unterschiede zwischen den beiden Messungen, oft mit unterschiedlichen Vorzeichen in den Kommunikationsformen. - Betrachten wir die Maße mit signifikanten Effekten des Erhebungszeitpunkts innerhalb der beiden Schichtgruppen (Tab. 33a), so findet sich keine auffällige Interaktion zwischen Erhebungszeitpunkt und Schicht. Betrachten wir die Maße

1) Für die Mädchen der Unterschicht sind jedoch die Differenzen auf der Kontrollvariablen so gering, daß die unterschiedlich gerichteten Differenzen in mündlichen und schriftlichen Experimentalleistungen durchaus zufällig sein können. Bei den Mittelschicht-Mädchen jedoch zeigen sich auf der Kontrollvariablen die größten Differenzen überhaupt und auf allen Explikations- und Verständnismaßen außer MIBEX hohe Differenzen mit unterschiedlichen Vorzeichen, so daß der bessere Mittelwert auf der Kontrollvariablen der in der Schule erhobenen Kinder sich nur in den mündlichen Maßen auswirkt. Allein in dieser Gruppe also hat die Nacherhebung nicht weiter erklärbares besondere Einflüsse gehabt; außerdem bietet die Vergleichsgruppe, die in der Schule erhobenen Mittelschicht-Mädchen, gemessen an ihrem hohen Wortschatzwert, nur Leistungen, die dem Durchschnitt der anderen Gruppe entsprechen. Daß für die Gruppe 3 sich keine Parallele der Experimental- und Kontrolleistungen herstellen ließ, kann also auf die Mittelschicht-Mädchen dieser Gruppe zurückgeführt werden.

mit signifikanten Effekten des Erhebungszeitpunkts innerhalb der 4 Experimentalgruppen (Tab. 34), so finden sich der häufigere Gebrauch von Worten und Verbphrasen pro Erklärung bei den in der Schule getesteten Kindern in allen Experimentalgruppen und Kommunikationsformen wieder. Aus diesem Grunde aber können wir nicht - wie bei den Maßen der Bedeutungserklärung - die Differenzen der sprachlichen Oberflächenmaße zu den zwischen den Experimentalgruppen das Vorzeichen wechselnde Differenzen auf der Kontrollvariablen (Tab. 36) in Beziehung setzen.

Die genannten sprachlichen Oberflächenmaße sind also von der Testsituation abhängig. Als Situationseinfluß aber kann der Erhebungszeitpunkt mit der Kommunikationsform verglichen werden und die Bedeutung der Nacherhebung an der der Kommunikationsform eingeschätzt werden. Tatsächlich zeigt sich nun auf jedem Niveau der Gruppenaufteilung - in Tab. 33, 34 und 35 -, daß die Unterschiede zwischen den Kommunikationsformen größer sind als die zwischen den Erhebungsarten; in der Gesamtgruppe etwa ist die Anzahl der Worte pro Antwort in der Erhebung in der Schule wie im Elternhaus im Mündlichen doppelt so hoch wie im Schriftlichen, während die Erhebung in der Schule die Anzahl der Worte pro Antwort im Mündlichen wie im Schriftlichen gegenüber der Erhebung im Elternhaus nur um das anderthalbfache vergrößert. Trotz Signifikanz ist der Effekt der Nacherhebung im Vergleich zum Effekt der Kommunikationsform von geringerem Gewicht.

Bei den Maßen des Erklärungsstils ergeben sich bis auf ein Maß im Mündlichen (NP mit 1 NPE: Nominalphrasen mit einer Nominalphrasen-ergänzung pro Antwort, $p < .09$) keine signifikanten Unterschiede. Diese Maße hängen offensichtlich vom Stimulusmaterial und den individuellen konstanten Arten der Aufgabenlösung eher ab als von der Testsituation. Das bestätigt sich auch beim Vergleich der Differenzen⁰¹ des signifikanten Maßes zwischen den Experimentalgruppen (Tab. 34): Der Quotient "NP mit 1NPE" schwankt mit den Experimentalgruppen und

Erhebungen, nicht aber mit den Kommunikationsformen. Die Differenzen der Erhebungsart werden hier also weniger durch die Kommunikationsformen relativiert, eher durch Differenzen zwischen den Experimentalgruppen. Zu den Differenzen der Kontrollvariablen kann keine Beziehung hergestellt werden.

Bei den Fehlermaßen ergeben sich wiederum nur in der schriftlichen Messung signifikante Effekte der Erhebungsart (Tab. 33): Schreibfehler und Gesamtfehler (Schreibfehler, die den größten Anteil stellen, Selektions- und Grammatikfehler pro Wort) treten beim Test im Elternhaus seltener auf. Innerhalb der Experimentalgruppen (Tab. 34) aber entsprechen die Differenzen der Erhebungen wieder - wie bei den Maßen der aktiven Erklärung und wiederum mit Ausnahme der Gruppe 3 - den Differenzen der Kontrollvariablen (Tab. 36); in der Gruppe 1 machen die in der Schule getesteten Kinder weniger Fehler, in der Gruppe 4 die im Elternhaus getesteten Kinder; beide sind auch im passiven Wortschatz den Kindern der gleichen Experimentalgruppe überlegen.

Daß die Unterschiede der Erhebungsarten bei den Fehlern, nicht aber bei den Maßen der sprachlichen Oberfläche und des Erklärungsstils mit Unterschieden der Kontrollvariablen zusammenhängen, ist aus dem Charakter der Maße verständlich. Die sprachliche Oberfläche hängt von spontan wahrgenommenen Besonderheiten der Situation ab; auch theoretisch läßt sich kein sinnvoller Zusammenhang mit Testleistungen wie dem passiven Wortschatztest denken: gute Wortkenntnis kann sich ebenso in einer wort- und verbarmen Erklärung ausdrücken, wie in einer wort- und verbreichen. Die Fehlermaße bilden dagegen - ebenso wie die Maße der aktiven Worterklärung - situationsunabhängige Fertigkeiten ab, die mit einem individuell konstanten Leistungsniveau zusammenhängen. Da der Wortschatztest seinerseits stark mit den Schulnoten korreliert (Tab. 21), ist die Abhängigkeit der notenähnlichen Fehlerquotienten von ihm nicht verwunderlich.

Wie lassen sich die verschiedenen Wirkungen der Nacherhebung zusammenbringen? Zuerst hat sie im Wesentlichen nur die schriftlichen

Leistungen beeinflusst, offensichtlich konnten von der Testtechnik her nur in der schriftlichen Messung die Unterschiede der sozialen Situation Einfluß gewinnen. Zum zweiten hat der schriftliche Test im Elternhaus die Anzahl der aktiven Erklärungen gesteigert, was zugleich auf eine bessere Gesamtleistung hinausläuft, und den auf die Anzahl der Antworten relativierten Gebrauch von Worten und Verben wie die Zahl von Fehlern pro Wort verringert. Insgesamt hat also die Nacherhebung im Elternhaus auf im Verhältnis zur Qualität der Worterklärung relativ periphere Mittel der Worterklärung einen disziplinierenden und konzentrierenden Einfluß ausgeübt; dadurch wurde zwar die Gesamtleistung verbessert, die Genauigkeit und Güte der gegebenen Erklärungen aber blieb unberührt.

Wir haben außerdem versucht, die Bedeutung der auftretenden Unterschiede einzuschätzen. Die Erhebungsunterschiede bei den Maßen der sprachlichen Oberflächenstruktur waren im Vergleich zu den Unterschieden der Kommunikationsform relativ gering; die Erhebungsunterschiede bei dem einzig signifikanten Maß des Erklärungsstils wurden durch Unterschiede der Experimentalgruppen relativiert. Die Erhebungsunterschiede bei den Maßen der aktiven Worterklärung und den Fehlermaßen konnten weitgehend auf Differenzen der Kontrollvariablen zurückgeführt werden. Gegen den letzten Versuch, die Erhebungseinflüsse zu relativieren, läßt sich natürlich einwenden, daß Ausgangsbedingungen, die für alle Maße bestanden haben, zur Erklärung von Einflüssen bei einzelnen Maßen herangezogen werden. Da aber die Kontrollvariable WST mit allen Maßen der aktiven Worterklärung ähnlich stark korreliert (Tab. 108), ließ sich zwischen den Maßen bezüglich ihrer Abhängigkeit von der Kontrollvariablen nicht differenzieren.

3. Kurze Wiederholung der Hypothesen der ersten und zweiten

Messung

Die Mittelschichtkinder erreichen erstens im passiven Wortschatz bessere Leistungen als die Unterschichtkinder. Dies gölt zweitens auch unabhängig von gemessener Intelligenz. (erste Messung)

Auch bei Kontrolle der Leistung im standardisierten passiven Wortschatztest geben die Mittelschichtkinder im aktiven Test bessere Antworten als die Unterschichtkinder. (zweite Messung)

In der zweiten Messung sind die Testbedingungen für Schichtunterschiede gegenüber der ersten Messung verschärft durch die Kontrolle des Wortschatztest-z-Wertes. Die erste Messung spiegelt in einer Stichprobe in der sozialen Realität vorhandene Unterschiede der sprachlichen Fähigkeiten zwischen Kindern beider Schichten. Die zweite Messung dagegen überprüft, ob die Kinder der Mittelschicht, obwohl sie über gleich gute passive Wortkenntnis verfügen, ihren Wortschatz aktiv besser ausschöpfen. Nur bei Kontrolle des passiven Wortschatzes können Unterschiede zwischen den Schichten im aktiven Wortschatz aus schichtspezifischen Arten des Gebrauchs von Sprache erklärt werden. Trotzdem können Unterschiede der Sprache zwischen beiden Schichten bestehen und ohne die experimentelle oder statistische Kontrolle von Faktoren gemessen werden.

Auf eine formale Ähnlichkeit zwischen der zweiten Hypothese der ersten und der Hypothese der zweiten Messung soll abschließend hingewiesen werden. Beide zielen auf Erklärungen durch Strategien des Sprachgebrauchs, insofern sie Faktoren ausschließen, die durch die schichtspezifische Verteilung von Fähigkeiten oder durch schichtspezifische Umweltanregungen erklärt werden könnten: die Intelligenz bzw. den passiven Wortschatz. Die Hypothese der zweiten Messung ist darüber hinaus auf eine Theorie des Sprachgebrauchs insofern bezogen, als sie unmittelbar eine aktive Sprachleistung mißt. Die zweite Hypothese der ersten Messung dagegen bezieht sich auf den passiven Wortschatz, der als ein langfristiges Produkt des aktiven Sprachgebrauchs gesehen werden muß, wenn man Schichtunterschiede durch Strategien des Sprachgebrauchs erklären will (s. Kapitel I und IV, 3.1.).

4. Ergebnisse der ersten Messung

=====

4.1. Ergebnisse der ersten Messung für den gesamten Test.

Die erste Messung unseres Experiments - die Anwendung des standardisierten passiven Wortschatztests (WST 5-6) - gibt, im Vergleich mit der zweiten Messung Auskunft über die "natürlichen" Unterschiede in der Wortkenntnis zwischen den Kindern verschiedener sozialer Schichten und das Ausmaß experimentell schon ausgeschalteter sozialer Einflüsse. Getestet wurden insgesamt 11 Schulklassen des 4. Schuljahres (N=329).

Die Ergebnisse wurden mit Varianzanalysen geprüft. In der einfachen Varianzanalyse mit 5 Schichten zeigt sich für die z-Werte wie die Roh-Werte des Wortschatztests ein hochsignifikanter Schichteffekt (Tab. 37). Die Mittelwertvergleiche ergeben ein deutliches Ansteigen der Leistung von der Unteren Unterschicht bis zur Mittleren Mittelschicht und einen starken Bruch zwischen der Unteren und Oberen Unterschicht sowie der Oberen Unterschicht und der Unteren Mittelschicht; die Obere Mittelschicht fällt wieder ab. Auffällig ist ferner die Ähnlichkeit der Standardabweichungen in den beiden unteren und den beiden mittleren Schichten und die extrem große Standardabweichung in der Oberen Mittelschicht. Ein Test der Inhomogenität der Varianzen erreichte jedoch nicht das 5%-Niveau ($F_{\max}(5,114)=1.998$; kritischer Werte auf dem 5%-Niveau für FG 5,60 = 2.04). Kinder aus den beiden großen Schichtblöcken unterscheiden sich also in der Wortkenntnis. In der Oberen Mittelschicht jedoch scheint die Bedeutung sprachlicher Kenntnisse für eine auf sozialen Aufstieg zielende Sozialisation zu schwinden oder die allgemein geringere Präokkupation der Eltern mit Erziehungsaufgaben auch zu vergleichsweise schlechteren Wortschatzleistungen der Kinder zu führen.

Die Zwei-Weg-Varianzanalyse mit 5 Schichten und Geschlecht als Faktoren deckt weiterhin einen signifikanten Geschlechtseffekt, aber keine signifikante Schicht-Geschlechts-Interaktion auf (Tab. 38).

Im Widerspruch zu den meisten bekannten Hypothesen und Ergebnissen (McCarthy 1954) sind die Jungen in allen 5 Schichtgruppen den Mädchen überlegen. Weiterhin aber unterscheiden sich auch die Profile der beiden Geschlechtsgruppen. Bei den Mädchen steigen die Leistungen über alle 5 Schichten kontinuierlich an, während bei den Jungen die Kluft zwischen den beiden Schichtblöcken sich noch vergrößert und innerhalb der unteren Straten die Untere Unterschicht, innerhalb der mittleren die Mittlere Mittelschicht herausragt. Die Leistungen der Jungen der Mittleren Mittelschicht stehen in besonders starkem Kontrast zu denen der Mädchen der gleichen Schicht. - Die Drei-Weg-Varianzanalyse mit dichotomer Schicht, Geschlecht und Schule (Tab. 39) ergibt keinen signifikanten Effekt für die Schulen und keine signifikanten Interaktionen. Allerdings kehrt sich der Geschlechtsunterschied in der Unterschichtgruppe der Schule Süd und der Mittelschichtgruppe der Schule Goldstein zugunsten der Mädchen um, bleibt aber insgesamt signifikant zugunsten der Jungen. Zur Übersicht sind die Mittelwertdifferenzen der ersten Messung noch einmal für zwei Schichten nach Geschlechtern zusammengefaßt. Die Trennung zwischen den beiden Schichten hier entspricht dem größten Mittelwertabstand in der Einteilung in 5 Schichten (zwischen "Oberer Unterschicht" und "Unterer Mittelschicht") (Tab. 40).

Die Abhängigkeit von sozialer Schicht teilt der Wortschatztest mit vielen anderen Test- und Schulleistungen. Wie stark jedoch ist diese Abhängigkeit trotz kontrollierter gemessener Intelligenz? Wir hatten die zweite Hypothese aufgestellt, daß auch unabhängig von gemessener Intelligenz Kinder der Mittelschicht eine bessere Leistung im passivon Wortschatztest zeigen würden als Kinder der Unterschicht. - Wie stark hängt ferner der Wortschatztest seinerseits mit Schulleistungen zusammen, unabhängig von Einflüssen der gemessenen Intelligenz? Wie stark schließlich sind diese nur auf den Wortschatz rückführbaren Zusammenhänge im Vergleich zum auf die gemessene Intelligenz unabhängig vom Wortschatz zurückführbaren Zusammenhang? Diese Fragen sollen anhand zweier, mit Hilfe multipler Verfahren gewonnener Ergebnisse beantwortet werden.

Auf der einen Seite ist der Einfluß der Sozialschicht der Eltern auf die Wortschatzleistung der Kinder, unabhängig von den Leistungen der Kinder in den Intelligenztests, größer, als auf jeden anderen Intelligenzsubtest einzeln, unabhängig von allen anderen und dem Wortschatztest. Dies belegt eine Kovarianzanalyse mit einer in fünf Straten differenzierten Schichteinteilung als Faktor und den neun Intelligenzsubtests und den Wortschatztest als Variablen. Die Varianz in der Schichteinteilung sagt die von seiner Kovarianz mit den Intelligenzsubtests unabhängige Varianz des Wortschatztests stärker voraus, als die von der Kovarianz mit den jeweils übrigen Tests unabhängige Varianz jedes Subtests (Tab. 41). Unabhängig von Schichtunterschieden der Intelligenz, verfügen die Kinder der Mittelschicht über einen größeren Wortschatz als die Kinder der Unterschicht; damit wird auch die schärfere Fassung unserer Hypothese der ersten Messung bestätigt. Auf der anderen Seite zeigt der Wortschatztest die höchste spezifische Voraussagekraft für jede einzelne der Schulnoten in den Leistungsfächern. Die von der Kovarianz mit Intelligenz unabhängige Varianz in den Wortschatzleistungen erklärt die Varianz der Schulnoten besser als die von der Kovarianz mit den übrigen Intelligenztests unabhängige Varianz jedes Einzeltests. Dies zeigt sich daran, daß der Wortschatztest für jede Schulnote den höchsten Rohregressionskoeffizienten aufweist (Tab. 42).

Beide Ergebnisse sagen nicht, daß Schicht auf Intelligenz und Intelligenz auf Noten keinen Einfluß hat, sondern nur, daß von den verglichenen Tests der Wortschatz in beide Richtungen der wichtigste spezifische Vermittler ist. Dafür bietet sich eine eher statistische und eine eher theoretische Erklärung an. Die neun Subtests des P-S-B bilden einen zusammenhängenden Test, so daß zwischen den Subtests höhere Interkorrelationen als mit dem Wortschatztest und damit eine geringere von der Kovarianz mit den anderen unabhängige Wirkung jedes Intelligenzsubtests zu erwarten wäre. Tatsächlich sind in der Stichprobe die Interkorrelationen der Intelligenzsubtests nicht höher (Median bei $r = .29$) als die Korrelationen der Subtests mit dem Wortschatztest (Median bei $r = .30$) (siehe Tab. 22). Deshalb muß die unterschiedliche Voraussage-

kraft der Tests auf die Schulnote doch theoretisch erklärt werden. Der Wortschatztest ist als verbaler Testeerstens stärker Umwelt- und Lerneinflüssen ausgesetzt, als das in P-S-B, das in der Hauptsache nicht verbale, "basale" Intelligenz messende Subtests umfaßt. Zweitens haben verbale Fähigkeiten in der Interaktion zwischen Schüler und Lehrer einen viel augenfälligeren Charakter als nicht-verbale; "intelligent" ist, besonders in einer sprachlich vermittelten Lehrsituation, zunächst einmal jemand, der sich gewandt und abwechslungsreich ausdrückt. Den beiden Erklärungen entspricht als weiterer Beleg, daß die verbalen Subtests des P-S-B nach dem Wortschatztest einmal dem Schichteinfluß am stärksten spezifisch ausgesetzt sind (Tab. 41, Test 1/2) und zweitens auch die nächststärkste spezifische Aussagekraft für die Schulnoten haben (Tab. 42, Test 1/2, 5, 6 für Deutsch Schriftlich).

Auf dem Hintergrund dieser durch die multivariaten Verfahren der Kovarianz- und Regressionsanalyse belegten herausragenden Bedeutung des Wortschatztests gewinnen die durch die einfache Varianzanalyse statistisch belegte Schichtunterschiede der Wortschatzleistung stärkeres Gewicht. Sie betreffen nicht, wie vorwiegend die Intelligenzsubtests, gemessene Unterschiede einer theoretisch konstruierten Dimension, sondern eine in sozialen Lebenszusammenhängen geförderten und wirksamen Kennzeichens für "Intelligenz" und "Leistung". In dem zirkulären sozialen Auslese- und Sozialisationsprozeß (Rolff 1967:18), in dem unterschiedlich objektive Sozallagen sich in unterschiedlichen Leistungen und diese wieder in unterschiedlichem Sozialstatus niederschlagen, muß der Wortschatz als wichtigstes Mittelglied angesehen werden. In der Kausalkette, die von der Ausgangsposition "Schicht" über kognitive und sprachliche Fähigkeiten zu Leistungen in der Schule führt, ist er im Vergleich zu den verschiedenen Bereichen gemessener Intelligenz der wichtigste Vermittler.

4.2. Ergebnisse der ersten Messung für einzelne Worte und Wortgruppen

Die Schichtdifferenzen im Gesamttest legen die Vermutung nahe, daß bestimmte Worte für die Schichtdifferenzen verantwortlich sind; sie müßten sich unter einem theoretischen Gesichtspunkt zusammenfassen lassen, der einen Schichtunterschied plausibel macht. Beispielsweise könnten die Worte, die Intentionen ausdrücken (Oevermann 1970), oder die, die extravertierte schwerzugängliche Bereiche definieren für die Mittelschicht leichter erklärbar sein.

Tab. 43 gibt eine Übersicht über die Ergebnisse der ersten Messung mit der Wortaufteilung der zweiten Messung; angegeben ist die Zahl der Kinder, die die richtigen der Antwortalternativen angekreuzt hatten, für die Gesamtgruppe, die Mittel- und die Unterschichtgruppe in Prozentzahlen der jeweiligen Basis, sowie die Schichtdifferenzen dieser Prozentzahlen. Die üblichen Hypothesen, die die Überlegenheit der Mittelschicht-Kinder behaupten, lassen sich zwar an einzelnen Worten demonstrieren (z.B. für "widerpenstig"), aber es lassen sich kaum "mittelschichtgünstige" Worte unter einer einheitlichen Hypothese zusammenfassen. Aus Gründen der Übersichtlichkeit sind die Worte in Tab. 44 noch einmal für jede Wortgruppe in der Reihenfolge nach den Schichtdifferenzen aufgeführt. Auch in dieser Tabelle lassen sich weder sinnvolle theoretische Wortgruppen bilden noch irgendein größerer Einschnitt bezüglich der prozentualen Schichtdifferenzen in der Rangfolge der Differenzen feststellen. Wenn überhaupt, so zeigen intentionale Worte (wie "Frevel", "bedächtig", "bescheiden") die geringeren, und rein objektbezogene Worte (wie "Echse", "Grotte", "Düne", "Traktor", "Trog", "Wolkenkratzer") die größten Differenzen zugunsten der Mittelschicht.

Da der Versuch scheiterte, induktiv Gruppierungen zu bilden, wurde eine Klassifizierung auch für die erste Messung übernommen, die zuerst für die zweite Messung ohne Kenntnis der Ergebnisse der ersten Messung entwickelt worden war. (siehe Abschnitt 5.4.2.).

In ihr wurden die Worte nach zwei zueinander querliegenden theoretischen Dimensionen klassifiziert: der extraverbalen Erklärbarkeit und dem intentionalen Sinngehalt. Die zweite Dimension wurde nur ergänzend hinzugenommen, so daß die Gruppen allein in der ersten benannt wurden. Die Gruppe I enthält Worte, die ohne weiteres durch Zeigen aufweisbar oder Gesten darstellbar sind ("extraverbal"); die Gruppe II Worte, für die dies nur in einem zuvor bekannten Situationskontext möglich ist ("extraverbal Kontext"); die Gruppe III Worte, für die eine extraverbale Kennzeichnung zwar noch möglich, aber kaum verständniserleichternd wäre ("fast verbal"); die Gruppe IV Worte, die nur noch in verbalen Kontexten erläutert werden können ("rein verbal"). Diese Kennzeichnung wurde für jedes Wort in Tab. 44 eingetragen. Auch hier zeigt der erste Überblick keinerlei Zusammenhang mit den Schichtdifferenzen. Zur genaueren Untersuchung wurden erstens die Mittelwerte der Wortgruppen extraverbaler Explizierbarkeit für Gesamtgruppe und Schichtdifferenzen und zweitens die Korrelation der extraverbalen Explizierbarkeit, als Vier-Punkte-Skala betrachtet, mit den Schichtdifferenzen errechnet.

In Tab. 45 (Abb.6) sind die Schichtdifferenzen der Prozentzahlen der richtigen Antworten gemittelt für insgesamt 16 verschiedene Wortgruppen angegeben. Auch hier sind alle Schichtdifferenzen zugunsten der Mittelschicht, die unterschichtgünstigen Wortgruppen die mit der geringeren Schichtdifferenz. Da alle Zahlen Prozentsätze darstellen, hängt jedoch im Vergleich der Wortgruppen die Höhe der Schichtdifferenz mit den richtigen Antworten in der Gesamtgruppe zusammen: im unteren und oberen Bereich sind geringere, im mittleren größere absolute Differenzen wahrscheinlich; diese Eigenart des Maßes kann Schichtunterschiede neben den Besonderheiten der Wortgruppen bedingt haben. Unter dieser Einschränkung müssen die Schichtunterschiede zwischen den Wortgruppen extraverbaler Explizierbarkeit und den Wortgruppen des Tests verglichen werden.

Die Wortgruppen extraverbaler Explizierbarkeit zeigen nur innerhalb der Gruppe A (ungerade, gerade und gesamt) eine kontinuierliche Abnahme der Schichtdifferenzen, zugleich jedoch auch eine Abnahme der prozentualen richtigen Antworten, die die Abnahme der Schichtdifferenzen mitbedingt haben kann. Innerhalb der Gruppe B zeigt sich überhaupt keine Zunahme der Schichtdifferenzen und nur eine minimale Zunahme der richtigen Antworten insgesamt mit zunehmender extraverbaler Explizierbarkeit der Worte. Aus dem Rahmen fällt die "extraverbale" Wortgruppe der Form "B ungerade": Trotz eines hohen Prozentsatzes richtiger Antworten ergibt sich nur eine minimale Schichtdifferenz. Die Erklärung liegt in dem (wohl zufälligen) Zusammentreffen der negativen Differenzen mit den Worten der Gruppe I (Tab. 44). Für beide Testformen zusammen hält sich das Ergebnis durch, daß in der "rein verbalen" Wortgruppe die Schichtdifferenzen am geringsten und in den übrigen Gruppen etwa gleich sind (Tab. 45, Spalte 7); dieser Unterschied läßt sich kaum mehr aus dem Prozentsatz richtiger Antworten in der Gesamtgruppe erklären, der nur noch gering schwankt. Bei den "rein verbalen" Worten also holen die Unterschichtkinder ihren Rückstand merklich auf, obwohl sie auch hier noch von den Mittelschichtkindern übertroffen werden.

Bei der Zusammenfassung der vier Gruppen extraverbaler Explizierbarkeit zeigen sich in den Testformen starke Unterschiede bei den Schichtdifferenzen, geringere bei den prozentualen Anteilen der richtigen Antworten. Da die Schichtdifferenzen der konstruierten Wortgruppen einer Testform auf dem fast gleichen Niveau der prozentualen richtigen Antworten verglichen werden, belegen unterschiedliche Schichtdifferenzen zwischen den Testformen tatsächlich einen schichtspezifischen Charakter der Testformen. Daß die Schichtdifferenzen bei diesem Vergleich tatsächlich von dem Anteil richtiger Antworten unabhängig sind, zeigt die Gegenüberstellung der Spalten 1 und 3: die größten mittleren Schichtdifferenzen tauchen in der Testform mit dem niedersten, die geringsten in der Testform mit dem höchsten mittleren Anteil richtiger Antworten auf. Die Unterschiede der Testformen in der ersten Messung können mit den

Unterschieden der Experimentalgruppen (und damit der Testformen) in der zweiten Messung in Verbindung gebracht werden, was wir in Abschnitt 5.3.5. versuchen wollen. An dieser Stelle soll die Besonderheit der Testformen und ihr Zusammenhang mit den Wortgruppen extraveraler Explizierbarkeit in der ersten Messung weiter verfolgt werden.

Die Gruppenbildung nach der extraveralen Explizierbarkeit der Worte läßt sich auch als eine Skala verstehen, auf der jedem Wort ein Wert zwischen 1 und 4 für "Extraverbalität" gegeben wurde. Genauer und kürzer als der Vergleich der Mittelwerte können dann Korrelationen den Zusammenhang zwischen "Extraverbalität" und Bekanntheitsgrad (Prozent der richtigen Antworten) bzw. Schichtdifferenzen der Bekanntheit darstellen. Wir vermuten, daß mit zunehmender "Verbalität" die Prozentsätze richtiger Antworten geringer (1) und die Schichtdifferenzen größer (2) werden.

Tab. 46, Zeile 2 belegt die erste Hypothese für die Wortgruppe A gut, für "B gerade" etwas und für "B ungerade" überhaupt nicht. Zeile 3 widerlegt die zweite Hypothese für die Wortgruppe A genügend, für "B gerade" etwas und bestätigt sie leicht für "B ungerade". Für diese letzte Wortgruppe gilt wieder die schon bei den Mittelwertdifferenzen referierte Beobachtung: die positive Beziehung der "Verbalität" mit den Differenzen zugunsten der Mittelschicht läßt sich auf das massive Zusammentreffen der "extraverbalen" Worte mit den von der Unterschicht besser beantworteten erklären. Für alle Wortgruppen zusammen aber bleibt ein leichter Zusammenhang von "Verbalität" und relativ besseren Leistungen der Unterschicht bestehen, entgegen der Hypothese. Dies Ergebnis ist um so erstaunlicher, als negative Korrelationen in Zeile 2 und - nicht gefundene - positive in Zeile 3 möglicherweise ein Artefakt sein könnten. Die Abfolge der Worte in der Testform hängt, wie man in Tab. 43a-d und Tab. 44 sehen kann, mit der Gruppeneinteilung zusammen (Worte der Gruppe 1 sind häufiger am Anfang, Worte der Gruppe 4 häufiger am Ende der Testformen A und B zu finden); ebenso aber mit der Schwierigkeit der Worte (Anger/Bargmann/Hylla 1954 Test-Beiheft, S. 10) und darüber vermittelt hypothetisch auch mit den Schichtdifferenzen, so daß die schwierigen und mittelschichtgünstigeren Worte am Ende stehen müßten. Zeile 4 jedoch zeigt, daß

die Prozentzahlen richtiger Antworten mit den Schichtdifferenzen in der Testform A positiv korrelieren (in der Gruppe "B ungerade" negativ aus dem schon genannten Grund), so daß die insgesamt schwierigeren Worte die für die Unterschicht relativ leichteren sind.¹⁾ Die Tatsache, daß die Unterschicht in den schwierigeren Worten der Testform A der Mittelschicht weniger unterlegen ist als in den leichteren kann also an zwei (allerdings interkorrelierenden: Zeile 2) Maßstäben der "Wortschwierigkeit" abgelesen werden: der Skala der "Verbalität" und den Prozentzahlen richtiger Antworten in der Gesamtgruppe.

Die 4 Testgruppen unterscheiden sich bei allen 3 Arten der Korrelationen (Zeile 2-4) auf die gleiche Weise. Das Muster positiver und negativer Vorzeichen wie die absolute Höhe der Korrelationen stimmt für jede Testgruppe zwischen den Zeilen überein. Zwischen den Testformen bestehen folgende Unterschiede: (1) bezüglich der Schichtdifferenzen (Zeile 1), (2) bezüglich des Zusammenhangs der Schichtdifferenzen mit dem Schwierigkeitsgrad der Worte (Zeile 4), (3) bezüglich des Zusammenhangs der extraverbalen Explizierbarkeit mit den Schichtdifferenzen (Zeile 3), (4) bezüglich des Zusammenhangs der Wortschwierigkeit mit der extraverbalen Explizierbarkeit (Zeile 2). Die Ergebnisse der beiden zuletzt genannten Zeilen können auch zu einer gewissen Validierung der Klassifikation der Werte nach extraverbalen Explizierbarkeit herangezogen werden; sie ist befriedigend jedoch nur mit der Testform A. Aus diesem Grund ist für Zeile 2 und 3 auch die Korrelation mit den Worten beider Testformen errechnet worden: je schwieriger die Worte (d.h. je geringer der Anteil der richtigen Antworten), desto eher wurden sie der Gruppe 4 zugeschlagen ($r=.28$), und je eher sie der Wortgruppe 4 zugerechnet wurden, desto geringer war die Unterlegenheit der Unterschichtkinder ($r=-.15$).

1)-----
Dies kann kein Artefakt sein, denn allein durch die Messung in Prozentzahlen müßten die größten Differenzen sich bei den mittleren, die geringsten bei den extremen Prozentzahlen richtiger Antworten finden, so daß eine Korrelation von 0 zu erwarten wäre. Im übrigen läßt sich aus Tab. 45 ersehen, daß die durchschnittlichen Prozentzahlen für die Schichtdifferenzen nicht durch die durchschnittlichen Prozentzahlen für die richtigen Antworten bedingt sind (s.o.)

Wir wollen die Ergebnisse der ersten Messung kurz zusammenfassen und diskutieren. Der passive Wortschatz der Kinder - gemessen im Frankfurter Wortschatztest - hängt hochsignifikant mit der sozialen Schichtzugehörigkeit der Eltern zusammen. Dieser Zusammenhang bleibt auch bei statistischer Kontrolle der Intelligenz signifikant. Von allen Intelligenzzubtests besitzt der Wortschatztest die höchste spezifische Voraussagekraft. Die allgemein festgestellten Schichtunterschiede haben jedoch den verschiedenen Versuchen der Aufhellung durch schichtdiskriminierende Einzelworte oder theoretisch faßbare Wortgruppen weitgehend widerstanden; allenfalls läßt sich belegen, daß die Unterschichtkinder in den extraverbal schwer zugänglichen Worten den Mittelschichtkindern weniger unterlegen sind. Stattdessen ergaben sich - nicht interpretierbare - Unterschiede zwischen den Testformen, die wir in der zweiten Messung wiederfinden werden. Die Schichtunterschiede lassen sich nicht auf immanent sprachliche Besonderheiten der Worte zurückführen, die sinnvoll mit schichtspezifischen Sozialisationsbedingungen verknüpft werden könnten.

Daß alle Worte den Unterschicht-Kindern weniger bekannt sind, könnte durch schichtspezifisch verteilte verbale Fähigkeiten, durch Unterschiede umweltlicher Lernanregung zwischen den Schichten, die sich auch in einer geringeren sprachlichen Realitätserfassung äußert, oder durch schichtspezifische Strategien des Sprachgebrauchs erklärt werden. Da auch bei Kontrolle der Intelligenz der Wortschatz der Mittelschichtkinder dem der Unterschichtkinder überlegen ist, scheidet die erste Erklärung aus. Zwischen einer Erklärung durch die Umweltanregung und durch Strategien des Sprachgebrauchs ist empirisch schwerer zu unterscheiden. Daß der Wortschatz in noch stärkerem Maße als die allgemeine Intelligenz durch Umweltanregungen beeinflusst wird, scheint eine plausible Annahme. Sie könnte erklären, warum Unterschiede des Wortschatzes zwischen Kindern verschiedener Schichten trotz kontrollierter Intelligenz bestehen bleiben; jedoch dürfte es schwer sein, die nur für Unterschiede des Wortschatzes verantwortlichen Anregungen konkret zu benennen. Genau an dieser Stelle aber führt eine Erklärung durch

schichtspezifische Arten des Sprachgebrauchs weiter; Arten des Sprachgebrauchs sind durch typische Sozialstrukturen und nicht durch Intelligenz bestimmt. Allerdings muß auch eine Erklärung durch Strategien des Sprachgebrauchs mit einer zusätzlichen Annahme arbeiten: da Unterschiede im passiven Wortschatz erklärt werden sollen, müssen sie als langfristige und intelligenzunabhängige Wirkungen des aktiven Sprachgebrauchs angesehen werden. Da beide Theorien eine zusätzliche Annahme einführen müssen, wollen wir die Entscheidung offenlassen.

5. Ergebnisse der zweiten Messung: Bedeutungserklärung

5.1. Zu den in der zweiten Messung verwandten Prüfverfahren.

Die Wahl der Prüfverfahren für Schichtunterschiede ergibt sich aus dem Aufbau des Experiments (siehe Kapitel II.1.) und dem Meßniveau der konstruierten Variablen (siehe 1.2.). Das Experiment ist aus Gruppen von auf dem passiven Wortschatz parallelisierten Paaren von Vpn. des gleichen Geschlechts aus der Unter- und Mittelschicht aufgebaut. Die den Variablen zugrundeliegende Verständnis- und Explikationsbewertung hat lediglich Ordinal-Niveau; aber schon mit den bei der Bildung der Maße verwandten Rechenverfahren haben wir Intervallskalen-Niveau angenommen. Streng genommen, sind jedoch nur die Maße GATEX und RATEX über dem ordinalen (nämlich auf dem Ratio-Skalen-)Niveau. Weiterhin haben wir die Verteilungsform unserer Testvariablen nicht geprüft. Dem Experimentalaufbau und dem Meßniveau voll angemessen sind daher nur "nicht-parametrische" Verfahren, die parallelisierte Gruppen auf ordinalen Variablen vergleichen: der Vorzeichen-Test und - als eine Erweiterung - der Wilcoxon-Test, den wir durchgehend verwandt haben (Siegel 1956). Der Wilcoxon-Test beruht auf der Rangfolge der absoluten Differenzen zwischen den Paaren. Paare mit Null-Differenzen werden von der Prüfung ausgeschlossen. Die Ränge der positiven, die Hypothese bestätigenden, und der negativen Differenzen werden summiert. Wie immer die positiven bzw. die negativen Rangsummen der Differenzen ausfallen, ihre Gesamtsumme muß konstant $P \times (P+1)/2$ sein, wobei "P" die Anzahl der Paare ist. Für jede Kombination positiver und negativer Rangdifferenzsummen kann daher, bei gegebener Anzahl der Paare, die Wahrscheinlichkeit ihrer kombinierten Werte zwischen 0 und + bzw. - $P \times (P+1)/2$ bestimmt werden. Geprüft wird die Wahrscheinlichkeit der Summe der kleineren Differenzen, die bei einer gerichteten Hypothese die Rangsumme der Differenzen gegen die Hypothese sein müßte. Übertrifft diese letzte Rangsumme die der Differenzen für die theoretische Hypothese, so muß der Test zweiseitig geführt werden.

Aus dieser kurzen Beschreibung werden folgende Vorzüge des Wilcoxon-Tests deutlich. Erstens verrechnet der Wilcoxon-Test nicht die absoluten Werte der zu prüfenden Variablen, sondern die Rangfolge der Differenzen zwischen den Paaren. Aus diesem Grunde setzt er nur ordinales Meßniveau voraus, anders als die Varianzanalyse, die prinzipiell Intervallskalen-Meßniveau verlangt. Zweitens benötigt er für die Prüfung von Unterschieden innerhalb der gegebenen Gruppe keine Annahmen über die Grundgesamtheit, aus der die Gruppe herausgegriffen wurde. Aus diesem Grunde zählt er zu den "nichtparametrischen" Tests und in dieser Hinsicht ist er einfacher als das "parametrische", d.h. mit Annahmen über die Grundgesamtheit arbeitende, Verfahren der Varianzanalyse.

Auf der anderen Seite erlaubt der Wilcoxon-Test, nur die Prüfung des Effekts, nach dem die Paare zusammengestellt wurden, in unserem Fall: der dichotomen Schicht. Da wir auch die weiteren dem Experimentalaufbau zugrundeliegenden Variablen: Geschlecht, Kommunikationsform, Experimentalform prüfen wollten, müßten wir das weitergehende "parametrische" Verfahren der Varianzanalyse verwenden. Sie bringt jedoch zusätzliche Information nur aufgrund weiterer Annahmen, die diskutiert werden müssen. Vorausgeschickt werden muß, daß es sich um ein varianzanalytisches Modell mit "festen Effekten" handelt, da alle in der zweiten Messung geprüften Effekte einander ausschließen und die logischen Möglichkeiten erschöpfen (Hays 1963).

Die Variananalyse setzt (1) Messung auf der Intervallskalenebene voraus und beruht auf folgenden statistischen Annahmen: (2) Normalverteilung der Testvariablen, (3) Homogenität der Varianzen in den Zellen, (4) Unabhängigkeit der Irrtumsanteile der Messungen für jedes beliebige Paar von Vpn. (1) Das Intervallskalen-Niveau ist zwar im strengen Sinne für unsere Maße nicht gegeben, kann jedoch - mit dem gleichen pragmatischen Recht wie etwa bei der Notengebung - angenommen werden. (2) Die Normalverteilung der Testvariablen in der Experimentalgruppe ist zwar nicht eigens überprüft worden; es ist aber unwahrscheinlich, daß die Variablen stark von ihr abweichen, etwa bimodal verteilt sind. Zudem kann diese Bedingung für

"feste Effekte" ohne ernste Folgen für die Schlüsse verletzt werden. (3) Die Homogenität der Varianzen muß für die einzelnen zu untersuchenden Gruppen dargestellt werden. Solange die Anzahl der Vpn. innerhalb der Zellen gleich ist, darf zudem auch diese Bedingung ohne Konsequenzen für die statistischen Folgerungen verletzt werden. (4) Die Unabhängigkeit der Irrtumsanteile der Messungen schließlich ist in Analysen mit der Kommunikationsform als Faktor problematisch, weil durch die im Mündlichen und Schriftlichen wiederholten Messungen der gleichen Vpn. deren Leistungen voneinander abhängen können (Hays 1963:455). Wenn man bei einem Experimentalplan mit wiederholten Messungen eine normale Varianzanalyse rechnet, werden die Faktoren ohne wiederholte Messungen über-, die Faktoren mit wiederholten Messungen unterschätzt. Aus diesem Grunde wurden - neben der dem Experimentalplan nicht ganz angemessenen üblichen Form der Varianzanalyse - erstens Varianzanalysen für die mündliche und schriftliche Messung getrennt und zweitens Varianzanalysen mit "wiederholten Messungen", d.h. mit Berücksichtigung der Gleichheit der Vpn.-Gruppe in den beiden Kommunikationsformen, gerechnet. Auch auf diese Bedingung kann genauer nur an der Stelle der jeweiligen Analyse (5.3.4. und 5.6.1.) eingegangen werden.

Der Vergleich der beiden Prüfverfahren - des Wilcoxon-Tests und der Varianzanalyse - zeigt, daß im strengen Sinne nur der Wilcoxon-Test für die Prüfung des Schichtunterschieds angemessen ist, die Varianzanalyse jedoch die Prüfung weiterer Effekte erlaubt. Wir werden daher zuerst die Ergebnisse des Schichtvergleichs nach dem Wilcoxon-Test und dann nach der Varianzanalyse darstellen, um dann zu den Ergebnissen der Varianzanalyse zu den übrigen Variablen überzugehen.¹⁾

¹⁾-----
Die Wilcoxon-Tests wurden mit einem Programm von Dr. U. Oevermann gerechnet, die Varianzanalysen mit dem Programmpaket MANOVA. Die Rechenarbeiten wurden im Deutschen Rechenzentrum (DRZ) in Darmstadt durchgeführt. Als weitere Programme wurden Programme des DRZ verwandt: für Korrelationen MISZ von Dr. F. Gebhardt, für Faktorenanalysen PAFA von P. Schnell und F. Gebhardt, für Häufigkeitszählungen HAUF von P. Schnell und L. Pöcker.

5.2. Einflüsse der dichotomen Schichteinteilung auf die Gesamtleistung der zweiten Messung nach dem Wilcoxon-Test.

Die verschiedenen Maße unseres Experiments wurden mit dem Wilcoxon-Test insgesamt dreimal überprüft, für stärker zusammengefaßte Experimentalleistungen oder größere Gruppen von Paaren. (1) Erstens wurden die Einzelwerte jeder Vpn. für jedes Wort innerhalb jeder Experimentalgruppe und jeder Kommunikationsform geprüft und nur die ursprüngliche Verständnis- und Explikationsbewertung berücksichtigt; dabei konnten maximal 8 (bzw. 9 in Experimentalgruppe 4) Paare gebildet werden. (2) Dann wurden die aus der ursprünglichen Bewertung gebildeten Gesamtmaße jeder Vpn. innerhalb jeder Experimentalgruppe und jeder Kommunikationsform geprüft; dabei konnten wiederum maximal 8 bzw. 9 Paare gebildet werden. (3) Schließlich wurden die Gesamtmaße jeder Vpn. ohne Trennung der Experimentalgruppen, nicht aber der Kommunikationsform geprüft; dabei konnten maximal 33 Paare geprüft werden. In allen drei Fällen wurde also, anders als bei einigen der Varianzanalysen, die Trennung von mündlicher und schriftlicher Messung beibehalten (siehe 5.3.). Jedoch haben die Wilcoxon-Tests nicht in allen drei Fällen die gleiche Funktion. Während sie in den beiden zuletzt genannten Fällen (2 und 3) Schichtunterschiede in der aktiven Worterklärung prüfen sollen, sollen sie im zuerst genannten Fall (1) schichtspezifische Bereiche des Wortschatzes aufdecken. Worte, die von Kindern der Unter- oder der Mittelschicht besser erklärt werden. Wir werden daher zuerst die allgemeinen Ergebnisse darstellen (5.2.1.), sie dann mit den Ergebnissen von Wilcoxon-Tests mit anderen Kontrollvariablen als dem Wortschatztest (5.2.2.) und den Ergebnissen der Varianzanalyse (5.3.2.) vergleichen, um dann erst die Ergebnisse zu einzelnen Worten darzustellen (5.4.1.).

5.2.1. Schichtunterschiede bei Kontrolle der Leistung im passiven Wortschatztest.

Die Ergebnisse der Wilcoxon-Tests innerhalb der einzelnen Experimentalgruppen wie der gesamten mündlichen und schriftlichen Messung sind für die positiven Gesamtmaße in Tab. 47, für die negativen Gesamtmaße in Tab. 48 dargestellt¹⁾. Innerhalb der Experimentalgruppen tauchen so gut wie keine Schichtdifferenzen auf; zuweilen sind sie - wie in der Wortgruppe "A gerade" bei den Maßen PRIAT und MIBVE - in der mündlichen und schriftlichen Messung einander entgegengesetzt, d.h. durch die Vpn.-Gruppen, nicht durch das allgemeine Merkmal der Schichtzugehörigkeit bedingt. Ist dies negative Ergebnis durch die Beschränkung auf jeweils nur 8 (bzw. 9) Paare bedingt, so daß bei einer Zusammenfassung aller 33 Paare doch noch signifikante Schichtunterschiede auftreten?

Die Wilcoxon-Tests der Gesamtmaße für die gesamte Experimentalgruppe belegen das Fehlen von Schichtunterschieden so eindrucksvoll wie die Tests innerhalb der Experimentalgruppen. Von den Tests mit den gesamten 33 Paaren zeigt kein einziges Maß im Mündlichen und im Schriftlichen auch nur annähernd signifikante, ähnliche Ergebnisse. Bei auffallend vielen der Tests dagegen liegt die Anzahl der Paare, in denen der Mittelschichtpartner dem Unterschichtpartner überlegen ist, ziemlich nah an der halben Zahl der überhaupt getesteten Paare. Die Haupthypothese unseres Experiments, daß die Mittelschichtkinder den Unterschichtkindern in der aktiven Worterklärung überlegen sind, auch wenn in der passiven Wortkenntnis gleich gute Kinder beider Schichten verglichen werden, kann also nicht bestätigt werden. Es

75-----

Die Ergebnisse der Wilcoxon-Tests sind für die negativen und positiven Gesamtmaße getrennt dargestellt. Die Richtung des Maßes wurde jedoch bei der Richtung der Differenzen berücksichtigt. Bei den positiven Maßen ist also in den Tabellen als Nenner des Quotienten die Anzahl der Paare angegeben, in denen die Mittelschicht einen größeren, bei den negativen, in denen die Mittelschicht einen kleineren Wert hatte, so daß in beiden Fällen die gleiche Hypothese getestet wurde und beide Tabellen, 47 wie 48, auf die gleiche Weise gelesen werden können.

bestehen keine Unterschiede zwischen den beiden Schichten, die sich auf unabhängig von sprachlichen Kenntnissen wirksame schichtspezifische Arten des Sprachgebrauchs zurückführen lassen. Die Theorie der linguistischen Kodes findet keine Bestätigung. Wir müssen nach Ursachen für das Scheitern der Theorie und alternativen Erklärungen suchen. Eine Ursache könnten Fehler bei der Wahl der Kontrollvariablen gewesen sein. Die passive Wortkenntnis könnte zwar theoretisch korrekt zur Kontrolle ausgewählt, aber ungünstig operationalisiert worden sein, oder es könnte überhaupt eine theoretisch nicht angemessene Leistung zur Kontrolle ausgewählt worden sein. Beide Möglichkeiten sollen im folgenden anhand von Wilcoxon-Tests mit alternativen Kontrollvariablen geprüft werden.

5.2.2. Schichtunterschiede bei Kontrolle verschiedener Intelligenz- und Schulleistungen.

Die dritte Serie der Wilcoxon-Tests, mit denen Gesamtmaßen jedes Individuums aller Experimentalgruppen für die mündliche und schriftliche Messung getrennt, wurde für die negativen Gesamtmaße auch noch mit anderen Kontrollvariablen als dem Wortschatztest-z-Wert gerechnet. Die Kontrolle des passiven Wortschatzes war notwendig, um evtl. auftretende Unterschiede in der aktiven Worterklärung zwischen den Schichten eindeutig auf schichtspezifische Strategien des Sprachgebrauchs zurückzuführen. Die alternativen Kontrollvariablen haben wir in zwei Gruppen eingeteilt, je nach dem ob sie den Wortschatztest als Kontrolle der passiven Wortkenntnis unmittelbar oder nur vermittelt ersetzen können; unter diesem Gesichtspunkt haben wir die verbalen Testleistungen und die Schulnoten den nicht-verbalen Testleistungen gegenübergestellt. Mit der ersten Gruppe von Kontrollvariablen vergleichen wir lediglich alternative Operationalisierungen der zu kontrollierenden passiven Wortkenntnis; dies gilt für die verbalen Testleistungen natürlich noch mehr als für die Schulnoten. Mit der zweiten Gruppe von Kontrollvariablen prüfen wir jedoch eine neue aus der Bernsteinschen Theorie abgeleitete Hypothese: Wenn wir mit Bernstein (1960) annehmen, daß in der

Unterschicht die nichtverbalen Intelligenz-Testleistungen, besonders im oberen Intelligenzbereich, zunehmend über den verbalen Testleistungen liegen, so müßte eine Kontrolle nicht-verbaler statt verbaler Intelligenztestleistungen bei der Untersuchung verbaler Leistungen die Unterschicht gegenüber der Mittelschicht benachteiligen. Wir können also bei der Kontrolle der nicht-verbalen Intelligenz eher eine Überlegenheit der Mittelschichtkinder erwarten als bei der Kontrolle der verbalen Intelligenz. - Von den beiden Gruppen alternativer Kontrollvariablen korreliert die erste natürlich stärker mit der ursprünglichen Kontrollvariable WST; wir werden zu jeder weiteren Kontrollvariablen ihre Korrelation mit dem WST angeben (siehe Tab. 21).

Zwei Besonderheiten dieser zusätzlichen Tests müssen beachtet werden. Erstens konnten die Paare für die übrigen Kontrollvariablen nicht wie für den Wortschatz-Wert aus dem gesamten ursprünglichen Sample ausgewählt werden, sondern nur aus dem schon unter Berücksichtigung des Wortschatz-Werts als Kontrollvariablen gebildeten Experiment-sample. Die Möglichkeit der unabhängigen Prüfung weiterer Einflüsse ist dadurch stark eingeschränkt, da immer die gleichen 66 Vpn. in wechselnden Kombinationen von 33 Paaren verglichen werden.¹⁾ Die Differenzen der Paare auf den anderen Kontrollvariablen sind wahrscheinlich größer als auf dem Wortschatz-Wert. Zweitens können sich innerhalb des Samples signifikante Schichtunterschiede auf den weiteren Kontrollvariablen ergeben haben (siehe Tab. 19 und 20). Dies ist für den TEST 5 und die Schulnoten in Heimatkunde ($p < .12$) auch tatsächlich der Fall: beide Male ist die Mittelschicht der Unterschicht überlegen, so daß bei dieser Kontrollvariablen die Mittelschicht begünstigt ist.

Bei Kontrolle der zweiten Gruppe von Variablen: GESINT (.28), NONVER (.04), TEST 3/4 (.21 bzw. .16), TEST 8 (.12) (Tab. 50)

¹⁾ Unter dieser Überlegung sind die durchweg negativen Ergebnisse dieses Abschnitts (Tab. 49 und 50) erstaunlich. Die Vergleiche der 66 Vpn. in insgesamt 10 Paarkombinationen vergrößern die Chance, daß zufällig Unterschiede auftreten.

zeigen sich wenige auch nur unter dem 10%-Niveau signifikante Differenzen, und nur ein über mündlich und schriftlich signifikantes Ergebnis: auf dem Mittelwert der A-Kategorisierung für Verständnis sind bei Kontrolle der nicht-verbale Intelligenz (NONVER) die Mittelschichtkinder den Unterschichtkindern überlegen. Für die aus Bernsteins Theorie abgeleitete zusätzliche Hypothese, daß in verbalen Leistungen Kinder der Mittelschicht Kindern der Unterschicht stärker bei Kontrolle nicht-verbaler als bei Kontrolle verbaler Fähigkeiten überlegen sind, findet sich keine Bestätigung.

Bei Kontrolle der ersten Gruppe von Variablen: VERBIT (.59), TEST 1/2 (.58), TEST 5 (.33), DEUMUE (.50), HEIMKU (.36) (Tab. 49) ergeben sich überhaupt keine im Mündlichen wie im Schriftlichen unter dem 10%-Niveau signifikanten Unterschiede. Dies ist besonders für die beiden Schulnoten ein interessantes Ergebnis. Sie sind wichtig als Indikator für die in der sozialen Realität wirksame Einschätzung verbaler Fähigkeiten, im Gegensatz zu der "objektivierten" Messung in den Tests. Jeder Unterricht beruht auf verbaler Verständigung, so daß auch in die Benotung die Fähigkeit des verbalen Ausdrucks sehr stark als für den Lehrer bewußtes oder unbewußtes Kriterium einfließen wird; insofern sind alle Schulnoten, nicht nur die der sprachlichen Fächer bedeutsam. Daß auch bei Kontrolle der Schulleistungen keine Unterschiede zwischen den Kindern beider Schichten in der aktiven Erklärung auftreten, gibt folgenden Hinweis. Wenn der Lehrer tatsächlich spezifische sprachliche Fähigkeiten der Mittelschichtkinder in bessere Schulnoten ummünzt, müßte man vermuten, daß die Unterschichtkinder gleiche Schulnoten nur durch außergewöhnliche sprachliche Anstrengungen erreichen, und bei einer Parallelisierung nach Schulnoten bessere Erklärungen der Unterschichtkinder erwarten. Daß sie nicht auftreten, spricht dafür, daß die Lehrer sprachliche Fähigkeiten entweder in einer für beide Schichten gleichen Weise oder überhaupt nicht bei der Notengebung berücksichtigen.

Auch die Berücksichtigung anderer Kontrollvariablen als des passiven Wortschatztests kann an dem negativen Hauptergebnis innerhalb unserer Untersuchungsgruppe von 33 Paaren nichts ändern. Auch bei alternativen Operationalisierungen der Wortkenntnis ergeben sich keine Schichtdifferenzen. Insbesondere lassen sich zwei theoretisch sinnvolle Vermutungen nicht bestätigen: eine Kontrolle der nichtverbalen Intelligenz führt nicht zu einer Überlegenheit der Mittelschichtkinder und eine Kontrolle der Schulnoten führt nicht zu einer Überlegenheit der Unterschichtkinder. Die Suche nach den Ursachen für das Scheitern der Theorie der linguistischen Codes wird daher auf eine genauere Untersuchung der Variablen der aktiven Worterklärung in Abhängigkeit von dichotomer Schicht (Abschnitt 5.4.) und auf eine genauere Untersuchung der sozialen Bedingungsvariablen (Abschnitte 5.6. bis 5.8.) verschoben. Zuvor aber müssen die Ergebnisse der Varianzanalyse zum Einfluß der sozialen Schicht mit denen des Wilcoxon-Tests verglichen und die Ergebnisse der übrigen im Experimentalaufbau verwandten Variablen dargestellt werden.

5.3. Einflüsse der dichotomen Schichteinteilung, des Geschlechts, der Kommunikationsformen und der Experimental- bzw. Wortgruppen auf die Gesamtleistung der zweiten Messung nach der Varianzanalyse.

5.3.1. Überblick über die Einzeleffekte und Interaktionen der Faktoren: Varianzanalysen aller Experimentalfaktoren.

Die Varianzanalysen sind im Vergleich zu den Wilcoxon-Tests - wie in Abschnitt 5.1. erläutert - dem Aufbau und dem Meßniveau des Experiments weniger angemessen: Sie setzen im Grunde Intervallskalennmessung voraus und nehmen eine Normalverteilung der Maße in der Grundgesamtheit an. Wird die Varianzanalyse auf alle Faktoren unseres Experiments - dichotome Schicht, Geschlecht, Kommunikationsform und Experimental- bzw. Wortgruppen - angewandt, so tritt ferner das Problem der Unabhängigkeit der Fehleranteile der in den beiden Kommunikationsformen wiederholten Messungen auf. Als pragmatischen Kompromiss haben wir hier einmal die Möglichkeit zwischen mündlicher und schriftlicher Messung korrelierter Werte wegen ihres geringen quantitativen Gewichts ignoriert und statt der korrekten Varianzanalyse mit "wiederholten Messungen" auf dem Faktor Kommunikationsform eine normale Vier-Weg-Varianzanalyse gewählt, und zweitens zwei Drei-Weg-Varianzanalysen für die mündliche und schriftliche Messung getrennt gerechnet.

Die Vier-Weg-Varianzanalyse mit Schicht, Geschlecht, Kommunikations- und Wortgruppen als Faktoren (Tab. 52) bzw. die beiden entsprechenden Drei-Weg-Varianzanalysen innerhalb der Kommunikationsformen (Tab. 53) stellen zunächst einen Vorblick auf alle folgenden Analysen einzelner Faktoren dar; in ihnen sind alle Informationen zusammengefaßt, die in den folgenden Analysen genauer betrachtet werden. Die für beide Arten von Varianzanalysen identischen Mittelwerte und Standardabweichungen der Gruppen sind in Tab. 51 dargestellt.

Das auffälligste Ergebnis ist das völlige Fehlen signifikanter Schicht- und das fast völlige Fehlen signifikanter Geschlechtsdifferenzen (bis auf die Maße MIBVE: $p < .044$ und MICVE: $p < .022$). Auf der anderen Seite ergeben sich signifikante Wortgruppen- und signifikante Wortgruppen-Kommunikationsform-Interaktionseffekte auf durchweg denselben Variablen. Auf den ersten Blick finden wir statt der gesuchten "inhaltlichen" nur "formale" Effekte. Die beiden der Vier-Weg-Varianzanalyse entsprechenden, für die mündlichen und schriftlichen Maße getrennten Varianzanalysen bestätigen das Ergebnis. Zwischen den Schichten und jetzt auch zwischen den Geschlechtern ergibt sich keine einzige signifikante Differenz, während der Experimentalgruppeneffekt auf den gleichen Variablen bis auf PEXAT wieder auftaucht. Weiterhin bringt, wenn wir für den Vergleich der "inhaltlichen" und "formalen" Effekte den weiteren Analysen vorgreifen, der Vergleich der Experimentalgruppen unter Ausschaltung der Schicht- und Geschlechtsdifferenzierung für mündlich und schriftlich getrennt auf fast allen Maßen hochsignifikante Effekte (Tab. 62), während umgekehrt der Vergleich der Schichten und Geschlechter unter Nichtbeachtung der Experimentalgruppen keine Schichtunterschiede, und Geschlechtsunterschiede nur auf den beiden schon in der Vier-Weg-Varianzanalyse signifikanten Variablen zeigt (Tab. 57). Soweit geben die beiden Arten von Varianzanalysen mit allen Faktoren einen Überblick über die in den folgenden Abschnitten genauer zu analysierenden Einzeleffekte.

Die spezifischen Aussagen dieser Analysen beziehen sich jedoch auf die Interaktion der Einzeleffekte. Hierzu zeigt sich, daß in den Drei-Weg-Varianzanalysen kein einziger Interaktionseffekt signifikant ist und kaum F-Werte über 1 auftreten. Ebenso zeigen in der Vier-Weg-Varianzanalyse keine von den Interaktionen zweiter und dritter Ordnung und nur eine von den Interaktionen erster Ordnung signifikante Effekte auf irgendeinem Maß; die Darstellung der F-Tests für diese Effekte wurde daher in Tab. 52 einer besseren Übersichtlichkeit geopfert. Für die Interaktion von Kommunikationsform und Wortgruppen findet sich bei mehreren Maßen ein signifikanter Effekt. Insgesamt also können, nach den Ergebnissen zu den Interaktionseffekten, die verschiedenen Einzeleffekte in den folgenden Abschnitten isoliert betrachtet werden. Der eine signifikante Interaktionseffekt ist jedoch

gleichen Vpn.-Gruppe mit dem Index der Experimentalgruppe gekennzeichnet sind. Zudem entsprechen die Unterschiede zwischen den Experimentalgruppen auf der Experimentalleistung Unterschieden auf der Kontrollvariablen. Für die Gruppe 1 (Wortschatztest-z-Wert 4.35) und 2 (Wortschatztest-z-Wert 5.24) hält sich das Muster der über Kreuz vertauschten Werte selbst auf den Variablen durch, für die der untersuchte Interaktionseffekt nicht signifikant ist. Für die Gruppe 3 (Wortschatztest-z-Wert 4.79) und (Wortschatztest-z-Wert 4.72) gibt es zwar auch in vielen Zellen über Kreuz vertauschte Werte; doch sind hier die Leistungsdifferenzen zwischen den Gruppen relativ gering (entsprechend den geringen Unterschieden auf der Kontrollvariablen) und die schriftlichen Leistungen unabhängig von der Wortform im allgemeinen besser. Insgesamt also ist der Interaktionseffekt zwar durch die Experimentalanordnung (d.h. die Kontrollvariable und die Wortgruppenzuweisung) bedingt, die sich aber stärker innerhalb der Gruppen mit der Testform A bemerkbar macht. - Ein zweiter Hinweis dafür, daß der Interaktionseffekt "WK" und Wortgruppeneffekt der Tab. 52 bzw. der Experimentalgruppeneffekt der Tab. 53 sich aus den gleichen Bedingungen ergeben, liegt in der Tatsache, daß beide Effekte auf den gleichen Variablen signifikant auftreten, nicht nur innerhalb Tab. 52, sondern auch beim Vergleich der Tab. 52 mit Tab. 53, 55 und 62.¹⁾

¹⁾-----
Die einzige Ausnahme bildet PEXAT, die allein in Tab. 52 wohl zufällig unter dem 5%-Niveau signifikante Wortgruppenunterschiede zeigt, sonst aber in keiner der Varianzanalysen.

5.3.2. Einflüsse der dichotomen Schichteinteilung.

Bei der Trennung von Explikation und Verständnis hatten wir angenommen, daß die Unterschicht in der Explikation schlechter, im Verständnis aber beide Schichten gleich und entsprechend die durchschnittlichen Differenzen für die Unterschicht größer seien. Zur Überprüfung stehen uns, neben den beiden schon dargestellten Analysen in Tab. 52 und 53, zwei Varianzanalysen mit dichotomer Schicht und den Experimentalgruppen in der mündlichen bzw. schriftlichen Messung (Tab. 55) sowie eine Varianzanalyse mit Schicht, Geschlecht und Kommunikationsform (Tab. 57) zur Verfügung. Da in den Varianzanalysen mit allen Faktoren die Experimentalgruppe sich als der hervorragende Effekt erwiesen hat, haben wir den Schichteffekt nicht mehr allein, sondern nur noch in einer Zwei-Weg-Varianzanalyse mit den Experimentalgruppen getestet.

Keine der genannten Hypothesen wird in irgendeiner der durchgeführten Analysen auf irgendeinem Maß signifikant bestätigt, weder bei den für mündlich und schriftlich getrennt durchgeführten Varianzanalysen (Tab. 53, 55), noch bei den Varianzanalysen mit der Kommunikationsform als Faktor (Tab. 52, 57). Ferner ergeben sich auch hier keine signifikanten Interaktionen der dichotomen Schicht, weder mit dem Faktor Geschlecht (Tab. 57) noch mit den Experimentalgruppen (Tab. 54). Gleichsam hinter diesen grundsätzlich negativen Ergebnissen lohnt es sich jedoch, nach der Richtung und dem Strukturbild der Differenzen bei allen Maßen zu suchen.

Schon in der nicht die Kommunikationsformen trennenden Vier-Weg-Varianzanalyse (Tab. 52) schält sich eine sönnvolle Struktur stärker schichtkritischer Maße heraus. Während für kein Maß der B- und C-Kategorisierung ein F-Wert über 1 auftaucht, ergeben sich für die Maße der A-Kategorisierung überall Signifikanzen unter .30 und in den meisten Fällen zwischen .10 und .15. Auf der einen Seite - der Maße, die noch relativ schichtkritisch sind (MIAVE, MIAEX, GATEX, RATEX, S12VE, S12EX, S13VE, S13EX) - wurde die Leistung in einem festen quantitativen Rahmen bewertet, der durch die Anzahl der Worte in den Wortforme

vorgegeben war. Auf der anderen Seite - der Maße, die sich als völlig schichtneutral erwiesen (MIBEX, MICEX, MIBVE, MICVE, PEXAT, PVEAT) - wurde die Leistung nur innerhalb des von der Vp. selbst gesetzten Rahmens versuchter Antworten bewertet. Wir können in der ersten Gruppe die Maße der Faktoren I und III der Faktorenanalyse der Maße (Tab. 15) bzw. die Maße der Antworthäufigkeit und -güte (Tab. 7), in der zweiten Gruppe die Maße der Faktoren II und IV bzw. die Maße der Antwortgüte wiedererkennen. Unter Bedingungen der Aufgabenwahl, die sie sich selber setzen nach dem Maß ihrer bewußt gewordenen verbalen Erfahrung oder ihrer Risikofreude, erklären Kinder beider Schichten sprachlich gefaßte Konzepte gleich gut; unter Bedingungen, in denen die Aufgabenmenge von außen festgesetzt ist und das Akzeptieren dieser Menge selber bewertet, ein Zurückstecken also sanktioniert wird, zeigen sich Unterschiede zwischen den Kindern beider Schichten.

Die für jede Kommunikationsform spezifischen Varianzanalysen (Tab. 53) reproduzieren dies Ergebnis zwar für die mündliche Messung mit leicht verminderten Signifikanzen, nicht aber für die schriftliche Messung: hier ergab sich für kein einziges Maß mehr ein F-Wert über 1. Offensichtlich stellt in der Unterschichtsguppe die Aufforderung zur schriftlichen Definition für die weniger ihrer Kenntnisse gewissen und zurückhaltenderen Kinder einen Ansporn zu stärkerer Konzentration und hartnäckigerem Versuchen dar.¹⁾ In schriftlichen Äußerungen ist ja die direkte Sanktionsmöglichkeit eines zumindest vorgestellten zuhörenden Partners stark eingeschränkt und in jedem Fall bis zum von der Vp. selbst festgelegten Abschluß der Erklärung aufgehoben. Der Vergleich der Varianzanalyse mit Schicht, Geschlecht und Kommunikationsform (Tab. 57) mit den beiden nach Kommunikationsform getrennten Varianzanalysen mit Schicht und Experimentalgruppen

1) -----
Außerdem findet sich dies Ergebnis in den Unterschieden der schichtspezifischen Interkorrelation wieder. Während in der Mittelschicht die Maße der A- mit denen der B- und C-Kategorisierung im Mündlichen wie Schriftlichen etwa gleich hoch interkorrelieren, sind in der Unterschicht dieselben Interkorrelationen im Schriftlichen etwa gleich hoch wie in der Mittelschicht, im Mündlichen jedoch niedriger (siehe Abschnitt 1.2.5.).

als Faktoren (Tab. 55) zeigt noch einmal dieses Muster: die relativ signifikanten Effekte auf den Maßen der Antwortgüte bleiben bei der gesonderten Betrachtung der Kommunikationsformen nur für die mündliche Messung bestehen.

Jenseits der konventionellen Signifikanzniveaus konnte also durch die F-Tests ein Muster von Schichtunterschieden festgestellt werden, das auf der Dimension der Kommunikationsform und der Maß-Art variiert. Dieses Muster kann an den Differenzen der Maße für die beiden Schichtgruppen insgesamt abgelesen werden, die in Tab. 54 und 56 wiedergegeben sind. Es müßte sich auch bei den Schichtdifferenzen innerhalb der Geschlechts- wie innerhalb der Experimentalgruppen wiederfinden: (1) Die Maße der Antworthäufigkeit und -güte müssen Differenzen zugunsten der Mittelschicht aufweisen, (2) die Differenzen zugunsten der Mittelschicht müssen bei den Maßen der A-Kategorisierung größer sein als bei den Maßen der B- und C-Kategorisierung, (3) innerhalb der Maße der Antworthäufigkeit und -güte müssen die Schichtdifferenzen der schriftlichen Maße kleiner sein als die der mündlichen. Diese drei Annahmen wollen wir innerhalb der Geschlechts- und Experimentalgruppen prüfen.

Der Vergleich der Schichten innerhalb der Geschlechter (Tab. 56) zeigt nur minimale absolute Differenzen. Immerhin ist innerhalb beider Geschlechtsgruppen die Mittelschicht auf den Maßen der A-Kategorisierung sowie GATEX und RATEX überlegen, die Annahme (1) also weitgehend erfüllt. Auf den Maßen der B- und C-Kategorisierung zeichnet sich, wenn auch minimal, eine interessante Umkehrung ab: Während die Jungen der Mittelschicht, nach ihrer besseren Leistung in den A-Maßen, jetzt gegenüber ihren Geschlechtsgenossen aus der Unterschicht zurückfallen, halten die Mädchen der Mittelschicht auch in dieser strengeren und, wie gezeigt, der Unterschicht günstigeren Bewertung ihre Überlegenheit durch - jedoch ebenfalls nur gegenüber ihren Geschlechtsgenossinnen. Denn, wie ein Vergleich mit den schichtinternen Geschlechtsdifferenzen ergibt, liegt dies nicht an einer Überlegenheit der Mädchen über die Jungen innerhalb der

Mittelschicht. Vielmehr gilt die allgemeine Überlegenheit der Jungen für die Unterschicht im stärkeren Maße. Die Vermutung, daß die Sozialisation der Mädchen und die Sozialisation in der Mittelschicht in gleicher Weise sprachliche Fähigkeiten fördern und beide Effekte kumulativ wirken können, findet hier keine Bestätigung. Die Jungen der Unterschicht, für die man den schlechtesten Platz erwarten würde, sind auf der B- und C-Kategorisierung meist die beste Gruppe, während die Mädchen der Mittelschicht, für die man den besten Platz erwarten würde, hinter den Jungen der Mittelschicht liegen.

Der Vergleich der Schichten innerhalb der Experimentalgruppen (Tab. 54 sowie Abb. 6) zeigt, daß die Annahme (1) durchweg in allen Gruppen, außer einigen Maßen der schriftlichen Messung der Gruppen 2 und 3, erfüllt ist. Die Annahme (2) ist ebenfalls in allen Gruppen, außer der Gruppe 1 wiederum auf den schriftlichen Maßen, erfüllt. Die Annahme (3) gilt für alle Experimentalgruppen außer der Gruppe 4. Das allgemeine Muster reproduziert sich in allen Experimentalgruppen im mündlichen vollständig, im schriftlichen teilweise; das entspricht genau dem in den Signifikanzniveaus vorgezeichneten Muster, denn bei den schriftlichen Maßen war wegen der geringeren Signifikanzen auch eine geringere Stabilität der Schichtdifferenzen zu erwarten. - Bezüglich der Größe der Schichtdifferenzen sticht die Experimentalgruppe 4 von den übrigen Gruppen ab. In ihr zeigen sich die größten Differenzen zugunsten der Mittelschicht, sowohl auf der A-Kategorisierung, wie auch der B- und C-Kategorisierung, auf der in allen übrigen Gruppen (mit der nur im Schriftlichen geltenden Ausnahme der Gruppe 1) die Unterschicht die Mittelschicht übertrifft. Eine Erklärung dafür ist weder in der Testform (Gruppe 3 zeigt die gegenteiligen Tendenzen) noch in der Kontrollvariablen zu finden.

Wir können also die Überlegenheit der Mittelschichtkinder in der mündlichen Messung, solange die Selbsteinschätzung der Wortkenntnis nicht mitbewertet wird, und die Gleichwertigkeit der Erklärungsleistungen von Kindern beider Schichten in den übrigen Fällen trotz der fehlenden Signifikanz als ein hinlängliches stabiles Muster ansehen. Offensichtlich sind für Unterschiede zwischen den Schichten

nur die Kommunikationsform und die Selbsteinschätzung der Vp. (A. vs. B- und C-Kategorisierung), nicht jedoch die Trennung unserer beiden Dimensionen "Verständnis-Explikation" relevant. Zwischen diesen beiden Gruppen von Maßen variieren die Schichtunterschiede nur gering; aber für alle Wort- und Experimentalgruppen zusammen (Tab. 54) sind die Differenzen auf den B- und C-Maßen der Explikation zugunsten der Unter-, auf den B- und C-Maßen des Verständnisses zugunsten der Mittelschicht. Dies den Erwartungen widersprechende Ergebnis legt nahe, die beiden Dimensionen der Maße im Hinblick auf ihre "Schichtanfälligkeit" zu betrachten.

Die Auswertung nach "Verständnis", die konzipiert war, um vermutliche Defizite der Unterschichtkinder in der bündigen Darbietung bei tatsächlich gleichwertiger Kenntnis des Gemeinten herauszuschälen, scheint eher den Mittelschichtkindern entgegenzukommen. Andererseits werden die Intentionen der Trennung "Verständnis-Explikation", folgt man dem Muster der Differenzen auf den Variablen, offensichtlich von der Differenzierung der drei Kategorisierungsarten übernommen, nicht mehr innerhalb jedes einzelnen Wortes, sondern für die Gesamtheit der Worte einer Wortgruppe. Jedes Kind trifft bei jedem Wort die Entscheidung, ob es das Wort "verstanden" hat und beginnt dann zu "explizieren". Wenn die vom Kind selbständig getroffene Entscheidung - wie in der B- und C-Kategorisierung - in die Bewertung eingeht, wird sie um die unterschiedlichen Verständnisausgangspunkte bereinigt und die Fähigkeit der Explikation, so wie sie jedes Kind bei seinem individuellen Verständnis realisieren kann, beurteilt. In diesem Sinne können wir die Gleichheit beider Schichtgruppen auf den Maßen der B- und C-Kategorisierung und die Überlegenheit der Mittelschichtkinder auf den Maßen der A-Kategorisierung so interpretieren, als hätten die Kinder beider Schichten über ein

gleich gutes Verständnis verfügt, die Mittelschichtkinder aber über ihr Verständnis hinausgehende Explikationen angestrebt. Zwar hat die Trennung "Verständnis-Explikation" nicht gleich gute implizite Konzepte der Unterschichtkinder herausholen, aber die Differenzierung der Kategorisierungsarten über das tatsächliche Verständnis überschüssige Explikationen bei den Mittelschichtkindern nachweisen können. Auch rein von der Logik der Maße her ist diese stärkere Betonung der Differenzierung nach den Kategorisierungsarten als nach den qualitativen Dimensionen geboten; Verständnis und Explikation mußten von der Auswertung her enger zusammenhängen als die durch die Wirkung einer relativ unabhängigen Variablen, Anzahl der Antworten, voneinander getrennten Arten der Kategorisierung (siehe Tab. 9).

Die Überlegenheit der Kinder aus der Mittelschicht auf den Maßen der A-Kategorisierung in der mündlichen Messung kann zudem mit den Überlegungen interpretiert werden, die hinter den Hypothesen des Experiments stehen (siehe 3.). Wenn die Kinder der Mittelschicht mit größerem Risiko als die Kinder der Unterschicht Erklärungen von Wortbedeutungen anstreben, ohne tatsächlich bessere Erklärungen zu geben, so haben wir ein Indiz für eine schichtspezifische Art des Sprachgebrauchs, das mit einem größeren aktiven trotz Gleichheit des passiven Wortschatzes vergleichbar ist. Auch daß die Unterschiede zwischen den Maßen nur in der mündlichen Messung auftreten, kann innerhalb der Theorie der linguistischen Codes interpretiert werden: Da - im Falle der Erklärung von Wortbedeutungen - die schriftliche Form das Kind zum Bedenken der eigenen Wortkenntnis zwingt, können schichtspezifische Strategien risikvoller Erklärung sich eher im Mündlichen als im Schriftlichen durchschlagen.

Wieweit läßt dieses durch die Varianzanalyse aufgedeckte Muster zwischen verschiedenen Maßen variierender Schichtdifferenzen sich nachträglich auch in den Wilcoxon-Test entdecken? Da in keinem der beiden Prüfverfahren signifikante Unterschiede auftauchen, kann

sich der Vergleich zwischen ihnen nur auf die Art und Weise beziehen, in der die Einzelwerte der Vpn. verrechnet werden. Während für die Varianzanalyse das Mittel der absoluten Werte für jede Gruppe errechnet wird, werden beim Wilcoxon-Test Paare von Vpn. aus beiden Gruppen verglichen und die Differenzen nach ihrer absoluten Rangfolge bewertet: Unterschiede der Mittelwertdifferenzen zwischen den Maßen müssen sich in Unterschieden der für die Mittelschicht sprechenden Paare wiederfinden, d.h. in einem sehr vergrößerten Maß. Solche Unterschiede zwischen den Maßen lassen sich bei den Wilcoxon-Tests für die mündliche Messung (Tab. 47 und 48) noch finden: bei der Zusammenfassung aller Experimentalgruppen liegt die Anzahl der für die Mittelschicht günstigen Paare bei den Mittelwerten der A-Kategorisierung über denen der B- und C-Kategorisierung. Wir finden also auch bei den den Wilcoxon-Tests zugrundeliegenden gröberen Berechnungen noch Anhaltspunkte für die Variation der Schichtdifferenzen zwischen den Maßen der mündlichen Messung.

Zusammenfassung: Auch die Varianzanalyse konnte, sowenig wie der Wilcoxon-Test, die Signifikanz von Unterschieden zwischen den beiden Schichtgruppen bei kontrolliertem passiven Wortschatz aufdecken. Durch ihre genauere Verrechnung der Werte konnte sie jedoch minimale Variationen der Mittelwert-Schichtdifferenzen zwischen den Maßen aufzeigen. In der mündlichen Messung zeigen die Mittelschichtkinder auf den Maßen bessere Leistungen, die die Antworthäufigkeit und die Antwortgüte bewerten; in der schriftlichen Messung verschwinden auch diese minimalen Schichtunterschiede. Hatten wir ursprünglich angenommen, daß in der Explikation die Kinder der Mittelschicht den Kindern der Unterschicht überlegen und im Verständnis Kinder beider Schichten gleich gut waren, so zeigte sich statt dessen, daß die Kinder der Mittelschicht risikofreudiger Erklärungen wagen, ohne tatsächlich bessere Erklärungen zu geben.

5.3.3. Einflüsse des Faktors Geschlecht.

Wenn auch widersprüchliche Ergebnisse vorliegen, so nimmt man allgemein eine Überlegenheit der Mädchen in sprachlichen Leistungen an. Die größere sprachliche Gewandtheit soll ein besonderer Fall der allgemein größeren sozialen Konformität der Mädchen sein, nicht unbedingt ein Ausdruck differenzierterer Kategorisierung; sie betrifft die sozial-konventionelle Seite der Sprache, nicht ihre kognitiv-instrumentelle. Für unser Experiment, in dem Explikation und Verständnis getrennt und absoluter Umfang wie kognitive Durchgliederung des Wortschatzes in den Maßen unterschiedlich gewichtet wurden, müssen wir daher (1) eine Überlegenheit der Mädchen in der Explikation und eine Gleichwertigkeit der Geschlechter im Verständnis und (2) eine Überlegenheit der Mädchen auf den Maßen der A-, nicht aber der B- und C-Kategorisierung annehmen.

In der Vier-Weg-Varianzanalyse (Tab. 52) zeigen sich Geschlechtseffekte auf den Maßen der B- und C-Kategorisierung - signifikant unter dem 5%-Niveau für die Verständnismaße, knapp über ihm für die Explikationsmaße. Jedoch bleiben für kein einziges Maß signifikante Differenzen, wenn man die entsprechenden Drei-Weg-Varianzanalysen ohne die Kommunikationsform (Tab. 53) heranzieht. Die Drei-Weg-Varianzanalyse mit Schicht, Geschlecht und Kommunikationsform (Tab. 57) zeigt wieder die gleichen Effekte auf den Maßen der C- und, etwas geringer, der B-Kategorisierung. Wir finden also im Mündlichen und Schriftlichen gleiche Differenzen auf den Maßen der Antwortgüte, die jedoch nur bei Zusammenfassung der Kommunikationsformen signifikant werden. Zudem werden in den hier dargestellten Varianzanalysen mit der Kommunikationsform als Faktor die innerhalb der Geschlechtsgruppen wiederholten Messungen der gleichen Vp. gegen eine Irrtumsvarianz wie bei unabhängigen Messungen getestet, so daß der Geschlechtseffekt leicht überschätzt wird. Wenn wir anhand der Mittelwerte die Richtung der Differenzen betrachten (Tab. 58), so zeigt sich die Überlegenheit der Jungen und eine Unabhängigkeit der Geschlechtsunterschiede von der Schichtzugehörigkeit: der Schicht-Geschlechtsinteraktionseffekt (Tab. 57) erzielt auf keinem Maß (außer PRIAT) einen F-Wert über 1; wenn überhaupt, übertreffen die Jungen die Mädchen in der Unterschicht

stärker. Im Vergleich mit der Kontrollvariablen setzen sich in der Mittel- wie in der Unterschicht die Jungen trotz einer schlechteren Ausgangslage gegen die Mädchen durch. Der Schluß indes, die Überlegenheit der Jungen würde wegen ihrer relativ geringeren passiven Wortkenntnis unterschätzt, ist an dieser Stelle voreilig; die Leistungen der Jungen und Mädchen werden ja in verschiedenen Experimentalgruppen verglichen, die sich bezüglich der Geschlechtsdifferenzen auf der Kontrollvariablen und den erfragten Worterklärungen unterscheiden. Den strengsten Prüfstein für die "Echtheit" der Geschlechtsdifferenz stellt daher der Vergleich mit dem Experimentalgruppeneffekt dar; ein starker Geschlechtseinfluß müßte sich in allen Gruppen abzeichnen.

Wenn wir die beiden für mündlich und schriftlich getrennten Varianzanalysen mit Geschlecht und Experimentalgruppe vergleichen, so ist ein signifikanter Schichteffekt nicht zu erwarten, da wir ja (wie in der Varianzanalyse in Tab. 53) den nur bei der Zusammenfassung signifikanten Effekt aufteilen. Dennoch kann mit diesen beiden Analysen die Stärke des Geschlechtsunterschieds in mündlicher und schriftlicher Messung verglichen und an der Stärke des Experimentalgruppeneffekts, für den ja die gleiche Teilung gilt, gemessen werden. Auch in Tab. 59 ist kein Geschlechtseffekt unter dem 5%-Niveau signifikant; nur im Mündlichen zeigen die Verständnismaße der B- und C-Kategorisierung noch unter dem 10%-Niveau signifikante Effekte; wenn überhaupt, setzt sich also die Überlegenheit der Jungen im Mündlichen durch. Fernerhin ergeben sich hochsignifikante Experimentalgruppeneffekte auf den Maßen der A-Kategorisierung (MIAEX, MIAVE, GATEX, RATEX) mit extrem hohen F -Werten im Schriftlichen; der Einfluß der Experimentalgruppen und des Geschlechtseffektors beziehen sich also auf unterschiedliche Aspekte der aktiven Worterklärung. Schließlich finden sich für die Interaktion von Experimentalgruppen und Geschlecht ein nahezu signifikanter Effekt bei MICVES, bei den übrigen Variablen jedoch nur F -Werte unter oder nahe bei 1; ein Wechsel der Richtung der Differenzen zwischen den Experimentalgruppen ist also nicht zu erwarten. Dennoch wollen wir die Mittelwertdifferenzen zwischen den Geschlechtsgruppen in den Experimentalgruppen und Kommunikationsformen betrachten und mit den Differenzen auf der Kontrollvariablen vergleichen (Tab. 58).

Zunächst zeigt sich eine weitgehende Übereinstimmung der Geschlechtsdifferenzen zwischen Mündlich und Schriftlich für jedes Maß; das gilt sowohl für die Zusammenfassung aller Experimentalgruppen wie für jede einzeln und besonders für die Maße der Antwortgüte, bei denen ja am ehesten signifikante Geschlechtseffekte auftraten. Die Gleichheit der Geschlechtsdifferenzen in den beiden Kommunikationsformen ist ein erstes Argument für einen tatsächlichen Einfluß des Geschlechtsfaktors. - Wenn wir ferner die Geschlechtsdifferenzen, global für alle Maße, mit den Geschlechtsdifferenzen auf der Kontrollvariablen innerhalb jeder Experimentalgruppe vergleichen ergibt sich schematisch folgendes Bild:

<u>Experimentalgruppe</u>	<u>Mittelwert auf der</u> <u>Kontrollvariablen</u>			<u>Leistungen in der</u> <u>zweiten Messung</u>		
1	J	>	M	J	>>	M
2	J	<	M	J	>	M
3	J	>	M	J	≪	M
4	J	≪	M	J	>	M

Nur in der 1. Gruppe stimmen beide Differenzen überein, während man in den übrigen Gruppen von der Kontrollvariablen das gegenteilige Ergebnis für die Experimentalmessung hätte erwarten sollen. Für eine stärker quantitative Betrachtung sind die Differenzen auf beiden Seiten in der 3. Gruppe die geringsten, so daß wir ihren Beitrag für die allgemeinen Geschlechtsunterschiede ignorieren können. In der 1. Gruppe andererseits, in der die Jungen eine bessere Ausgangsposition auf der Kontrollvariablen aufweisen, sind die Differenzen der Experimentalmessung, so viel größer als in der 2. und 4. Gruppe, die eine schlechtere Ausgangslage der Jungen aufweisen, daß man aus dieser Diskrepanz erschließen kann, wieweit die Überlegenheit der Jungen in den Gruppen 2 und 4 tatsächlich unterschätzt wird. Die Überlegenheit der Jungen in der aktiven Worterklärung trotz einer schlechteren Ausgangsposition auf der Kontrollvariablen innerhalb aller Experimentalgruppen außer der Gruppe 3 ist ein zweites Argument für die tatsächliche Bedeutung des Geschlechtsfaktors für die aktive Worterklärung.

Wir können die Ergebnisse dieses Abschnittes kurz zusammenfassen. Die Jungen sind den Mädchen auf den Maßen der Erklärungsgüte (B- und C-Kategorisierung) generell überlegen. Einerseits ist dieser Unterschied nur dann ein statistisch signifikanter Effekt, wenn in einer nicht ganz korrekten Anwendung der Varianzanalyse die beiden Kommunikationsformen zusammengenommen werden. Andererseits tritt die Überlegenheit der Jungen in beiden Kommunikationsformen und in so gut wie allen Experimentalgruppen auf und setzt sich gegen ihre Unterlegenheit auf der Kontrollvariablen durch, so daß wir an Unterschieden zwischen beiden Geschlechtsgruppen in der aktiven Worterklärung festhalten werden.

Unsere Hypothesen allerdings wurden dabei nicht bestätigt: Entgegen der Erwartung besserer Leistungen der Mädchen auf den A-Maßen und gleicher Leistung der Geschlechter auf den B- und C-Maßen, ergaben sich auf den ersten keine signifikanten, und auf den zweiten Unterschiede zugunsten der Jungen. Entgegen der Erwartung besserer Leistungen der Mädchen in der Explikation und gleicher Leistung der Geschlechter im Verständnis, ergaben sich kaum Unterschiede in der Explikation, aber fast signifikante im Verständnis zugunsten der Jungen. In beiden Fällen wird unsere Vermutung nicht bestätigt, für die Mädchen habe die sozialkonventionelle Seite der Sprache relativ größere Bedeutung als die kognitiv-instrumentelle. Auf der anderen Seite sind die Ergebnisse nicht deutlich genug, um für die Jungen wegen ihrer besseren Leistungen im Verständnis und den Maßen der B- und C-Kategorisierung eine größere Bedeutung der kognitiv-instrumentellen Seite der Sprache zu belegen. Nachträglich erscheint die Verbindung der beiden theoretischen Sprachfunktionen mit unseren Meßdimensionen nicht mehr einleuchtend. Vermutlich lassen sich die überall recht widersprüchlichen Ergebnisse von Geschlechtsunterschieden auf Sprachvariablen nur noch beschreibend registrieren, und nicht mehr zu angenommenen Unterschieden in sprachlichen Fähigkeiten oder bevorzugten sprachlichen Funktionen in theoretische Beziehung setzen.

5.3.4. Einflüsse der Kommunikationsform

Bis jetzt haben wir mit den Schicht- und Geschlechtseffekten Einflüsse untersucht, die an die Experimentalaufgabe in gewissem Sinne von außen herangetragen wurden. In den beiden folgenden Abschnitten sollen mit den Effekten der Kommunikationsform (5.3.4.) und der Wortgruppen (5.3.5.) Einflüsse untersucht werden, die mit der Experimentalaufgabe unmittelbar verbunden sind. - Die Einführung des Faktors "Kommunikationsform" lag nahe, weil Sprechen und Schreiben ganz andere Anforderungen an Planung und Konzentration der Vpn. stellen: Während beim Sprechen ein Konzept ungeplant und gleichsam assoziativ ausgebreitet wird, zwingt das Schreiben zu einer Entwicklung des Konzepts zunächst "im Kopf" und anschließend einer planvollen Wiedergabe (siehe dazu u.a. Wygotsky 1969). Die Experimentalleistungen müßten sich in der schriftlichen gegenüber der mündlichen Messung verbessern. Der Wechsel zwischen den Kommunikationsformen verlangt einen "Kode-Wechsel", den die Kinder beider Schichten vermutlich nicht in der gleichen Weise bewältigen (Bernstein 1965). Die schriftliche Messung stellt höhere Ansprüche an die verbale Planung als die mündliche Messung und zugleich erzwingt sie eine höhere Leistung. Stellt man den ersten Aspekt heraus, so müßte man eine stärkere Leistungssteigerung der Mittelschichtkinder in der schriftlichen Messung voraussagen; stellt man den zweiten Aspekt heraus, so müßte man eine stärkere Leistungssteigerung der Unterschichtkinder in der schriftlichen Messung voraussagen.

Die beiden Kommunikationsformen wurden in wiederholten Messungen der gleichen Vpn. erhoben. Es ergaben sich nur minimale Mittelwertunterschiede zwischen der mündlichen und schriftlichen Messung auf allen Maßen (Tab. 8 und 56). In der Unterschicht sind die Leistungen der schriftlichen Messung für jedes Maß besser als in der mündlichen Messung; in der Mittelschicht finden wir Steigerung wie Rückgang der Leistung in der schriftlichen Messung, durchweg sind die Unterschiede geringer als in der Unterschicht (Tab. 56, Spalte 12 und 13). In den beiden bisher dargestellten Varianzanalysen mit der Kommunikationsform als Faktor (Tab. 52) und 57) ergab sich für kein Maß ein

signifikanter Effekt der Kommunikationsform, sowie für kein Maß eine signifikante Interaktion zwischen Schicht und Kommunikationsform. Beide Analysen wurden aber ohne Berücksichtigung der wiederholten Messungen gerechnet, so als stammten die mündlichen und schriftlichen Erklärungen einer Vp. von zwei Vpn. Deshalb sollen kurz Unterschiede zwischen Varianzanalysen mit und ohne Berücksichtigung von wiederholten Messungen am Beispiel einer Ein-Weg-Varianzanalyse der Kommunikationsform erläutert und an einigen Maßen die korrekte Form der Varianzanalyse gerechnet werden.

In den bisherigen Analysen wird der Varianzanteil der Kommunikationsform, der ja nur zwischen den Leistungen jeder einzelnen Vp. entstanden sein kann, gegen die Varianzanteile getestet, die auch durch die Leistungsunterschiede zwischen den Vpn. entstanden sind; die Möglichkeit, daß signifikante Unterschiede der Kommunikationsform auftreten können, wird dadurch verringert. In der Varianzanalyse mit Berücksichtigung der wiederholten Messungen dagegen, wird die durch Unterschiede innerhalb der Leistungen jeder Vp. verursachte Varianz aufgeteilt in systematische, durch die Experimentalbedingung "Kommunikationsform" produzierte, und residuale Varianz und die erste gegen die letzte getestet; die durch Unterschiede zwischen den Vpn. verursachte Varianz, die in der Regel entschieden größer ist als die Varianz innerhalb der Vpn., bleibt dabei völlig außer Betracht (Winer 1962:109). In beiden Formen ist die Varianz, die für den Faktor "Kommunikationsform" errechnet wird, die gleiche; ohne Berücksichtigung der wiederholten Messungen wird sie jedoch gegen eine Fehlervarianz getestet, die der residualen Varianz innerhalb der Vpn., erweitert um die Varianz zwischen den Vpn., in der Varianzanalyse mit wiederholten Messungen entspricht. In der Varianzanalyse ohne Berücksichtigung der wiederholten Messungen ist es daher viel weniger wahrscheinlich, daß ein wiederholt gemessener Effekt signifikant wird, als in einer Varianzanalyse mit wiederholten Messungen

In Tab. 60 sind Varianzanalysen mit Maßen aus allen drei der theoretisch konzipierten Maßdimensionen (siehe Tab. 7) zusammengefaßt. Auch bei Verwendung des korrekten statistischen Prüfmodells können wir nicht den geringsten Effekt der Kommunikationsform feststellen; zudem gilt dies für die Häufigkeit und die Güte der Erklärungen in der gleichen Weise. Ein weiteres Indiz für einen insgesamt schwachen Effekt der Kommunikationsform sind die geringen und inkonsistenten Unterschiede innerhalb der Experimentalgruppen (Tab. 61 und 63). Wenn wir uns allerdings - aus Gründen, die bei der Behandlung der Experimentalgruppen (5.3.5.) ausgeführt werden - auf die Gruppe 3 und 4 beschränken, so ergibt sich ein nahezu signifikanter (bei MIAEX sogar auf dem 5%-Niveau signifikanter) Effekt der Kommunikationsform.

Welche Erklärungen lassen sich für das Versagen der auf den ersten Blick so suggestiven Hypothese einer konzentrierenden Wirkung der schriftlichen Erklärung finden? Wir glauben, daß die vermutlichen Ursachen für eine unterschiedliche Wirkung der Kommunikationsform durch unsere Experimentalanordnung nicht mehr wirksam werden konnten. Bei den mündlichen Erklärungen scheint die akustische Isolierung von der Umwelt durch die Gegensprechanlage die Überprüfung des spontanen Ausdruckswillens mindestens ebenso stark gefördert zu haben wie der Zwang zur Niederschrift. Außerdem scheint, für Erklärungen von Wortbedeutungen wie vielleicht auch allgemein, durch die mündliche Kommunikationsform eine lockere und assoziative Darstellung nur dann provoziert zu werden, wenn sie tatsächlich an einen Gesprächspartner gerichtet werden können; solange dies nicht der Fall ist, scheinen Unterschiede der Kommunikationsform gegenüber der Einheitlichkeit der Erklärungsaufgabe irrelevant.

5.3.5. Einflüsse der Experimentalgruppen

Die Einführung des Faktors "Wort- bzw. Experimentalgruppen" ergab sich aus der Anlage des Experiments (siehe II.3.1.); ein theoretischer Unterschied zwischen den Wortgruppen ist nicht faßbar. In diesem Abschnitt soll der Einfluß der Experimentalgruppen dargestellt und nach Erklärungen gesucht werden. Abschließend soll der Effekt der Experimentalgruppen mit den drei übrigen in der Varianzanalyse geprüften Effekten verglichen werden.

Die Varianzanalysen in Tab. 52, 53, 59 und 62 zeigen alle den gleichen Effekt von Wort- bzw. Experimentalgruppen. Daß der zusätzliche Wortgruppen-Kommunikationsform-Interaktionseffekt in Tab. 52 einen Experimentalgruppeneffekt verbirgt, wurde schon in Abschnitt 5.3.1. erläutert. Alle genannten Effekte tauchen bei den Maßen der A-Kategorisierung und bei den Maßen der Anzahl (GATEX und RATEX) auf, seltener - und wenn, weit weniger stark - bei den Maßen der B- und C-Kategorisierung. Da der Unterschied zwischen A- und B- bzw. C-Kategorisierung in der unterschiedlichen Relativierung der Leistung - auf die möglichen und gegebenen bzw. "richtigen" (Kategorie 1-8) Antworten - liegt, muß auch der Unterschied zwischen den Experimentalgruppen allein auf unterschiedlicher Bekanntheit, nicht aber auf unterschiedlich schwieriger Explizierbarkeit der Worte in den einzelnen Wortgruppen beruhen. Die Unterschiede zwischen den Wortgruppen resultieren also aus der "objektiven" Tatsache unterschiedlicher Anzahlen gegebener Antworten, nicht aber aus Unterschieden in der differenzierten "subjektiven" Bewertung der gegebenen Antworten. Tab. 61 zeigt die unterschiedlichen Anzahlen der gegebenen Antworten, Tab. 5 die unterschiedlichen Prozent der "k.A." in den Wortgruppen bzw. den Experimentalgruppen. Unabhängig von der Testform, die sie mündlich oder schriftlich zu beantworten hatte, gibt die erste Experimentalgruppe am häufigsten, die zweite am zweithäufigsten keine Antwort; mit einigem Abstand folgen die vierte und die dritte Experimentalgruppe. Da die 1. und 2. wie die 3. und 4. Gruppe insgesamt die gleichen Worte beantwortet haben, kann man den Experimentalgruppeneffekt zum einen Teil der unterschiedlichen Schwierigkeit der Worte, die in

Wortgruppen oder Testformen zusammengefaßt sind, zum anderen Teil den unterschiedlichen Fähigkeiten der Individuen zuschreiben. Diese beiden Einflüsse sollen gegeneinander abgewägt werden (siehe Tab. 63).

Die unterschiedliche Schwierigkeit der beiden Testformen A und B deuten sich schon in der ersten Messung an. Nun sind die beiden Testformen aus jeweils zwei statistisch gleichwertigen Aufgaben zusammengesetzt und korrelieren für Kinder des 5. Schuljahres $r = .927$ (Anger/Bargmann/Hylla: Testbeiheft). Die Diskrepanz zu unseren Ergebnissen in der ersten Messung kann zweierlei bedeuten: entweder ist der Test, der für das 5. und 6. Schuljahr bestimmt ist, für die Kinder des 4. Schuljahres zu schwer, aber je nach Testform in unterschiedlichem Grade; oder, wahrscheinlicher, der Test, der Anfang der 50er Jahre konstruiert wurde (Anger/Bargmann, 1954) enthält in der Form A mehr Worte, die mittlerweile veraltet sind. Die Diskrepanz zu unseren Ergebnissen der zweiten Messung (Tab. 63, Spalte 5 und 6, Zeile 2-14) scheint uns zu groß, um aus den schon in der ersten Messung auftauchenden Differenzen erklärt werden zu können. Offensichtlich hat zusätzlich die Umwandlung in einen aktiven Test die beiden Testformen unterschiedlich erschwert. Wie weiter aus dem Vergleich der Experimentalgruppenunterschiede zwischen den Maßen der A- und der B- und C-Kategorisierung hervorgeht (Zeile 7-14 vs. Zeile 15-18), waren die Worte der Testform A zwar weniger bekannt, aber sie wurden nicht schlechter expliziert und bewertet, wenn sie bekannt waren. Durch die Aufforderung zur freien Definition sind die Worte der Testform A den Kindern stärker entfremdet worden, aber die bekannten Worte beider Testformen waren für die Kinder nicht unterschiedlich schwer zu erklären. Ein weiterer Hinweis dafür ist es, daß die Anzahl der Worte, auf die nur zwei oder weniger Kinder (in der mündlichen oder der schriftlichen Messung) geantwortet haben, in der Testform A 26, in der Testform B 12 beträgt (siehe Abschnitt 5.4.6.).

Die bei der Aufteilung der Ergebnisse der ersten Messung nach den Wortgruppen der zweiten Messung auftauchenden noch größeren Unterschiede (Tab. 63, Zeile 1, Spalte 1-4) müssen nicht die Zuverlässigkeit der Testform beeinträchtigen; vielmehr waren sie, auch wenn die steigende Schwierigkeit der Worte durch die alternerende Aufteilung berücksichtigt ist, bei einer Halbierung der Testform zu erwarten. Sie sollen hier lediglich ein Maßstab für die Wortgruppen der zweiten Messung darstellen. Nun entsprechen die Rangfolgen der Schwierigkeit der Wortgruppen in der ersten und in der zweiten Messung (Tab. 62, Zeile 1-5) einander nur insofern, als in beiden Fällen die Testformen A die schwierigeren sind. In die Wortgruppenmittelwerte der zweiten Messung sind dabei Erklärungen von jeweils zwei Experimentalgruppen eingegangen, ohne Beachtung der Kommunikationsformen. Trotz des besseren Gruppenmittelwertes auf der Kontrollvariablen geben die Gruppen 1 und 2 auf die Worte der Form A weniger und schlechtere Antworten; dies zeigt sich deutlich beim Vergleich der gesamten Testformen in den Spalten 5 und 6. Innerhalb der Wortgruppen ist für die Gruppen 1 und 2 "A gerade" nach den Maßen der A-Kategorisierung, "A ungerade" nach den Maßen der B- und C-Kategorisierung die schwierigere Wortgruppe. Für die Gruppen 3 und 4 hingegen findet sich so gut wie kein Unterschied zwischen den Wortgruppen. Wo die Wortgruppen sich aber unterscheiden, innerhalb der Testform A, sind die Unterschiede zwischen den verschiedenen Wortgruppen der gleichen Gruppe von Vpn. geringer als zwischen den gleichen Wortgruppen verschiedener Gruppen von Vpn. Deshalb müssen wir zur Untersuchung der zweiten möglichen Ursache hinter den Experimentalgruppeneffekten übergehen: den unterschiedlichen Fähigkeiten der Vpn.

Sie können aus den Gruppenmittelwerten auf der Kontrollvariablen abgelesen werden. Nun stimmt deren Rangfolge mit der der Leistungen der Gruppen im aktiven Wortschatztest lediglich für das "Schlußlicht", die Gruppe 4, überein. Allerdings kann die entschieden größere Schwierigkeit der Testform A (siehe Spalte 5 und 6), die ja auch in der ersten Messung auftritt, das Absinken der Gruppe 1 von dem ersten auf den 3. Rangplatz erklären. Der gleiche Einfluß

schlägt die, gemessen am Gruppenmittelwert auf der Kontrollvariablen, sowieso schon schwache Gruppe 1 weiter ab: die Abstände innerhalb der beiden Gruppen, die die Testform A bekamen, sind auf allen Variablen, außer dem Maß GATEX, nahezu ebensgroß wie die zwischen der Gruppe 2 und den beiden Gruppen, die die Testform B bekamen; sie spiegeln die extreme Differenz der Gruppen 1 und 2 auf der Kontrollvariablen. Man kann also annehmen, daß die Gruppe 2 trotz der relativ besseren Ausgangslage wegen der schwierigeren Testform noch hinter den Gruppen mit der Testform B zurückblieb, und bei Gruppe 1 die schwierigere Testform und die schlechtere Ausgangslage zusammen das schlechteste Ergebnis bewirkten. -

Die Ergebnisse zum Einfluß der Kommunikationsform (5.3.4.) und der Experimentalgruppen (5.3.5.) sollen zusammengefaßt werden. Zwischen den Kommunikationsformen ergaben sich - auch bei Anwendung des korrekten Prüfmodells - keine signifikanten Unterschiede. Die Experimentalgruppenunterschiede können auf unterschiedliche Schwierigkeit der Worte oder unterschiedliche Fähigkeiten der Vpn. zurückgeführt werden. Sie tauchen nur bei den Maßen der A-, nicht aber der B- und C-Kategorisierung auf. Die Differenz zwischen den beiden Gruppen, die im Erklärungstest die Testform A und den beiden, die die Testform B bekamen, spiegelt wahrscheinlich eine unterschiedliche Schwierigkeit der Testformen, die teils schon im passiven Test sich zeigte, teils bei der Umwandlung in einen aktiven Test sich ergeben hat. Innerhalb der beiden Gruppen mit der Testform A waren Leistungsunterschiede im wesentlichen durch Unterschiede auf der Kontrollvariablen bedingt. Innerhalb der Gruppen mit der Testform B ergaben sich weder Unterschiede in den Experimentalleistungen noch auf der Kontrollvariablen, wohl aber Unterschiede zwischen den Kommunikationsformen: beide Gruppen waren im Schriftlichen besser; Differenzen zwischen den Wortgruppen zeigten sich nicht.

Die Ergebnisse der beiden Gruppen mit der Testform B entsprechen also durchaus den Erwartungen. Da beide auch auf der Kontrollvariablen fast gleiche Werte zeigen, haben wir für sie noch einmal gesondert den Effekt der Kommunikationsform überprüft (Tab. 60). Es zeigt sich, daß der Effekt der Kommunikationsform hier nahezu Signifikanz erreicht. - Die Ergebnisse der beiden Gruppen mit der Testform A hingegen geben trotz aller Interpretationsschwierigkeiten wegen der Unterschiede auf der Kontrollvariablen einen wichtigen Hinweis über die Wirkung der Testtechnik. Die Erschwernis der Testform A gegenüber der Testform B beim Übergang von passiven multiple-choice zum aktiven Erklärungstest liegt, wie gezeigt, in der geringeren Bekanntheit, weniger der geringeren Erklärbarkeit der Worte. Der entscheidende Unterschied zwischen beiden Testtechniken - multiple-choice und aktive Erklärung - betrifft demnach weniger die Fähigkeit, bekannte Worte zu erklären, als die Wortkenntnis überhaupt. Man kann sich vorstellen, daß manche Vpn. im multiple-choice Test Wortkenntnis erfolgreich vortäuschen können, während andere über den Anforderungen der aktiven Erklärung tatsächliche Kenntnis nicht mehr ausweisen können. Eine solche Erklärung widerspricht der üblichen Auffassung der beiden Testtechniken, indem sie die "aktiven" Elemente des passiven Wortschatztests - die ja nicht nur im reinen, sondern auch durch semantische Hypothesen angeleiteten Raten liegen -, und die Wichtigkeit des Umfangs des "passiven" Vokabulars für die aktive Erklärung herausarbeitet. Sie zeigt die Schwierigkeit eindeutiger Zuordnungen von Techniken und theoretischen Konzepten. Aus diesem Grunde sind wir auf sie trotz der schwachen empirischen Basis eingegangen. -

Wenn wir die Häufigkeit signifikanter Effekte der Experimentalgruppen mit der Häufigkeit signifikanter Effekte der Faktoren Schicht, Geschlecht und Kommunikationsform vergleichen (d.h. Tab. 62 mit Tab. 55, 59 und 60), so haben unbestreitbar "Experimentalgruppe" den größten, "Geschlecht" nur schwache und "Kommunikationsform" wie "Schicht" keinen Einfluß auf die Maße der zweiten Messung. Die "Experimentalgruppen" beeinflussen ausschließlich die Maße der

A-Kategorisierung (Antworthäufigkeit und -güte), die Geschlechtszugehörigkeit ausschließlich die Maße der B- und C-Kategorisierung (Antwortgüte). Wenn Schicht und Kommunikationsform überhaupt F-Werte über 1 zeigen, so Schicht bei den Maßen der A-Kategorisierung im Mündlichen und Kommunikationsform bei MIBEX (für alle 4 Experimentalgruppen). Wenn wir die durch den Faktor Experimentalgruppen erklärte Varianz (geschätztes Omega-Quadrat, Hays 1963:382) des Maßes MIAEX errechnen, so liegt sie bei 11% für die mündliche und 25% für die schriftliche Messung. Diese starke Abhängigkeit der aktiven Worterklärung von den Besonderheiten der Wortgruppen und der Vpn. bei gleichzeitiger Bedeutungslosigkeit der dichotomen Sozialschicht ist ein für das soziologische Vorurteil enttäuschendes Ergebnis. Insgesamt bestätigt die Varianzanalyse also nicht nur die bezüglich des Schichteinflusses negativen Ergebnisse der Wilcoxon-Tests, sondern streicht zusätzlich die Wichtigkeit nicht-sozialer Faktoren heraus.

Die folgenden Analysen suchen dies negative Hauptergebnis zu modifizieren, entweder durch eine genauere Betrachtung der abhängigen Variablen, d.h. eine Analyse von Einzelleistungen der zweiten Messung (5.4.) oder durch eine Differenzierung der unabhängigen Variablen, d.h. des sozialen Hintergrunds der Vpn. (5.6. und 5.7.). Da dem Faktor "Experimentalgruppen" eine so große Bedeutung bei der Erklärung der Maße der Gesamtleistung zukommt, sind bei Einzelleistungen kaum noch Schichtunterschiede zu erwarten. Von den beiden Möglichkeiten, die Ausgangshypothese weiterzuführen, wird die zweite die erfolgversprechendere sein.

5.4. Einflüsse der dichotomen Schichteinteilung auf Einzelleistungen
===== oder besondere Aspekte der zweiten Messung. =====

Die bisherigen Analysen konnten keinen bedeutsamen Unterschied zwischen den beiden Schichtgruppen aufdecken. Alle überprüften Maße bezogen sich jedoch auf die gesamte Testleistung. Es ist durchaus möglich, daß sie zu undifferenziert sind, um Besonderheiten der beiden Schichten in der Erklärungsweise oder in der Struktur der Wortkenntnis aufzudecken. Dies soll in den folgenden sechs Analysen versucht werden. Zuerst suchen wir nach Gemeinsamkeiten von Stichworten, die von der Unter- bzw. der Mittelschicht besser erklärt werden, dann nach Unterschieden zwischen beiden Schichten in unabhängig von den Ergebnissen nach dem Grad extravverbaler Explizierbarkeit konstruierten Wortgruppen, und drittens nach Unterschieden in der schichtspezifischen Rangfolge der Erklärungsschwierigkeit der einzelnen Worte. Diese drei Analysen beziehen sich alle auf die Erklärungen einzelner Stimulusworte und versuchen sie in unterschiedlicher Weise zusammenzufassen, um schichtspezifische Strukturen der Wortkenntnis zu finden; natürlich ist diese Suche durch die in den beiden Testformen des WST getroffene Auswahl von Worten eingeschränkt. Die vierte, fünfte und sechste Analyse dagegen beziehen sich wieder auf die Gesamtleistung und versuchen einen bestimmten Aspekt der Worterklärung, der global schon in einigen Gesamtmaßen erfaßt war, genauer herauszuarbeiten: die Differenz zwischen Verständnis und Explikation, die Verteilung der Antworten auf die Kategorien der Bewertung und die Häufigkeit falscher Antworten. Nach dem signifikanten Experimentalgruppeneffekt auf die Gesamtleistung müssen wir Wortgruppen und Vpn.-Gruppen auch bei den folgenden sechs Analysen beachten.

Die Einzelanalysen, die wir im folgenden darstellen werden, sind ebenso wie die vorangegangene Globalanalyse geeignet, Arten des schichtspezifischen Sprachgebrauchs aufzudecken. Denn erstens beziehen sie sich auf die gleiche Experimentalanordnung wie die Globalanalyse: Schichtunterschiede müssen trotz Kontrolle des

schichtabhängigen passiven Wortschatzes auftreten. Zweitens können sie von ihrem Gegenstand her als Untersuchung nicht nur von Merkmalen der Sprache, sondern Merkmalen des Sprachgebrauchs verstanden werden: Die ersten drei Analysen untersuchen Unterschiede des Lexikons, d.h. die von den Kindern der beiden Schichten jeweils besonders gut erklärten Worte (5.4.1. bis 5.4.3.); sie können als inhaltliches Resultat von Einstellungen zur Sprache angesehen werden. Noch direkter beziehen sich die drei übrigen Analysen auf Strategien des Sprachgebrauchs, da sie Einstellungen zur Sprache messen: die Ausschöpfung des Verständnisses in der Explikation (5.4.4.), die Wahl zwischen verschiedenen Formen der Erklärung (5.4.5.), die Entscheidung zwischen einem Verzicht auf oder einem Versuch von Erklärungen (5.4.6.).

Wenn in den Einzelanalysen sich ebensowenig wie in den Globalanalysen Belege für schichtspezifische Arten des Sprachgebrauchs finden, so scheint dies für alternative Theorien zu sprechen, die die Gleichheit der Leistungen und ihre Unabhängigkeit von sozialen Faktoren voraussagen. Zwei derartige Theorien sind denkbar: eine "entwicklungspsychologische", die die Gleichheit der Leistungen durch den allen Kindern gemeinsamen Entwicklungsstand, und eine "sprachlogische", die sie durch die verbindliche Struktur des Lexikons erklärt. Beide Erklärungen gewinnen umso mehr Suggestivkraft, je detaillierter die Analysen sind, in denen eine volle Übereinstimmung zwischen beiden Schichtgruppen nachgewiesen werden kann. Beide Erklärungen sind jedoch durch die experimentelle Anordnung von Anfang begünstigt: Da durch die Kontrolle des passiven Wortschatzes nachgewiesen schichtspezifische Einflüsse (siehe 4.1.) ausgeschaltet worden sind, kann anhand der Übereinstimmung der Leistungen in beiden Schichtgruppen ein entwicklungsbedingtes Muster herausgearbeitet, nicht aber seine Unabhängigkeit von sozialen Faktoren nachgewiesen werden. Das Experiment ist angelegt, durch schichtspezifische Arten des Sprachgebrauchs verursachte Unterschiede aufzudecken, d.h. Unterschiede zwischen Schichtgruppen, die sich trotz kontrollierter Kenntnis und Fähigkeiten ergeben; einheitliche Muster, die bei derartigen Kontrollen auftreten, können durchaus von den kontrollierten und - wie wir wissen - schichtspezifisch verteilten Faktoren abhängig sein.

5.4.1. Schichtunterschiede auf einzelnen Worten.

Wie schon bei der ersten Messung sollen auch die Ergebnisse des aktiven Erklärungstests an den einzelnen Stimulus-Wörtern verfolgt werden. Anders aber als im standardisierten passiven Wortschatztest, bei dem bestimmte Worte gesucht wurden, die für die global gemessenen Schichtunterschiede in besonderem Maße Ursache gewesen sein mögen, sollen jetzt einzelne Worte ausfindig gemacht werden, die trotz der generellen Gleichwertigkeit der beiden Schichten in der Erklärungsleistung dennoch für die Kinder aus der einen oder der anderen Schicht leichter zugänglich waren. Einer schichtspezifischen Struktur des Wortschatzes wäre nachgewiesen, wenn sich für die mittelschicht- bzw. unterschichtgünstigen Worte gemeinsame Merkmale ergäben.

In Tab. 64a-d sind die Mittelwerte der Bewertung nach der A-Kategorisierung in allen Wortgruppen für die mündliche bzw. schriftliche Messung der jeweiligen Experimentalgruppe angegeben, fernerhin die Differenzen und die Differenzen der Differenzen zwischen Verständnis und Explikation für beide Schichten; in Tab. 65a-d sind auf die gleiche Weise die Mittelwerte nach der B-Kategorisierung angegeben. Tab. 66a-d faßt, parallel dazu, die Ergebnisse des Vergleichs der auf der Wortschatzleistung gleichen Paare beider Schichten zusammen. Der Wilcoxon-Test wurde für jedes Wort dreimal durchgeführt, für jede der drei Kategorisierungsarten. Bei der A-Kategorisierung ("mit k.A.") wurden alle 8 Paare einer Experimentalgruppe berücksichtigt; ausgeschlossen wurden nur die Paare, bei denen beide Partner den gleichen Score (meistens "9", öfter aber auch einen anderen Skalenwert) hatten. Bei der B- und C-Kategorisierung ("ohne k.A." bzw. "ohne 9") jedoch wurden alle Paare ausgeschlossen, bei denen nur ein Partner eine "0" hatte; da die "0" in diesen beiden Kategorisierungen immer die Bedeutung "k.A." hatte, hätte bei der Errechnung der Differenz in einem Paar mit einer "0" das Vorzeichen den verkehrten Sinn bekommen. Der Wilcoxon-Test der B-Kategorisierung unterschied sich von den Tests der C-Kategorisierung nur noch durch die Einbeziehung derjenigen Paare, bei denen ein Partner eine definitiv falsche Ant-

wort ("9" für "falsch", nicht für "k.A."), der andere aber eine Antwort zwischen "1" und "8" gegeben hat. Wie ein Überblick über die Tab. 66 zeigt, sind diese Paare äußerst selten, der Unterschied zwischen den beiden Kategorisierungen also minimal. Auf der anderen Seite ist der Unterschied zwischen der A- und B- bzw. C-Kategorisierung beträchtlich: sehr viele Paare fallen aus, weil ein Partner keine Antwort gegeben hat. Aus diesem Grunde müssen wir die Tests mit der B- und C-Kategorisierung ignorieren und uns auf die Ergebnisse mit der A-Kategorisierung beschränken.

Auch innerhalb der A-Kategorisierung jedoch fallen sehr viele Paare aus, selbst bei Wörtern, die eine relativ hohe Anzahl von Antworten in beiden Schichten hervorgehört haben (siehe Tab. 64, eingerahmte Spalten). Es muß also auch eine Reihe von Gleichbewertungen der Paare gegeben haben, die nicht auf einer Verweigerung der Antwort bei beiden Partnern beruht. Auf der anderen Seite muß die Verweigerung der Antwort häufiger bei beiden Partnern zugleich sich gefunden haben als nur bei einem. Auch übertrifft die Anzahl der Paarbildungen (Tab. 66) die jeweils höhere Anzahl der Antworten aus einer der beiden Schichten (siehe Tab. 65) nicht stark, so daß die Partner aus der "schweigsameren" Schicht, die geantwortet haben, meist auf einen Partner getroffen sind, der ebenfalls eine Erklärung gegeben hat (siehe dazu auch die Gegenüberstellung der Antworten der Paare im Anhang). Schließlich zeigt der Vergleich der Worte untereinander, daß für die weitgehend unbekanntesten Worte, meist im Mündlichen sowohl wie im Schriftlichen, sich kaum Paare bilden ließen; dies wäre ja schon möglich, wenn nur ein Kind irgendeine Erklärung versucht. - Obwohl also in der A-Kategorisierung jede Vp. berücksichtigt wurde und sich von ihrem Partner hätte abheben können, wurden relativ wenig ungleiche Paare verglichen. Die geringe Anzahl der tatsächlich durchgeführten Signifikanzprüfungen ist für sich schon ein Ergebnis: es deutet auf die weitgehende Gleichheit der Schichten.

In dieser Ausgangslage ist die Suche nach schichtdiskriminierenden Worten mit Hilfe des Wilcoxon-Tests schon sehr stark zu Erfolglosigkeit verurteilt. Wir haben es dennoch versucht, an den Ergebnissen des Wilcoxon-Tests, mit Hilfe des Mittelwertvergleichs und anhand der

Differenzen zwischen Verständnis und Explikation. Bei allen drei Versuchen haben wir eine Übereinstimmung zwischen der mündlichen und schriftlichen Messung als Mindestbedingung angesetzt. (1) Bei den Wilcoxon-Tests haben wir alle Worte, die ein signifikantes Ergebnis im Mündlichen und Schriftlichen in Verständnis oder Explikation unterdem 30%-Niveau (und bei weniger als vier Paaren alle Paare) in einer der beiden Richtungen¹⁾ hatten, ausgesucht: Nach diesem Kriterium gab es fünf Worte, die in der Mittelschicht ("Staupe", "gescheit", "Ampulle", "Mailand", "Fachwerk") und zwei die in der Unterschicht ("vergnügt", "Armbrust") jeweils sowohl in Verständnis und in Explikation besser bewertet wurden; die Parallele beider Dimensionen, die ja nicht als Auswahlkriterium festgelegt war, erklärt sich aus ihrem Zusammenhang bei der Auswertung. Ein einheitlicher Gesichtspunkt, der die Worte als jeweils schichtspezifische Gruppe hätte zusammenfassen können, ließ sich jedoch nicht finden. (2) Ferner haben wir alle Worte, bei denen sich im Mündlichen und Schriftlichen in Verständnis oder Explikation Mittelwertdifferenzen in der gleichen Richtung über 2.00 ergaben, herausgeschrieben. Hier konnten wir nur bei Verständnis je ein Wort zugunsten der Mittel- ("Fachwerk") und zugunsten der Unterschicht ("vergnügt") finden. Diesen zwei Worten, die über die mündliche und schriftliche Messung durchgängig Unterschiede für eine der beiden Schichten zeigten, stehen jedoch vier Worte gegenüber, die zwar beide Male die gleiche Höhe Differenzen, jedoch in der entgegengesetzten Richtung, aufweisen ("Kobra", "Dukaten", "Raspel", "unbeholfen"). Nicht nur das angewandte Kriterium, sondern die Suche nach einheitlich schichtdiskriminierenden Worten überhaupt wird dadurch entwertet. (3) Schließlich haben wir alle Worte ausgesucht, bei denen im Mündlichen wie im Schriftlichen die Differenzen zwischen Verständnis und Explikation bei einer Schicht um mehr als .50 größer waren als bei der anderen (siehe Tab. 64 a-d, Spalte 13 und 14).

¹⁾-----
Die zweiseitigen Tests haben wir für diesen Zweck wieder als einseitige betrachtet, d.h. das Signifikanzniveau wieder halbiert, da sonst die Bedingungen für beide Schichten ungleich gewesen wären.

Größere derartige Differenzen für die Mittelschicht zeigten sich bei 7 Wörtern ("erquicken", "Künstler", "Traktor", "Säuseln", "gescheit", "Puma", "Nordpol"), für die Unterschicht die 3 Wörter ("Bambus", "Armbrust", "Raspel"). Entgegen der Erwartung zeigen sich große Differenzen also eher in der Mittelschicht als in der Unterschicht (siehe auch unten, 5.4.4.). - Wie sich schon bei den Wilcoxon-Tests andeutete, verlief die Suche nach schichtdiskriminierenden Worten erfolglos. Nach gemeinsamen Merkmalen unter- bzw. mittelschichtgünstiger Worte brauchen wir gar nicht mehr zu fragen. Stattdessen spricht die Vielzahl der im Wilcoxon-Test nicht-testbaren Worte, die geringe Anzahl der vergleichbaren Paare und die Ausgewogenheit der Vergleiche für die Leistungsgleichheit der Vpn. bei der aktiven Worterklärung jedes Stimulus-Worts.

5.4.2. Schichtunterschiede innerhalb der Wortgruppen extraverbaler Explizierbarkeit.

Die induktive Suche nach Besonderheiten unter- und mittelschichtgünstiger Worte scheiterte schon an der minimalen Zahl solcher Worte. Dennoch bleibt es möglich, durch eine Zusammenfassung aller Worte nach theoretischen Merkmalen unabhängig von den Ergebnissen der aktiven Erklärung schichtspezifische Bereiche der Wortkenntnis aufzudecken. Die Leistungen beider Schichtgruppen in nach dem Grad extraverbaler Explizierbarkeit zusammengefaßten Wortgruppen sollen im folgenden dargestellt werden.

Klassifikation der Stimulus-Worte nach extraverbaler Explizierbarkeit. Die 30 Stimulus-Worte jeder der 4 Wortgruppen wurden von 3 voneinander unabhängigen Bewertern in vier Stufen der Erklärungsschwierigkeit eingeteilt, zunächst nach zwei unter den Bewertern abgesprochenen Kriterien: dem intentionalen Sinngehalt und der extraverbaler Explizierbarkeit der Worte; diese Klassifikation war unabhängig von den Ergebnissen der 1. und 2. Messung und ging von der gemeinsam festgelegten Definition der Stimulus-Worte aus, die auch bei der Bewertung der Erklärungen verwandt wurde. Die Übereinstimmung zwischen den 3 Bewertern wurde zwar nicht formell geprüft, war jedoch zufriedenstellend. Die endgültige Klassifikation wurde in längeren gemeinsamen Diskussionen der 3 Bewerter festgelegt. Dabei zeigt sich, daß das Kriterium der extraverbaler Explizierbarkeit zur Charakterisierung der endgültigen Gruppen ausreicht (siehe Tab. 68). Definitionen der vier Wortgruppen wurden in Abschnitt 4.2. gegeben.

Als eine gewisse Validierung der Klassifikation können die Mittelwerte der für die 2. Messung gebildeten Wortgruppen in der 1. Messung (Tab. 45, Abb. 6) sowie die Korrelationen der als Vier-Punkte-Skala verstandenen Klassifikation mit den Prozentzahlen richtiger Antworten in der Gesamtgruppe (Tab. 46, Zeile 2) angesehen werden. In allen vier Wortgruppen sind in der 1. Messung die Worte der Gruppe I am leichtesten, die Worte der Gruppe IV am schwersten richtig zu beantworten. Allerdings sind die Worte der Gruppe II

schwieriger als die der Gruppe III. Das mag die allgemein recht niedrigen Korrelationen zwischen der extraverbale Explizierbarkeit und der Schwierigkeit der Beantwortung der Worte erklären; die Korrelation ist in der Wortgruppe am niedrigsten, in der die Mittelwerte der Wortgruppen extraverbaler Explizierbarkeit am ehesten zusammenliegen ("B" ungerade"). Wir können daraus schließen, daß die Klassifikation in den beiden extremen Gruppen schwierige und leichter erklärbare Worte gut auseinanderhält, in den beiden mittleren Gruppen jedoch nichts mit der Verfügbarkeit der Worte im passiven Wortschatz zu tun hat. - Die Schichtdifferenzen der 1. Messung wurden in Abschnitt 4.2. ausführlich analysiert: sie sind in der Gruppe IV geringer als in den übrigen Gruppen; bei den rein verbalen Worten sind die Unterschichtkinder am wenigsten unterlegen.

Für die 2. Messung hatten wir angenommen, daß bei der aktiven Wortklärung die Kinder der Mittelschicht in den stärker "verbalen" Wortgruppen überlegen sind; entsprechend müßten - bei dem allgemein negativen Ergebnis - die Kinder der Unterschicht bei den stärker "extraverbalen" Wortgruppen überlegen sein. Tab. 69 zeigt die Mittelwertdifferenzen zweier Maße; die Schichtdifferenzen sind in der mündlichen und schriftlichen Messung sehr häufig ungleich, werden aber für beide Messungen zusammengefaßt, so daß die Wortgruppen der Testform A von den Experimentalgruppen 1 und 2, die Wortgruppen der Testform B von den Experimentalgruppen 3 und 4, d.h. von 32 (34) Vpn. (bzw. 16 (17) Paaren in den Tests) beantwortet wurden; Tab. 70 zeigt die Ergebnisse der entsprechenden Vorzeichen- und Wilcoxon-Tests. Die Mittelwerte zeigen zwar eine schwache Tendenz in Richtung der Voraussage, die aber in den Wortgruppen nicht gleich bleibt. Keiner der Vorzeichen- und nur drei der - weniger strengen - Wilcoxon-Tests zeigen signifikante Unterschiede zugunsten der Mittelschicht. Kurzum, auch bei einer Aufteilung der Stichworte nach dem Grade der extraverbale Explizierbarkeit ergeben sich keine signifikanten Unterschiede zwischen den beiden Schichtgruppen.

Auf der anderen Seite lohnt es sich, die Unterschiede der Leistungen

der Gesamtgruppe zwischen den Wortgruppen zu betrachten (Tab. 71, Abb. 8). Zwischen der mündlichen und schriftlichen Messung zeigt sich innerhalb jeder der Experimentalwortgruppen eine weitgehende Ähnlichkeit der Rangfolgen der Wortgruppen extraverbaler Explizierbarkeit die allerdings nicht der prognostizierten Rangfolge entspricht. Wie in der ersten Messung kann innerhalb der Wortgruppe "B ungerade" am wenigsten zwischen den Gruppen extraverbaler Explizierbarkeit unterschieden werden. Wie in der ersten Messung sind die Worte der Gruppe II meist schwieriger zu erklären als die Worte der Gruppe III. Die Ähnlichkeit der Mittelwerte bzw. der Profile zwischen der mündlichen und schriftlichen Messung - jeder der Mittelwerte ist aus den Bewertungen der Erklärungen der 66 Vpn. zu nur 6 bis 10 Stichworten gebildet - zeigt die Bedeutung, die die Besonderheit der Stichworte für die Erklärungen für die aktive Worterklärung in beiden Schichtgruppen gleichermaßen hat; sie ist zugleich ein positives Indiz für die Reliabilität der Bewertungen aller Erklärungen.

Wie die induktive Suche nach schichtdiskriminierenden Worten ist auch die Suche mit Hilfe nach theoretischen Merkmalen zusammengefaßter Wortgruppen ergebnislos verlaufen; sie hat jedoch zusätzlich die Abhängigkeit der aktiven Erklärung beider Schichtgruppen von dem Charakter der Stimulus-Worte erwiesen. Wie schon in der ersten Messung, konnte auch in der zweiten Messung das Globalergebnis nicht anhand der Ergebnisse zu einzelnen Worten differenziert werden. Wenn überhaupt Schichtspezifität des Wortschatzes in der aktiven Worterklärung noch durchschlagen kann, dann in Unterschieden der relativen Schwierigkeit der Worte zwischen den Schichtgruppen. Diese Frage soll im nächsten Abschnitt anhand von Rangkorrelationen untersucht werden.

5.4.3. Schichtspezifität und Reliabilität des Testmaterials.

Die Untersuchung von Schichtunterschieden in der aktiven Worterklärung ist bis jetzt negativ verlaufen: Weder fanden wir Unterschiede in der Gesamtleistung zwischen Kindern aus der Unter- und Mittelschicht, noch durchgängige Schichtdifferenzen bei einzelnen Worten. Das Fehlen signifikanter Schichtunterschiede auf allen oder einzelnen Wörtern einer Testform könnte jedoch geringe, aber systematische Unterschiede zwischen den Schichtgruppen in der Rangfolge der Schwierigkeit der Worte verdecken. Anders gesagt: Gibt es trotz fehlender Niveauunterschiede Strukturunterschiede des Wortschatzes? Wenn Strukturunterschiede nachgewiesen werden können, so würde das Ergebnis, daß keine Schichtunterschiede im aktiven Wortschatztest auftreten, durch die schichtspezifische Zugänglichkeit des Testmaterials erschüttert. Wenn Strukturunterschiede nicht nachgewiesen werden können, so würde damit nicht nur die negative Tendenz unserer Ergebnisse erneut bestätigt, sondern auch der Verdacht der Schichtdiskriminierung gerade gegen verbale Tests entkräftet.

Um diese Frage zu klären, haben wir die Mittelwerte der Bewertungen der Einzelworte für die Unterschicht-, die Mittelschicht- und die Gesamtgruppe sowie der Schichtdifferenzen innerhalb jeder Wortgruppe und Kommunikationsform in eine Rangfolge gebracht (siehe Tab. 64 a-d: Kategorisierung A; Tab. 65a-d: Kategorisierung B) und die verschiedenen Rangfolgen miteinander korreliert. Die gleichen Rechnungen wurden auch mit den Prozentzahlen richtiger Antworten auf die Einzelwerte der ersten Messung durchgeführt. Durch den Vergleich der Rangfolgen zwischen erster und zweiter Messung sowie zwischen mündlichen und schriftlichen Maßen der zweiten Messung ergab sich zusätzlich die Möglichkeit, die Reliabilität des aktiven Wortschatzes noch einmal zu beurteilen. Deshalb sollen die Eigenart der Maße und einige Hypothesen über Beziehungen zwischen ihnen kurz erörtert werden.¹⁾

¹⁾-----
In Abschnitt 1.2.5. haben wir Unterschiede zwischen den beiden Schichtgruppen in der Beziehungsstruktur der Maße der zweiten Messung dargestellt. Ebenso wie in diesem Abschnitt handelte es sich dort um Strukturunterschiede der Bedeutungserklärung unabhängig vom absoluten Leistungsniveau, bei Kontrolle des passiven Wortschatzes. Anders als in diesem Abschnitt handelte es sich um die Beziehung zwischen Aspekten der gesamten Erklärungsleistung, nicht um die Beziehung zwischen der Erklärung einzelner Worte. Aus diesem Grunde haben wir die Ergebnisse des Abschnitts 1.2.5. im methodischen Teil der Darstellung belassen.

Alle im folgenden dargestellten Korrelationen¹⁾ verbinden die verschiedenen Messungen unseres Experiments über die Mittelwerte aller jeweils untersuchten Vpn. für einzelne Worte, nicht aber die mittlere Bewertung aller erklärten Worte für einzelne Vpn. (wie in Tab. 14). Die Maße MIAEX und MIAVE ("A-Kategorisierung") sowie MIBEX und MIBVE ("B-Kategorisierung") wurden hier nicht (wie in Tab. 14) für die mündliche und schriftliche Leistung jeder Vp. miteinander korreliert; sondern für jedes Wort in der Unterschicht-, Mittelschicht- und Gesamtgruppe. Die mittleren Leistungen der genannten Gruppen - beginnend mit der besten Leistung - und die mittleren Schichtdifferenzen - beginnend mit der größten Schichtdifferenz zugunsten der Mittelschicht - wurden in eine Rangfolge gebracht. Im Mittelwert der A-Kategorisierung sind alle Vpn. der jeweiligen Gruppe berücksichtigt; Vpn., die keine oder falsche Erklärungen gaben, erhielten den schlechtesten Skalenwert "9". Im Mittelwert der B-Kategorisierung sind nur diejenigen Vpn. berücksichtigt, die eine Erklärung des jeweiligen Stimulus-Worts versucht haben; hat keine Vp. einer Gruppe eine Erklärung gegeben, so wurde für das Wort der "Skalenwert "8" als Mittelwert angenommen; eindeutig falsche Erklärungen wurden jedoch mit dem Wert "9" verrechnet. Eine zutreffende Erklärung nur einer Vp. läßt das Stimuluswort in der A-Kategorisierung wegen der fehlenden Antworten der anderen Vp. als recht schwer, in der B-Kategorisierung als recht leicht erscheinen.

Diese Berechnungsverfahren orientieren sich an den theoretischen Charakteristika der verschiedenen Messungen. Die erste Messung bewertet die Bekanntheit der Worte, die A-Kategorisierung der zweiten Messung die Bekanntheit und Explizierbarkeit und die B-Kategorisierung der zweiten Messung allein die Explizierbarkeit. Im Vergleich

¹⁾-----
Obwohl wir die Bewertungen der aktiven Messung wie intervallskalierte Daten verarbeitet haben, haben wir hier mit Rang- und nicht mit Maßkorrelationen gerechnet. Die Rangkorrelation entspricht unserer Fragestellung (Wieweit ähneln sich zwei Rangreihen?) unmittelbarer als die Maßkorrelation (Wie gut läßt sich die Stellung auf einer Variablen durch die Stellung auf der anderen voraussagen?).

zwischen der ersten und zweiten Messung sind daher höhere Korrelationen für die A- als für die B-Kategorisierung und für die Verständnis- als für die Explikationsmaße zu erwarten. Schließlich ist in der B-Kategorisierung eine geringere Korrelation zwischen der mündlichen und schriftlichen Messung als in der A-Kategorisierung zu erwarten, zumal die Berechnung der Wortschwierigkeit nach der B-Kategorisierung für eine ganze Gruppe von den Antworten einer Vp. abhängen kann.

Die Analyse der Rangkorrelationen (Tab. 72: ausführliche Darstellung; Tab. 73: Zusammenfassung)¹⁾ soll also zwei Zielen dienen: der Prüfung der Reliabilität und der Prüfung der Schichtenspezifität des Testmaterials. Wir wollen die letzte Frage zuerst kurz für die erste Messung, dann beide Fragen ausführlich für die zweite Messung behandeln. -

Schichtenspezifität des Testmaterials in der ersten Messung. Zwei Hypothesen über den schichtspezifischen Charakter verbaler Tests lassen sich an unserem Material überprüfen: (1) Unterschichtkinder sind mit anderen Wörtern gut vertraut als Mittelschichtkinder; d.h. die Rangreihen der Schwierigkeit der Worte sind in Unter- und Mittelschicht nicht dieselben. (2) Je unbekannter und schwieriger ein Wort überhaupt wird, desto unbekannter und schwieriger wird es besonders für Unterschichtkinder; d.h. die Schwierigkeit eines Wortes in der Gesamtgruppe korreliert hoch mit den Schichtdifferenzen. Beide Vermutungen werden nicht bestätigt (Tab. 72, Spalte 1 und 2); lediglich in den Wortgruppen der Testform A finden sich mäßige Beziehungen zwischen Wortschwierigkeit und Schichtdifferenz. Dieses Ergebnis ist umso erstaunlicher, als wir beide Testformen des passiven multiple-choice-Tests noch einmal wie in der zweiten Messung

1) -----
Außer den in Tab. 72 bzw. 73 aufgeführten Korrelationen wurden folgende berechnet: (1) zwischen den Rangfolgen für die Gesamtgruppen der ersten und zweiten Messung sowie (2) der Gesamtgruppen der mündlichen und schriftlichen zweiten Messung; beide unterscheiden sich kaum von den entsprechenden Korrelationen innerhalb der Schichtgruppen (Tab. 68, Spalte 3-10 bzw. 15-18). (3) zwischen V und E innerhalb der Schichtgruppen und Kommunikationsformen; sie haben für die A-Kategorisierung einen Durchschnitt von .95, für die B-Kategorisierung von .88 - ein Reflex des durch die Messung bedingten engen Zusammenhangs zwischen V und E. (4) Korrelationen zwischen der A- und B-Kategorisierung innerhalb der Schichtgruppen; sie haben einen Durchschnitt von .76.

(d.h. für die erste Messung willkürlich) aufgeteilt haben. In allen vier Wortgruppen korrelieren die Schwierigkeitsrangfolgen der Worte in Unter- und Mittelschicht miteinander fast so hoch wie man sie üblicherweise bei Reliabilitätsmessungen von Tests erwartet. (Allerdings steht die unterschiedliche Schichtanfälligkeit der beiden Testformen A und B in einem gewissen Widerspruch zur hohen Korrelation beider Testformen, die Anger/Bargmann/Hylla (1954:11) im Testbeiheft berichten.)

Reliabilität des Testmaterials in der zweiten Messung. Wir können die Korrelationen der Rangfolgen zwischen der ersten und zweiten Messung als eine Art "externe Validierung", die zwischen der mündlichen und schriftlichen zweiten Messung als eine Reliabilitätsprüfung ansehen, - Da die zu den gleichen Stimulus-Worten verlangten Leistungen der ersten und zweiten Messung: passive und aktive Wortkenntnis, doch recht unterschiedlich sind, waren hier keine hohen Korrelationen zu erwarten. Dennoch ergaben sich (Tab. 72, Spalte 3-10 und entsprechende Spalten in Tab. 73) mittelhohe Korrelationen mit einem Durchschnitt von .59 und einer Reichweite von .42 bis .76 bei der A-Kategorisierung und einem Durchschnitt von .49 und einer Reichweite zwischen .25 und .73 bei der B-Kategorisierung. Die Beziehungen zwischen Wortrangfolgen der ersten und zweiten Messung sind für die Unterschichtgruppe nur minimal geringer als für die Mittelschichtgruppe. Die Beziehung der ersten Messung ist eher höher zu den Explikationsmaßen der zweiten Messung, nicht, wie man erwartet hätte, zu den Verständnismaßen. Gemessen an dem "äußeren" Kriterium des passiven Verständnistests gibt die Verständnisbewertung den relativen Schwierigkeitsgrad der Worte nicht besser wieder als die Explikationsbewertung. Auf der anderen Seite ist, wie zu erwarten, die Beziehung der ersten Messung zu den Maßen der A-Kategorisierung im Durchschnitt höher als zu den Maßen der B-Kategorisierung. Vergleicht man die einzelnen Korrelationen der A-Kategorisierung mit den entsprechenden der B-Kategorisierung, so fällt allerdings auf, daß 7 von den 16 Korrelationen für die Verständnismaße (siehe Tab. 72, 2.1.) und 5 von den 16 Korrelationen für die Explikationsmaße (siehe Tab. 72, 2.2.) in der B-Kategori-

sierung die entsprechenden Korrelationen der A-Kategorisierung übertreffen; sie finden sich - mit einer Ausnahme - alle unter den Korrelationen der Testform B und vorwiegend unter den Korrelationen der Mittelschichtgruppe: die Korrelation für die Mittelschicht innerhalb der Testform B sind sämtlich in der B- höher als in der A-Kategorisierung.

Zusammenfassung von Tab. 72, Spalte 3-10, nach Testformen und Kategorisierungen

		<u>Kateg.-A</u>	<u>Kateg.-B</u>
Testform A	D	.62	.42
	R	49-76	25-61
Testform B	D	.55	.56
	R	42-70	30-73

In der Testform A korreliert die Wortrangfolge nach Kategorisierung A mit der der ersten Messung höher als mit der Kategorisierung B, in der Testform B dagegen die Rangfolge nach Kategorisierung B höher als nach Kategorisierung A. Wenn man die Besonderheiten beider Kategorisierungsarten in Betracht zieht, heißt das: Die Testform B enthält leichter zugängliche Worte, die aber oft falsch und nicht besser als die Worte der Testform A erklärt werden; oder: Die Testform B stellt an die Wortkenntnis geringere, an die Wort-erklärung aber zumindest gleiche Anforderungen als die Testform A. Dies Ergebnis fügt sich zu zwei weiteren Ergebnissen. Erstens zeigte sich in Testform A ein höherer Anteil von Antwortverweigerungen, die gegebenen Erklärungen waren jedoch nicht schlechter als die der Testform B (siehe 5.3.4. Tab. 63). Zweitens wurden in Testform B nahezu doppelt so viel falsche Erklärungen wie in Testform A gegeben (siehe 5.4.6.).

Da die zu den gleichen Stimulus-Worten verlangten Leistungen der mündlichen und schriftlichen zweiten Messung sich wenig unterscheiden, können wir die Korrelation zwischen ihnen als zusätzlichen Reliabilitätstest ansehen (Tab. 72, Spalte 15-18). Es ergeben sich befriedigende Korrelationen für die Kategorisierung A mit einem Durchschnitt von .80 und einer Reichweite von 81 bis 88, und - wie zu erwarten - geringere, aber immer noch recht gute Korrelationen für die Kategorisierung B mit einem Durchschnitt von .47 und einer Reichweite von .02 bis .68. Die geringeren Korrelationen in der

B-Kategorisierung ergeben sich aus der geringeren Stabilität extrem schwieriger Worte über die mündliche und schriftliche Messung. Die Korrelationen der B-Kategorisierung zwischen mündlicher und schriftlicher Messung sind jedoch immer noch fast genau so hoch wie die zwischen der ersten und zweiten Messung der B- und auch der A-Kategorisierung. Wenn man zwischen Verständnis und Explikationsmaßen und zwischen der Unterschicht- und Mittelschichtgruppe vergleichen will, so sind die Rangfolgen der Wortschwierigkeit für die Verständnismaße und für die Unterschichtgruppe stabiler. Die hohen Korrelationen der mündlichen und schriftlichen Messung in beiden Schichtgruppen nach der A-Kategorisierung bewirken, daß die Korrelationen der Schichtdifferenzen zwischen mündlicher und schriftlicher Messung (Spalte 23, 24) nur noch um 0 schwanken können.

Schichtspezifität des Testmaterials in der zweiten Messung.

Die beiden Hypothesen, die wir schon an der ersten Messung überprüft haben, lassen sich auch hier wieder testen, wiederum mit negativem Ergebnis: (1) von einer unterschiedlichen Rangfolge der Wortschwierigkeit in beiden Schichten kann keine Rede sein (Spalte 19-22). Sowohl in der A- wie in der B-Kategorisierung korrelieren die Rangfolgen zwischen den Schichten höher als innerhalb jeder der beiden Schichten zwischen mündlicher und schriftlicher Messung; dies gilt für die Verständnis- wie die Explikationsmaße (Vergleich von Spalten 19-22 mit 15-18). (2) Von einer Benachteiligung der Unterschicht durch die schwierigeren Worte kann ebenfalls keine Rede sein. Es gibt keine Korrelationen zwischen der Schwierigkeit des Wortes in der Gesamtgruppe und den Schichtdifferenzen innerhalb der zweiten Messung (Spalte 27-30) - eine direkte Widerlegung unserer Hypothese. Außerdem sind die Korrelationen zwischen mündlicher und schriftlicher Messung innerhalb der Schichtgruppen und zwischen den Schichten innerhalb der mündlichen bzw. schriftlichen Messung so hoch, daß die Schichtdifferenzen zwischen mündlicher und schriftlicher Messung kaum noch korrelieren können (Spalte 23, 24). Unter- und Mittelschichtgruppe zeigen in der mündlichen wie schriftlichen Aufgabe bei jedem Wort weitgehend ähnliche Leistungen, so daß die Schichtunterschiede sich mehr oder minder zufällig ergeben und ihre Größe zwischen mündlicher und schriftlicher Messung bei den meisten Worten nicht konstant bleibt - eine indirekte Widerlegung unserer Hypothese. Schließlich sind zwar die

Korrelationen innerhalb der Schichtgruppen zwischen erster und zweiter Messung nicht groß genug, um geringe Korrelationen der Schichtdifferenzen zwischen erster und zweiter Messung zu erzwingen; dennoch aber schwanken diese Korrelationen um 0 (Spalten 11-14) - eine weitere indirekte Widerlegung unserer Hypothese. - Alle drei aufgeführten, unseren Vermutungen widersprechenden Ergebnisse zeigen sich sowohl in der A- als auch in der B-Kategorisierung, wo sich fast durchweg sogar ein negativer Trend zeigt. Er ist jedoch (wie eine Inspektion der Tab. 65a-d zeigt) wiederum durch die geringere Stabilität der schwierigeren Einzelwörter in der B-Kategorisierung bedingt. Wenn etwa auf ein Stichwort in der mündlichen Aufgabe nur ein Kind der Mittelschichtgruppe, in der schriftlichen Aufgabe nur ein Kind der Unterschichtgruppe geantwortet haben, so ergibt sich im mündlichen eine für die Mittelschicht extrem günstige, im schriftlichen extrem ungünstige Schichtdifferenz, d.h. eine Tendenz zu negativen Korrelationen.¹⁾

Die Ergebnisse dieses Abschnittes können in zwei Sätzen zusammengefaßt werden. Erstens bleibt die Rangfolge der Wortschwierigkeit gleich unter welchen Meßbedingungen ziemlich konstant - Test und Auswertung sind also verläßlich. Zweitens zeigen sich weder in der ersten noch in der zweiten Messung Unterschiede der Rangfolgen der Schwierigkeit der Worte zwischen den Schichten oder Zusammenhänge zwischen der allgemeinen Schwierigkeit eines Wortes und schlechtem Abschneiden der Unterschicht - Test und Auswertung sind also nicht schichtdiskriminierend. Nicht nur Niveau-, sondern auch Strukturunterschiede des Wortschatzes zwischen Unter- und Mittelschicht konnten also nicht nachgewiesen werden; die Schwierigkeit der aktiven Erklärung steigt über die Worte des Tests für Kinder beider Schichten in der gleichen Weise. Angesichts der bekannten Vorwürfe der Schichtdiskriminierung gerade durch verbale Tests ist das ein überraschendes Ergebnis.

¹⁾-----
Als Beispiele für eine solche Tendenz seien genannt: A ungerade: "Java", "Objekt", "Frevel"; A gerade: "Sympathie", "Verhängnis", B ungerade: "Schotter", "Olymp", B gerade: "Kluft", "Export", "Fjord".

5.4.4. Schicht- und Wortunterschiede bei den Differenzen zwischen Verständnis und Explikation

Bis jetzt haben wir das absolute Niveau der Schichtgruppen auf den Maßen des Verständnisses und der Explikation verglichen; dabei ergaben sich keinerlei signifikante Differenzen. Da wir aber die durchschnittliche Differenz jeder Vp. zwischen diesen beiden Dimensionen errechnet haben, können wir die Schichtgruppen auch noch auf den durch diese Differenz definierten Maßen vergleichen. Sie sollten es erlauben, jenseits der absoluten Leistungen des Verständnisses und der Explikation qualitative Unterschiede in der internen Struktur der Wortkenntnis aufzuspüren. Die Annahme geringerer Differenzen zwischen Verständnis und Explikation in der Mittelschicht ist aus den gleichen Überlegungen wie die Grundhypothesen des Experiments begründet (siehe 3.). Die "Verständnis"-Bewertung kann mit der des passiven Repertoires an Konzepten verglichen werden, während in der "Explikations"-Skala seine kommunikative Aktivierung beurteilt wird. Wenn wir die mit dem elaborierten Sprachgebrauch verbundene Einstellung zur Versprachlichung außersprachlicher Bedeutungen auf die Explikation sprachlicher Bedeutungen selber übertragen, so müßten Kinder der Mittelschicht ihr Wortverständnis besser und häufiger explizieren, auch wenn sie über kein besseres Wortverständnis als die Kinder der Unterschicht verfügen. Die Differenz zwischen Verständnis und Explikation müßte deshalb in der Mittelschicht geringer sein.

Bei der Interpretation von Ergebnissen der Differenzmaße müssen wir davon ausgehen, dass weder auf den Explikations- noch den Verständnismaßen signifikante Schichtunterschiede gefunden werden konnten. Schichteffekte können sich nur durch die Addition oder gar Gegenläufigkeit noch zufälliger Differenz auf dem Explikations- und dem zugehörigen Verständnismaß ergeben haben. In Tab. 74 sind die Differenzmaße aller 32 Zellen des Experiments so zusammengestellt, daß man für jede Differenz das zugehörige absolute Niveau im Verständnis und in der Explikation kontrollieren kann; Tab. 77 enthält die Signifikanzprüfung der Differenzmaße aus allen nach dem Experimentalplan möglichen Varianzanalysen. Der Überblick der Tab. 77

zeigt zuerst auf den beiden Maßen DIBVE und DICVE einen Experimental- bzw. Wortgruppeneffekt. Er ist in der Vier-Weg-Varianzanalyse hochsignifikant. Bei der Trennung der Kommunikationsformen bleibt der Gruppeneffekt für DICVE im Mündlichen hochsignifikant, während er im Schriftlichen zwischen dem 5- und 10%-Niveau schwankt; das gleiche Muster mit etwas verminderten Signifikanzen gilt für DIBVE. Der Überblick der Tab. 77 zeigt zweitens auf den beiden Maßen DIBVE und DICVE im Mündlichen einen auf dem 10%-Niveau signifikanten Wortgruppen-Schicht-Interaktionseffekt, im Schriftlichen einen fast auf dem 10%-Niveau signifikanten Schichteffekt.¹⁾

Da wir von den Differenzmaßen Aufschluß über schichtspezifische Strukturen des Wortschatzes erwarten, wollen wir die Schichteffekte nicht wegen ihrer geringen Signifikanz und wegen des Überwiegens der Wortgruppeneffekte sofort ignorieren. Nach dem negativen Hauptergebnis sollte jeder Hinweis auf dennoch bestehende Unterschiede zwischen den Schichten verfolgt werden, um es entweder inhaltlich zu differenzieren oder mit weiteren Belegen zu unterstützen. Wir wollen anhand der Mittelwertvergleiche die Stichhaltigkeit der Schichteffekte eingehend prüfen und erst danach untersuchen, wie die Experimentalgruppeneffekte erklärt werden können.

¹⁾-----
In dem Maß DIAVE wird, im Unterschied zu den Maßen DIBVE und DICVE, die gegebene Anzahl der Antworten GATEX berücksichtigt; aus diesem Grund korrelieren DIAVE und GATEX sehr hoch (Tab. 12, siehe I.1.2.4.). Für die in diesem Abschnitt untersuchte Struktur der Wortkenntnis ist jedoch der Umfang der Wortkenntnis unwesentlich. Aus diesem Grunde haben wir DIAVE in Tab. 77 nicht mehr aufgenommen. - Bei DIAVE finden wir keine Schichteffekte, sowie im Schriftlichen, nicht aber im Mündlichen Experimentalgruppeneffekte (Tab. 54) - im Vergleich zu DIBVE und DICVE also eine Vertauschung der Signifikanz der Experimentalgruppeneffekte zwischen den Kommunikationsformen. Auch GATEX variiert zwischen den Experimentalgruppen im Mündlichen viel stärker als im Schriftlichen (Tab. 54, sowie die unterschiedlichen F-Werte in Tab. 55). Die Vertauschung der Experimentalgruppeneffekte zwischen den Kommunikationsformen beruht daher auf den Ergebnissen für GATEX.

Betrachten wir die Schichtgruppen des ganzen Experiments (Tab. 56), so zeigt die Mittelschicht auf allen Maßen in beiden Kommunikationsformen die größeren Differenzen zwischen Verständnis und Explikation; dies bleibt im wesentlichen auch so, wenn man die Schichten innerhalb der beiden Geschlechtsgruppen vergleicht. Entgegen unserer Hypothese finden sich also insgesamt die größeren Differenzen in der Mittelschicht. Können wir die größeren Differenzen der Mittelschicht auch in den einzelnen Wort- bzw. Experimentalgruppen wiederfinden?

Schon die beiden Varianzanalysen mit Schicht und Experimentalgruppen als Effekten (Tab. 62) deuten auf eine geringe Konsistenz der Schichtunterschiede innerhalb der Wort- bzw. Experimentalgruppen: wir finden zwar im Mündlichen einen schwach signifikanten Schichteffekt, im Schriftlichen aber einen schwach signifikanten Schicht-Experimentalgruppen-Interaktionseffekt, so daß wir im Schriftlichen einen Wechsel der Richtung der Schichtdifferenzen zwischen den Wortgruppen erwarten müssen. Wir können die Schichtunterschiede innerhalb der Wortgruppen in Tab. 76 und Abb. 10 verfolgen. Im Mündlichen hat auf allen Differenzmaßen die Mittelschicht in den Wortgruppen, die allgemein höhere Differenzmaße aufweisen ("A ungerade" und "B gerade") die höheren Differenzen zwischen Verständnis und Explikation, während die Unterschicht in den Wortgruppen die größeren Differenzen zeigt, in denen allgemein weniger zwischen Explikation und Verständnis unterschieden werden konnte ("A gerade" und "B ungerade"). Im Schriftlichen zeigt die Mittelschicht auf allen Differenzmaßen und in allen Wortgruppen (mit einer minimalen Ausnahme) die größeren Differenzen. Die Mittelschicht zeigt also in "A ungerade" und "B gerade" die größeren Differenzen, unabhängig von der Kommunikationsform, während die Richtung der Schichtdifferenz in den beiden anderen Wortgruppen von der Kommunikationsform abhängt. Die Mittelwertvergleiche zeigen also eine Abhängigkeit der Schichtunterschiede von der Kommunikationsform und der Wortgruppen; sie bieten nicht den gesuchten Anhaltspunkt für eine Interpretation der Schichtunterschiede trotz ihrer geringen Signifikanz. Strukturunterschiede der Wortkenntnis zwischen den Schichten lassen sich anhand der Differenzen zwischen Verständnis und Explikation nicht beschreiben.

Der zweite der eingangs erwähnten Einflüsse auf die Differenzmaße, der Wort- bzw. Experimentalgruppeneffekt, kann sich aus unterschiedlichen Leistungen der Vpn. oder unterschiedlichen Schwierigkeiten der Wortgruppen ergeben haben. Wir müssen zunächst diese Frage klären. In der Vierweg-Varianzanalyse aller Experimentalfaktoren (Tab. 77) können wir den Wortgruppeneffekt unmittelbar, den Experimentalgruppeneffekt in der Form des Wortgruppen-Kommunikationsform-Interaktionseffekts überprüfen. Der signifikante Wortgruppeneffekt auf beiden Differenzmaßen und die F-Werte unter 1 für den Wortgruppen-Kommunikationsform-Interaktionseffekt auf beiden Differenzmaßen belegen deutlich die größere Wichtigkeit des Unterschieds zwischen den Wortgruppen gegenüber dem Unterschied zwischen den Vpn.-Gruppen. Sie läßt sich veranschaulichen, wenn wir die Zusammenfassung der Differenzen nach Experimentalgruppen mit der Zusammenfassung der Differenzen nach Wortgruppen vergleichen (Tab. 76 und Abb. 9 für die Gesamtgruppe, Tab. 76 und Abb. 10 für die Mittel- und Unterschicht).

Wie aus Tab. 76 und Abb. 9 eindeutig hervorgeht, sind in der Gesamtgruppe die Wortgruppen die entscheidende Ursache für die Höhe der Differenzen.¹⁾ Obwohl die Vpn. vertauscht wurden, wiederholt sich das Profil der Differenzen in den Wortgruppen im Mündlichen und Schriftlichen sehr genau, während bei der Zusammenfassung nach Experimentalgruppen alle Unterschiede sich nivellieren. Die größten Unterschiede liegen zwischen den Wortgruppen der Testform A, während sie in der Testform B geringer, aber ebenso unabhängig von der Experimentalgruppenzugehörigkeit wie in der Testform A sind.²⁾ Innerhalb der Experimentalgruppen auf der anderen Seite durchkreuzen sich die Einflüsse der Wortgruppen fast völlig, so daß die Unterschiede in

¹⁾ -----
Daß die Differenzmaße von der Wortgruppe und nicht von der Vpn.-Gruppe abhängen, deutet in einer ähnlichen Weise auf semantische Eigenschaften der Worte jeder Gruppe wie die Ähnlichkeit der Profile der Explikationsbewertungen zwischen mündlicher und schriftlicher Messung innerhalb jeder der Wortgruppen, nicht aber innerhalb der Experimentalgruppen (siehe Tab. 2, (2) und (3), Abb. 4).

²⁾ Die größere Unterschiedlichkeit der Wortgruppen der Testform A auf den Differenzmaßen hat eine Parallele in ihrer größeren Unterschiedlichkeit auf der durch die "k.A." gemessenen Erklärungsschwierigkeit (siehe Abschnitt 5.3.4.); dies gilt gerade für die Maße DIBVE und DICVE, in die die Anzahl der Antworten nicht eingeht.

den Gruppen 1 und 2 - der Testform A - und der Gruppen 3 und 4 - der Testform B - auf dem gleichen Niveau liegen. Denselben Sachverhalt wie in der Gesamtgruppe können wir auch in den beiden Schichtgruppen wiederfinden. Wie aus Tab. 76 und Abb. 10 hervorgeht, sind die Profile der beiden Schichtgruppen bei einer Zusammenfassung nach Wortgruppen ausgeprägter als bei einer Zusammenfassung nach Experimentalgruppen; weiterhin ähneln sich die Profile beider Schichtgruppen im ersten Fall stärker als im letzten. Wir können an Tab. 76 und Abb. 10 fernerhin an einem Beispiel sehen, daß die Wortgruppe viel stärker als die Schichtzugehörigkeit die Differenzmaße bestimmt. Obwohl zwischen den Wortgruppen "A gerade" und "A ungerade" die Richtung der Schichtdifferenz auf DIBVE und DICVE sich umkehrt, sind in beiden Schichtgruppen DIBVE und DICVE in der Wortgruppe "A gerade" größer als in der Wortgruppe "A ungerade".

Tatsache ist, daß die Differenzen weder von der Schicht, noch von den Experimentalgruppen, sondern allein von den Wortgruppen abhängen, d.h. sie werden von den Besonderheiten der Stimuluswörter präformiert, noch ehe eine empirisch wirksame Bedingung die Erklärungen auf die eine oder andere Art beeinflussen kann. Wenn wir uns aber die Eigenarten der Maßkonstruktion ansehen, wird deutlich, warum die Differenzen durch die Besonderheiten der Stimulusworte bestimmt sind. Die Differenzen beziehen sich auf die Werte ein und derselben Worterklärung auf zwei Skalen, von denen der eine nur dadurch gewonnen werden konnte, daß über die in der ersten Skala berücksichtigten expliziten Informationen hinaus die in ihnen impliziten Aussagen, die die Vp. für die gegebene Erklärung im Kopf gehabt haben muß, vom Auswerter rückerschlossen und der Vp. gleichsam gutgeschrieben wurden. Bei der Explikation galt nur, was gesagt oder gesprochen war; beim Verständnis auch noch, was man daraus herauslesen konnte. Der Auswerter arbeitete mit den

Regeln semantischer Implikation. Nun wird in der Semantiktheorie (Katz/Foder 1963) die Bedeutung eines Wortes als eine Hierarchie von semantischen Implikationen allgemeiner Kategorien - "semantic marker" - und zusätzlicher wortspezifischer Bedeutungsartikel - "distinguisher" - dargestellt. Wenn eine solche Bedeutungsstruktur eines einzelnen Wortes das Regelsystem ist, mit dem der Auswerter arbeitet und mit dem er besonders seine Verständnisbewertungen gewinnt, dann ergeben sich zwei Möglichkeiten für die Unterschiedlichkeit der Worte, die direkt die Verständnisbewertung erleichtern: die Worte können sich durch eine hohe Anzahl von hierarchischen Stufen und Verflechtungen innerhalb des semantischen "Phrasenbaums" und durch ein besonders günstiges Verhältnis der in allgemeinen Kategorien zu den residualen "Distinguishers" repräsentierbaren Bedeutungsartikel ausgezeichnet haben. Diese Vermutung ist zwar keineswegs für die vorliegenden Worte durch eine wenigstens ad hoc konstruierte Semantik bestätigt worden, aber die Konstanz der Differenzen innerhalb der Wortgruppen bei sich ändernden experimentellen Bedingungen drängt den Gedanken der unbemerkten Wirksamkeit semantischer "Phrasenbäume" bei der Auswertung geradezu auf. Ganz entsprechend hat ja auch eine derartige Semantiktheorie viel mehr Plausibilität für sich als Programm denn als penibel durchgeführte Einzelanalyse.

5.4.5. Schichtunterschiede in der Verteilung der Antworten auf die Kategorien der Auswertung.

Wir haben im vorangehenden Abschnitt die durchschnittlichen Unterschiede für Schicht- und Wortgruppen zwischen den Bewertungen für Verständnis und Explikation analysiert. In diesem Abschnitt sollen die Verteilung der Bewertungen auf die 9 Verständnis- und die 9 Explikationskategorien, aus denen sich die Differenzen ergeben haben, wiederum innerhalb der Schicht- und Wortgruppen untersucht werden. Ein gleiches Gesamtmaß für beide Schichtgruppen in Explikation oder Verständnis kann sich aus unterschiedlichen Häufigkeiten der Verständnis bzw. Explikationskategorien in den beiden Schichten

ergeben haben. Die Häufigkeitsprofile können mit der Theorie schichtspezifischer Strategien des Sprachgebrauchs unmittelbar verbunden werden: sie spiegeln die Wahl zwischen verschiedenen Erklärungsarten und damit Einstellungen zur Sprache wider. Mit den Häufigkeitsverteilungen der Kategorien lassen sich, wie mit den Differenzen zwischen Verständnis und Explikation Schichtunterschiede unabhängig vom absoluten Leistungsniveau untersuchen. Sie bieten also wiederum eine Prüfmöglichkeit für die Grundhypothesen unseres Experiments (siehe 3.): eine Theorie schichtspezifischer Strategien des Sprachgebrauchs müßte Unterschiede in der Häufigkeitsverteilung der Kategorien zwischen den Schichten trotz gleicher absoluter Erklärungsleistungen voraussagen.

Auch anhand der Häufigkeitsprofile der Explikations- und Verständniskategorien können wir jedoch keine vom absoluten Leistungsniveau unabhängige Unterschiede im Sprachgebrauch der beiden Schichten belegen. Die Profile der Verständniskategorien (Abb. 2) ähneln sich in der Unterschicht und der Mittelschicht verblüffend,¹⁾ die Profile der Explikationskategorien (Abb. 3) weitgehend. Insbesondere finden wir keine Bevorzugung von Klassen-Definitionen in der Mittel- und Beispiels-Definitionen in der Unterschicht (Siller 1957, siehe I. 2.2.1.). Die Korrelationen der Rangfolgen der Explikationskategorien zwischen der Unterschicht und der Mittelschicht betragen für die mündliche Messung .83, für die schriftliche Messung .95 (Tab. 2). Die Gleichförmigkeit der Profile ist durch die Gleichheit der Durchschnittswerte keineswegs bestimmt; durch die Gleichheit der Durchschnittsmaße sind die Möglichkeiten der Verteilung der Antworten auf die 9 Bewertungsstufen zwar eingeschränkt,

¹⁾-----
Innerhalb der Verständnisbewertung (Tab. 4, Abb. 2) wurde von den wegen der U-Form der Verteilung allein interessanten mittleren Kategorien die Kategorie 5 ("Verwechslung") in beiden Kommunikationsformen von beiden Schichten am häufigsten gewählt. Zudem scheint die Unterschicht Kategorie 5 dort zu bevorzugen, wo die Mittelschicht Kategorie 4 ("Hierarchisierungsfehler") wählt; denn bei den übrigen Kategorien zeigen sich auch keine minimalen Unterschiede zwischen den Schichten.

aber Bewertungsunterschiede ausgleichende Profilalternativen nicht ausgeschlossen. Die Stabilität der Profile in beiden Schichten ist daher ein Ergebnis, das über das negative Hauptergebnis hinausführt. Es ist ein weiterer Beleg gegen die Theorie schichtenspezifischer Strategien des Sprachgebrauchs und lenkt unsere Aufmerksamkeit wiederum auf alternative Erklärungen: entweder auf entwicklungspsychologische Bedingungen, die für Kinder beider Schichten gelten, oder - besonders nach den Ergebnissen zu den Differenzen zwischen Explikation und Verständnis - auf Besonderheiten des im Test verwandten sprachlichen Materials. Für beide Erklärungen müßte ein "Sinn", eine "Logik" hinter den Häufigkeitsverteilungen der Kategorien gefunden werden. Wir wollen zuerst eine solche "Logik" in der Verteilung der Kategorien für alle Gruppen und dann anhand der Verteilungen der Kategorien in den einzelnen Wortgruppen Indizien für eine der beiden Erklärungen suchen; wir beschränken die weitere Darstellung auf die Explikationsbewertung.

Die Gleichartigkeit des mündlichen und schriftlichen Profils der Explikationskategorien in den Schichtgruppen (Abb. 3), ist ein Indiz, daß die Häufigkeit einzelner Kategorien sich aus ihrer Besonderheit ergeben hat. Diese Besonderheit können wir einerseits auf Abgrenzungsprobleme bei der Auswertung (siehe 1.1.4.) und andererseits auf den allen Kindern gemeinsamen Erklärungsstil zurückführen. Kategorie 4 ("Beschreibung der wesentlichen Funktion, Bedingung etc.") ist in allen Gruppen - beiden Schicht- und beiden Geschlechtsgruppen - in der schriftlichen Messung am seltensten gewählt worden und in der mündlichen Messung unter den drei am wenigsten gewählten Kategorien. Sie beschreibt jedoch eine klar umgrenzte und leicht identifizierbare Art der Explikation, die - zumal in schriftlichen Äußerungen - durch Normen des konventionellen Sprachgebrauchs abgewertet ist. Wir glauben deshalb nicht,

75-----

Wie wir schon in Abschnitt 1.1.4. dargelegt haben, sind die mündlichen und schriftlichen Profile auch in den beiden Geschlechtsgruppen gleich; siehe dazu die Korrelationen in Tab. 2.

daß die Seltenheit dieser Kategorie ein Ergebnis von Schwierigkeiten der Auswertung ist. Kategorie 2 ("Sehr prägnante Umschreibung der Bedeutung") steht zwischen zwei recht detailliert beschriebenen Kategorien, ohne selber in Unterklassen aufgeteilt zu sein; zudem ist ihre Beschreibung rein wertend und nicht an identifizierbaren Merkmalen orientiert. Nachträglich glauben wir, daß ihre Seltenheit in allen betrachteten Gruppen und Messungen sich auch aus ihrer Unschärfe ergibt. Kategorie 7 ("Beschreibung einer peripheren Bedeutungskomponente") und 8 ("Vage Assoziation") sind relativ genau beschrieben, so daß ihre Seltenheit sich wohl aus dem Erklärungsstil der Kinder ergibt; vermutlich wäre es besser gewesen, die fünf Unterklassen beider Kategorien in einer Kategorie zusammenzubringen. Sehr häufig treten die Kategorien 3 ("Beschreibung der wesentlichen Bedeutung") und 6 ("Richtige beispielhafte Verwendung") auf; beide sind genau spezifiziert, ihre Häufigkeit kann sich nicht aus Schwierigkeiten der Auswertung ergeben haben. Insgesamt scheinen eher der tatsächlich vorherrschenden Erklärungen der Kinder als Schwierigkeiten der Auswertung für die Häufigkeiten der einzelnen Kategorien verantwortlich zu sein.

Die Kategorien 1, 3 und 5 lassen sich als unterschiedlich gut gelungene Versuche zusammenfassen, mit Oberbegriffen und Differenzierungen Wortbedeutungen zu explizieren. Die Kategorien 2 ("Prägnante Umschreibung"), 4 ("Beschreibung der wesentlichen Funktion") und 6 ("Richtige beispielhafte Verwendung") sind Alternativformen kategorialer Explikation; bemerkenswert ist, daß 10jährige Kinder hier die Beispiele gegenüber den Funktionsumschreibungen sehr stark bevorzugen, obwohl die Funktionsumschreibungen doch als die "reifere" Form angesehen werden muß. Wir können die beiden Gruppen von Kategorien als "analytische" und "deskriptive" bezeichnen.

Mit Hilfe der Gegenüberstellung "analytischer" und "deskriptiver" Kategorien läßt sich die für beide Schichtgruppen gültige "Logik" dieser Verteilungen beschreiben: Die Kinder unserer Untersuchungsgruppe (4. Schuljahr) sind in der Entwicklung "analytischer"

Explikationen von zu partiellen oder zu allgemeinen Erklärungen (Kategorie 5) bereits weitgehend abgekommen, ohne jedoch schon durchweg Synonyme und exakte Kategorisierungen (Kategorie 1) geben zu können; wenn sie "analytische" Erklärungen anstreben, bleiben sie meist zwischen beiden Extremen (d.h. bei Kategorie 3) stecken. In der Entwicklung "deskriptiver" Explikationen sind die Kinder unserer Untersuchungsgruppe jedoch noch weniger fortgeschritten; hier bevorzugen sie eindeutig "Beispiele" (Kategorie 6) vor funktionalen (Kategorie 4) oder gar exakten (Kategorie 2) Umschreibungen. Vermutlich werden die besseren "deskriptiven" Erklärungen nicht gewählt, weil sie den besseren "analytischen" an Deutlichkeit und Akzeptabilität unterlegen sind; wer ein Wort gut kennt, wird es auch gleich "analytisch" explizieren, d.h. Kategorie 1 oder 3 statt Kategorie 2 oder 4 wählen. Umgekehrt werden die schlechteren "analytischen" Erklärungen nicht gewählt, weil sie den schlechteren "deskriptiven" an Farbigkeit und Kommunikationseffizienz unterlegen sind; wer ein Wort nur ungefährkennt, wird es lieber mit einem Beispiel als mit zu allgemeinen oder zu partiellen Kategorisierungen erklären, d.h. statt Kategorie 5 Kategorie 6 wählen. Auf diese Weise kann man für jeden Stand des Wortverständnisses die Worterklärungen als Alternativen zwischen den beiden allgemeinen Arten "analytischer" und "deskriptiver" Erklärung ansehen. Eine solche "Logik" der Verteilung der Explikationskategorien¹⁾ ließ sich evtl. durch eine Auszählung der vorwiegend zu den Explikations-

1)-----
1) Wir haben die "Logik" der Verteilungen der Explikationskategorien aus der Sicht des Kindes beschrieben, von dem eine Worterklärung verlangt wird; wir hätten sie genau so gut aus der Sicht des Auswerters beschreiben können, der die Worterklärungen bewerten muß. Ein illustratives Beispiel soll genügen. Statt: "Wer ein Wort gut kennt, wird es auch gleich 'analytisch' erklären" hätten wir ebenso schreiben können: "Wo eine Erklärung gutes Wortverständnis ersehen läßt, wird sie wahrscheinlich mit einer 'analytischen' Kategorie bewertet werden". Beides läuft auf eine "Wahl", des Kindes oder des Auswerters, von Kategorie 1 an Stelle von Kategorie 2 hinaus.

kategorien gehörenden Verständniskategorien verifizieren. Wir wollen jetzt versuchen, anhand der Profile der Explikationsbewertungen in den einzelnen Wortgruppen zwischen der "entwicklungspsychologischen" und der "sprachlogischen" Erklärung zu entscheiden. Deutliche Profilunterschiede zwischen den Wortgruppen sprechen für eine "sprachlogische", geringe für eine "entwicklungspsychologische" Erklärung.

Betrachtet man die Profile der Explikationsbewertungen in den 4 Wortgruppen (Tab. 3, Abb. 4) als Ganzes, so stimmen die Profile der beiden Wortgruppen B weitgehend überein; sie sind auffällig markanter als die stärker nivellierten Profile der Wortgruppen A, die jedoch intern wenig Gemeinsamkeiten zeigen. Betrachtet man die Profile in den 4 Experimentalgruppen (Tab. 5, Abb. 5) als Ganzes, so finden wir Differenzen in der Ausgeprägtheit, nicht aber der Struktur zwischen den jeweils zwei Experimentalgruppen mit der gleichen Testform und kaum Unterschiede innerhalb der zwei Experimentalgruppen mit der gleichen Testform. Aus den globalen Profilvergleichen lassen sich also ebenso viel Gründe für die "sprachlogische" wie die "entwicklungspsychologische" Erklärung entnehmen. Jedoch kann die Stabilität der Profile in der Gesamtgruppe um so schwerer aufrechterhalten werden, je weiter man die Gruppen aufteilt. Insofern liegt in unserem Verfahren ein "bias" zugunsten der "sprachlogischen" Erklärung; die für die "entwicklungspsychologische" Erklärung sprechenden Belege aus dem Vergleich der Experimentalgruppen (Tab. 5, Abb. 5) müßten stärker gewichtet werden als die für die "sprachlogische" Erklärung sprechenden Belege aus dem Vergleich der Wortgruppen.

Eine genauere Betrachtung der einzelnen Profile in den vier Wortgruppen, die sich an der zuvor erläuterten "Logik" der Verteilung orientiert, bringt zwar insgesamt mehr Belege für die "sprachlogische" Erklärung, steht aber auch unter dem genannten "bias". Einerseits ragt die Kategorie 4 in allen vier Wortgruppen hervor. Andererseits aber verliert die Kategorie 6 sowohl in der Wortgruppe "A gerade" wie "B gerade" ihre Dominanz gegenüber der Kategorie 5 und ist die Kategorie 1 in der Wortgruppe "A ungerade" im Gegensatz

zu allen drei übrigen Wortgruppen nur minimal besetzt. Alle diese genannten Tendenzen für Einzelkategorien stimmen in beiden Kommunikationsformen überein. Ein weiteres Entscheidungskriterium findet sich in den Rangkorrelationen der Häufigkeiten der Kategorien: Die Korrelationen für verschiedene Vpn. mit der gleichen Wortgruppe sind höher als die Korrelationen der gleichen Vpn. mit verschiedenen Wortgruppen (siehe Tab. 2 (2) und (3)). Auch das spricht eher für die "sprachlogische" Erklärung. - Insgesamt jedoch sind die Belege für die "sprachlogische" Erklärung, angesichts des dargestellten "bias", nicht stark genug. Wir wollen also nach wie vor auch die "entwicklungspsychologische" Erklärung zulassen. Abschließend wollen wir noch einmal wiederholen, daß die Stabilität der Häufigkeitsprofile der Kategorien keine triviale Tatsache ist. Nach den negativen Ergebnissen für die Gesamtleistungen hätten Unterschiede der Kategorien-Profile ein Indiz für schichtspezifische Arten des Sprachgebrauchs sein können. Daß jedoch auch die Kategorien-Profile zwischen den Schichtgruppen nicht unterscheiden, ist ein weiteres Indiz für die Abhängigkeit der Testleistungen vom Entwicklungsstand der Vpn. oder von den Besonderheiten der Stimulusworte.

5.4.6. Schichtunterschiede der Anzahl falscher Antworten

Obwohl die Varianzanalysen auf keinem Maß der zweiten Messung unter den Konventionellen Schwellen signifikante Schichtunterschiede herausarbeiten konnte, so zeigt sich doch auf den Maßen, die Erklärungs-bereitschaft und Worterklärung zugleich maßen (A-Kategorisierung), eine leichte Überlegenheit der Mittelschicht; auf den Maßen, die Worterklärung unabhängig von der Erklärungs-bereitschaft prüfen (B- und C-Kategorisierung), konnte eine Überlegenheit der Mittelschichtkinder auch nur der Tendenz nach nicht mehr nachgewiesen werden (siehe 5.3.2.). Obwohl die Mittelschichtkinder die Worte nicht besser erklären können, sind sie zu größeren Risiken bereit. Aus dem gleichen Grund aber, aus dem sie mehr Antworten geben, müßten sie auch mehr falsche Antworten geben. Wenn nicht nur die Anzahl, sondern auch der Anteil der falschen an den gegebenen Antworten bei Kindern aus der Mittelschicht größer ist als bei Kindern aus der Unterschicht, so ließe sich Selbsttäuschung über die eigene Wortkenntnis als ein zusätzlicher unabhängiger Faktor der größeren Antwortfreudigkeit der Mittelschichtkinder nachweisen. Anders als bei den Maßen der A-Kategorisierung, die die Risikobereitschaft positiv bewerteten, soll der Vergleich des Anteils der falschen an den gegebenen Antworten die negativen Bedingungen größerer Risikobereitschaft prüfen.

Um diese Frage zu klären, haben wir in jeder Wortgruppe und Kommunikationsform alle diejenigen Antworten ausgezählt, die für einen mißglückten Explikationsversuch, nicht für eine fehlende Explikation den Skalenwert "9" erhielten (siehe dazu auch Tab. 3: Verteilung der Explikationswerte auf "K.A.", "9", und zwischen "1" und "8"). Die Anzahl der Fehler (Tab. 79, Zeile 1) ist immer auf die mögliche Gesamtzahl von Antworten in einer Wortgruppe (30) bezogen; sie mißt, entsprechend der Kategorisierung A, falsches Wortverständnis unabhängig von der Erklärungs-bereitschaft einer Gruppe. Der Anteil der Fehler (Tab. 75, Zeile 3) ist jedoch auf die von der jeweiligen Gruppe gegebenen Erklärungen bezogen; er mißt, entsprechend der

Kategorisierung B, falsches Wortverständnis im Rahmen der Erklärungsbereitschaft einer Gruppe.¹⁾ Der Vergleich der Fehleranteile zwischen den Schichten zeigt, in welcher Schicht Unkenntnis des eigenen Wortschatzes zusätzlich zur gegebenen Risikobereitschaft einer Gruppe zu falschen Antworten geführt hat.

Tab. 80 gibt einen Überblick über die Stimulusworte, die falsche Antworten hervorgerufen haben; die falschen Antworten sind nicht vollständig zitiert, sondern nur die Fehler, die vermutlich zur falschen Antwort geführt haben. Zugleich ist versucht worden, die falschen Antworten nach der Art der zugrundeliegenden Fehler zu klassifizieren. Die vorherrschende Ursache von falschen Antworten ist, sowohl für jede einzelne als auch alle Wortgruppen zusammen, "Lesefehler bzw. Lautassoziation"; "Konzeptmißverständnis" ist lediglich eine Restkategorie, da sich die Zwischenschritte, die ein Kind zu einer für den erwachsenen Auswerter nicht verständlichen Erklärung geführt haben, natürlich nicht rekonstruieren lassen. Da sich die Schichtunterschiede der beiden Fehlerkategorien in allen Wortgruppen (Tab. 81) nicht unterscheiden, werden wir im folgenden nur noch die undifferenzierten Anzahlen und Anteile der Fehler berücksichtigen.²⁾

1) -----
Die Anzahl der gegebenen falschen Antworten entspricht der Differenz zwischen den Maßen GATEX und RATEX; sie werden in der A- und C-Kategorisierung wie eine nicht gegebene Antwort bewertet, sind in den Gesamtmaßen der A-Kategorisierung jedoch mit, in der C-Kategorisierung ohne Einfluß auf die Gesamtbewertung ein Vp.; allein in der B-Kategorisierung werden die falschen als gegebene Antworten gewertet und in der Gesamtbewertung berücksichtigt. Der Anteil der falsch gegebenen Antworten entspricht dem Maß PRIAT, das auch in den verschiedenen Varianzanalysen geprüft wurde; es ergaben sich für PRIAT jedoch weder ein signifikanter Schicht-, noch Geschlechts- oder Experimentalgruppeneffekt.

2) Zudem zeigte sich entgegen unseren Erwartungen kein Überwiegen der "Lesefehler bzw. Lautassoziationen" in der mündlichen, der "Konzeptmißverständnisse" in der schriftlichen Messung. Auch unter dem Gesichtspunkt der Kommunikationsform können also beide Fehlerarten zusammengelegt werden (Tab. 82).

Für alle Wortgruppen zusammen zeigt sich zwar in der mündlichen wie schriftlichen Messung eine größere Fehleranzahl in der Mittelschicht, jedoch ist die Differenz der Fehleranteile nur noch minimal und allein in der mündlichen, nicht mehr der schriftlichen zugunsten der Mittelschicht (Tab. 79). Innerhalb der Wortgruppen und Kommunikationsformen wechseln die Schichtdifferenzen der Fehleranzahlen und -anteile gleichsinnig, aber unsystematisch; Schichteinflüsse scheinen daher nur sekundär den beherrschenden Einfluß der Wortgruppen zu durchbrechen. Anders gesagt: Auf ihrem Entwicklungsstand können die meisten der etwa 10jährigen Kinder der Versuchung zur Abschweifung, die in bestimmten Worten liegt, noch nicht widerstehen - die Unterschichtkinder vielleicht etwas besser. Weitere Hinweise für den dominanten Einfluß von Entwicklungsstand und Wortbeschaffenheit und den nur modifizierenden Einfluß sozialer Schichtzugehörigkeit sollen kurz aufgezählt werden. Zunächst fällt die Gleichartigkeit vieler den falschen Erklärungen zugrundeliegender Mißverständnisse, besonders bei "Lesefehlern bzw. Lautassoziationen" auf (Tab. 80). Zweitens haben einige Worte besonders viele falsche Erklärungen hervorgerufen, sehr viele Worte jedoch überhaupt keine (Tab. 83); die falschen Erklärungen auf die extrem oft mißverstandenen Worte verteilen sich gleichmäßig auf beide Schichten (Tab. 84). Drittens haben einige Vpn. besonders viele falsche Erklärungen gegeben, recht viele jedoch nur eine oder keine (Tab. 85); die Vpn., die am häufigsten falsche Antworten gegeben haben, verteilen sich wiederum gleichmäßig auf beide Schichten (Tab. 86). Viertens haben die Wortgruppen der Testform B in beiden Schichten eine höhere Anzahl, nicht aber einen höheren Anteil von falschen Erklärungen als die Wortgruppen der Testform A hervorgerufen. Für beide Schichten hatte die Verwandlung des passiven multiple-choice-Test in einen aktiven Erklärungstest ähnliche Verfremdungseffekte: In der Testform A werden weniger Worte erklärt mit einem geringfügig größeren Fehleranteil, in der Testform B tauchen absolut mehr falsche Erklärungen auf (siehe Abschnitt 5.3.5.).

Die Untersuchung der Schichtunterschiede von falschen Erklärungen ist insgesamt also negativ verlaufen. Ganz wie die Berücksichtigung der Erklärungsbereitschaft bei den Maßen der A-Kategorisierung den Mittelschichtkindern nur eine sehr geringe Überlegenheit ermöglicht

hat, so zeigen sie die negativen Bedingungen von Erklärungsbereitschaft, Unkenntnis über den eigenen Wortschatz, nur etwas stärker als die Unterschichtkinder. Anzahl und Anteil der falschen Erklärungen sind offensichtlich nicht durch schichtspezifische Strategien des Sprachgebrauchs bedingt; vielmehr hängen sie innerhalb unserer Gruppe, in der von der Sozialschicht beeinflusste passive Wortschatz kontrolliert war, von der Beschaffenheit der Worte bzw. dem Entwicklungsstandes des Kindes ab.

5.4.7. Bewertung der Einzelanalysen.

Alle sechs Analysen dieses Abschnitts konnten keine Schichtunterschiede nachweisen; jede von ihnen aber bekräftigte den Eindruck, daß die Leistungen beider Schichtgruppen stark von der Besonderheit der Stimulus-Worte oder der Vpn. abhängig sind. Die Wilcoxon-Tests für einzelne Worte (5.4.1.) zeigten, daß beide Schichten auf die gleichen Worte keine oder gleich gute Erklärungen geben. Auch die Zusammenfassung der Einzelworte nach dem Grad extravverbaler Explizierbarkeit (5.4.2.) erbrachte, daß beide Schichten die einzelnen Wortgruppen in der gleichen Weise besser oder schlechter erklären können. Schließlich unterschied sich auch die Rangfolge der Schwierigkeit aller Worte (5.4.3.) in keiner Weise zwischen den Schichten. Ähnliche Ergebnisse wie bei der Betrachtung der Einzelworte zeigten sich auch bei der Untersuchung einzelner Aspekte der Gesamtleistung. Die Differenzen zwischen Verständnis und Explikation unterscheiden so wenig wie beide Dimensionen für sich zwischen den Schichtgruppen, sind jedoch von den Wortgruppen des Tests abhängig und können auf Besonderheiten der Worte hinweisen, die unbemerkt und durchgängig die Auswertung geleitet haben (5.4.4.). Die Häufigkeitsprofile der Verständnis- und Explikationsbewertungen sind in den Schichtgruppen und Kommunikationsformen erstaunlich stabil, so daß sie von den Vpn. oder, wie die Differenzen, von den Wortgruppen abhängen (5.4.5.). Die falschen Antworten scheinen schließlich von den Versuchungen des Assoziationsbereichs und der Phonetik bestimmter Worte und der Impulsivität bestimmter Vpn. abzuhängen (5.4.6.).

Wie wir zu Beginn dieses Abschnittes erläutert haben, hätten alle Analysen dieses Abschnittes auch die Theorie der Strategien schichtspezifischen Sprachgebrauchs unterstützen können; für alle untersuchten Variablen galten durch die Kontrolle des passiven Wortschatzes die gleichen strengen Prüfbedingungen wie für die Gesamtleistung. Die negativen Ergebnisse der Einzelanalysen unterstreichen also noch das für die Theorie schichtspezifischen Sprachgebrauchs negative Ergebnis für die Gesamtleistung.¹⁾

Auf der anderen Seite sind alle in diesem Abschnitt untersuchten Variablen entweder von den Vpn. oder den Wortgruppen abhängig: wir finden "entwicklungspsychologisch" oder "sprachlogisch" bedingte Muster. Da wir jedoch durch die Kontrolle des passiven Wortschatzes nicht direkt auf schichtspezifische Arten des Sprachgebrauchs verweisende Einflüsse der sozialen Schicht in unserer Experimentalgruppe bereits ausgeschaltet haben, können die Ergebnisse der Einzelanalysen nicht die Unabhängigkeit dieser Muster von sozialen Einflüssen nachweisen. Insbesondere können sie nicht die Annahme schichtspezifischer Vokabulare - die allerdings in einer Theorie des schichtspezifischen Sprachgebrauchs nur von untergeordnetem Interesse ist - widerlegen. Für die Überprüfung einer solchen Annahme hätte es weder der Beschränkung auf die Worte eines standardisierten Tests noch der experimentellen Kontrolle des passiven Wortschatzes bedurft. Kurz: die Ergebnisse dieses Abschnitts schließen einen schichtspezifischen Wortschatz nicht aus, sie falsifizieren aber die Annahme eines schichtspezifischen Gebrauchs des Wortschatzes.

1)-----
In dieses negative Gesamtbild fügen sich auch die Ergebnisse zur Beziehungsstruktur der Maße, die in Abschnitt 1.2.4. und 1.2.5. dargestellt sind: Es finden sich viel stärkere Übereinstimmungen als Differenzen zwischen den beiden Schichtgruppen in der Beziehungsstruktur der Maße.

5.5. Zusammenfassung der Ergebnisse des dichotomen Schichtvergleichs

Bevor wir zur Untersuchung differenzierter sozialer Faktoren übergehen, wollen wir die Ergebnisse des dichotomen Schichtvergleichs noch einmal kurz und thesenartig zusammenfassen:

1. Auf allen Maßen und in beiden Prüfverfahren - dem Wicoxon-Test und der Varianzanalyse - zeigten sich durchweg keine signifikanten Schichtunterschiede in der Gesamtleistung der zweiten Messung. Bei Kontrolle des passiven Wortschatzes ergeben sich zwischen Kindern der Unter- und Mittelschicht also keine Unterschiede in der aktiven Worterklärung. Schichtunterschiede des Sprachgebrauchs konnten daher nicht nachgewiesen werden; Schichtunterschiede in der Sprache sind dadurch aber nicht ausgeschlossen und nach den Ergebnissen der ersten Messung nach wie vor gegeben.
2. In der Varianzanalyse konnte, unterhalb des konventionellen Signifikanzniveaus, ein Muster stärker und schwächer schichtdiskriminierender Maße aufgedeckt werden, an dem sich auch die Sensibilität unserer Maßdimensionen für Schichtunterschiede ablesen läßt:

Tendenziell in der Mittelschicht bessere Leistungen zeigten sich auf allen Maßen, die durch die vorgegebene Anzahl der Worte in den Testformen definiert waren ("A-Kategorisierung"); —, überhaupt keine Unterschiede oder Trends mehr zeigten sich auf den Maßen, die durch die von der Vpn. selbst gewählten Anzahl von Antworten definiert waren ("B- und C-Kategorisierung"). Dies Muster zeigt sich in der mündlichen Messung, nicht mehr in der schriftlichen Messung. Insgesamt also scheint die schriftliche Darstellung eine, wenn auch geringe, disziplinierende Wirkung für die Kinder aller Schichten gehabt zu haben, während die mündliche Erklärung die Kinder der Mittelschicht eher dazu verleitet hat, alle ihnen im Test vorgegebenen Worte zu erklären.

Die Trennung von Verständnis und Explikation, die mögliche Defizite der Unterschichtkinder in der sprachlichen Vermittlung trotz guter Kenntnisse aufdecken sollte, kommt eher den Mittelschichtkindern entgegen; sie sind den Unterschichtkindern auf den Verständnismaßen eher überlegen als auf den Explikationsmaßen. Die Intention dieser Trennung läßt sich noch am besten an der der Kategorisierungen verfolgen, insofern die A-Kategorisierung bei den Mittelschichtkindern über ihr Verständnis hinausgehende Explikationen erfaßt.

Die Maße der Differenz zwischen Explikation und Verständnis sind entsprechend eher in der Mittelschicht größer. Viel stärker aber als zwischen den Schichten unterscheiden sie sich zwischen den Wortgruppen, und zwar unabhängig von den Experimentalgruppen und Kommunikationsformen.

Die starke Abhängigkeit der Differenzmaße von den Wortgruppen läßt sich am besten aus der Problematik verstehen, in der Auswertung zwischen Verständnis und Explikation zu differenzieren.

3. Wilcoxon-Tests mit anderen Kontrollvariablen als dem passiven Wortschatztest ergeben ebenfalls keine signifikanten Schichtunterschiede der Gesamtleistungen. Auch bei Kontrolle der verbalen Intelligenz, der nicht-verbalen Intelligenz oder der Schulnoten unterschieden sich die Kinder unserer Untersuchungsgruppe nicht in den aktiven Worterklärungen.
4. Auch bei verschiedenen Einzeluntersuchungen innerhalb des Experimentalplans, d.h. bei Kontrolle des passiven Wortschatzes, konnten keine Unterschiede zwischen den beiden Schichtgruppen aufgedeckt werden. Die Wilcoxon-Tests einzelner Worte und Wortgruppen ergaben keine signifikanten Schichteinflüsse, die sich mit Hypothesen über schichtspezifische Sozialisationsbedingungen hätten zusammenfassen lassen. Auch die Rangfolge der Schwierigkeit der Worte stimmt in beiden Schichten überein, ein "bias" des Testmaterials konnte nicht nachgewiesen werden. Ebensowenig zeigen sich Unterschiede zwischen den Schichten bei direkt auf den Sprachgebrauch bezogenen Variablen: der Differenz zwischen Verständnis und Explikation, der Verteilung der Erklärungskategorien und der Anzahl falscher Antworten. Alle genannten Variablen hingen jedoch mit den Wortgruppen oder den Experimentalgruppen zusammen. Auch wenn die Ergebnisse sich als entwicklungspsychologisch oder sprachlogisch bedingte Muster auffassen lassen, ist ihre Unabhängigkeit von sozialen Faktoren keineswegs erwiesen, da ein Großteil der Schichteinflüsse bei ihrer Prüfung ausgeschaltet war.

5.6. Unabhängig von dichotomer Schicht erfaßte Einflüsse
=====
der gesellschaftlichen Lage der Eltern auf die Lei-
=====
stungen des Kindes.
=====

Die Experimentalgruppe der zweiten Messung wurden gebildet, indem Paare von einem Unter- und einem Mittelschichtkind mit gleicher Leistung im passiven Wortschatztest aus der Gesamtstichprobe ausgewählt wurden. Die im passiven Wortverständnis parallelisierten Schichtgruppen wurden dann in ihrer aktiven Worterklärungsleistung verglichen. Die Hypothese der Überlegenheit der Mittelschichtkinder bei dem dichotomen Schichtvergleich, der dem Experiment zugrundelag, wurde jedoch nicht bestätigt. Außerdem aber hatten wir Indikatoren der sozialen Herkunft auf andere, natürlich mit der Dichotomisierung von Schicht zusammenhängende Arten erfaßt. Wenn in folgenden, trotz negativen Ergebnissen für die dichotome Schicht, nach Unterschieden in der aktiven Worterklärung zwischen Gruppen gesucht wird, die nach solchen Merkmalen gebildet sind, so müssen zuvor die Konsequenzen für den Experimentalplan sowie die möglichen Ziele einer solchen detaillierten Nachforschung erläutert werden.

Der Experimentalplan kann im folgenden nicht mehr aus innerhalb des Faktors Schicht parallelisierten Gruppen bestehen, die jeweils dem gleichen Geschlecht angehören und die gleichen Stimulusworte erklärt haben. Vielmehr ordnen wir den Vpn., die ursprünglich für eine Prüfung dichotomer Schicht bei kontrolliertem passiven Wortschatz ausgewählt wurden, Gruppen nicht von Anfang berücksichtigter Merkmale zu. Nicht mehr beachtet werden dabei die Faktoren Geschlecht und Experimental- bzw. Wortgruppe; es entstehen ungleiche Zellenhäufigkeiten. Nicht mehr ausgeglichen sind dadurch die Gruppenmittelwerte auf der Kontrollvariablen; so daß scheinbare Effekte auftauchen können. Da manche Merkmale - vor allem die Arbeitsplätze des Vaters und der Mutter - nicht für alle Vpn. des

ursprünglichen Experimentalplans angegeben wurden, verringert sich die Anzahl der Stichprobe I; in diesen Fällen, wo Ergebnisse der Stichprobe I mit den Ergebnissen des dichotomen Schichtvergleichs von der Vpn.-Zahl wie den Werten der Kontrollvariablen her sowieso nicht mehr vergleichbar sind, stützen wir uns auf die größere Stichprobe II (siehe Tab. 17). Aber auch in dieser größeren Stichprobe gilt, daß die untersuchten Einflüsse mit denen des Geschlechts und - bei dem Charakter der Maße und den schon dargestellten Ergebnissen (siehe 5.3.) wichtiger - der Experimental- bzw. Wortgruppen vermischt sind. Auch hier gilt, daß die Vpn. nicht nach den untersuchten Merkmalen ausgesucht wurden und die Gruppen sich auf der Kontrollvariablen unterscheiden können.

Daß der passive Wortschatz nicht mehr kontrolliert ist, vermindert die theoretische Aussagekraft der folgenden gegenüber den vorangegangenen Analysen der Gesamt- (5.2. und 5.3.) und Einzelleistungen (5.4.): Schichtunterschiede können nicht mehr ohne zusätzliche Überlegungen auf schichtspezifische Strategien des Sprachgebrauchs bezogen werden. Um die folgenden mit den vorangegangenen Analysen vergleichbar zu machen, werden wir für ihre einzelnen Gruppen die Wortschatz-Werte angeben; erst wenn sich unabhängig von der Variation der Wortschatzwerte (WST) Gruppendifferenzen ergeben, können sie mit Hilfe der Theorie der linguistischen Codes interpretieren.

Die Hypothesen, die wir zu Anfang der folgenden Abschnitte erörtern werden, sind daher nicht mehr direkt auf die Theorie der linguistischen Codes bezogen. Zudem ist ihr Spielraum durch das negative Hauptergebnis eingeschränkt. Soweit die untersuchten Einflüsse mit der Schichtdichotomie übereinstimmende Differenzierungen aufweisen - wie die weitere Aufgliederung der Schicht und die Arbeitsplatzkategorisierung des Vaters - können nur noch Unterschiede innerhalb der Schichtgruppen gesucht werden. Ziel aller folgenden Darstellungen trotz ihrer Beschränkungen soll sein, das negative Hauptergebnis zu differenzieren: Entweder läßt es sich auch noch mit Hilfe feinerer Indikatoren untermauern oder aber anhand widersprechender, aber soziologisch genauerer Ergebnisse kritisch modifizieren. Am Ende der folgenden Darstellung wollen wir entscheiden können, welche der beiden Richtungen der Interpretation angemessener ist.

5.6.1. Einflüsse einer weiter gegliederten Schichteinteilung.

Die Schichtzugehörigkeit der Vpn. wurde durch eine Kombination der Indikatoren "gegenwärtiger Beruf des Vaters", "Einkommen der Familie", "Ausbildung der Mutter", "Ausbildung des Vaters" festgelegt. Zuerst wurden fünf Straten gebildet, von denen dann jeweils die beiden extremen als die "Unter-" bzw. "Mittelschicht" unseres Experiments galten; die fünfte Schichtgruppe (die UMS) wurde für die Auswahl des Experimentalsamples (siehe II.2.) nicht berücksichtigt. Im folgenden sollen die Einflüsse dieser ursprünglichen Schichtgruppierung auf die Erklärungsleistungen untersucht werden. Sie ergibt sich durch eine erneute Aufteilung der globalen Gruppierung von Unter- und Mittelschicht in insgesamt vier Gruppen und enthält wie diese die Trennung von manueller und nichtmanueller Arbeit - jetzt zwischen den beiden oberen und unteren Schichten.¹⁾ Neue Ergebnisse nach dem negativen Hauptresultat können sich daher nur ergeben durch (1) extreme Unterschiede innerhalb der globalen Schichten und (2) durch eine Vergrößerung der sozialen Distanz beim Vergleich zwischen den globalen Schichten. Für beides sollen kurz mögliche Hypothesen erwogen werden.

(1) Innerhalb der Unterschicht ist nur eine Hypothese der Überlegenheit der OUS über die UUS plausibel. Innerhalb der Mittelschicht jedoch kann man eine Zunahme verbaler Leistungen mit sozialer Schicht vermuten, ebenso aber auch eine Abnahme: Die MMS kann um der Realisierung noch gegebener Mobilitätschancen willen gedrängt sein, die für die soziale Bewertung besonders wichtige Fähigkeit des verbalen Ausdrucks überzubetonen (Labov 1966, 1970); oder in der OMS könnte materielle Sätturiertheit zu geringeren persönlichen Aufwendungen der Eltern führen, so daß die verbalen Fähigkeiten der OMS-Kinder gegenüber den umhegteren UMS-Kinder abfallen (einen Hinweis dafür geben die Ergebnisse der ersten Messung, siehe 4.2. und Tab. 37).

¹⁾-----
Eine detaillierte Auflistung der in die Bestimmung der 4 Schichten: UUS, OUS, MMS, OMS eingehenden Indikatoren wurde für die Vpn. unseres Experiments in 2.1. gegeben, sowie in den Tab. 24, 26, 27, 28.

(2) Wegen des negativen Ausgangsergebnisses kann nur noch ein Vergleich zwischen Gruppen aus beiden Globalschichten neue Informationen liefern: Wir vermuten eine Unterlegenheit der UUS gegenüber der MMS oder der OMS oder beiden Mittelschichtgruppen zusammen. Ein solches Ergebnis würde es erlauben, den Widerspruch zwischen unseren negativen Ergebnissen und der in anderen Untersuchungen aufgedeckten Überlegenheit von Mittelschicht- über Unterschichtkinder in sprachlichen Leistungen auf unterschiedliche soziale Distanz zwischen den Schichtgruppen in den jeweiligen Experimentalgruppen zu beziehen. - Keine der hier diskutierten Hypothesen ist direkt aus der Theorie der linguistischen Kodes abgeleitet. -

Zur Überprüfung des Einflusses einer mehrstufigen Schichteinteilung bieten sich zwei Verfahren an: die Varianz- und die Regressionsanalyse. Beide Verfahren setzen für die abhängigen Variablen Messung auf dem Intervallskalenniveau voraus; inwiefern diese Bedingung für unsere Maße zutrifft, ist in 2.1. und 5.1. diskutiert worden. Die Regressionsanalyse setzt zudem für die unabhängige Variable Intervallskalenniveau voraus. Die fünf Schichten des Filterindex wurden innerhalb der gesamten ursprünglichen Untersuchungsgruppe, für die beantwortete Fragebogen vorlagen (N=294), z-transformiert; die vier in unserem Experiment vertretenen Schichtgruppen gingen mit diesen z-Werten in die Regressionsanalyse ein. Da die Regressionsanalyse eine lineare Beziehung zwischen abhängiger und unabhängiger Variabler zugrundelegt, ist sie ungeeignet, Unterschiede zwischen den extremen und den mittleren Gruppen herauszuarbeiten. Die Regressionsanalyse ist im Grunde - nach dem Wilcoxon-Test und der Varianzanalyse - ein drittes Prüfverfahren für den dichotomen Schichtvergleich. Mit der Varianzanalyse kann man dagegen extreme und mittlere Gruppen vergleichen und die zuvor skizzierten Hypo-

thesen überprüfen. Aus diesem Grunde werden wir die Ergebnisse der Regressionsanalyse nur kurz streifen, um dann die Anwendungsbedingungen und die Ergebnisse der Varianzanalyse ausführlich zu referieren.

Die Regressionsanalysen wurden mit für beide Kommunikationsformen getrennten Maßen durchgeführt. Die Korrelationen des Filterindex der sozialen Schicht mit den Maßen der zweiten Messung (Tab. 87) erreichen nirgends das 5%-Signifikanzniveau; es wäre nutzlos, mit diesen Koeffizienten die Nullhypothese für lineare Regression zu testen. Unterschiede in der Höhe der Korrelationskoeffizienten zwischen den beiden Stichproben (N=65 und N=83) und zwischen den Maß-Gruppen zeigen sich so gut wie nicht. Allein in der schriftlichen Messung für N=65 finden sich für die 6 Explikationsmaße und die 7 Maße der Erklärungsgüte Korrelationen noch unter dem allgemeinen niedrigen Niveau. Dies entspricht den Ergebnissen der Varianzanalysen, daß eine geringfügige und in keinem Fall signifikante Überlegenheit der Mittelschichtkinder bei den Explikations- und Gütemaßen nur in der mündlichen Messung auftaucht, die Berücksichtigung von "Verständnis" und Antworthäufigkeit also die Mittelschichtkinder begünstigt (siehe 5.3.2.).

Für die Anwendung der Varianzanalyse muß in diesem Abschnitt - wie in den beiden folgenden Abschnitten - zweierlei beachtet werden. Erstens ergeben sich anders als in den bisherigen Varianzanalysen ungleiche Häufigkeiten in den Zellen. In diesem Fall wird die Bedingung gleicher Varianzen der Irrtümer in den Zellen wichtig (Hays 1963:379). Wir haben diese Bedingungen in der Experimentalstichprobe I mit Hilfe des Hartley-Tests überprüft, der annäherungsweise auch im

Fälle ungleicher Anzahlen in den Zellen verwandt werden kann (Winer 1962:92 und 240); die $F(\max)$ -Werte sind in der letzten Zeile von Tab. 24 dargestellt. Leider ergeben sich für fast alle Explikations- und Verständnismaße hochsignifikante Unterschiede zwischen der größten und der kleinsten Zellenvarianz. In fast allen Fällen aber auch ist die größte Varianz in der mündlichen, die kleinste in der schriftlichen Messung der UUS. Dieser Unterschied zwischen den beiden Messungen in der UUS ergibt sich - wie eine Inspektion der Werte der 5 Vpn. zeigt - daraus, daß Vpn. 322 in der mündlichen Messung keine einzige Erklärung gegeben hat. Mittelwerte, Standardabweichungen und Varianzen von Maßen, bei denen die Zahl der gegebenen Antworten ins Gewicht fällt, wurden für die mündliche Messung der UUS daher ohne Vpn. 322 gesondert errechnet; jedoch unterscheiden sich, auch wenn man diese veränderten Varianzen der Berechnung der $F(\max)$ -Werte zugrundelegt, die Varianzen zwischen den vier Schichtgruppen hochsignifikant. Wie sich aus dem Vergleich der Standardabweichungen in Tab. 88 aber entnehmen läßt, sind die Varianzen beider Messungen der UUS einerseits und der OUS, MMS und OMS andererseits recht homogen. Trotz der Verletzung der Bedingung homogener Varianzen in allen 4 Schichtgruppen haben wir deshalb Varianzanalysen gerechnet. Die rechnerische Behandlung der ungleichen Zellenhäufigkeiten können wir erst nach der Klärung der weiteren Anwendungsbedingungen und der Vorstellung des verwandten Rechenprogramms erläutern.

Eine zweite Anwendungsbedingung muß in den folgenden - wie in allen vorangegangenen - Varianzanalysen beachtet werden, gewinnt aber durch die Ungleichheit und geringe Größe der Zellen stärkere Bedeutung; die Unabhängigkeit der Fehleranteile der Messungen zwischen den Zellen bzw. die Wiederholung der Messungen. In dieser und zwei der folgenden Analysen (5.6.3. und 5.6.4.) wird ein Faktor von geringerem Interesse, Kommunikationsform, innerhalb jeder Ebene eines anderen Faktors von vordringlichem Interesse, Schicht, Arbeitsplatz des Vaters bzw. der Mutter, wiederholt gemessen. Wir haben eine zweifaktorielle Experimentalanlage mit Wiederholung der Messungen auf einem Faktor; wir wollen erläutern, inwiefern die normale Zwei-Faktor-Varianzanalyse für diese Experimentalanlage unangemessen ist.

Rechnen wir für die gegebenen Experimentalanlage eine normale Varianzanalyse, so wird jeder Wert als der Wert einer neuen Vp. betrachtet, auch wenn er durch wiederholte Messungen der gleichen Vo. gewonnen wurde. Die Unterschiede zwischen den Ebenen des mit gleichen Vpn. gemessenen Faktors, die nur durch die entsprechende Experimentalmanipulation, nicht durch Unterschiede zwischen Vpn. bedingt sind, werden gegen die gesamte Irrtums-Varianz getestet und folglich unterschätzt (siehe 5.3.4.). Andererseits tritt auf jeder Ebene des mit neuen Vpn. gemessenen Faktors jede Vp. mit einer der Anzahl der Ebenen des anderen Faktors entsprechenden Anzahl von Werten auf; der Effekt des mit neuen Vpn. gemessenen Faktors wird wiederum gegen die gesamte Irrtumsvarianz getestet, die sich auf die Gesamtzahl der Werte und nicht auf die viel kleinere Anzahl der Vpn. bezieht. Folglich können Besonderheiten einzelner Vpn. leicht als Unterschiede zwischen Faktor-Ebenen analysiert, der Effekt also leicht überschätzt werden.¹⁾ Aus diesen Gründen müssen wir in der gegebenen Experimentalanlage eine sog. Varianzanalyse mit wiederholten Messungen rechnen. In ihr wird die gesamte Irrtums-Varianz aufgeteilt in die durch

15)-----

Wenn die wiederholten Messungen wie Messungen neuer Vpn. aufgefaßt werden, so kann die Zwischen-Gruppen-Varianz beim F-Test gegenüber der im Vergleich zu der verdoppelten Anzahl der Vpn. geringeren Irrtumsvarianz sehr schnell als ein Vielfaches erscheinen. Dagegen wird die Zwischen-Gruppen-Varianz beim F-Test auf eine größere Irrtums-Varianz bezogen, wenn die durch die Wiederholung der Messungen verursachte geringe Varianz ausgeklammert und die verbleibende Irrtumsvarianz auf die tatsächliche Anzahl der Vpn. bezogen wird. In beiden Fällen bleibt die Zwischen-Gruppen-Varianz dieselbe, aber die Irrtumsvarianz ändert sich, sowohl durch die Aufteilung der Quadratsumme wie der Freiheitsgrade (als Beispiel etwa Tab. 90). Für unseren Fall von nur zwei wiederholten Messungen ist die skizzierte Fehlermöglichkeit noch gering; da wir aber vor allem an dem Faktor ohne wiederholte Messungen interessiert sind, haben wir sie - besonders bei den ungleichen Zellenhäufigkeiten - dennoch überprüft. Drastisch veranschaulichen läßt sich die Fehlermöglichkeit für den Fall mehrerer wiederholter Messungen mit nur einer Vp. auf jedem Niveau des nicht wiederholt gemessenen Faktors. Wenn diese Vpn. zufällig sehr unterschiedliche Fähigkeiten haben, so ergeben sich, wenn wir jede der wiederholten Messungen wie die Messungen einer neuen Vpn. behandeln, hochsignifikante Unterschiede zwischen den Ebenen des nicht wiederholt gemessenen Faktors, die allein aus den Unterschieden zwischen den Vpn. herrühren.

die Unterschiede zwischen den Vpn. verursachte Varianz und durch die Wiederholung der Messung an den gleichen Vpn. verursachte Varianz. Der Faktor, auf dessen Ebenen verschiedene Gruppen gemessen werden, wird gegen die erste Varianz, der Faktor, auf dessen Ebenen die Messungen wiederholt werden, und die Interaktion wird gegen die zweite Varianz getestet (Winer 1962:304). Gegenüber einer normalen Zwei-Weg-Varianzanalyse, bei der beide Faktoren und die Interaktion gegen die gesamte Irrtumsvarianz getestet werden, werden daher in der Analyse mit wiederholten Messungen die Testbedingungen für den mit zwischen den Ebenen wechselnden Vpn. gemessenen Faktor verschärft, für den Faktor mit wiederholten Messungen der Vpn. erleichtert (Winer 1962:299).

Diese Argumentation gilt im Prinzip für alle mehrfaktoriellen Varianzanalysen unseres Experiments. Da die bisherigen Untersuchungen sich innerhalb des ursprünglichen Experimentaldesigns bewegten und daher auch mit nichtparametrischen Verfahren überprüft werden konnten, da sie sich ferner auf für jede Ebene eines Faktors recht große ($N=33$) und vor allem gleich große Gruppen bezogen, haben wir uns mit den üblichen Mehr-Faktor-Varianzanalysen begnügt. Für die folgenden Untersuchungen jedoch, bei denen zahlreiche, kleine und ungleiche Gruppen verglichen werden, und bei denen die Varianzanalyse das einzig verwendbare Prüfverfahren war, haben wir folgenden Weg für die statistische Prüfung eingeschlagen. Zunächst wurde auch hier die übliche Zwei-Faktor-Varianzanalyse mit dem Programm MANOVA in einem Rechenzentrum gerechnet, und zwar mit allen Variablen der zweiten Messung des Wortschatzexperiments und der Experimentalgruppe I ($N=66$); für Variable mit signifikanten Schichtunterschieden

sind diese Ergebnisse tabellarisch dargestellt. Zur strikteren Prüfung aber haben wir für diejenigen Variablen, die im ersten Schritt signifikante Schichtunterschiede aufwiesen, Varianzanalysen mit wiederholten Messungen für die Experimentalstichproben I und II mit einem Tischrechner gerechnet.

Das Programm MANOVA "gibt die exakten Lösungen (für univariate Varianzanalysen) sowohl im orthogonalen wie nicht-orthogonalen Falle" (Beiheft), d.h. für den Fall gleicher Anzahlen wie ungleicher oder nicht proportionaler Anzahlen in den Zellen werden die Quadratsummen der Faktoren in gleicher Weise errechnet. In unserem Falle sind die Zellenhäufigkeiten innerhalb der Kommunikationsformen proportional; trotz der ungleichen Zellenhäufigkeiten haben wir also noch ein orthogonales Design. In einem solchen Falle ist die Lösung nach den kleinsten Quadraten mit der normalen Lösung einer Varianzanalyse mit gleichen Häufigkeiten identisch, mit der selbstverständlichen Ausnahme, daß die Quadrate der Zellenmittelwerte durch die jeweilige Anzahl der Vpn. in den Zellen dividiert werden müssen, (siehe für den Fall einer Varianzanalyse mit wiederholten Messungen: Winer 1962, Vergleich zwischen S. 307, gleiche Zellenhäufigkeiten, und S. 376, ungleiche Zellenhäufigkeiten). Wir können also die Lösungen des Programms MANOVA übernehmen und für Rechenarbeiten mit dem Tischrechner die von Winer (1962: 374 ff.) angegebene, für den Fall proportionaler Designs vereinfachte Lösung nach den kleinsten Quadraten verwenden.

Die Mittelwerte verschiedener Maße der zweiten Messung sind in Tab. 88 für die Experimentalstichprobe I, in Tab. 89 für die Experimentalstichprobe II dargestellt. Die Varianzanalysen sind in Tab. 90 zusammengefaßt: nur bei den Maßen, die in Tab. 90 (a) genannt sind, ergaben sich signifikante Schichtunterschiede; bei keinem Maß fanden sich signifikante Effekte der Kommunikationsform (von den 23 getesteten Variablen ergab sich bei nur einer ein F-Wert über 1) oder der Interaktion. Trotz der nicht ganz angemessenen Form der Varianzanalysen wie in Tab. 90 (a), können wir weitere Effekte ausschließen. Weitere Varianzanalysen mit Berücksichtigung des Faktors "Versuchspersonen" sind in Tab. 90 (b) dargestellt.

Die Aussagen der Mittelwertvergleiche und Varianzanalysen lassen sich folgendermaßen zusammenfassen. (1) Auf den Maßen, die die gegebene Zahl der Antworten mitberücksichtigen (A-Kategorisierung) zeigen sich keine signifikanten Schichtunterschiede. Ein konsistentes, aber schwaches Ansteigen der Leistung mit der Sozialschicht, das unabhängig vom Signifikanzniveau ein interessantes Ergebnis ist, findet sich nur in der mündlichen Messung, bei der Mittelung beider Messungen aber auch in der Gesamtmessung. (2) Auf den Maßen, die nur die Güte der gegebenen Antworten berücksichtigt, (B- und C-Kategorisierung), ergeben sich für die Explikationsmaße keine, für die Verständnis- und Differenzmaße aber signifikante Effekte (Tab. 90). Diese Effekte verfehlen auch in einer Varianzanalyse mit Berücksichtigung der wiederholten Messungen nur knapp das 5%-Signifikanzniveau. Die Inspektion der Schichtunterschiede zeigt folgendes Muster: Während bei den Verständnismaßen die UUS gegenüber allen übrigen Schichten abfällt, ergeben sich die größeren Differenzen in den mittleren, die geringeren in den extremen Schichten und die größten regelmäßig in der MMS; dies gilt für die B- wie für die C-Kategorisierung in gleicher Weise und ist deshalb grafisch nur einmal - für die B-Kategorisierung (Abb. 11) - herausgestellt. Eine Erwartung, die wir von der ersten Messung hätten übernehmen können, daß nämlich die OMS gegenüber der MMS abfällt, hat sich hier nicht bestätigt; sondern eher deren Gegenteil. Offensichtlich ist die MMS der OMS im passiven Wortschatz, die OMS aber der MMS in den aktiven Erklärungen von Worten geringfügig überlegen. Innerhalb der Unterschicht jedoch ist die OUS der UUS in der aktiven Worterklärung wie in der passiven Wortkenntnis überlegen.

Vor der Interpretation der Ergebnisse müssen zwei theoretisch irrelevante Einflußmöglichkeiten erläutert werden. Erstens: Wie wir schon bei der Prüfung der Homogenität der Varianzen dargestellt haben, hat ein Mädchen aus der UUS (Vpn. 322) im mündlichen (nicht aber im schriftlichen) keine einzige Antwort gegeben; daher hat es den schlechtestmöglichen Durchschnittswert (9 bzw. 8) auch in der B- und C-Kategorisierung, in der nur die gegebenen Antworten berücksichtigt werden, erhalten und die Mittel-

werte der B- und C-Kategorisierung für die UUS sehr stark gedrückt (sowie die Varianzen vergrößert). Obwohl die Verweigerung jeder mündlichen Worterklärung bei einer Vpn. der UUS für sich ein Ergebnis darstellt, wurden die Mittelwerte der UUS in der Experimentalgruppe I ohne diese Vpn. errechnet und in Tab. 88 und Abb. 11 gesondert dargestellt. Mit diesem neu errechneten Wert verringert sich der Abstand der UUS von den übrigen Schichtgruppen auf den Explikations- und Verständnismaßen erheblich, auf den Differenzmaßen jedoch nur geringfügig. Das zeigt sich auch in den Ergebnissen der Varianzanalysen (Tab. 90). Während bei einer Nichtberücksichtigung der Vpn. 322 für Verständnis der Schichteffekt nicht einmal mehr einen F-Wert von 1 erreicht, ist er für das Differenzmaß zwar nicht mehr signifikant, aber nicht unerheblich.

Zweitens: Wieweit spiegeln die Unterschiede zwischen den Schichtgruppen in der zweiten Messung Unterschiede auf der Kontrollvariablen wider? Wie in Abschnitt 2.1.4. (Tab. 24 und 25) dargestellt wurde, sind innerhalb der dichotomen Unterschicht die Vpn. mit besseren Leistungen auf der Kontrollvariablen aus der OUS, mit schlechteren jedoch aus der UUS gewählt worden. Bei der hohen Korrelation zwischen der Kontrollvariablen und den Leistungen der zweiten Messung (siehe Tab. 108), können wir Unterschiede zwischen der UUS und den übrigen drei Schichtgruppen nicht mehr interpretieren, Unterschiede also, die lediglich das Profil der Schichtgruppen auf der Kontrollvariablen nachzeichnen. Genau dies aber ist für die Verständnismaße der Ball, während die Differenzmaße unabhängig von der Kontrollvariablen variieren. Kurz: Der Schichteffekt auf dem Verständnismaß kann auf die besonderen Bedingungen des Experiments zurückgeführt werden, der Schichteffekt auf den Differenzmaßen aber scheint den Versuch einer theoretischen Erklärung wert zu sein.

Die Differenzmaße fassen die Variation in der Explikation und im Verständnis zusammen; da diese beiden Dimensionen bei der Auswertung ja nicht unabhängig voneinander festgelegt wurden, sind von der Logik der Maße DIBVE und DICVE bei den schlechtesten Explikationsbewertungen die größten Differenzen zu erwarten.

Unterschiede zwischen den Differenzen können also nur interpretiert werden, wenn sie nicht linear mit der entsprechenden Explikationsbewertung zusammenhängen. Dies ist bei den oben dargestellten Ergebnissen nicht der Fall; die Variationen der Schichtmittelwerte für Explikation bewegen sich zudem in einem so eng begrenzten Bereich, daß die Variation der Differenzen nicht durch die Logik der Maßkonstruktion beeinflußt worden sein kann, sondern eine Erklärung verlangt.

Auch aus einem zweiten Grunde scheint uns die Interpretation der Differenzmaße sinnvoll: das beschriebene Schichtprofil (Tab. 88 und 89; Abb. 11) hält sich bei allen Aufgliederungen der Daten: Sowohl in allen drei Kategorisierungsformen, wie in mündlicher und schriftlicher Messung, wie bei der Aufteilung der 4 Schichtgruppen nach Geschlechtern¹⁾, wie auch der Zusammenfassung der Experimentalgruppen nach Wortgruppen (Tab. 91)²⁾; weiterhin treten bei den Differenzen als den einzigen Maßen keine inhomogene Varianzen in den 4 Schichtgruppen auf (siehe Tab. 88). Wie lassen sich also die Schichtunterschiede der Differenzmaße erklären?

1) -----
Die Mittelwerte der Maße für beide Geschlechter in den 4 Schichtgruppen wurden nicht mehr gesondert dargestellt. Die Unterschiede zwischen den Geschlechtern auf den Verständnis- und Explikationsmaßen wurden nämlich schon beim dichotomen Schichtvergleich in Abschnitt 5.4. dargestellt. Über die Ergebnisse des dichotomen Schichtvergleichs hinaus zeigt die differenzierte Schichteinteilung innerhalb der Geschlechter, daß (1) die Überlegenheit der Jungen auf den Verständnis- und Explikationsmaßen für alle vier Straten gleichermaßen zutrifft und die Schichtprofile innerhalb der Geschlechtsgruppen weitgehend gleich sind - dies ist ein weiterer Beleg für das Fehlen von Interaktionen zwischen Schicht und Geschlecht und die Echtheit der Geschlechtsdifferenzen - und daß (2) weder die absolute Höhe noch die Schichtprofile der Differenzmaße sich zwischen den Geschlechtsgruppen unterscheiden - dies führt zurück zu den möglichen Erklärungen für die Schichtprofile der Differenzmaße.

2) Da sich die Differenzmaße zwischen den Wortgruppen sehr stark unterscheiden (s. 5.4.4.) sind in Tab. 91 alle Experimentalgruppen, die die gleichen Worte (Testform A bzw. B) erklärt haben, zusammengefaßt, unabhängig von der Kommunikationsform; d.h. Experimentalgruppen 1 und 2 sowie 3 und 4. Die Mittelwerte für jede Vpn. wie für jede Gruppe beziehen sich daher auf 60 (nicht mehr 30) Worte.

Ursprünglich hatten wir angenommen, daß die Kinder der Unterschicht die Wortbedeutungen schlechter erklären, sie aber gleich gut verstehen würden als die Kinder der Mittelschicht und daß aus diesem Grund die Differenzen zwischen Verständnis und Explikation in der Unterschicht größer sein würden. Entgegen dieser Hypothese waren jedoch beim Vergleich zweier Schichtgruppen die Differenzen in der Mittelschicht größer als in der Unterschicht (siehe 5.4.4.). Beim Vergleich von vier Schichtgruppen finden sich die geringeren Differenzen in den beiden extremen Schichtgruppen - der UUS und der OMS -, die größeren Differenzen in den beiden mittleren Schichtgruppen der OUS und der MMS; auch dies Ergebnis hatten wir nicht vorausgesagt. Wir wollen versuchen, die ursprünglichen theoretischen Überlegungen diesen Ergebnissen entsprechend zu modifizieren.

Versuchen wir zuerst die hohen Differenzen in der MMS, der zahlenmäßig stärksten der vier Schichtgruppen, zu erklären, um dann die Brauchbarkeit dieser Erklärung für die OUS zu prüfen, für die sich ja fast gleich hohe Differenzen ergaben. Man kann annehmen, daß in der MMS nicht nur die vorherrschende Form der Sozialbeziehungen Explizitheit der Worterklärungen fordert, sondern auch eine quasi-didaktische Einstellung der Eltern zur Erklärung von sprachlicher Bedeutungen vorherrscht.¹⁾ Die Kinder der MMS können dann bei der Erklärung ihrer vorsprachlichen Ausdrucksintentionen die Hilfe des Anderen eher erwarten als in der Unterschicht, das Risiko kommunikativer Mängel um eines breiteren Lernertrages willen in ihre Äußerungen einkalkulieren. Wenn wir unsere punktuellen Messungen genetisch interpretieren, so ist die Differenz zwischen Verständnis und Explikation das Vehikel eines sozial vermittelten Lernprozesses. Große Differenzen werden nun nicht als das Ergebnis einer

75-----

Die Theorie der linguistischen Codes wird mit dieser Erweiterung von einer Theorie des Wissenschaftlers zu einer Theorie des Handelnden; Einstellungen, die sonst von der Theorie rekonstruiert wurden z.B. die zur Explikation von Bedeutungen, werden als dem Handelnden bewußt und verfügbar angesehen.

für die Unterschicht typischen Strategie eingeschränkter Verbalisierung, sondern als das Ergebnis in der Mittelschicht eher sozial gestützter risikovoller Erklärungsweisen angesehen. Bei einem relativ besseren Verständnis könnten wir eine relativ schlechtere Explikation und damit eine größere Differenz erwarten, entgegen unseren ursprünglichen Annahmen gerade in den die sprachliche Sozialisation forcierenden sozialen Milieus.

In dieser Hinsicht können wir die OUS, der im wesentlichen Facharbeiter und einfache Angestellte und Beamte angehören, und die MMS, zu der mittlere und gehobene Angestellte und Beamte gezählt wurden, zusammenrücken. Sei es durch die Aufnahme populärpädagogischer Theorien, sei es durch realistische Chancen des Aufstiegs durch die Anpassung an die Mittelschichtnormen, oder sei es gar durch einen Trend zur universalen Mittelschichtkultur - die OUS hebt sich von der UUS ab und übernimmt mittelschichtspezifisches sprachliches Erziehungsverhalten (Indizien für diese Annahme z.B. bei Grimm 1966). Auf der anderen Seite wird dies gerade in der OMS am ehesten in Frage gestellt und wegen fehlenden sozialen Aufstiegs- oder Konformitätsdrucks schlicht überflüssig werden können. Die Sonderstellung der beiden extremen Schichtgruppen erklärt sich aus entgegengesetzten Gründen: Während die Eltern der UUS, die in unserer Stichprobe im wesentlichen un- und angelernte Arbeiter mit Volksschulbildung waren, das sprachlich wenig explizierte Erziehungsverfahren des restingierten Code beibehalten haben, können die Eltern der OMS, die Selbständige, Freiberufliche oder Höhere Beamte mit großteils (auch bei der Mutter) abgeschlossenem Hochschulstudium waren, von den Gefahren eines elaborierten Sprachgebrauchs, wie Redundanz und Ungenauigkeit sich schon wieder distanzieren, ohne auf ein differenziertes Erziehungsverhalten zu verzichten.

Welche Rechtfertigung gibt es für die Vermutung eines ähnlichen Erziehungsverhaltens in den beiden mittleren Schichten, obwohl sie durch das Kriterium "manuelle oder nicht-manuelle Arbeit" und durch die in unserem Experimentalplan nicht berücksichtigte UMS voneinander getrennt sind? Vielleicht ist der Status als Angestellter oder

Beamter und sogar Facharbeiter wegen der diesen Gruppen gemeinsamen im Vergleich zum un- oder angelernten Arbeiter relativ großen sozialen Sicherheit doch der entscheidende Faktor. Es ist auch wahrscheinlich, daß dieser Status relativer allgemeiner sozialer Sicherheit zu einer relativ stärkeren Hinwendung zu familialen Belangen und damit zu der quasi-didaktischen Einstellung zur Erziehung führt, auf der unser Erklärungsversuch ruht.¹⁾ Wenn diese Erklärung für die unterschiedlichen Differenzmaße beim Vergleich von vier Schichten akzeptabel ist, so kann man mit ihr ebenso die größeren Differenzmaße der Mittelschicht beim dichotomen Schichtvergleich erklären.

Zusammenfassung: Beim Vergleich der aktiven Worterklärung von Kindern aus vier sozialen Schichten ergeben sich keine Unterschiede, solange Erklärungshäufigkeit und -güte gemessen werden. Berücksichtigt man nur die Erklärungsgüte, so zeigen sich geringe oder nicht-signifikante Unterschiede der Maße für Verständnis und der Maße für Explikation, aber durchweg große Unterschiede der Differenzen zwischen Verständnis und Explikation. Die Differenzen sind in der OUS und MMS - der größten Gruppe - größer als in der UUS und OMS. Wir glauben, dies Ergebnis auf die größere bewußte Bemühung der Eltern dieser Schichten um sprachliche Sozialisation ihrer Kinder zurückführen zu können, die - entweder als augenblicklichen Zurückbleiben, aber potentiell mit um so größeren Erfolgen verbundenes Zwischenstadium eines sozialen Lernprozesses oder als dauerhaft negatives Ergebnis - zu gemessenen am Verständnis geringen Explikationsleistungen der Kinder führen.

¹⁾-----
Diese Interpretation könnte zudem die (nicht-signifikante) Rangfolge der Schichten auf den Explikations- und Verständnismaßen aller drei Kategorisierungsarten erklären: Auch dort werden die beiden, ungefähr gleich guten mittleren Schichten von der OMS übertroffen, sind ihrerseits aber besser als die UUS.

Diese Interpretation beruht auf einer Übertragung der Theorie der linguistischen Kodes auf quasi-didaktischen Sprachgebrauch. Sie stimmt mit der ursprünglichen Voraussage der Theorie des linguistischen Kodes nicht überein, daß die Differenz zwischen Verständnis und Explikation in der Mittelschicht kleiner ist. Sie würde statt dessen gerade in den sozialen Milieus große Differenzen zwischen Verständnis und Explikation voraussagen, in denen eine bewußte Einstellung zur allgemeinen und zur sprachlichen Sozialisation vorherrscht.

5.6.2. Unterschiedliche Einflüsse der einzelnen Indikatoren sozialer Schicht.

Während wir im vorangegangenen Abschnitt die Einflüsse einer mehrstufigen Einteilung der sozialen Schichten der Eltern auf die aktiven Erklärungen der Kinder untersucht haben, sollen in diesem die Bedeutung der Einzelnen in den Filterindex eingegangenen Dimensionen für die aktive Worterklärung analysiert werden. Während zuvor "soziale Schicht" in der Vertikalen weiter differenziert, sollen jetzt die horizontalen Aspekte des Filterindex einzeln betrachtet werden.

Vier Dimensionen wurden im Filterindex integriert, wobei die Reihenfolge zugleich das Gewicht der Dimension für den Filterindex bestimmte: Der Beruf des Vaters (9 Merkmale) (siehe Tab. 26), die Ausbildung des Vaters (9 Merkmale) (siehe Tab. 27), die Ausbildung der Mutter (9 Merkmale) (siehe Tab. 27) und das Familiennettoeinkommen (4 Merkmale) (siehe Tab. 28). Sie wurden für alle Fälle aus der Gesamterhebung, für die die entsprechende Information vorlag, z-transformiert, so daß die für die Korrelations- bzw. Regressionsanalyse notwendige Bedingung der Intervallskalenmessung wenigstens annähernd erfüllt ist. Eine Charakterisierung unserer Untersuchungsgruppe auf diesen Merkmalen wurde in Abschnitt 2.1.4. gegeben.

Nach den geringen Korrelationen des gesamten Filterindex mit den Maßen der zweiten Messung (Tab. 87) sind sehr hohe Korrelationen einzelner Dimensionen nicht zu erwarten, da diese mit dem gesamten Filterindex ziemlich hoch korrelieren (Tab. 21). Da jedoch im Filterindex ein Teil der spezifischen Varianz jeder der Dimensionen ignoriert wird und jede Dimension mehr Merkmale enthält als der Filterindex, sind Korrelationen dieser Dimensionen mit den Maßen der zweiten Messung nicht von vornherein ausgeschlossen. Dies gilt um so mehr, je schwächer die Dimensionen im Filterindex gewichtet wurden: Um so größer nämlich wird der Teil der Varianz einer Dimension, der nicht in den Filterindex aufgenommen wurde.

Keine auch nur annähernd signifikanten Korrelationen mit den Maßen der zweiten Messung zeigen der gegenwärtige Beruf des Vaters, die Ausbildung des Vaters und das Einkommen. Nur die Ausbildung der Mutter korreliert in einer Höhe über .20 signifikant mit den Maßen der zweiten Messung (Tab. 92); dies ist umso aufschlußreicher, als AUSBMU mit der Kontrollvariablen (Wortschatz erster Messung) überhaupt nicht korreliert (siehe Tab. 21). Während AUSBMU auf die Güte der Antworten ebensowenig wie die anderen Dimensionen einen Einfluß hat, zeigen sich für die Häufigkeit der Antworten und die Maße der A-Kategorisierung in beiden Kommunikationsformen signifikante Korrelationen. Allein durch die Gewichtung der Dimensionen im Filterindex hätte das negative Ergebnis für den Filterindex eher für EINKOM als für AUSBMU noch hohe Korrelationen zugelassen. Das Fehlen jeglicher Korrelationen bei EINKOM spricht für die inhaltliche Bedeutung der Korrelationen bei AUSBMU.

Betrachtet man die Korrelationen von AUSBMU in den beiden Schichtgruppen, so finden wir in der Mittelschicht bei den gleichen Maßen noch höhere signifikante Korrelationen als in der Gesamtgruppe, in der Unterschicht jedoch nur bei den Maßen der Häufigkeit Korrelationen, die annähernd so hoch wie die entsprechenden Korrelationen der Gesamtgruppe, aber nicht mehr signifikant sind. Dieser Unterschied der Korrelationen zwischen den beiden Schichtgruppen ist zu einem Teil nur scheinbar, da AUSBMU mit der Kontrollvariablen in der Mittelschicht positiv, in der Unterschicht negativ korreliert (siehe Tab. 21). Partial-Korrelationen zwischen AUSBMU und den wichtigsten Maßen der drei Dimensionen der Maßkonstruktion, ohne Einfluß der Kontrollvariablen, unterscheiden sich nur noch geringfügig zwischen den Schichten (Tab. 93). Der noch verbleibende Unterschied kann auch aus der fast dreifach größeren Varianz der Variable AUSBMU in der Mittelschicht herrühren (Tab. 92); die Varianzen der Maße der zweiten Messung sind in beiden Schichtgruppen weitgehend gleich (siehe die Standardabweichungen in Tab. 8). Die Unterschiede der Varianz sowie der Mittelwerte von AUSBMU treffen ihrerseits natürlich eine soziale Realität: Da die Ausbildung der Mutter nicht unmittelbar den Status der Familie festlegt, haben die Mütter in Mittelschichtfamilien die verschiedensten Bildungsabschlüsse, aber in so gut

wie keiner Unterschichtsfamilie hat die Mutter mehr als Volksschulabschluß mit Lehre (Tab. 27).

In den drei übrigen Dimensionen des Filterindex tauchen auch bei einer getrennten Betrachtung der beiden Schichtgruppen kaum nennenswerte Korrelationen mit den Maßen der zweiten Messung auf (Tab. 94). Auffällig ist der gegenläufige Einfluß des Einkommens auf die Anwohnhäufigkeiten in den beiden Schichten.

Den gemeinsamen Einfluß aller vier Schichtindikatoren und die spezifische Voraussagekraft jedes einzelnen unabhängig von den jeweils anderen kann eine multiple Regressionsanalyse darstellen: Sie gibt an, wieviel Varianz der abhängigen Variablen (Kriterium) durch eine Reihe von unabhängigen Variablen (Prädiktoren) erklärt wird (R^2), und um wieviel Einheiten sich die abhängige Variable durch die Veränderung einer einzelnen unabhängigen Variablen verändert, wenn alle übrigen Variablen konstant bleiben (Roh-Regressions-Koeffizient), bzw. welche relative Bedeutung ein Prädiktor unabhängig von den übrigen hat (Beta-Gewicht).¹⁾ Die Signifikanz des R^2 kann, genau wie in einer Varianzanalyse, mit einem F-Test überprüft werden. Die Ergebnisse einer multiplen Regressionsanalyse mit den vier Schichtindikatoren als Prädiktoren und verschiedenen Maßen der drei theoretisch konzipierten Maßgruppen sind in Tab. 95 dargestellt. Da in einer multiplen Regressionsanalyse nur die Vpn. berücksichtigt werden, für die auf allen Variablen Daten vorliegen, und die hier dargestellte Analyse Teile einer Analyse mit 21 Prä-

1) -----
Der Roh-Regressions-Koeffizient, multipliziert mit der Standardabweichung des Prädiktors und dividiert durch die Standard-Abweichung des Kriteriums, ergibt das Beta-Gewicht. Da im Falle der Schichtindikatoren die Standardabweichungen der Prädiktoren sich nur geringfügig unterscheiden, können die Roh-Regressions-Koeffizienten die gleiche Information geben wie die Beta-Gewichte. Da in den noch folgenden multiplen Regressionsanalysen Prädiktoren mit stark variierenden Standardabweichungen und in allen multiplen Regressionsanalysen Kriterien mit stark variierenden Standardabweichungen auftauchen, haben wir in allen Tabellen die Beta-Gewichte angegeben.

diktoren ist, mußten insgesamt 6 Vpn. ausgeschieden werden. Weder die Balance der beiden Schichtgruppen noch der Kontrollvariablen-Mittelwerte konnten aber dadurch gestört werden.

Bevor wir anhand der multiplen Regressionsanalyse die spezifische Bedeutung der einzelnen Schichtindikatoren darstellen können, müssen wir ihre Interkorrelationen untersuchen, d.h. das bei der multiplen Regressionsanalyse entscheidende Hindernis der Multikollinearität der Prädiktoren diskutieren. Durch die Multikollinearität tendieren die Werte der Prädiktoren dahin, in einem begrenzten Bereich des n -dimensionalen Raum im Extremfall auf einer Ebene im $(n-1)$ -dimensionalen Raum zu liegen. Die Regressionsebene wird dadurch unstabil, alle möglichen Ebenen im n -dimensionalen Raum erfüllen das Kriterium der kleinsten Quadrate. Bei zwei voll miteinander korrelierenden Prädiktoren etwa ist eine Aussage über den relativen Einfluß beider Prädiktoren überhaupt nicht möglich, weil die Werte nur auf einer Ebene mit der Prädiktoren-Kombination und dem Kriterium als Ordinaten liegen, in der sich zwar eine Regressionslinie von Prädiktoren-Kombinationen auf das Kriterium bestimmen läßt, auf die sich aber jede beliebige Regressionsebene nach dem Kriterium der kleinsten Quadrate legen läßt (Wonnacott/Wonnacott 1971:294-296). Die Berechnung des spezifischen Einflusses einzelner Prädiktoren wird dadurch unmöglich.¹⁾

Die Interkorrelationen der 4 Statusindikatoren liegen nun zwischen .50 und .69 (Tab. 22); die Anwendung der multiplen Regressionsanalyse zur Bestimmung der spezifischen Voraussagekraft der Prädiktoren scheint dadurch in Frage gestellt. Wir glauben jedoch, daß das skizzierte statistische Argument nur dann sinnvoll ist, wenn von der Logik der Sache her alle Kombinationen von Prädiktorenwer-

1)-----
Dieser Einwand gilt, sofern die multiple Regressionsanalyse die spezifische Voraussagekraft einzelner Prädiktoren darstellen, nicht aber sofern sie die Voraussagbarkeit des Kriteriums überhaupt prüfen soll.

ten möglich sind; anders gesagt: Wenn in einer anderen oder größeren Stichprobe mit den gleichen Prädiktoren das Fehlen von Multikollinearität erwartet werden kann. Genau diese Annahme trifft jedoch für Indikatoren sozialer Lage nicht zu: Es gibt mittlere Angestellte, aber keine ungelernten Arbeiter mit abgeschlossener Hochschulbildung; selbst in der fünfmal größeren Gesamtstichprobe, aus der unser Experiment gezogen wurde, liegen die Interkorrelationen der Schichtindikatoren auf dem genau gleichen Niveau (zwischen .49 und .70; siehe Tab. 22). Unter diesen Umständen müssen wir entweder eine gewisse Kollinearität in Kauf nehmen oder auf die Einschätzung der spezifischen Voraussagekraft von Variablen ganz verzichten. Wir haben uns für das erste aus zwei Gründen entschieden. Erstens sind die Interkorrelationen der Schicht-Variablen ja nur mittelhoch. Zweitens können wir die Gefahr, daß durch Multikollinearität der Prädiktoren eine instabile Neigung der Regresseionsebene und damit eine zufällige Verteilung der Beta-Gewichte verabsolutiert wird, durch den Vergleich der wiederholten Messungen mit denselben Vpn. kontrollieren: In der mündlichen und schriftlichen Messung ergab sich, mit geringen Ausnahmen für GEBERV, die gleiche Struktur von spezifischer Voraussagekraft der einzelnen Schichtindikatoren (Tab. 95). Wir gehen daher über zur Darstellung der Ergebnisse.

Für die Maße der Erklärungsgüte und -häufigkeit und für die Maße der Erklärungshäufigkeit ergeben sich durchweg signifikante oder fast signifikante R^2 ; für die Maße der Erklärungsgüte jedoch nur ein unter dem 5%-Niveau signifikantes R^2 .¹⁾

7) -----
Der quadrierte multiple Regressions-Koeffizient gibt die maximale Voraussage des Kriteriums durch die Prädiktoren in den gegebenen Daten, keine Schätzung des Zusammenhangs der Prädiktoren mit dem Kriterium unabhängig von den gegebenen Daten. Wir haben ihn in Tab. 95 dargestellt, um durch Vergleiche mit dem einfachen Determinationskoeffizienten r^2 die Wichtigkeit der einzelnen Variablen abzuwägen.

In beiden Kommunikationsformen finden sich also signifikante multiple Regressionen der vier Schichtindikatoren genau bei den Maßen, die schon signifikante einfache Korrelationen mit der Ausbildung der Mutter aufwiesen: Die schriftlichen Maße für die Erklärungsgüte und -häufigkeit haben niedrigere F-Werte, die für die Erklärungsgüte höhere F-Werte als die entsprechenden Maße der mündlichen Messung, m.a.W. zwischen den beiden Gruppen von Maßen differenziert die Voraussage der Schichtindikatoren in der schriftlichen Messung weniger als in der mündlichen Messung.¹⁾ Trotzdem bleibt erklärungsbedürftig, daß sich hier nach den negativen Ergebnissen der Varianz- und der Regressionsanalysen mit dem Filterindex, der ja eine Kombination der hier verwandten Prädiktoren ist, und nach den - mit Ausnahme der Ausbildung der Mutter - negativen Ergebnissen der Korrelationen der Einzeldimensionen des Filterindex signifikante Effekte von Variablen der sozialen Schicht ergeben.

Für diese Erklärung müssen wir uns zunächst noch einmal die Konstruktionsweise des Filterindex vergegenwärtigen. Durch die Bildung des Filterindex wurde ein Teil der Varianz der Maße, nämlich alle Variationen der "späteren" Indikatoren, die "unter" dem vergleichbaren Niveau der "vorangegangenen" Indikatoren lag, ignoriert; so kann z.B. bei einem relativ hohen Beruf des Vaters die Ausbildung des Vaters und die der Mutter relativ stark schwanken. Während der Filterindex diese Varianz nicht mehr berücksichtigt, ist sie in den Korrelationen der vier Schichtindikatoren, die in die multiple Regressionsanalyse eingehen, enthalten und kann in jeder der vier Dimensionen zur Varianz der Maße der zweiten Messung beigetragen haben. Zudem sind, wie die Vorzeichen der Beta-Gewichte der Schichtindikatoren zeigen, im Filterindex offensichtlich widerstreitende Einflüsse nivelliert worden. Für die weitergehende Prüfung

¹⁾-----
Wir finden hier wiederum die Tatsache, daß in der mündlichen Messung sich eher Schichtdifferenzen finden als in der schriftlichen (siehe 5.3.2., 5.3.4. und 5.5.).

der Frage, wie durch die Verwendung der einzelnen Schichtindikatoren anstelle des Filterindex ein signifikanter Schichteffekt zustande gekommen ist, müssen wir daher die Beta-Gewichte und die einfachen Korrelationen der Statusindikatoren mit den Maßen der zweiten Messung heranziehen.

Bei allen Maßen der Erklärungsgüte und -häufigkeit, d.h. bei den Maßen mit vorwiegend signifikanten R^2 , hat die Ausbildung der Mutter die weitaus größte spezifische Voraussagekraft; während sie durchweg positiv die aktive Worterklärung beeinflusst, übt die Ausbildung des Vaters und das Einkommen einen negativen und der Beruf des Vaters einen nur geringen, über die Kommunikationsformen oft nicht konsistenten Einfluß aus. Bei den Maßen der Erklärungsgüte, d.h. den Maßen mit vorwiegend nicht-signifikantem R^2 , müssen die Unterschiede der Voraussagekraft zwischen den Prädiktoren geringer sein; aber auch hier zeigt die Ausbildung der Mutter durchgehend positiven und in der mündlichen Messung auch den größten Einfluß, während der negative Einfluß der Ausbildung des Vaters in der schriftlichen Messung etwas stärker ist. Durch die Bank also hängen die Leistungen der zweiten Messung bei konstanter Ausbildung des Vaters und konstantem Einkommen mit der Ausbildung der Mutter positiv zusammen, während die Ausbildung des Vaters und in geringerem Maße das Einkommen die Leistungen der zweiten Messung auf jedem Ausbildungsniveau der Mutter negativ beeinflussen.

Für eine weitere Analyse dieses Sachverhalts müssen wir an die einfachen Korrelationen der Schichtindikatoren erinnern. Mit den Leistungen der zweiten Messung korrelieren die Ausbildung des Vaters und das Einkommen überhaupt nicht, die Ausbildung der Mutter korreliert mit den Maßen der Erklärungsgüte und -häufigkeit sowie den Maßen der Erklärungshäufigkeit (Tab. 92). Die durch den einfachen Einfluß der Ausbildung der Mutter erklärte Varianz der Maße der Erklärungshäufigkeit (r^2) beträgt etwa 15%, die durch die multiple Regression erklärte (R^2) zwischen 15% und 25%. Die Ausbildung des Vaters und das Einkommen vergrößern also, obwohl sie nicht mit den Maßen der zweiten Messung korrelieren, die insgesamt erklärte Varianz; dabei ist ihre spezifische Voraussagekraft negativ. Für

den positiven Zusammenhang zwischen Ausbildung der Mutter und der aktiven Worterklärung des Kindes haben sie den Status von "Suppressor"-Variablen; der Teil der Varianz der Ausbildung der Mutter, der mit den beiden anderen Indikatoren zusammenhängt, korreliert seinerseits nicht mit der zweiten Messung. Auf dem indirekten Wege der Ausschaltung nicht erklärungskräftiger Varianz der Ausbildung der Mutter beeinflussen auch Ausbildung des Vaters und Familieneinkommen die Leistungen der zweiten Messung. Gibt es für diesen statistisch aufgedeckten Sachverhalt eine soziologische Interpretation?

Wir können annehmen, daß die durch die multiple Regression zusätzlich "erklärte" Varianz durch die Diskrepanz zwischen der Ausbildung des Vaters und des Familiennettoeinkommens und der Ausbildung der Mutter bedingt ist. Zunächst einmal - das zeigen die einfachen Korrelationen - hängt die Ausbildung der Mutter und kein anderer Statusindikator positiv mit der aktiven Worterklärung des Kindes zusammen. Weiterhin aber - das zeigt die multiple Regressionsanalyse - wird bei konstanter Ausbildung der Mutter die aktive Worterklärung des Kindes um so schlechter, je höher die Ausbildung des Vaters und das Familiennettoeinkommen sind. Die Ausbildung der Mutter wird um so wichtiger, je höher sie im Vergleich zu der des Vaters und dem Familiennettoeinkommen ist. Die multiple Regression belegt ein soziologisch sinnvolles Substitutionsverhältnis zwischen der Ausbildung der Mutter und den übrigen Statusindikatoren: Solange die Ausbildung der Mutter nicht von den übrigen Indikatoren absticht, beeinflußt sie das Lernen des Kindes nur geringfügig. Sobald die Mutter aber dem Vater in der Ausbildung überlegen ist und die Familie materiell schlechter gestellt ist, als die Ausbildung der Mutter erwarten ließe, versucht die Mutter, die durch die Ausbildung des Vaters und die materielle Lage der Familie verminderten Lernchancen des Kindes auszugleichen, indem sie ihre Ausbildung in die Waagschale wirft. Diese Interpretation bezieht sich auf einen objektiv wirksamen Zusammenhang, dem nicht notwendig Erkenntnisse und Schlußfolgerungen der Handelnden entsprechen müssen.

Wir wollen die durch eine genauere Betrachtung des Filterindex der sozialen Schicht in diesem und dem vorangegangenen Abschnitt gewonnenen Ergebnisse kurz zusammenfassen. Bei der Aufteilung des Filterindex in 4 Schichtgruppen zeigten sich zwischen den Schichten im allgemeinen keine Unterschiede, mit der einen Ausnahme, daß die Differenzmaße der B- und C-Kategorisierung in den beiden mittleren Schichtgruppen am größten sind. Bei der Aufteilung des Filterindex in seine einzelnen Dimensionen ergaben sich für drei Dimensionen keine Zusammenhänge mit der aktiven Worterklärung. Allein die Ausbildung der Mutter zeigt einen signifikanten Einfluß auf die Maße der Häufigkeit und der A-Kategorisierung; dies gilt auch bei einer getrennten Betrachtung der beiden Schichten. Der Einfluß der Ausbildung der Mutter verstärkt sich noch - wie die multiple Regressionsanalyse zeigt - wenn sie von der Ausbildung des Vaters und dem Familiennettoeinkommen positiv absticht. Die Gegenläufigkeit der Einflüsse der Ausbildung des Vaters und der Mutter für die Erklärungsleistungen des Kindes, die im Filterindex neutralisiert wird, scheint uns das wichtigste Ergebnis beider Abschnitte.

Die Differenzierung des Filterindex in mehrere Straten hat in der Hauptsache zwar zu einer Bestätigung des negativen Hauptergebnisses geführt, aber auch zu einer Modifikation: Die schichtspezifische Verteilung der Differenzmaße - wenn auch empirisch nicht von großem Gewicht - führte uns zu einer Übertragung der Theorie der linguistischen Codes auf quasi-didaktischen Sprachgebrauch. Die getrennte Betrachtung der einzelnen Dimensionen des Filterindex erzwang eine weitere Modifikation: Die im Filterindex erst an dritter Stelle berücksichtigte Ausbildung der Mutter hängt signifikant mit den aktiven Worterklärungen des Kindes zusammen; sie steht zudem zur Ausbildung des Vaters und dem Familieneinkommen, als einem Indikator der materiellen Lernumgebung, gleichsam in einem Substitutionsverhältnis, das durch die Zusammenfassung der Dimensionen im Filterindex verdeckt war. Es bietet sich die Vermutung an, daß gerade die Mutter didaktisch auf die Explikation sprachlicher Bedeutungen dringt und daß dies um so stärker geschieht, je höher die Ausbildung der Mutter ist und je weniger der Vater von seiner

Ausbildung her in der Lage ist, in der Familie Lehrfunktion zu übernehmen. Die Ausbildung der Mutter ist vielleicht weniger für ihre Berufstätigkeit bedeutsam, als für die Aufnahme von Rollenelementen eines Lehrers in die Familienrolle der Mutter. Wir wollen diese Vermutungen in Abschnitt 5.6.5. weiter verfolgen.

5.6.3. Einflüsse der Arbeitsplatzsituation des Vaters auf die Wortschatzleistung der ersten und zweiten Messung.

Die Qualität des Arbeitsplatzes: Gegenstand, Überwachung und Entscheidungsspielraum der Arbeit, beeinflussen neben vielen anderen Orientierungen: Selbsteinschätzung, politische Einstellungen etc. auch Auffassungen der Eltern von der Erziehung (Kohn 1969).

In diesem Abschnitt soll untersucht werden, wieweit sich dieser Einfluß des Arbeitsplatzes des Vaters einen Schritt weiter bis zu den verbalen Leistungen des Kindes durchschlägt, die - wie sich in der Gesamtstichprobe gezeigt hat (siehe Bosse u.a. 1971:264 ff.) - eher als nicht-verbale Fähigkeiten von den mit der Qualität des Arbeitsplatzes zusammenhängenden Erziehungsweisen der Eltern geprägt sein können. Zugleich sollen die Arten des Arbeitsplatzes einen genaueren Indikator für "subkulturelle Milieus" als die Schichtindikatoren ergeben und die mit Hilfe der Schichtindikatoren gewonnenen Ergebnisse differenzieren. Wir werden zuerst die Ergebnisse des standardisierten Wortschatztests (erste Messung) berichten, die vielleicht Anhaltspunkte für die Interpretation der Ergebnisse der zweiten Messung liefern, und dann die Ergebnisse der zweiten Messung darstellen.

In der ersten Messung hatte sich ein kontinuierliches Ansteigen der Wortkenntnis mit 5 sozialen Schichten ergeben - sieht man vom leichten Rückgang in der OMS ab (Abschnitt 4.1., Tab. 41). Da die Arbeitsplatzkategorien sehr eng mit der sozialen Schicht zusammenhängen, findet sich auch bei ihnen ein kontinuierliches Ansteigen (Tab. 29, Abb. 12), jedoch mit einigen Unterbrechungen. Die Varianzanalyse der Wortschatzleistungen in der ersten Messung mit den Arbeitsplatzkategorien des Vaters als Faktor ergibt einen hochsignifikanten F-Wert ($F(13,292) = 3,02, p < .001$).

Innerhalb der vorwiegend der Unterschicht zugeordneten Arbeitsplatzkategorien, zeigen die Gruppen 3, 5 und, ganz besonders, 6 bessere passive Wortkenntnis: Sehen wir von der Kategorie 4 ("Qualifiziertes Handwerk") ab, so hängen - unabhängig vom Wirtschaftssektor - unspezialisierte Tätigkeiten des Vaters (Kategorie 1 und 2)

mit schwächeren, spezialisierte (Kategorie 3 und 5) mit besseren Wortschatzleistungen des Kindes zusammen; die Spezialisierung und nicht der Wirtschaftssektor ist ausschlaggebend. Offenbar trägt sie zu einer stärkeren Einbeziehung des Vaters in die familiäre Erziehung bei, wodurch sich die Kommunikation des Kindes mit beiden Eltern erweitert; wenn dies für die spezialisierten Tätigkeiten in der Industrie (Facharbeiter) und der Dienstleistung gilt (Kategorie 3 und 5), dann um so stärker für die Kontrolltätigkeiten in der Industrie (Kategorie 6 - vorwiegend Vorarbeiter und Meister), die schon von der Substanz der Arbeit her eine stärkere verbale Orientierung verlangen, die sich leicht von der Führung der Arbeitsuntergebenen auf die Führung von Kindern übertragen läßt. Die Leistung der Kinder von Vätern mit industriellen Kontrolltätigkeiten ragt so stark aus den übrigen der Unterschicht zuzuordnenden Gruppen heraus, daß wir sie nicht nur den schwächeren Gruppen 1, 2 und 4, sondern auch den relativ besseren 3 und 5 gegenüberstellen können: die Mittelwertdifferenz zwischen den zusammengefaßten Gruppen 1 bis 5 und der Gruppe 6 beträgt 1.37 (WST-z-Werte); sie ist jedoch, nach einem Scheffé-Test, nicht signifikant. Die kritische Stellung, die diese Gruppe in der betrieblichen Hierarchie (Dahrendorf 1963) wie am Schnittpunkt der gesamtgesellschaftlichen Schichtentrennung einnimmt, scheint für die Leistung der Kinder in diesem Fall eine der Mittelschichtsozialisation vergleichbare Wirkung zu haben.

Innerhalb der vorwiegend der Mittelschicht zugeordneten Arbeitsplatzkategorien zeigen die Gruppen 7, 8, 11 und 13 relativ schlechtere Leistungen. Bei kleinen und mittleren Selbständigen (Kategorie 7) führt - wahrscheinlich ähnlich wie beim "qualifizierten Handwerk in Industrie und Handelsgewerbe" (Kategorie 4) - stärkere Belastung des Vaters durch Berufstätigkeit (wie nicht deklarierte Mitarbeit der Mutter) zu geringeren Kommunikationsmöglichkeiten des Kindes mit beiden Eltern und zu verminderter verbaler Anregung. Auffällig ist, daß Kinder von Vätern aus Bürokratien mit flexiblem Karrieremuster (Kategorie 9 und 12) den Kindern von Vätern aus

Bürokratien mit fixem Karrieremuster (Kategorie 8 und 11) unabhängig von der Qualifikationsstufe überlegen sind; der Unterschied beträgt .68 (WST-z-Werte), ist im Scheffé-Test jedoch nicht signifikant. Dies Ergebnis ist verwunderlich, da man eher bei Vätern aus verbeamteten Bürokratien, deren Fortkommen mehr vom Senioritätsprinzip als von der Leistungskonkurrenz abhängt, eine intensivere Beschäftigung mit den Kindern erwartet hätte. Dies Ergebnis ist um so erstaunlicher, als die Kinder von Vätern mit "nichtöffentlichen Führungstätigkeiten" (Kategorie 14), die zumindest den hochqualifizierten Tätigkeiten in Bürokratien mit flexiblen Karrieremustern bezüglich der stärkeren Gewichtung des Berufs gegenüber dem Familienleben nahestehen, wie erwartet vergleichsweise geringere Wortkenntnis aufweisen. Das sehr gute Abschneiden der Kinder von Vätern in "hochqualifizierten" Tätigkeiten der Lebensfürsorge und der Ausbildung (Kategorie 13, vorwiegend Ärzte und Lehrer) ist dagegen leicht erklärbar: die berufliche Tätigkeit verlangt hier eine kommunikative Orientierung, zudem läßt sie den meisten Fällen viel Raum für das Familienleben. Negative Wirkungen beruflichen Leistungs- und Karrieredrucks wie bei den "nichtöffentlichen Führungskräften" sind hier gerade nicht zu erwarten; vielmehr sind die Berufe des Lehrers und des verbeamteten, wie frei praktizierenden, Arztes typischerweise karriereentlastet. Dies erklärt den gewaltigen Unterschied zwischen den beiden Gruppen (13 und 14), deren materielle Bedingungen wohl vergleichbar sind (1.49 WST-z-Werte, nicht signifikant). Der Vergleich der Gruppen 13 und 14 zeigt zudem, daß das Abfallen der OMS in der Analyse der Wortschatzleistung nach 5 Schichten auf die Kinder der "Führungskräfte" zurückgeführt werden muß.

Neben einzelnen Gruppen können auch zusammengefaßte Gruppen verglichen werden; da sich die stärksten Unterschiede zwischen den Arbeitsplatzkategorien im Bereich der Mittelschicht ergeben (Kategorien 6-14), werden wir vor allem hier sinnvolle Vergleichsmöglichkeiten suchen. Der Vergleich aller extrem guten mit allen extrem schwachen Gruppen (6, 9, 13 vs. 7, 11, 14) läßt sich theoretisch

nur schwer begründen, ergibt aber auch statistisch keinen signifikanten Scheffé-Test (1.24 WST-z-Werte, n.s.). Umgekehrt lassen sich die Gruppen 8, 11, 13, wo Berufserfolge eher auf sozialer Fähigkeit zurückgeführt wird und die berufliche Laufbahn relativ klar vorstrukturiert ist, und die Gruppen 7, 9, 12, 14, wo Erfolg eher individuellem Leistungsstreben zugeschrieben wird und die Berufskarriere relativ offen ist, als "bürokratische" und "unternehmerische" (Miller/Swanson 1962) vergleichen; es ergibt sich jedoch nur ein minimaler Unterschied. Eine Vielzahl von weiteren Gruppenvergleichen innerhalb der Mittelschicht, die entweder von den gegebenen Leistungen der Einzelgruppen oder von theoretischen Überlegungen ausgingen, haben keine signifikanten Scheffé-Tests erbracht. Lediglich Vergleiche, die sich über die ganze Breite der Arbeitsplatzkategorien erstreckten (Gruppen 1-5 und 7 vs. 6 und 8-13 bzw. 6 und 8-14), ergaben auf dem 5%-Niveau signifikante Scheffé-Tests. Da sie aber zugleich (durch die Gruppeneinsparung noch verschärfte) Schichtunterschiede darstellen, die ja schon in Abschnitt 4.1. behandelt wurden, so müssen wir zusammenfassen, daß es nicht gelungen ist, den insgesamt hochsignifikanten Effekt der Arbeitsplatzkategorien auf signifikante Mittelwertunterschiede zwischen den Gruppen zurückzuführen, die vorwiegend der gleichen Schicht zuzuordnen sind. Die Unterschiede zwischen den Gruppen, die wir interpretiert haben, tragen nicht signifikant zum Arbeitsplatz-Effekt bei; trotzdem bleiben sie als Illustration möglicher subkultureller Einflüsse wichtig, die jedoch den globalen Schichteffekten untergeordnet werden müssen.

In der zweiten Messung standen uns zwei Experimentalstichproben zur Verfügung (N=63 und N=78). In den beiden Stichproben wurden einige Vpn. wegen fehlender Angaben nicht berücksichtigt. Wir werden uns in der Hauptsache auf die Stichprobe II (N=78) beziehen. Sie hat größere Anzahlen in den Zellen und in ihr können die mündlichen und schriftlichen Messungen der Vpn. 400 und 110 in der

jeweiligen Kategorie des Arbeitsplatzes berücksichtigt werden.¹⁾ Zudem unterscheiden sich die beiden Experimentalstichproben auf den Mittelwerten der Kontrollvariablen für die 14 Arbeitsplatzkategorien kaum (siehe Tab. 29 und Abb. 12) - mit Ausnahme der Kategorien 2 und 5, denen die zwei Vpn. 110 und 400 angehören. Schließlich - das sei vorweggenommen - unterscheiden sich auch die Ergebnisse der beiden Experimentalstichproben kaum - wiederum mit Ausnahme der beiden Gruppen 2 und 5.

Die Überlegungen zur Anwendung von Prüfverfahren für den differenzierten Schichtvergleich (siehe Abschnitt 5.6.1.) gelten auch für den Arbeitsplatz des Vaters. Wir haben auch hier im Prinzip die gleiche Vorgehensweise gewählt. Allerdings haben wir auf die Prüfung der Homogenität der Varianzen verzichtet; sie ist bei der Vielzahl der - gemessen an der Gesamtstichprobe - relativ willkürlich schwächer oder stärker besetzten Zellen, deren Varianz - wie ein Vergleich von Tab. 29 mit Tab. 96 zeigt - schon durch die Parallelisierung stark schwanken muß, unwahrscheinlich und eine Prüfung wäre bei zwei Zellen mit einer Besetzung von nur einer Vp. und einer Varianz von 0 ohnehin negativ verlaufen. Unter diesen Umständen hat die Varianzanalyse nicht mehr die Funktion, Effekte zu prüfen, sondern sie kann nur noch auf die Variablen mit den am stärksten schwankenden Mittelwerten hinweisen.

¹⁾-----
In der Experimentalstichprobe I wurden die schriftlichen Messungen der Vpn. 400 als mündliche Messung zu der schriftlichen Messung von Vpn. 110 hinzugenommen (siehe Abschnitt 2.1.). Da Vp. 400 zur Arbeitsplatzkategorie 2, Vp. 110 zur Arbeitsplatzkategorie 5 gehört, ergaben sich für die Kategorie 2: 3 mündliche und 2 schriftliche, für die Kategorie 5: 3 mündliche und 4 schriftliche Messungen. In der Experimentalstichprobe II hingegen wurde die schriftliche Messung von Vp. 400 als schriftliche Messung belassen, ebenso wie die anderen Messungen der beiden Vpn. Da in den Gruppen der Arbeitsplatzkategorien sowieso Vpn. mit verschiedenen Testformen zusammengefaßt sind, fällt die Unterschiedlichkeit der Testformen bei der mündlichen und schriftlichen Messung von Vp. 110, die anfänglich zur Ersetzung ihrer mündlichen Messung durch die schriftliche Messung von Vp. 400 geführt hat, nicht mehr ins Gewicht.

Die Mittelwerte einer Reihe von Variablen in den 14 nach der Arbeitsplatzkategorie des Vaters gebildeten Gruppen von Vpn. sind für die Experimentalstichprobe I in Tab. 96, für die Experimentalstichprobe II in Tab. 97 wiedergegeben. Eine Varianzanalyse mit Kommunikationsform und Arbeitsplatz als Faktoren, also ohne Berücksichtigung des Faktors Versuchspersonen ("wiederholte Messungen"), ergab keinerlei unter 10% signifikante Effekte für die Interaktion und für die Kommunikationsform, wohl aber signifikante Effekte für den Arbeitsplatz des Vaters auf den folgenden Variablen: MIBEX ($F=1.698; p<.073$), DIBEX ($F=2.117; p<.019$), PEXAT ($F=1.914; p<.037$), Rangsummen bzw. mittlerer Rangplatz ($F=2.417; p<.007$). Zur Verdeutlichung für diese Variablen sind die Mittelwerte in Abb. 13 dargestellt; da die Profile für beide Experimentalstichproben nahezu identisch sind, haben wir uns auf die Experimentalstichprobe II beschränkt.

Vor der Interpretation der Ergebnisse der zweiten Messung sollen die Besonderheiten der Experimentalauswahl kurz rekapituliert werden (siehe Abschnitt 2.1.). Erstens hat sich der Zusammenhang zwischen Arbeitsplatzkategorien und dichotomer Sozialschicht durch den Fortfall der UMS noch verschärft; allein die Väter von Vpn. der Gruppe 6 gehören beiden Schichten an - dies gilt für beide Experimentalstichproben. Zweitens sind, unabhängig vom durch die Parallelisierung bedingten Ausgleich der Schichtunterschiede auf der Kontrollvariablen, aus den Gruppen 8, 9, 12, 13 und 14 besonders schwache Vpn., aus den Gruppen 10 und 11 besonders gute Vpn. ausgewählt worden. Da die passive Wortkenntnis der ersten Messung einen starken Einfluß auf die Leistungen in der aktiven Wortklärung der zweiten Messung hat (siehe Tab. 108), sind die Mittelwerte der Kontrollvariablen (Wortschatz erster Messung) für die 14 Gruppen der Experimentalstichprobe II der Abb. 13 zum Vergleich beigegeben.

Ein Profilvergleich zwischen der Kontrollvariablen und den in der zweiten Messung signifikanten Variablen MIBEX, PEXAT und RAGSU

(durchschnittlicher Rangplatz) zeigt nun sehr weitgehende Ähnlichkeiten, besonders innerhalb des der Mittelschicht zuzuordnenden Bereichs von Arbeitsplatzkategorien 7-14, in dem die größten Schwankungen der Kontrollvariablen wie der Experimentalmessung auftreten.¹⁾ Lediglich die signifikanten Arbeitsplatzunterschiede auf dem Differenzmaß DICVE können nicht durch die Kontrollvariablen bedingt sein. Innerhalb der der Unterschicht zuzuordnenden Arbeitsplatzkategorien zeigen sich hier aber nur geringfügige Unterschiede. Von der Konstruktion der Maße her kann zudem das Maß DIBVE umso größer werden, je schlechter die Explikationsbewertung ist. Abb. 13 zeigt aber eine auffällige Abhängigkeit der Größe der Differenzen von der Güte der Explikation im Bereich der Mittelschicht (bis auf Gruppe 7) und nur geringe Schwankungen im Bereich der Unterschicht, so daß auch die Unterschiede des Differenzmaßes DIBVE nicht inhaltlich interpretiert werden können.

Wir haben gesehen, daß die möglicherweise signifikanten Arbeitplatzeffekte auf die Bedingungen der Kontrollvariablen oder die Logik der Maßkonstruktion zurückgeführt werden können. Hat die Qualität des väterlichen Arbeitsplatzes, wenn schon nicht auf einzelne Maße, so doch auf die Struktur der Maße der zweiten Messung - also unabhängig von den Bedingungen der Kontrollvariablen - einen Einfluß? Wir haben versucht, einen solchen Einfluß zu überprüfen, indem wir die wichtigsten Maße aus den drei der Maßkonstruktion zugrundeliegenden Dimensionen miteinander in jeder der 14 Kategorien verglichen haben (siehe Abb. 14); MDAEX für die Maße, die Antwortgüte und Antworthäufigkeit bewerten, MIBEX für die Maße, die nur die Antwortgüte bewerten und GATEX für die Maße, die nur die Antworthäufigkeit bewerten.

¹⁾-----
Allein bei den durchschnittlichen Rangplätzen gibt es geringe Abweichungen des Profils vom Profil der Kontrollvariablen: die Gruppen 4 und 13 zeigen bessere durchschnittliche Rangplätze als man von den Werten der Kontrollvariable erwarten würde.

Die Betrachtung der Struktur dieser Maße zeigt folgendes: Innerhalb der der Unterschicht zugehörigen Arbeitsplatzkategorien 1-5 werden vergleichbare Gesamtleistungen der aktiven Worterklärung (MIAEX) entweder durch besonders zahlreiche (GATEX) oder besonders gute (MIBEX) Erklärungen hervorgerufen; das gleiche gilt für die Arbeitsplatzkategorien 5-7. Innerhalb der der Mittelschicht zugehörigen Arbeitsplatzkategorien 9-14 hängt eine - mit der Kontrollvariablen variierende - Gesamtleistung (MIAEX) in der gleichen Weise mit zahlreichen wie guten Erklärungen zusammen. Besonders zahlreiche, aber relativ schlechte Worterklärungen geben die Kinder von Vätern aus dem Dienstleistungsgewerbe und dem Handwerk (Kategorien 2: Kellner, Pförtner, Boten (N=5); 4: Handwerk (N=8); 5: Kraftfahrer, Briefträger (N=6) und 7: (N=5)), während die Kinder von Industriearbeitern, gleich welcher Qualifikation: Hilfsarbeiter (Kategorie 1, N=5), Facharbeiter (Kategorie 3, N=11) oder Meister und Vorarbeiter (Kategorie 6, N=8) ihre Gesamtleistung stärker durch gute als durch zahlreiche Erklärungen erlangen.

Dieser Sachverhalt - so wenig er auch statistisch gesichert sein mag - läßt sich sinnvoll interpretieren. Zunächst scheinen innerhalb der Mittelschicht angehörenden Subkulturen die psychischen Bedingungen für Anworthäufigkeit und -güte - sagen wir Risiko-bereitschaft und korrekte Selbstwahrnehmung - gleichgewichtig sozialisiert zu werden. Ferner scheinen innerhalb der Unterschicht zuzurechnender oder ihr nahestehender Subkulturen diejenigen Väter, deren Arbeit entweder den Umgang mit Personen und daher den Einsatz von Sprache erfordert oder auch nur begleitend nahelegt und einen gewissen Entscheidungsspielraum offen läßt, ihre Kinder stärker im Sinne unkontrollierter Geschwätzigkeit verbal zu fördern, während die Väter, deren Arbeit in mehr oder minder spezialisierten Umgang mit Sachen oder in der Kontrolle solcher Arbeit (die jedoch - ganz abgesehen von Gesichtspunkten der zurückgelegten Karriere - die gleiche Einstellung verlangt) besteht, ihre Kinder mehr zu kontrolliert zurückhaltendem Sprachgebrauch erziehen.

Zusammenfassung: In der ersten Messung zeigten sich besonders gute Wortkenntnis innerhalb der Unterschicht bei den Kindern von Meistern und Vorarbeitern (Kategorie 6), sowie besonders schlechte Wortkenntnis innerhalb der Mittelschicht bei Kindern von Vätern, die in Bürokratien mit flexiblem Karrieremuster arbeiten, sowie bei Kindern von Führungskräften. In der zweiten Messung zeigten sich keine Unterschiede zwischen den Arbeitsplatzkategorien, wenn man die unterschiedlichen Ausgangslagen der Kinder auf der Kontrollvariablen berücksichtigte; dies ist für das soziologische Vorurteil besonders enttäuschendes Ergebnis, da es zugleich den offensichtlich besonders starken Einfluß der - an dieser Stelle soziologisch nicht weiter erklärbaren - passiven Wortkenntnis unabhängig von mutmaßlichen Einflüssen der Arbeit des Vaters belegt. Nur bei einer Betrachtung der Struktur der Variablen der zweiten Messung ergaben sich sozialisations-theoretisch sinnvolle Einzelergebnisse. Insgesamt sind die Ergebnisse zum Einfluß des Arbeitsplatzes nicht zu einer Modifizierung der negativen Ergebnisse zum Einfluß der sozialen Schicht geeignet, sondern eher zu einer Bekräftigung.

5.6.4. Einflüsse der Arbeitsplatzsituation der Mutter auf die Experimentalleistung der zweiten Messung.

Für das Lernen und die Sozialisation des Kindes hat die Qualität des Arbeitsplatzes der Mutter - schon theoretisch - nicht den gleichen Einfluß wie der Arbeitsplatz des Vaters. Erfahrungen im Beruf schlagen sich nicht unmittelbar im Erziehungsverhalten durch, sondern nur über das Zwischenglied vom Beruf sowohl geprägter wie auf Dauer stabilisierter Wertorientierungen. Was aber die Stärke und Durchsetzungskraft von Werten und Normen betrifft, ist der Vater der Mutter in der typischen Familienstruktur überlegen. Da der Vater als der Ernährer der Familie sich auf Dauer mit seinem Beruf identifizieren muß, wird er auch - nicht notwendig ausdrücklich - feste Einstellungen aus dem Beruf in das Familienleben einbringen; da die Familie auf ihn als den Außenrepräsentanten angewiesen ist, wird sie die Werte und Normen des Vaters akzeptieren. Wenn man in dieser Art die Familie als ein von personenbezogenen Merkmalen zunächst unabhängiges Muster sozialer Prozesse ansieht, so sind Werte und Normen der Mutter - überspitzt gesagt - funktionslos: Sie wird nicht ihre eigenen Vorstellungen zur Geltung bringen, sondern die des Vaters übermitteln; in einer Studie über den Gebrauch gegenüber dem Kinde nicht gerechtfertigter physischer Gewalt z.B. wandte die Mutter dieses Erziehungsmittel in Abhängigkeit von Autoritarismus des Vaters an, aber unabhängig von eigenen autoritären Einstellungen (Hoffmann 1963).

Die Annahme, daß die Wertorientierungen der Mutter, weil sie die Erziehung der Kinder unmittelbar in den Händen halte, für die Sozialisation der Kinder wichtiger seien als die des Vaters, ist also viel weniger plausibel, als sie auf den ersten Blick scheinen mag. Gerade weil der Vater durch seine Berufstätigkeit den sozialen Status der Familie festlegt, können seine Erfahrungen außerhalb der Familie Ziele, Normen und Verhaltensweisen des Familienlebens bestimmen. Umgekehrt werden Werte und Verhalten der Mutter, gerade weil sie vorwiegend die inneren Angelegenheiten der Familie verwaltet, stärker von der Tatsache einer interferierenden Berufstätigkeit

überhaupt als von spezifischen beruflichen Erfahrungen beeinflusst. Erst wenn die Position der Mutter in der Familie sich der des Vaters annähert, kann - wie für den Vater - gefragt werden, ob spezifische berufliche Erfahrungen erzieherische Einstellungen prägen und ob sich die durch berufliche Erfahrungen gesättigten Einstellungen im Lernerfolg des Kindes niederschlagen.

Wir haben die Stärke und Durchsetzungskraft von Wertorientierungen als nicht überprüftes Zwischenglied zwischen beruflichen Erfahrungen der Eltern und Lernerfolgen des Kindes eingeführt, um die größere Bedeutung des Vaters in der in einer Richtung ablaufenden Bedingungsfolge zu erläutern. Im Falle der Mutter ist fernerhin der dargestellte Zusammenhang nicht nur schwächer, er kann auch in einer anderen Abfolge vorgestellt werden: Die Entscheidung für Berufstätigkeit und die Wahl eines bestimmten Arbeitsplatzes ebenso wie die Erziehung des Kindes können durch die gleichen zugrundeliegenden Wertorientierungen bestimmt sein, die Zusammenhänge zwischen Beruf und Wertorientierung sich also umkehren; dies ist um so wahrscheinlicher, je weniger die Berufstätigkeit der Mutter durch materielle Zwänge bestimmt ist. Die Konzeption, daß Berufs- wie Erziehungstätigkeit der Mutter in gleicher Weise von allgemeineren Motiven bestimmt ist, wird der Tatsache, daß die meisten Mütter nur früher berufstätig waren, besser gerecht als die Konzeption einer Abfolge in nur einer Kausalrichtung. Gleichgültig aber, wie wir den Zusammenhang zwischen dem Arbeitsplatz der Mutter und Wertorientierungen uns denken, kommen wir zu der Voraussage, daß die Berufstätigkeit in höher qualifizierten Positionen mit einer intensiveren Förderung des Kindes zusammenhängt, sei es weil qualifiziertere Tätigkeiten im Beruf den Wertvorstellungen der Mutter in der Familie ähnliches Gewicht geben wie denen des Vaters, sei es weil materielle Entlastung und aufgeklärte Ansichten die Mutter zur Berufstätigkeit ebenso wie zu einer am Kind orientierten Erziehung motivieren.

Unter dem Gesichtspunkt der Unterschiede zwischen manuellen und qualifizierten weiblichen Berufstätigkeiten sollen die Besonderheiten unserer Experimentalsamples rekapituliert werden (Abschnitt 2.1.). Die Häufigkeitsverteilung der Arten von Berufstätigkeit ist bei den Müttern unregelmäßiger als bei den Vätern; die meisten Mütter sind oder waren entweder in manuellen Hilfs- und Routinetätigkeiten (Kategorien 1 und 2) oder in qualifizierten Tätigkeiten in Bürokratien und der Lebensfürsorge tätig (Kategorien 7, 8, 9). Dem entspricht die Trennung zwischen Unter- und Mittelschicht, die jedoch nicht so eindeutig wie bei den Arbeitsplatzkategorien des Vaters ist; einige Frauen, die in typischen Berufen der Unterschicht gearbeitet haben, konnten später eine Familie gründen, die von uns zur Mittelschicht gerechnet wurde. Noch stärker wahrscheinlich hängt die Art der Berufstätigkeit der Mutter mit ihrer Ausbildung zusammen, d.h. einem Merkmal, das vor der Berufstätigkeit und der Familiengründung erworben wurde. Schließlich sind insgesamt die meisten Mütter nur früher berufstätig gewesen; dies gilt auch in allen einzelnen Kategorien mit Ausnahme der vorwiegend heute noch berufstätigen Mütter in hochqualifizierten Tätigkeiten (Kategorie 10), die zudem alle entweder eine Fachschul- oder Hochschulausbildung abgeschlossen haben. Aus diesen Gründen werden wir die Kategorien des Arbeitsplatzes der Mutter nicht - wie beim Vater - mit den beiden dichotomen Schichten identifizieren, sondern als einen zwar mit gegenwärtiger sozialer Schicht zusammenhängenden, aber biografisch frühere Schichtzugehörigkeit repräsentierenden Faktor ansehen.

Vergleichen wir Kinder gegenwärtig berufstätiger und gegenwärtig nicht berufstätiger Mütter (Tab. 98), so zeigt sich im passiven Wortschatztest, der Kontrollvariablen unseres Experiments, eine deutliche Überlegenheit der Kinder nicht berufstätiger Mütter, die sich sehr wahrscheinlich auch in der Gesamtstichprobe wiederfinden wird. Eine vergleichbare Überlegenheit der Kinder nicht berufstätiger Mütter in der Experimentalleistung findet sich jedoch nicht. Die Nichtbeachtung der Tatsache der Berufstätigkeit überhaupt kann also

den Einfluß der Arten von Berufstätigkeiten nicht sehr stark verzerrt haben. Die Kinder ohne Angabe zur Berufstätigkeit der Mutter sind die auf der Kontrollvariablen schwächsten unseres Experiments; in der Experimentalleistung jedoch sind sie den beiden anderen Gruppen zwar auf den Maßen der Erklärungshäufigkeit, nicht aber auf den Maßen der Erklärungsgüte unterlegen. Daß sie aus dem Vergleich der Arten von Berufstätigkeit herausfallen, kann die Varianz nur auf den Maßen der Erklärungshäufigkeit verringert haben - ganz abgesehen von der Frage, zu welcher der beiden anderen Gruppen diese Kinder zu rechnen wären.

Für die Untersuchung des Einflusses der Arten von Berufstätigkeiten werden wir wieder die beiden Experimentalstichproben sowie die im 5.6.1. skizzierten statistischen Verfahren heranziehen. Da sich zwischen den beiden Stichproben auf der Kontrollvariablen und in den Ergebnissen nur geringfügige Unterschiede ergeben, werden wir die Darstellung auf Stichprobe II konzentrieren. In den Tabellen der Stichprobe I ist die Kategorie 7 mit nur 1 Vp. gesondert aufgeführt; in den Tabellen der Stichprobe II wurden die Kategorien 7 und 8 (qualifizierten Tätigkeiten in Bürokratien mit flexiblen und mit fixem Karrieremuster) in einer Gruppe zusammengefaßt, da die Unterscheidung zwischen fixem und flexiblem Karrieremuster für verheiratete berufstätige Frauen nicht die gleiche Bedeutung hat wie für Männer. Da diese beiden Gruppen zwar unter Gesichtspunkten der Häufigkeit, nicht aber der zu prüfenden Ergebnisse zusammengelegt wurden, führten wir auch noch mit diesen zusammengelegten Gruppen Varianzanalysen durch. Sie sollen, da sich hier wie beim Arbeitsplatz des Vaters ungleiche Varianzen in den Zellen ergeben, wiederum nur Variablen mit größeren Mittelwertunterschieden herausgreifen, ohne Unterschiede statistisch streng zu prüfen.

Die Mittelwerte und Standardabweichungen für die 10 Kategorien des Arbeitsplatzes der Mutter in Stichprobe I stellt Tab. 99, die Mittelwerte für die 9 Kategorien des Arbeitsplatzes der Mutter in Stichprobe II Tab. 100 dar. Die F-Tests des Arbeitsplatzeffektes einer

Varianzanalyse verschiedener Maße ohne Berücksichtigung der wiederholten Messungen sind in Tab. 1o1 aufgeführt; verschiedene Varianzanalysen mit dem Maß MIAEX sind in Tab. 1o2 vollständig dargestellt. Zuerst soll die Struktur der signifikant zwischen den Arbeitsplätzen unterscheidenden Maße gefunden und dann die Unterschiede zwischen den einzelnen Kategorien untersucht werden.

In der Varianzanalyse mit dem Arbeitsplatz der Mutter und der Kommunikationsform als Effekten ohne die Berücksichtigung der wiederholten Messung (Tab. 1o1) ergaben sich für jede der drei Dimensionen der Maßkonstruktion signifikante Effekte des Arbeitsplatzes der Mutter: Für Maße der Erklärungsgüte, der Häufigkeit und der Erklärungsgüte wie -häufigkeit.¹⁾ Als eine dem Experimentalaufbau angemessenere Prüfung haben wir Varianzanalysen mit Berücksichtigung der wiederholten Messung gerechnet; um Aussagen auch für andere Variablen zu gewinnen, wählten wir eine Variable mit einem der höchsten F-Werte in der Varianzanalyse ohne Berücksichtigung der wiederholten Messungen, die zugleich den größten Informationsgehalt hat: MIAEX (Tab. 1o2). Für die Stichprobe I, in der Gruppe 7 und 8 getrennt sind, ergibt sich zwar kein signifikanter, aber noch ein recht großer F-Wert für den Effekt des Arbeitsplatzes der Mutter; für die Stichprobe II, in der Gruppe 7 und 8 zusammengelegt wurden, jedoch auch kein großer F-Wert mehr; offensichtlich hat die besonders gute Leistung der Vp. aus Gruppe 7, die ja durch ihre Stellung

1)-----
Alle Maße der Explikation zeigen signifikante F-Werte, Maße des Verständnisses jedoch nur in der A-Kategorisierung, alternierend mit den Differenzmaßen in der B-Kategorisierung und C-Kategorisierung. Das Alternieren signifikanter und nicht signifikanter Effekte von Verständnis- und Differenzmaßen zwischen der A- und B- bzw. C-Kategorisierung, ergibt sich aus der Konstruktion der Maße: Da Verständnis und Explikation hoch korrelieren, sind signifikante Effekte auf Verständnis-, Explikations- und Differenzmaßen zugleich sehr unwahrscheinlich; da in der A-Kategorisierung die für die Differenz irrelevante Anwohnhäufigkeit mitberücksichtigt wird, ergeben sich hier signifikante Effekte für das Verständnis, in der B- und C-Kategorisierung aber für die Differenz. Eine Interpretation sind daher höchstens die Differenzmaße der B- und C-Kategorisierung wert.

auf der Kontrollvariablen bedingt ist, einiges zu den signifikanten F-Werten aus Tab. 101 beigetragen. Zugleich taucht in beiden Varianzanalysen mit Berücksichtigung der wiederholten Messungen ein im Vergleich zu der Varianzanalyse ohne Berücksichtigung der wiederholten Messungen recht großer F-Wert für die Kommunikationsform auf. Ein Blick auf Tab. 99 und 100 zeigt, daß besonders in den Gruppen 6 und 4 größere Unterschiede zwischen mündlicher und schriftlicher Messung bestehen. Nach den Ergebnissen dieser Varianzanalysen muß wiederholt werden, daß die folgende Inspektion der Mittelwerte nur illustrativen Wert haben kann.

Für die Mittelwertvergleiche lassen sich die Kategorien nach sachlichen Gesichtspunkten des Arbeitsplatzes der Mutter und nach den Ergebnissen der Erklärungsleistung der Kinder zusammenfassen; eine Zusammenfassung der Mittelwerte von Variablen aus den drei Dimensionen der Maßkonstruktion sowie der Kontrollvariablen nach beiden Gesichtspunkten ist in Tab. 103 dargestellt. Die nach sachlichen Gesichtspunkten zusammengefaßten Gruppen weisen in den Ergebnissen nur geringe Differenzen auf, die in vielen Fällen Unterschieden der Kontrollvariablen entsprechen. Die Unterschiede zwischen den nachträglich gebildeten Gruppen sind eher noch stärker durch Unterschiede der Kontrollvariablen geprägt. Wie schon bei der Analyse des Arbeitsplatzes des Vaters finden wir hier statt soziologisch interpretierbarer Einflüsse eine allgemeine Abhängigkeit der aktiven Worterklärung von der passiven Wortkenntnis. Eine einzige Ausnahme gibt es von dieser Regel. Die Kinder von Müttern in hochqualifizierten Tätigkeiten (Kategorie 10) zeigen ihrer Stellung auf der Kontrollvariablen widersprechende Leistungen. Obwohl innerhalb dieser Gruppe Kinder mit relativ geringer aktiver Wortkenntnis für das Experiment ausgewählt wurden, geben sie sowohl die besten wie auch relativ zahlreiche aktive Worterklärungen. Dies Ergebnis entspricht genau unseren Erwartungen. In hochqualifizierten Berufen kann offensichtlich die Art der Berufstätigkeit wie die zugehörige Ausbildung der Mutter - alle Mütter dieser Gruppe haben abgeschlossene Fachschul- oder Hochschulbildung - für das verbale Lernen der

Kinder förderlich sein; zudem ist die Mehrzahl der Mütter dieser Gruppe noch heute berufstätig, so daß wir drei eng miteinander zusammenhängende Faktoren der Erklärung haben.

Der Vergleich zwischen den beiden am stärksten besetzten Gruppen der manuellen (Kategorie 1-3) und qualifizierten Tätigkeiten (Kategorie 7-9) bietet die Möglichkeit den Einfluß für die Unterschicht und für die Mittelschicht typischer Berufstätigkeiten von Frauen unabhängig von der Schichtzugehörigkeit ihrer späteren Familie gegenüberzustellen; die wenigsten dieser Mütter ja sind heute noch berufstätig (Tab. 31). Für die Gesamtleistung ergibt sich eine geringere Überlegenheit der Kinder von Müttern mit qualifizierter Berufstätigkeit, die einem geringfügig besseren Mittelwert auf der Kontrollvariablen entspricht; die bessere Gesamtleistung ergibt sich aber aus der größeren Antwortbereitschaft (GATEX) der Kinder in qualifizierten Tätigkeiten arbeitender Mütter. Ein Beleg für den Einfluß verschiedener Arten früherer, mit der gegenwärtigen zwar zusammenhängender, aber nicht unmittelbar identischen Schichtzugehörigkeit läßt sich aus diesen Unterschieden jedoch nicht gewinnen.

Die Ergebnisse zum Einfluß der Art der Berufstätigkeit der Mutter müssen wir ähnlich negativ wie für die Art der Berufstätigkeit des Vaters resümmieren. Auch hier wiegen durch die Kontrollvariable bedingte Unterschiede vor. Zusätzlich stellt sich die Schwierigkeit, zwischen Einfluß früherer und noch andauernder Berufstätigkeit zu unterscheiden. Unabhängig von der Kontrollvariablen hängt hochqualifizierte Berufstätigkeit der Mutter mit guten aktiven Worterklärungen des Kindes zusammen; die Art der Berufstätigkeit und Tatsache, daß sie gegenwärtig noch andauert, wirken aber hier zusammen. Innerhalb von Arten vorwiegend früherer Berufstätigkeit ließen sich, auch bei relativ geringeren Unterschieden auf der Kontrollvariablen, keine Unterschiede ausmachen.

5.6.5. Der Einfluß des Buchbesitz als eines Indikators der sozio-kulturellen Lebenswelt auf die Experimentalleistung der zweiten Messung.

Buchbesitz kann als ein Indikator für sozio-kulturelle Gewohnheiten des Elternhauses angesehen werden: Für die Berührung mit Gütern der allgemeinen Kultur und für den Umfang des in die Familie eingebrachten Umweltkontaktes; er kann nicht auf spezifische Orientierungen des Alltags und der Freizeit - etwa im Sinne bildungsbürgerlicher Vorurteile - hinweisen. Mit dem "Buchbesitz" verlassen wir bei unserer Untersuchung sozialer Einflüsse den Bereich außerfamilialer Faktoren und suchen nach Indikatoren der familialen, subkulturell typischen Lebenswelt. Über diese allgemeine Erwägungen hinaus kann "Buchbesitz" im Zusammenhang einer sozio-linguistischen Untersuchung zudem als ein Indikator der sprachlich vermittelten Anregung - gleich welchen Inhalts - in der Familie angesehen werden.¹⁾

Nach dem Buchbesitz wurde im Elternfragebogen in einer geschlossenen Frage mit 5 Vorgaben gefragt; die Antworten wurden innerhalb der Gesamtstichprobe (N=294) z-transformiert. Wie nicht anders zu erwarten, korreliert der Buchbesitz hoch mit der Schichtzugehörigkeit, sowohl in der Gesamtstichprobe wie in unseren Experimentalstichproben (siehe Tab. 21). Innerhalb der Unterschichtgruppe findet sich jedoch kein Zusammenhang mehr zwischen Schicht und Buchbesitz, sehr wohl aber innerhalb der Mittelschichtgruppe. Anders gesagt: Der Buchbesitz kann zwischen subkulturellen Milieus der Unterschicht nicht, aber sehr wohl subkulturellen Milieus der Mittelschicht differenzieren; dies stimmt mit Beobachtungen der Gesamt-

1) -----
Wenn auch diese spezifische Verwendung des "Buchbesitzes" in einem vordergründigen Sinn in einer sozio-linguistischen Untersuchung naheliegt, so ist er für die Sozialisationsforschung, zu der die sozio-linguistischen Untersuchungen ja beitragen sollen, jedoch ein allgemeinerer Indikator. "Buchbesitz" zeigt das "kulturelle Kapital" (Bourdieu/Passeron) eines Individuums oder sein "hidden curriculum" (Strodtbeck), d.h. alle Erfahrungen der herrschenden Kultur, die in der Schule vorausgesetzt werden und durch die sie selektiv wirkt, ohne selber zu selektieren.

stichprobe überein (Bosse u.a. 1971:108; Devermann/Krappmann/Kreppner 1972). Ferner korreliert der Buchbesitz in keiner unserer Experimentalgruppen mit der Kontrollvariablen, Wortschatztest erste Messung; Einflüsse des Buchbesitzes können also nicht durch Einflüsse der Kontrollvariablen abgedeckt sein (siehe Tab. 21).

Die Korrelationen des Buchbesitzes mit den Maßen der zweiten Messung sind in Tab. 104 dargestellt. Insgesamt hat der Buchbesitz einen geringen positiven, jedoch signifikanten Einfluß auf die Erklärungshäufigkeit und die Erklärungsgüte und -Häufigkeit, nicht aber auf die Erklärungsgüte allein. Bei Trennung der beiden Schichtgruppen findet sich in der Unterschicht überhaupt kein, in der Mittelschicht jedoch ein verstärkter Zusammenhang der gleichen Art wieder: Von den schriftlichen Variablen kann durch den Buchbesitz durchschnittlich fast ein Viertel der Varianz erklärt werden. Die Tatsache, daß die Erklärungsleistungen in den oberen Straten der Mittelschicht ansteigen, findet hier qualitativ eine Präzisierung und quantitativ eine Verstärkung. In der schriftlichen Messung der Mittelschicht kann der Buchbesitz als Prädiktor sogar mit der Kontrollvariablen WST konkurrieren. Die Korrelationskoeffizienten von BUCHBE mit den Maßen der Erklärungshäufigkeit und der Erklärungsgüte sind so hoch wie die entsprechenden Korrelationen des Wortschatztests (siehe Tab. 108).

Der Buchbesitz beeinflusst die gleichen Variablen wie die Ausbildung der Mutter: die Häufigkeit und die Güte der Erklärungen (siehe 5.6.2.). Aber noch weitere Gemeinsamkeiten ergeben sich zwischen diesen beiden Prädiktoren: Beide sind unabhängig von der Kontrollvariablen WST; beide stellen einen Einfluß dar, der in beiden Schichtgruppen wirksam ist, ohne in der Schichteinteilung schon voll erfaßt zu sein; beide stellen schließlich einen Einfluß dar, der in der Mittelschicht stärker ist als in der Unterschicht. Die Ausbildung der Mutter korreliert in der Gesamtgruppe (Stichprobe I: .64; Stichprobe II: .65) wie in der Mittelschichtgruppe (.62) sehr hoch mit dem Buchbesitz der Familie, in der Unterschichtgruppe je-

doch gar nicht (.08).¹⁾ Dort wo beide Einflüsse, Ausbildung der Mutter und Buchbesitz als Indikator kultureller Gewohnheiten, zusammengehen, nämlich in der Mittelschicht, haben sie auch den stärkerem Einfluß auf die Leistungen des Kindes.

Diese recht weitgehenden Beziehungen zwischen der Ausbildung der Mutter und dem Buchbesitz der Familie legt die Vermutung nahe, daß die Mutter gerade über ihre Ausbildung, ihre Stellung in der Familie und ihren Einfluß auf die Bildung des Kindes gewinnt. Mit Bourdieu (1973) gesprochen, ist das "kulturelle Kapital" der Mutter das "kulturelle Kapital" der Familie und damit natürlich auch der Kinder. Je höher die Ausbildung der Mutter, desto besser steuert sie das sozio-kulturelle Anregungspotential in der Familie - vom Kinobesuch über die Ferienfahrten bis zum Thema des Tischgesprächs -, das zum Lernerfolg der Kinder beiträgt. Wir haben gesehen, daß der Beruf der Mutter kaum auf die Erklärungsleistungen des Kindes einen Einfluß hat - es sei denn in den höchsten Qualifikationsgruppen, wo der Einfluß des Berufs mit dem der Ausbildung zusammenfällt. Wir haben analysiert, daß durch ihren Beruf die Mutter kaum innerfamiliären Einfluß gewinnt und bezüglich des Berufs gleichsam im Schatten des Vaters bleibt. Wir sehen jetzt, daß die Ausbildung - und wenn er der Ausbildung entspricht, auch der Beruf - der Mutter ihre Stellung in der Familie verschafft, indem die Mutter, als Ressort gleichsam, die kulturellen Gewohnheiten der Familie verwaltet.

Daß dieser Zusammenhang gerade in der Mittelschicht besonders deutlich ist, läßt sich aus verschiedenen Gründen erklären. Einmal muß die Ausbildung der Mutter einen gewissen Schwellenwert überschritten haben um in der Familie als kultureller Katalysator wirken zu

75-----
Zwar korreliert BUCHBE mit den übrigen Schichtindikatoren GEBERV, AUSBVA und EINKOM in der Unterschicht ähnlich niedrig (.02, .15, .08), in der Mittelschicht ähnlich positiv (.45, .40, .58) wie mit AUSBMU. Jedoch korreliert keiner der übrigen Schichtindikatoren einzeln mit den Experimentalleistungen.

können; die meisten Mütter der Unterschichtgruppe haben aber nur einen Volksschulabschluß ohne Lehre (Tab. 27). Zum zweiten muß ein gewisses materielles Niveau, und zum dritten ein gewisses Niveau außerfamilialer Kommunikationsbeziehungen der Familie gegeben sein, um die genannten kulturellen Gewohnheiten innerhalb der Familie überhaupt zum Verhandlungsgegenstand werden zu lassen. Beides ist aber in der Mittelschicht wahrscheinlicher als in der Unterschicht.

Wir können von dieser Interpretation aus zurückgreifen auf die Interpretation der Unterschiede zwischen vier Schichtgruppen (5.6.1.) und des Einflusses der Ausbildung der Mutter (5.6.2.), besonders bei gleichzeitiger hochqualifizierter Berufstätigkeit der Mutter (5.6.4.). Wir hatten die größeren Differenzen zwischen Verständnis und Explikation bei den Kindern der mittleren Straten als das Resultat eines dem Lernen der Kinder entgegenkommenden Sprachgebrauchs der Erwachsenen interpretiert und vermutet, daß besonders die Mutter bewußt auf Lernbedürfnisse der Kinder eingeht - je höher ihre Ausbildung ist, um so stärker. Die hohe Korrelation des Buchbesitzes mit der Ausbildung der Mutter und der relativ starke Einfluß beider Variablen auf die Erklärungsleistungen des Kindes unterstreichen diese Interpretation.

5.6.6. Abschließende Beurteilung des Einflusses unabhängig von dichotomer Schicht erfaßter sozialer Faktoren.

Zu Beginn des Abschnitts 5.6. hatten wir die Untersuchung weiterer sozialer Einflußfaktoren auf die Erklärungsleistungen der Kinder trotz der negativen Ergebnisse zum Einfluß dichotomer Schicht gerechtfertigt: sie sollte das negative Hauptergebnis entweder bekräftigen oder soziologisch gehaltvoll modifizieren.

Auf den ersten Blick herrschen die Ergebnisse vor, die das negative Hauptergebnis weiter bekräftigen. In der differenzierten Schichteinteilung ergaben sich sowenig signifikante Schichtunterschiede wie in der dichotomen. Die vier Dimensionen des Schichtindex zeigen auch einzeln keine Zusammenhänge mit den aktiven Erklärungsleistungen - mit Ausnahme der Ausbildung der Mutter. Die Leistungen der Kinder unterschieden sich weder nach dem Arbeitsplatz der Mutter noch dem Arbeitsplatz des Vaters; mehr noch: fast alle hier auftretenden Unterschiede ließen sich auf Unterschiede zwischen den Gruppen auf dem Mittelwert der Kontrollvariablen zurückführen. Trotz dieser negativen Zusatzergebnisse aber erlaubt die Integration der positiven Ergebnisse der vorangegangenen Analysen eine Modifikation des negativen Hauptergebnisses, die wir im folgenden erläutern wollen.

Wir gehen aus von dem u.E. wichtigsten Ergebnis des Abschnitts 5.6., daß die Ausbildung der Mutter mit den Leistungen des Kindes zusammenhängt, besonders wenn sie mit der Ausbildung des Vaters in Diskrepanz steht. Zwischen den beiden Dimensionen besteht eine Substitutionsbeziehung, die durch den Filterindex der sozialen Schicht verdeckt ist. Innerhalb der Familie - so kann man folgern - sind die Einflüsse des Vaters und der Mutter auf das Lernen des Kindes nicht synchronisiert. Entsprechend müssen wir die außerfamilialen Einflüsse danach trennen, ob sie durch den Vater oder die Mutter vermittelt werden.

Auf der Seite des Vaters ließen sich weder für den Beruf noch den Arbeitsplatz Einflüsse auf die Experimentalleistungen des Kindes aufweisen. Außerfamiliale Einflüsse, die durch den Vater vermittelt werden, konnten also nicht belegt werden. Auf der Seite der Mutter finden wir einen starken Einfluß der Ausbildung, der offensichtlich über die Verwaltung der kulturellen Interessen der Familie läuft. Wir finden auch hier keinen Einfluß des Arbeitsplatzes als eines außerfamilialen Einflusses. Jedoch fügen sich die Ausbildung der Mutter und der Buchbesitz als ein innerfamiliärer Einflußstrang zusammen, der zudem umso mehr an Bedeutung gewinnt, je niedriger die Ausbildung des Vaters ist. Wir können diese Charakteristik so zusammenfassen, daß die Statusindikatoren auf die Experimentalleistung nicht wegen ihres außerfamilialen Werts, sondern wegen ihrer innerfamiliären Konstellation Einfluß haben. Mit dieser Feststellung läßt sich zudem trennen, in welcher Hinsicht die Ergebnisse des Abschnittes 5.6. das negative Hauptergebnis bekräftigen und in welcher Hinsicht sie es modifizieren. Auf der einen Seite haben wir negative Ergebnisse für die dichotome und differenzierte Schichteinteilung, Beruf und Ausbildung des Vaters, das Einkommen und die Arbeitsplätze des Vaters und der Mutter; auf der anderen Seite positive Ergebnisse für die Ausbildung der Mutter, die Ausbildung der Mutter in Kombination mit dem Beruf des Vaters und die Ausbildung der Mutter in Kombination mit dem Buchbesitz als einem Indikator kultureller Gewohnheiten. Ganz allgemein können wir folgern, ist das negative Hauptergebnis auch dadurch bedingt, daß auf den verschiedenen Stufen "sozialer Schicht" die für sie grundlegenden und für Sozialisationsprozesse wichtigen Dimensionen zu stark variieren, um noch einen einheitlichen Einfluß des Globalfaktors "Schicht" auf Sozialisationsprozesse zuzulassen.

Neben der Modifikation des negativen Hauptergebnisses sollen die Ergebnisse einiger Detailanalysen festgehalten werden. Diesen Detailanalysen ist gemeinsam, daß sie nicht mit absoluten Werten, sondern Strukturen von Maßen arbeiten, so daß sie direkt in der Theorie linguistischer Codes interpretiert werden können (siehe 3.).

Die signifikant größeren Differenzen zwischen Verständnis und Explikation in den beiden mittleren Straten haben wir aus dem bewußt didaktischen Einsatz des im elaborierten Sprachgebrauch habituellen Gebrauchs von Sprache zur Explikation von Bedeutungen erklärt. Mit dieser Erweiterung der Theorie der linguistischen Kodes haben wir auch das Zusammentreffen der Ausbildung der Mutter und des Buchbesitzes als Prädiktoren der Experimentalvariablen zu erklären versucht. Eine weitere Detailanalyse erbrachte, daß innerhalb der Unterschicht die Kinder von Vätern, deren Arbeitsplatz den Umgang mit Personen erfordert, mehr durch zahlreiche, aber wenig treffende, die Kinder von Vätern, deren Arbeit mehr den spezialisierten Umgang mit Sachen erfordert, mehr durch wenige, aber treffende Erklärungen die fast gleichen Gesamtleistungen erreichen.

Schließlich zeigten die Ergebnisse dieses Abschnittes durchgängig, daß viele, auch signifikante, Unterschiede zwischen Vpn.-Gruppen auf Unterschiede der Kontrollvariablen WST reduziert werden könnten. Die aktive Worterklärung hängt recht deutlich vom passiven Wortverständnis ab. Dies ist ein Hinweis auf die Schärfe der Kontrolle von Schichteinflüssen in den Hauptanalysen, den wir auch in den hohen Korrelationen des WST mit den Maßen des Experiments wiederfinden werden (siehe 5.8.1.; Tab. 108).

5.7. Einflüsse der Kommunikations- und Kontrollweisen der Mutter auf die aktiven Worterklärungen des Kindes.

Bei der Untersuchung der Einflüsse auf die aktive Worterklärung haben wir zuerst die dichotome Schichteinteilung sowie die übrigen im Experimentalplan vorgesehenen Faktoren (5.2., 5.3., 5.4.) untersucht, und anschließend die im Vergleich zur dichotomen Sozialschicht differenzierten Indikatoren der sozialen Lage der Eltern hinzugezogen (5.6.). Im folgenden Abschnitt nun sollen Variable des innerfamilialen sozialen Verhaltens, nämlich die Kommunikations- und Kontrollweisen der Mutter geprüft werden. Erst im nächsten Abschnitt wollen wir dann zu Variablen, die auf das Kind bezogen sind übergehen: Die Intelligenzwerte und die Schulnoten.

Die Kontrolle des Kindes durch die Mutter konnte nicht als Verhalten erhoben werden, sondern nur in Form von Einstellungsfragen, die die innerfamiliale Kontrolle und die Bewertung der Sprache für die Kommunikation in Problemsituationen betreffen und schon in Untersuchungen der Bernsteinschen Forschungsgruppe in England als "Index of maternal Control and Communication (ICC)" verwandt worden waren. Der Elternfragebogen unserer Untersuchung (siehe Bosse u.a. 1971, 67-73 und 110-145) enthielt Fragen über die Bereitschaft der Mutter, auf vom Kind initiierte verbale Interaktion einzugehen ("Kommunikationsfrage"), über die Bedeutung der Sprache für die Mutter-Kind-Beziehung ("Sprachfrage") und über die Beteiligung des Kindes an einer Auseinandersetzung zwischen den Ehepartnern ("Nachbarnfrage"). Die Kommunikationsfrage wurde von Bernsteins Forschungsgruppe übernommen (Brandis/Henderson, 1970); ihre Übersetzung lautete:

"Kleine Kinder schwatzen oft eine ganze Menge. Als Ihr Kind noch nicht zur Schule ging, was haben Sie da gewöhnlich getan, wenn es bei folgenden Gelegenheiten anfang zu schwatzen?"

Die Gelegenheiten lassen sich in öffentlich und private unterteilen und sind in Tab. 106 wörtlich zitiert. Als Antwort waren vier Möglichkeiten vorgegeben, in folgender Reihenfolge:

Ich habe ihm gesagt, es soll still sein.
Ich habe ihm gesagt, es soll warten.
Ich habe ihm kurz und bündig geantwortet.
Ich habe mich mit ihm unterhalten.

Die Vorgaben wurden als Intervallskalen genommen und für jede der 7 Gelegenheiten innerhalb der Gesamtstichprobe (N=392) z-transformiert. - Die Sprachfrage wurde ebenfalls von Bernstein übernommen (Bernstein/Henderson, 1969); sie lautete in der Übersetzung:

Wenn Eltern nicht sprechen könnten, also stumm wären, um wieviel schwieriger wäre es dann ihrer Meinung nach für diese Eltern, mit kleinen Kindern, die noch nicht zur Schule gehen, die folgenden Dinge zu tun?

Die Problematik einer solchen Fragestellung (Vergleich von Situationen bezüglich eines Komparativs unter einer irrealen Bedingung) ist ausführlich diskutiert worden (Cook-Gumperz 1973). Vorgesehen wurden 13 Lebensbereiche, die nach Bernstein in "Skill area" und "person area" unterteilt werden können; sie sind sämtlich in Tab. 107 wörtlich zitiert. Als Antwort standen 5 Möglichkeiten zwischen "sehr viel schwieriger" und "einfach" zur Auswahl. Sie wurden, als Intervallskala, für jeden der 13 Bereiche innerhalb der Gesamtstichprobe z-transformiert. In der Nachbarnfrage wurden verschiedene Reaktionen der Eltern auf einen vor den Augen des Kindes entflammten Streit der Ehepartner erfragt, die in einem Index zwischen "autoritär" und "anti-autoritär" zusammengefaßt wurden. Alle drei Fragen sollten nur von der Mutter beantwortet werden.

Die Kommunikations- und Sprachfragen dienten in der Bernsteinschen Untersuchung zur Feststellung schichtspezifischer Erziehungsstile. Bernstein nahm an, daß bei Müttern der Mittelschicht eine größere Bereitschaft zu verbaler Interaktion besteht; ferner, daß Sprache für den interpersonalen Bereich in beiden Schichten stärker bewertet, die Zunahme der Bewertung gegenüber dem sachbezogenen Bereich aber in der Mittelschicht größer sei; Bernstein konnte beide Hypothesen bestätigen. In der Untersuchung, aus der unsere Experimentalgruppe genommen ist, konnten Bernsteins Hypothesen jedoch nur teilweise bestätigt werden. Eine Faktorenanalyse der Kommunikationsfragen ergab keine sinnvolle Faktorenstruktur und eine Faktorenanalyse der Sprachfragen konnte nur eine Einteilung der

Vorgaben unabhängig von der in "sachbezogene" (skill) und "personenbezogene" bestätigen. Für unsere Untersuchung, die den Einfluß der Erziehungsstile ihrerseits auf den Sprachgebrauch des Kindes prüfen will, muß festgehalten werden, daß die Sprach- und Kommunikationsfragen nicht als geprüft Instrument angesehen werden können und die Klassifizierung der Items nicht verbindlich ist.¹⁾

Wir müssen daher die einzelnen Items getrennt betrachten und können sie nur nachträglich - nach ihrem Einfluß auf die aktive Worterklärung - zusammengruppieren.

Für die einzelnen Kommunikationsfragen scheint uns die Hypothese sinnvoll, daß verbale Interaktionsbereitschaft der Mutter mit guten aktiven Worterklärungen des Kindes zusammenhängt. Für die einzelnen Sprachfragen ist es schwierig, einfache Zusammenhänge zu formulieren. Die Betonung der Sprache durch die Mutter kann ebensogut als gezielte Reaktion auf geringe verbale Fähigkeiten des Kindes wie als naturwüchsige Forderung verstanden werden. Da der zielorientierte Einsatz von Sprache selber schichtspezifisch sein kann, scheinen uns hier schichtspezifische Zusammenhänge wahrscheinlicher als allgemeine, ohne daß wir die Richtung für die einzelnen Fragen und Schichtgruppen voraussagen wollen.

Innerhalb der gesamten Experimentalgruppe - sowohl in der Stichprobe I (N=65) wie in der Stichprobe II (N=83) - ergeben sich nirgendwo größere (über .20) Korrelationen, die sich sowohl in beiden Kommunikationsformen wie in einer Gruppe gleichartiger Maße wiederfinden. Wir müssen also die Vermutung von schichtunabhängigen Einflüssen der Kommunikations- und Kontrollweisen der Mutter auf die aktiven Worterklärungen des Kindes aufgeben und nach schichtspezifischen Zusammenhängen suchen.

Bei den schichtspezifischen Korrelationen der Kommunikationsfragen (Tab. 106) fällt zunächst auf, daß in beiden Schichtgruppen nur die

¹⁾-----
Siehe dazu ausführlicher: Bosse u.a. (1971).

Kommunikationsbereitschaft der Mutter in stärkerem Maße mit aktiven Worterklärungsleistungen des Kindes zusammenhängt, allerdings für unterschiedliche Situationen: In der Unterschicht C 3 (Entspannung), in der Mittelschicht C 1 (Hausarbeit) und C 6 (Bus oder Straßenbahn). Die Zurückweisung des Kindes ("Ich habe ihm gesagt, es soll stille sein") zeigt nur sporadisch Zusammenhänge: Am ehesten noch in der Unterschicht und den Situationen C 2 (Straße) und C 6 (Bus oder Straßenbahn), in Situationen also, in denen ein aufdringliches Kind die Mutter als eine unfähige Erzieherin abstempeln könnte. Die Gegensätzlichkeit der Zusammenhänge in C 3 (Bus oder Straßenbahn) ist trotz z.T. kleinen und sporadischen Korrelationen bemerkenswert: Während in der Unterschicht die Gesprächsbereitschaft der Mutter in privaten Situationen mit guten Erklärungsleistungen des Kindes zusammenhängen, sind es in der Mittelschicht eher öffentliche Situationen.

Die schichtspezifischen Mittelwerte der Kommunikationsfragen (Tab. 105) unterscheiden sich nur geringfügig, die schichtspezifischen Standardabweichungen zeigen nahezu die gleichen Werte, so daß unterschiedliche schichtspezifische Varianzen kaum zu Unterschieden der Korrelationen beigetragen haben könnten. Wir wissen aber weiter aus Tab. 109, daß in beiden Schichtgruppen die Leistung im passiven Wortschatz hoch mit der aktiven Worterklärung korreliert und können Tab. 105 entnehmen, daß in der Unterschicht Kommunikationsverweigerungen der Mutter bei Frage C 2 und Kommunikationsbereitschaft bei der Frage C 3, in der Mittelschicht Kommunikationsbereitschaft bei der Frage C 5 mit dem passiven Wortschatz des Kindes positiv zusammenhängen. Können die beschriebenen Zusammenhänge zwischen Kommunikationsstil der Mutter und aktiver Worterklärung durch die Korrelation beider mit dem passiven Wortschatz produzierte Scheinkorrelationen sein? Wenn wir den passiven Wortschatztest in Partialkorrelationen kontrollieren, so verringern sich die Korrelationen nur minimal,¹⁾ so daß wir insgesamt von vereinzelt, aber

¹⁾-----
In der US: C 3 mit MIAEXM von .37 auf .29, mit MIAEXS von .47 auf .38, mit MIBEXM von .54 auf .48, mit MIBEXS von .49 auf .40; in der MS: C 6 mit MIAEXM von .48 auf .44, mit MIAEXS von .52 auf .48, mit GAEXM von .47 auf .42, mit GATEXS von .41 auf .39.

signifikanten Zusammenhängen zwischen Kommunikationsstil der Mutter und aktiver Worterklärung reden können.

Für die Sprachfragen (Tab. 107) zeigen sich durch die Bank noch weniger und noch schwächere schichtspezifische Korrelationen als für die Kommunikationsfragen. Generell hängt eine starke Betonung der Bedeutung von Sprache mit guten aktiven Erklärungsleistungen des Kindes zusammen: In der Unterschicht für S 1 (Aufgaben des Alltags) und in der Mittelschicht für S 3 (Wissen lassen, was man fühlt) und S 6 (Spiele veranstalten). Diese Zusammenhänge sind - wie ein Blick auf Tab. 105 zeigt - unabhängig von passiver Wortkenntnis.

Wegen der starken Bedeutung der Ausbildung der Mutter für die aktiven Erklärungsleistungen des Kindes, haben wir die Korrelationen der Ausbildung der Mutter mit den Kommunikations- und Sprachfragen, die ja auch von der Mutter beantwortet wurden, überprüft. Es ergeben sich keine bemerkenswerten schichtspezifischen Korrelationen für die Fragen, die größere schichtspezifische Korrelationen mit den Experimentalvariablen aufweisen.

Insgesamt sind die Ergebnisse zum Einfluß der Kontroll- und Kommunikationsweisen der Mutter auf die Worterklärungen des Kindes enttäuschend: In der Gesamtgruppe ergaben sich überhaupt keine, in den beiden Schichtgruppen nur sehr sporadische Zusammenhänge. Insbesondere fanden sich keine Beziehungen zwischen den Kommunikations- und Kontrollweisen der Mutter und der Ausbildung der Mutter, so daß die Ergebnisse dieses Abschnittes nicht zu den vorangegangenen in Beziehung gesetzt werden konnten.

5.8. Zusammenhänge zwischen Intelligenzleistungen, Schulnoten und aktiver Worterklärung.

Während bis jetzt die Bedeutung sozialer Faktoren für die Erklärungsleistungen des Kindes dargestellt wurde, werden im folgenden die Einflüsse individueller Eigenschaften der Vpn. untersucht. In einem deskriptiven Sinn gehen wir von der Darstellung "soziologischer" zu der "psychologischer" Variablen über und wollen entsprechende Hypothesen vergleichen. Diese Gegenüberstellung ergibt sich durch die Anordnung der Variablen in bezug auf eine punktuelle experimentelle Messung; sie soll nicht die Vielfalt von Interaktionen leugnen, die in einer theoretischen Analyse bedacht werden müssen.

Wir werden die Zusammenhänge zwischen Intelligenzleistungen bzw. Schulleistungen und aktiver Worterklärung zuerst anhand einfacher Korrelationen darstellen; dann anhand der multiplen Regressionen jeder der beiden Prädiktorengruppen auf die aktive Worterklärung. Aus den Ergebnissen dieser beiden Berechnungen läßt sich über die relative Bedeutung "psychologischer" und "soziologischer" Einflüsse noch nichts entnehmen; deshalb werden wir in einem weiteren Schritt multiple Regressionen mit der jeweiligen "psychologischen" Variablen-Gruppe und den vier Schichtindikatoren als Prädiktoren analysieren. Mit ihnen können die spezifischen Voraussagen aller Variablen der beiden Gruppen verglichen werden.

Auch die folgenden sind zusätzliche Analysen, die das Hauptergebnis modifizieren oder bekräftigen sollen; zum besseren Verständnis wollen wir sie kurz mit den anderen Zusatzanalysen vergleichen. Die einfachen und multiplen Regressionen beziehen sich nicht mehr - wie die meisten der Indikatoren der sozialen Lage in Abschnitt 5.6. - auf Untergruppen, die das ursprüngliche Experimental-Design durchbrechen, sondern auf die gesamte Experimentalgruppe. Wie für die Analysen der Einzelleistungen in Abschnitt 5.4. gilt daher, daß durch die Parallelisierung der Schichtgruppen ein Großteil des

Schichteinflusses schon ausgeschaltet ist: Aus geringen Korrelationen "soziologischer" und gleichzeitigen hohen Korrelationen "psychologischer" Variablen kann deshalb nicht auf die Irrelevanz sozialer Faktoren für die Experimentalleistung geschlossen werden.

5.8.1. Zusammenhänge zwischen Intelligenzleistungen und der aktiven Worterklärung.

Von der Ähnlichkeit der Meßinstrumente sind die höchsten Korrelationen mit der aktiven Worterklärung für die Kontrollvariable, Wortschatztest erste Messung, zu erwarten, geringere für die Intelligenzgesamtmaße (VERBIT, GESINT), wiederum geringere für die verbalen Einzel-Intelligenztests (TEST 1/2, TEST 5, TEST 6) und die geringsten für die nichtverbalen Einzeltests und die nichtverbale Gesamtintelligenz. Für die Gesamtgruppe zeigt sich folgendes Ergebnis (Tab. 108). Der Wortschatztest hängt mit allen Maßen der zweiten Messung etwa gleich stark zusammen, mit der Häufigkeit der Antworten jedoch merklich geringer. Da im ersten Wortschatztest eine Auswahl zwischen Vorgaben zu treffen war, das Raten über die eigene Kenntnis hinaus also, wenn überhaupt, belohnt wurde, hätte man eher erwartet, daß die Häufigkeit der Antworten in der zweiten Messung besonders stark mit dem ersten Wortschatztest korreliert.¹⁾ Wir hatten jedoch schon in Abschnitt 5.3.5. Anlaß zu der Vermutung, daß - paradoxerweise - die beiden Testtechniken - multiple-choice und freie Erklärung - eher die Wortkenntnis oder -unkenntnis unterschiedlich diagnostizieren als die Differenziertheit eines Wortverständnisses. - Wider Erwarten auch korreliert die aktive Worterklärung mit keinem der verbalen Einzeltests durchgängig signifikant; für die nichtverbalen Testteile des PSB fanden sich nicht einmal vereinzelte Korrelationen.

¹⁾-----
Unsere Korrelationen zwischen dem aktiven und passiven Wortschatz in einer etwa 10jährigen Bevölkerung liegen, mit ungefähr .40, unter den Korrelationen zwischen aktivem und passivem Wortschatz bei Hurlburt (1950) in einer 15jährigen (.62) und 17jährigen (.60) Population.

Innerhalb der beiden Schichtgruppen differenziert sich das Gesamtbild nur wenig (Tab. 109). Die Korrelationen des ersten Wortschatztests mit den Maßen der zweiten Messung sind deutlich geringer in der Unterschichtgruppe; das gilt vor allem für die Erklärungshäufigkeit. In der Mittelschichtgruppe tauchen ferner relativ hohe Korrelationen aller Maße der zweiten Messung, mit Ausnahme der Häufigkeit mit TEST 1/2 (allgemeine Wortkenntnis) auf, die stark genug sind, um noch bei der verbalen Gesamtintelligenz (VERBIT) höhere Korrelationen durchzusetzen. Insgesamt ergibt sich für die Annahme Oevermanns (1970), daß gerade die intelligenten Unterschichtkinder sich von elaborierten Arten des Sprachgebrauchs distanzieren, kein Anhaltspunkt.

Die Multiple Regression der Intelligenztests auf die einzelnen Maße der zweiten Messung (Tab. 110) soll wiederum - wie schon bei den Schichtindikatoren - in erster Linie über das Gewicht der einzelnen Tests unabhängig von den anderen und erst in zweiter Linie über die Bedeutung des gemeinsamen Einflusses der Einzeltests Auskunft geben. Das Problem der Multikollinearität der Prädiktoren taucht hier nicht auf: Die Interkorrelationen der Intelligenzsubtests sind sehr niedrig (Tab. 22).

Die 9 Intelligenzsubtests und der Wortschatztest, erste Messung, beeinflussen signifikant oder fast signifikant alle Maße der Erklärungshäufigkeit und -güte, von den Maßen der Erklärungsgüte vorwiegend die schriftlichen, keines von den Maßen der Erklärungshäufigkeit. Wie der gemeinsame Einfluß der Schichtindikatoren auf diejenigen Maße am stärksten war, die mit der Einzelvariablen "Ausbildung der Mutter" am höchsten korrelierten, variiert hier der gemeinsame Einfluß der Intelligenzsubtests auf die verschiedenen Maße der zweiten Messung mit dem einfachen Einfluß des Wortschatztests (Tab. 108). Wie bei den Schichtindikatoren wollen wir auch hier überprüfen, ob die multiple Regressionsanalyse zusätzliche positive oder negative Einflüsse der übrigen Intelligenzsubtests bei konstanter Wortschatzleistung aufdeckt.

Die Beta-Gewichte sind - was nach den Ergebnissen der einfachen Korrelation kaum überrascht - bei allen Maßen für den Wortschatztest am größten. Auffällig ist, daß als

nächststärkster Subtest der nicht-verbale TEST 3 (logisches Schließen anhand von Figuren) einen positiven spezifischen Einfluß auf die Maße der Erklärungsgüte und -häufigkeit ausübt. Durchgängig negativ, wenn auch geringen spezifischen Einfluß hat dagegen der verbale TEST 6 (Verstümmelte Wörter), der jedoch auch von sprachlichen Fähigkeiten relativ unabhängige Leistungen der Wahrnehmung prüft. Die rein verbalen Subtests 1/2 und 5 haben nur sehr geringe oder in den Kommunikationsformen entgegengesetzte Voraussagekraft. Allgemein läßt sich den Beta-Gewichten der 9 Intelligenztests nur schwer ein sinnvolles Muster entnehmen.

Wenn man die Interkorrelationen der Intelligenzsubtests (Tab. 22) heranzieht, so zeigt sich, daß die verbalen Subtests 1/2 und 5 wegen ihrer hohen Korrelationen mit dem Wortschatztest - bei gegebenem hohen Einfluß des Wortschatztests und geringem der verbalen Subtests auf die zweite Messung - keine positive spezifische Voraussagekraft mehr haben konnten. Die Korrelation des Subtest 6 mit dem Wortschatztest ist jedoch ähnlich hoch, so daß wir ihm die Funktion einer "Suppressor"-Variablen zusprechen müssen. Die Interkorrelationen der übrigen Intelligenzsubtests sind ähnlich unregelmäßig und gering wie ihre Beta-Gewichte in der multiplen Regressionsanalyse.

Insgesamt wird durch die multiple Regression der Intelligenzsubtests und des Wortschatztests die Varianz der Maße der zweiten Messung, die schon durch den Wortschatztest allein erklärt werden konnte, zwar noch einmal vergrößert - die r^2 (Tab. 108) schwanken zwischen 20 und 25%, die R^2 (Tab. 109) zwischen 25 und 30% -, jedoch nur in einer unsystematischen, nicht auf einzelne Subtests zurückführbare Weise. Wir resümmieren, daß die multiple Regressionsanalyse keine über die einfachen Korrelationen hinausgehenden Erkenntnisse erbracht hat. Beide Analysen bestätigen sowohl die Irrelevanz der einzelnen Subtests wie die Bedeutung des Wortschatztests, erste Messung, für die zweite Messung.

Die wider Erwarten geringen Korrelationen der verbalen Testteile des PSB mit den Maßen der zweiten Messung und ihre geringe spezifische Voraussagekraft für die zweite Messung muß abschließend festgehalten werden. Aufgrund der Tatsache, daß der Wortschatztest, erste Messung, als Kontrollvariable für die Bildung paralleler Gruppen benutzt wurde, können seine hohen Korrelationen mit der zweiten Messung nicht erklärt werden, sowenig wie die niedrigen Korrelationen der übrigen verbalen Tests mit der zweiten Messung. Wir erinnern daran, daß in der Experimentalgruppe der zweiten Messung die verbalen Testleistungen genauso hoch wie in der ursprünglichen Gesamtstichprobe (N=357) mit dem WST korrelierten (siehe 2.1., Tab. 21). An dieser Stelle wird deutlich, daß von allen möglichen Kontrollen der passiven Wortkenntnis der Wortschatztest bei weitem die schärfste Prüfung von Schichtunterschieden der aktiven Worterklärung ermöglicht hat. Es ist offen, wie der Schichtvergleich ausgefallen wäre, wenn die Vpn. beider Schichten nach der gleichen Leistung etwa auf der verbalen Gesamtintelligenz oder auf TEST 1/2 ausgewählt worden wären. Allerdings verweisen die Wilcoxon-Tests mit VERBIT und TEST 1/2 als Kontrolle darauf, daß - zumindest im gegebenen Sample - der Wechsel der Kontrollvariablen keine anderen Ergebnisse gebracht hätten (siehe 5.2.2., Tab. 49).

5.8.2. Zusammenhänge zwischen Schulnoten und der aktiven Worterklärung.

Zur Charakterisierung der Schulnoten muß zweierlei vorausgeschickt werden. Erstens sind die Interkorrelationen der Schulnoten in der Gesamtstichprobe wie in unserer Untersuchungsgruppe sehr hoch (siehe Tab. 22); eine Faktorenanalyse der Schulnoten in der Gesamtgruppe ergab, daß ein erster Faktor, der auf den vier "Kernfächern" Deutsch-Mündlich, Deutsch-Schriftlich, Heimatkunde und Rechnen hoch lädt, schon etwa 42% der Varianz aller Schulnoten abbaut (siehe Bosse u.a. 1971:96). Wo ferner die Schulnoten zwischen sozialen Schichten diskriminieren, tun sie es auf die gleiche Weise. Offensichtlich unterliegt die Notengebung der Lehrer einem "Halo-Effekt": Sie ruht auf einem allgemeinen Eindruck, der auf alle Fächer ausstrahlt.

Zweitens ist die Notengebung signifikant durch die Schichtzugehörigkeit des Schülers beeinflusst, selbst wenn man die gemessene Intelligenz des Schülers statistisch kontrolliert (siehe Bosse u.a. 1971:170). Die Notengebung der Lehrer ist z.T. durch intelligenzunabhängige Verhaltensmerkmale des Schülers bestimmt. Unter ihnen wird, da die Mitarbeit im Unterricht neben schriftlichen Arbeiten Grundlage der Benotung in allen Fächern ist, der Stil der verbalen Kommunikation einen wichtigen Platz einnehmen. Für die Leistungen des mündlichen Schulunterrichts kann die Aufgabe der zweiten Messung unseres Experiments: die Aufforderung, ein Stichwort zu erklären, geradezu ein Modell abgeben. Nach diesen Überlegungen können wir recht starke Korrelationen, zwischen den Schulnoten - und zwar allen - und den aktiven Worterklärungen erwarten.

Die Schulnoten wurden innerhalb jeder der 11 in der Gesamtstichprobe erfaßten Schulklassen z-transformiert, um unterschiedliche Basen und Differenzierungsweisen der Lehrer auszuschalten. Alle Klassen haben auf diese Weise den gleichen Mittelwert und die gleiche Streuung. Wir beschränken die Interpretation auf die vier Noten der "Kernfächer" und die zwei Verhaltensnoten Fleiß und Aufmerksamkeit. Für die übrigen Fächer (Religion, Kunst, Musik, Werken-Nadeln, Leibeserziehung) ergaben sich keine Korrelationen, so daß sie in den Tab. 111 und 112 nicht aufgeführt werden.

In der gesamten Experimentalstichprobe (Tab. 111) korrelieren fast alle Experimentalleistungen der zweiten Messung signifikant mit den Schulnoten in Aufmerksamkeit, Deutsch-Mündlich und Heimatkunde, nicht jedoch mit Fleiß, Deutsch-Schriftlich und Rechnen; mit der Rechen-Note hängt die Anwohnhäufigkeit in der mündlichen Messung sogar negativ, wenn auch nicht signifikant, zusammen. Die Vermutung, daß alle Leistungs- und Verhaltensnoten ähnlich stark mit der aktiven Worterklärung korrelieren, hat sich also nicht bestätigt. Zudem können die unterschiedlichen Korrelationen der Schulnoten mit der aktiven Worterklärung nicht auf Unterschiede zwischen den Korrelationen mit der Kontrollvariablen zurückgeführt werden (siehe Tab. 21). Stattdessen müssen wir nachträglich nach einer Dimension suchen, auf der sich die Wichtigkeit oder Unwichtigkeit bestimmter Schulnoten für die aktive Worterklärung verstehen läßt.

Die Schulnote für Aufmerksamkeit prämiert offenbar doch eine gewisse Konzentration und Selbstkontrolle, dafür sprechen sowohl die höheren Korrelationen in der schriftlichen Messung als auch das Fehlen von Korrelationen mit der reinen Anorthäufigkeit. Die Schulnote für Heimatkunde bewertet offenbar auch allgemeine Sachkenntnis; auf diese Weise jedenfalls ließe sich die Korrelation mit der aktiven Worterklärung verstehen. Die Korrelation der Schulnote für Deutsch-Mündlich wäre selbstverständlich, wenn nicht gleichzeitig für die Note in Deutsch-Schriftlich überhaupt keine Korrelationen aufträten. Im Hinblick auf ihre Bedeutung für die aktive Worterklärung müssen wir die Schulfächer offenbar danach trennen, ob sich nach außen darstellbare, sozial sichtbare und unmittelbar interesseabhängige oder nur individuell realisierbare, mit einer gewissen Abstraktion und Neugier-Frustration verbundene Leistungen verlangen. Der Gegensatz von Deutsch-Mündlich und Deutsch-Schriftlich ist auf dieser Linie verständlich, ebenso der Heimatkunde und Rechnen, ja sogar noch der von Aufmerksamkeit und Fleiß.

Betrachtet man die beiden Schichtgruppen getrennt (Tab. 112), so fällt zunächst auf, daß in der Mittelschicht die Korrelationen höher und häufiger sind, während in der Unterschicht fast nur noch Korrelationen mit den schriftlichen Maßen der zweiten Messung auftreten. Mit dieser Einschränkung finden sich in beiden Schichtgruppen die Zusammenhänge für die Note in Aufmerksamkeit, Deutsch-Mündlich und Heimatkunde mit den Maßen der zweiten Messung wieder; ebenso die negative Korrelation der Rechen-Note mit den mündlichen Maßen, jetzt sogar bei mehreren Maß-Gruppen.

In der Mittelschicht jedoch tauchen zusätzlich Zusammenhänge mit der Note für Deutsch-Schriftlich auf, für die in der Unterschicht sich aber nach wie vor keine Korrelationen ergeben. Selbst für die Fleiß-Note finden wir in der Mittelschicht geringe, aber in beiden Kommunikationsformen konsistente Korrelationen. Auch dieses Muster kann sich nicht aus unterschiedlichen Korrelationen der Schulnoten mit der Kontrollvariablen ergeben haben (siehe Tab. 21) - am wenigsten sein wichtigstes Teilstück: Die schichtspezifische Bedeutung der Note Deutsch-Schriftlich.

Eine Charakterisierung schichtspezifischer Ergebnisse kann so zusammengefaßt werden, daß in der Mittelschicht der vermutete durchgängige Zusammenhang aller Schulnoten - von der Rechen-Note abgesehen, die in beiden Schichten nur negativ mit der zweiten Messung korreliert - mit den Maßen der zweiten Messung gegeben ist, während in der Unterschicht - wie in der Gesamtgruppe, jedoch eingeschränkt auf die schriftliche Messung - nur diejenigen Schulnoten mit der aktiven Worterklärung zusammenhängen, die mutmaßlich spontanes Interesse und soziale Beteiligung honorieren. Die schichtspezifischen Ergebnisse verwischen also den für die Gesamtgruppe gewonnenen Eindruck.

Da sich mit den beiden Schulnoten für Deutsch, den beiden Kommunikationsformen unseres Experiments, - zumindest oberflächlich - vergleichbare Indikatoren verbaler Leistungen anbahnten, haben wir für sie die Korrelationen mit allen einzelnen Maßen der zweiten Messung in den beiden Schicht- und der Gesamtgruppe noch einmal gesondert dargestellt (Tab. 113). Es wird noch einmal sichtbar, daß für Deutsch-Mündlich gleichsam ein Interaktionseffekt zwischen Schicht und Kommunikationsform der zweiten Messung besteht, während Deutsch-Schriftlich überhaupt nur in der Mittelschicht mit den Experimentalleistungen zusammenhängt. Wie in Tab. 112 wird ferner sichtbar, daß die Zusammenhänge zwischen Deutsch-Mündlich und beiden Kommunikationsformen der zweiten Messung in der Mittelschicht stark auf den Maßen der Häufigkeit (GATEX, RATEX) beruhen, der gleiche Zusammenhang für die schriftliche Kommunikationsform in der Unterschicht aber vor allem die Maße der Explikationsgüte betrifft. In der Mittelschicht korrelieren alle "Kommunikationsformen" der Noten mit allen Kommunikationsformen des Experiments; in der Unterschicht allein mündliche Note und schriftliche Experimentalleistung.

Auch mit den Schulnoten als Prädiktoren wurden multiple Regressionen auf die Maße der zweiten Messung gerechnet; es wurden alle 11 erhobenen Schulnoten verwandt (Tab. 114). Wie schon zu Beginn dieses Abschnittes betont, sind die Interkorrelationen der Schulnoten sehr hoch, auch in unserer Experimentalgruppe (Tab. 22). Dem Fall der

Schichtindikatoren analoge Überlegungen zum Problem der Multikollinearität können hier jedoch nur schwer eingewandt werden, da Schulnoten auf die Differenzierung eines Leistungsprofils angelegt sind. Ferner ist die Struktur der Beta-Gewichte der Prädiktoren nicht so stabil über die Kommunikationsformen wie bei den Schichtindikatoren; deshalb werden wir uns auf Prädiktoren mit ähnlichen Beta-Gewichten in beiden Kommunikationsformen konzentrieren.

In der mündlichen Messung und bei den Maßen der Erklärungsgüte ergaben sich keinerlei signifikante multiple Regressionskoeffizienten; nur bei den dargestellten (und nicht bei einer Reihe noch geprüften) Maßen der Erklärungshäufigkeit und -güte für die schriftliche Messung finden sich annähernd signifikante R^2 . In recht scharfem Gegensatz zu den Intelligenztests können die Schulnoten also die Leistungen der zweiten Messung nicht signifikant voraussagen.

Die geringe Signifikanz der multiplen Regressionsanalysen mit den Schulnoten als Kriterien ist ein weiteres Argument, die Beta-Gewichte der einzelnen Prädiktoren vorsichtig zu betrachten. Immerhin zeigen - analog den einfachen Korrelationen - Deutsch-Mündlich und Heimatkunde positive, Deutsch-Schriftlich nur geringe und oft zwischen den Kommunikationsformen wechselnde und Rechnen schließlich durchweg negative Voraussagekraft. Die übrigen Beta-Gewichte sind sehr gering oder nicht konsistent. Im Vergleich zu den einfachen Korrelationen ist das interessanteste Ergebnis der multiplen Regression die negative Bedeutung der Rechen-Note bei Konstanz der übrigen Noten für die Leistungen der zweiten Messung; um so interessanter, als die Schulnoten sehr stark interkorrelieren und mit den Maßen der zweiten Messung im allgemeinen positiv zusammenhängen.

Da im nächsten Abschnitt die relative Bedeutung der einzelnen Tests und Schulnoten für die aktive Worterklärung anhand einer multiplen Regressionsanalyse untersucht werden soll, wollen wir zuvor die Ergebnisse für die Einzelbetrachtung der beiden Prädiktorengruppen zusammenfassen. In der Gesamt-, der Unterschicht- und der Mittelschichtgruppe finden sich hohe Korrelationen zwischen der Kontroll-

variablen, Wortschatztest erste Messung, und allen Maßen der zweiten Messung. Die Korrelationen der verbalen Intelligenztests sind dagegen viel geringer, als man von ihrer theoretischen Verwandtschaft mit dem Wortschatz her erwartet hätte. Auch in der multiplen Regression zeigt der Wortschatz die höchste spezifische Voraussagekraft. Von den Schulnoten korrelieren in der Unterschicht nur Aufmerksamkeit, Deutsch-Mündlich und Heimatkunde lediglich mit den schriftlichen Maßen der zweiten Messung; in der Mittelschicht dagegen alle Noten mit beiden Kommunikationsformen. Die Rechen-Note dagegen hängt in beiden Schichten negativ mit der Experimentalleistung zusammen und zeigt in der multiplen Regression als einzige negative Beta-Gewichte. Die multiple Regression der Tests auf die Maße der zweiten Messung zeigt durchweg signifikante, die der Schulnoten kaum signifikante R^2 . Von ihren Anwendungsbedingungen - keine Multikollinearität - wie von ihren Ergebnissen her - signifikante R^2 - kann die multiple Regressionsanalyse im Falle der Tests sehr wohl, im Fall der Schulnoten nur bedingt über das relative Gewicht der Prädiktoren Auskunft geben.

5.8.3. Die relative Bedeutung der Intelligenztests und der Schulnoten für die aktive Worterklärung.

Wir hatten gesehen, daß die Intelligenztests die Leistungen der zweiten Messung durchweg signifikant voraussagen können, nicht aber die Schulnoten. Andererseits ergaben sich die signifikanten multiplen Regressionen der Tests vorwiegend aus dem Wortschatztest und - durchgängig für verschiedene Maße und Kommunikationsformen - aus keinem anderen Subtest; während bei den Schulnoten zumindest Deutsch-Mündlich, Heimatkunde und Rechnen durchgängig stärkere Beta-Gewichte aufwiesen. Die multiple Regressionsanalyse mit den Tests und den Noten als Prädiktoren soll zeigen, ob die überragende Bedeutung des Wortschatztests bestehen bleibt, wenn die Noten kovariieren, und ob die relativ hohe spezifische Voraussagekraft einzelner Noten bestehen bleibt, wenn die Tests kovariieren. Die Gefahr der Multikol-

Linearität ist hier nicht stärker als bei den beiden entsprechenden vorangegangenen Analysen gegeben, da Tests und Noten in unserer Experimentalgruppe nur sehr schwach korrelieren (siehe Tab. 22).¹⁾

Wie aus Tab. 115 ersichtlich, behält der Wortschatztest seine Vorrangstellung auch, wenn Tests und Noten als Prädiktoren kovariieren. Allerdings bleiben, wenn auch schwächer und teilweise mit umgekehrten Vorzeichen durchsetzt, der relativ starke positive Einfluß von Deutsch-Mündlich und Heimatkunde sowie der negative Einfluß der Rechen-Note ebenfalls bestehen. Das Beta-Gewicht für Deutsch-Mündlich erreicht bei den Maßen S13EXS und MIBVEM nahezu das des Wortschatztests. Was wir erwartet hätten, daß die Beta-Gewichte der Schulnoten bei Kovariation der Intelligenztests "durcheinanderpurzelten", ist nicht oder nur in sehr geringem Maße aufgetreten. Als Resümee bleibt festzustellen, daß die multiple Regressionsanalyse mit Tests und Noten als Prädiktoren keine über die Ergebnisse der beiden einzelnen multiplen Regressionsanalysen hinausgehende Erkenntnisse gebracht hat, insbesondere keine Bedeutungslosigkeit des Wortschatztests bei Kovariation der Noten oder der Noten bei Kovariation der Tests.

1)-----
Vorausgeschickt sei, daß die multiple Regressionsanalyse die Zusammenhänge in den gegebenen Daten maximiert. Für den Vergleich zwischen Gruppen von Prädiktoren innerhalb der gleichen Daten ist dies jedoch unerheblich. Außerdem suchen wir nicht aus einer Vielzahl von Variablen diejenige Variablen-Kombination heraus, die am meisten Varianz voraussagen im Vergleich mit allen anderen möglichen Variablen-Kombinationen - wodurch die Gefahr des Herausstreichens zufälliger Zusammenhänge maximiert wird -, sondern wir vergleichen den Einfluß zuvor festgelegter Kombinationen von Prädiktoren.

5.8.4. Die relative Bedeutung von Intelligenztests bzw. Schulnoten und Schichtindikatoren für die aktive Worterklärung.

Wir haben bis jetzt den Einfluß individueller Leistungen und Fähigkeiten, gemessen in Intelligenztests und Schulnoten, auf die aktive Worterklärung dargestellt. Im folgenden soll anhand von multiplen Regressionsanalysen untersucht werden, ob diese beiden Gruppen von Prädiktoren, einzeln oder gemeinsam, ihre Voraussagekraft behalten, wenn die Merkmale der sozialen Schicht kovariieren; anders gesagt: Welche Art von Prädiktoren das größte Gewicht haben, wenn individuelle Leistungen und soziale Schichtzugehörigkeit als Prädiktoren kovariieren. Wir werden zuerst Tests und Schichtindikatoren, dann Noten und Schichtindikatoren und schließlich Tests, Noten und Schichtindikatoren vergleichen.

In der multiplen Regression der vier Schichtindikatoren auf die Maße der zweiten Messung hatten sich signifikante R^2 für die Maße der Erklärungshäufigkeit und -güte wie der Erklärungshäufigkeit ergeben; durchweg hatte die Ausbildung der Mutter positive, die Ausbildung des Vaters und das Einkommen negative spezifische Voraussagekraft (Tab. 95). In der multiplen Regression der zehn Intelligenzmaße auf die Maße der zweiten Messung ergaben sich ebenfalls für die Maße der Erklärungshäufigkeit und -güte signifikante - wegen der größeren Anzahl von Prädiktoren natürlich auch größere - R^2 , sowie signifikante R^2 für die Erklärungsgüte, nicht aber für die Erklärungshäufigkeit; durchweg zeigte nur der passive Wortschatztest eine große spezifische Voraussagekraft (Tab. 110).

Nach den Ergebnissen der multiplen Regression für beide Prädiktorengruppen einzeln wird die multiple Regression beider Gruppen zusammen in der Hauptsache den Einfluß der Ausbildung der Mutter mit dem des passiven Wortschatztest vergleichen. Die Interkorrelationen der Prädiktorengruppen (Tab. 22) sind für die Schichtindikatoren relativ hoch, für die Intelligenztests gering; die Schichtindikatoren korrelieren - in unserer ausgelesenen Experimentalgruppe - mit den Intelligenztests so gut wie nicht.

Beide Gruppen von Prädiktoren zusammen (Tab. 116) sagen wiederum die Maße der Erklärungshäufigkeit und -güte signifikant voraus, während die Maße der Erklärungsgüte in der schriftlichen Messung - übereinstimmend mit den Ergebnissen der multiplen Regression mit den Prädiktoren "Intelligenzmaße" -, und die Maße der Erklärungshäufigkeit in der mündlichen Messung - übereinstimmend mit den Ergebnissen der multiplen Regression mit den Prädiktoren "Schichtindikatoren" - vorausgesagt werden. Die höchsten Beta-Gewichte haben der Wortschatztest und die Ausbildung der Mutter; wie in den beiden Regressionsanalysen mit den getrennten Prädiktorengruppen zeigen sich anhand der Beta-Gewichte für die neun Intelligenzsubtests nur geringe und wenig konsistente Einflüsse, für die übrigen Schichtindikatoren aber starke negative spezifische Einflüsse der Ausbildung des Vaters und des Einkommens. Erstens also wiederholt diese multiple Regressionsanalyse die Ergebnisse der beiden vorangegangenen. Zweitens aber belegt sie, daß von den 5 überhaupt wichtigen Prädiktoren - Wortschatztest, Beruf des Vaters, Ausbildung des Vaters, Ausbildung der Mutter, Einkommen - die Ausbildung der Mutter ein fast ebenso großes spezifisches Gewicht hat wie der Wortschatztest.

Die einfachen Korrelationen des Wortschatztests mit den Maßen der zweiten Messung (Tab. 108) liegen bei den Maßen der Erklärungsgüte mit .45 über denen der Ausbildung der Mutter (Tab. 92); bei den Maßen der Erklärungshäufigkeit unter denen der Ausbildung der Mutter mit .50; bei den Maßen der Häufigkeit und Güte liegen die Korrelationen beider Variablen mit den Maßen der zweiten Messung etwa gleich bei .47: Jede der beiden Variablen kann also allein etwa halb so viel Varianz abbauen wie die 14 Prädiktoren insgesamt. Ganz analog hat der Wortschatztest bei den Maßen der Erklärungsgüte, die Ausbildung der Mutter bei den Maßen der Erklärungshäufigkeit das größere Beta-Gewicht. Die Korrelationen zwischen Wortschatztest, erste Messung, und Ausbildung der Mutter ist .00 (Tab. 22). Wir können also zusammenfassen, daß der Wortschatztest und die Ausbildung der Mutter unabhängig voneinander die größte spezifische Einflußkraft auf die Maße der zweiten Messung haben, jeder Prädiktor

mit besonderen Akzenten bei einzelnen Maßen. Noch wichtiger für die bisherigen Ergebnisse aber ist die Tatsache, daß die Ausbildung der Mutter nahezu so eng wie der Wortschatz mit den Maßen der zweiten Messung zusammenhängt, und daß ihre spezifische Voraussagekraft, die wir ja schon in Abschnitt 5.6.2. herausgestellt hatten, auch bestehen bleibt, wenn sie mit den Intelligenzsubtests und dem Wortschatztest als Prädiktor gleichsam konkurrieren muß.

Anders als die Tests konnten die Schulnoten die Maße der zweiten Messung kaum signifikant voraussagen. Einige Schulnoten zeigten jedoch in den beiden Kommunikationsformen und bei verschiedenen Maßen durchgängig hohe Beta-Gewichte. Auch in der multiplen Regressionsanalyse mit den Schulnoten und den Schichtindikatoren als Prädiktoren (Tab. 117) behalten die Note in Heimatkunde ihr positives, die Note in Rechnen ihr negatives Beta-Gewicht durchgängig bei; die Note für Deutsch-Mündlich jedoch nur bei den mündlichen Maßen der Erklärungshäufigkeit und -güte und den Maßen der Erklärungsgüte. Von den Schichtindikatoren auf der anderen Seite zeigt die Ausbildung der Mutter wiederum ein hohes positives, das Einkommen ein hohes negatives Beta-Gewicht, während die Ausbildung des Vaters geringere und nicht mehr ganz durchgängige negative Beta-Gewichte aufweist. Im Vergleich ist das Beta-Gewicht der Ausbildung der Mutter bei fast allen Maßen das höchste, mit der Ausnahme von MIBEXM, wo die Schulnote für Deutsch-Mündlich an ihre Stelle tritt; die Beta-Gewichte für Deutsch-Mündlich bei den Maßen der Erklärungsgüte sind überhaupt denen der Ausbildung der Mutter vergleichbar. Keine der Schulnoten jedoch zeigt annähernd ein Beta-Gewicht, das mit dem des Wortschatztests in der multiplen Regression der Tests und Schichtindikatoren vergleichbar wäre. Von den Tests und Noten ist also der Wortschatztest der wichtigste einzelne Prädiktor; damit wird bestätigt, was sich schon in der multiplen Regression der Tests und Noten (Tab. 115) ergab und was sich auch in der multiplen Regression der Tests, Noten und Schichtindikatoren (Tab. 118) wiederfinden wird.

Die multiplen Regressionen der Tests, Noten und Schichtindikatoren (Tab. 118) zeigen für die drei Gruppen von Maßen der zweiten Messung vergleichsweise starke Unterschiede in der Struktur der Beta-Gewichte, so daß die Ergebnisse für jede Maßgruppe besprochen werden sollen. Bei den Maßen der Erklärungshäufigkeit und -güte haben der Wortschatztest und die Ausbildung der Mutter die höchsten, etwa gleichen Beta-Gewichte - wie in den anderen Regressionsanalysen mit positiven Vorzeichen. Von den neun Intelligenzsubtests hat keiner herausragende Bedeutung - wie schon in der multiplen Regression der Tests allein (Tab. 110). Von den Schulnoten zeigen - wie in Tab. 114 - Heimatkunde positive und Rechnen negative Bedeutung. Von den Schichtindikatoren behält - wie in Tab. 95 - das Einkommen seine negative Bedeutung, während das negative Beta-Gewicht der Ausbildung des Vaters sehr gering wird. Bei den Maßen der Erklärungsgüte konkurriert der Wortschatztest nicht mehr mit der Ausbildung der Mutter, sondern mit der Schulnote für Deutsch-Mündlich als wichtigstem Prädiktor. Von den Tests gewinnt - im Vergleich zu Tab. 110 - der TEST 5 (Worteinfall) ein auffällig starkes negatives, sowie der nichtverbale Subtest 7 (Faltaufgaben) ein auffällig starkes positives Beta-Gewicht. Von den Schulnoten verlieren - im Vergleich zu Tab. 114 - die übrigen Kernfächer gegenüber Deutsch-Mündlich an Bedeutung. Von den Schichtindikatoren bleiben - im Vergleich zu Tab. 95 - die Vorzeichen der Beta-Gewichte erhalten, wenn auch ihre absolute Größe beträchtlich schwankt. Bei den Maßen der Erklärungshäufigkeit zeigt die Ausbildung der Mutter wiederum das höchste Beta-Gewicht, gefolgt von der Schulnote in Heimatkunde und erst dann vom Wortschatztest.

Was läßt sich aus den multiplen Regressionsanalysen in ihrer Gesamtheit entnehmen?¹⁾ Von den jeweiligen Faktor-Gruppen konnten die Schichtindikatoren und die Tests, nicht aber die Schulnoten die

1) -----
Für die beiden wichtigsten Kriterien MIAEX und MIBEX sind die multiplen Regressionen mit allen möglichen Prädiktor-Kombinationen in Tab. 119 und 120 synoptisch dargestellt worden.

Leistung der zweiten Messung signifikant voraussagen. Während aber von den Tests allein der Wortschatztest den signifikanten Zusammenhang bewirkt, waren es bei den Schichtindikatoren die Ausbildung der Mutter zusammen mit dem negativen spezifischen Einfluß der Ausbildung des Vaters und des Einkommens. Die größte spezifische Voraussagekraft in allen Analysen aber zeigten der Wortschatztest und die Ausbildung der Mutter. Die Bedeutung dieser beiden Prädiktoren soll im Zusammenhang mit unserem negativen Hauptergebnis der zweiten Messung diskutiert werden.

Erstens hat mit dem passiven Wortschatztest eine Variable auf die aktive Worterklärung den stärksten Einfluß, die selber signifikant mit sozialer Schicht zusammenhängt. Für die einwandfreie Prüfung von Hypothesen über den schichtspezifischen Gebrauch von Sprache war eine solche Experimentalanlage jedoch notwendig, so daß das negative Hauptergebnis nicht modifiziert werden kann. Zweitens kann, obwohl sich bei Kontrolle des passiven Wortschatzes in der aktiven Worterklärung keine Unterschiede zwischen den Schichten ergaben, durch eine einzelne Dimension der Schichtzugehörigkeit, die Ausbildung der Mutter und mit Hilfe der, jeweils unabhängig von den übrigen, gegenläufigen Einflüssen aller 4 Schichtdimensionen die aktive Worterklärung vorausgesagt werden. Dies Ergebnis impliziert eine Modifikation des mit Hilfe des Filterindex sozialer Schicht gewonnenen Hauptergebnisses: In ihm sind zum Teil widerstreitende Einflüsse einzelner Dimensionen sozialer Schicht nivelliert. Selbst bei einer starken Neutralisierung sozialer Einflüsse durch eine Kontrolle des passiven Wortschatzes, ergibt sich eine Beeinflussung der aktiven Worterklärung durch die im Globalindikator verwischten typischen Kombinationen einzelner Dimensionen sozialer Schicht, die eine signifikante Voraussage erlauben (siehe auch 5.6.6.).

6. Ergebnisse der zweiten Messung: Sprachliche Mittel der
=====
Erklärung.
=====

6.1. Ziele der Analyse der sprachlichen Mittel.

Bernsteins Theorie der linguistischen Kodes ist oft vorgeworfen worden, sie gebe lediglich eine wissenschaftliche Legitimation für Vorurteile und Stereotype der Mittelschicht. Die Konfrontation des "restringierten" Kodes mit dem "elaborierten" wiederhole herrschende Sprachnormen und sei eine wissenschaftlich schlecht getarnte Bewertung; die Verknüpfung der Kodes mit der Unterschicht bzw. der Mittelschicht bestätigte überhebliche Selbsteinschätzungen sowie so privilegierter Gruppen, die Verknüpfung der Kodes mit Lernerfolgen vulgäre Vorstellungen über den Zusammenhang von Sprache und Denken. Außerdem liefen die aus einer solchen Theorie gezogenen Konsequenzen auf eine Anpassung der Unterschichtkinder an die Mittelschichtkultur hinaus, in der sie genuine Interessen zu verleugnen lernten. Spezifischer läßt sich die Kritik an der Theorie der linguistischen Kodes als der Versuch verstehen, zwei verschwiegene Ziele und deren unbedachte Folgen ideologiekritisch herauszuarbeiten: Die Theorie der linguistischen Kodes diene einer Rechtfertigung mittelschichtspezifischer Sprachnormen, so daß vorschulische Programme auf eine Einübung in die Mittelschichtsprache hinauslaufen müßten, und sie fingiere die Abhängigkeit psychischer Kompetenzen von der Sprache, so daß eine Sprachförderung mit einer Intelligenzförderung gleichgesetzt werden könne.

Für die Diskussion dieser Einwände wollen wir drei Ebenen unterscheiden, auf denen Funktionen von Sprache angesiedelt sein können. Auf der ersten Ebene dient Sprache der Darstellung von Gruppenzugehörigkeiten, der impliziten Anerkennung von Normen; Zugehörigkeiten und Normen können situationsspezifisch und gruppenspezifisch sein. Sprache ist hier ein Zeichen für etwas - von der betreffenden

Sprachform unabhängig darstellbares - Andres und grundsätzlich durch andere Medien, Kleidung oder Gestik, ersetzbar. Auf der zweiten Ebene läßt sich Sprache in Beziehung setzen zu Fertigkeiten sozialen Handelns und Prozessen sozialer Interaktion. Die Definition sozialer Situationen, die Klärung eigener und das Verstehen fremder Absichten setzt - weder immer noch ausschließlich, aber vorwiegend - einen kontrollierten Gebrauch der Sprache voraus. Auf der dritten Ebene schließlich kann die Sprache als Bedingung individueller psychischer, vor allem kognitiver Kompetenzen angesehen werden, so wie etwa die Whorf-Hypothese eine innerpsychische Kausalbeziehung zwischen Sprache und Denken annimmt. Nur auf der zweiten Ebene ist die für die Theorie der linguistischen Kodes grundlegende Trennung zwischen Sprache und Arten des Sprachgebrauchs möglich. Die Kritik ignoriert jedoch diese Trennung und faßt die Theorie der linguistischen Kodes als Theorie schichtspezifischer Sprachnormen und als Theorie über den Zusammenhang von Sprechen und Denken auf. In beiden Hinsichten verfehlt sie die Ebene des sozialen Handelns, in die Arten des Sprachgebrauchs eingebettet sind. Diese Behauptung bedarf einer längeren Begründung.

Auf den beiden Ebenen, auf die die Kritik die linguistischen Kodes abdrängt, kann sie nur als objektive, vom aktuellen sozialen Handeln grundsätzlich unabhängige Struktur angesehen werden - sei es im Sinn der hochsprachlichen Grammatik, situationstypischer Register oder gruppenspezifische Dialekte. Von diesen objektiven sprachlichen Strukturen müssen die Arten des Sprachgebrauchs getrennt werden, auch wenn sie meist - mit den wichtigen Ausnahmen von Verzögerungsphänomenen und fehlerhaftem Sprachgebrauch - durch Kategorien von "Grammatiken" operationalisiert wurden. "Sprachgebrauch" meint den von der Wahrnehmung sozialer Situationen und Rollen gesteuerten Einsatz von Sprache (oder anderer kommunikativer Medien) im sozialen Handeln. Durch Gewohnheiten der Handelnden wie Standardisierungen von Situationen und Rollen ist der Sprachgebrauch in der Regel festgelegt - genau aus diesem Grunde ist er ja soziologisch faßbar -, aber grundsätzlich kann er individuell reflektiert und intersubjektiv verhandelt werden.

Auf der ersten Ebene - Sprache als sozial normiertes Zeichen - läßt sich eine Trennung von objektiver Struktur und Anwendung nur schwer denken. Es gibt Zeichen mit sozial typischem Gehalt und den Prozeß ihrer Übernahme durch Ego und ihrer Anerkennung durch Alter; aber es gibt keinen sozial standardisierten Gebrauch jenseits vom sozial standardisierten Inhalt. Wohl ist die auf den Empfänger hin kalkulierte Manipulation jeder Art von Erkennungsmerkmalen - sei es Sprache, Kleidung oder Gestik - möglich; in diesem Fall aber kann der Empfänger die Absicht wie den Inhalt der Täuschung selber wieder prüfen, es wiederholt sich der soziale Prozeß auf einer nächsten Ebene, was Goffman (1970) die "fundamentale Assymetrie" in der alltäglichen Selbstdarstellung nennt.

Ebensowenig kann auf der dritten Ebene - Sprache als Bedingung psychischer Fähigkeiten - zwischen objektiver Struktur und Gebrauch unterschieden werden. Zunächst ist "Art des Sprachgebrauchs" als individualpsychologischer Begriff entweder sinnlos oder mit außersprachlichen Fähigkeiten wie Gedächtnis, Intelligenz, Wahrnehmungstempo gleichzusetzen. Darüber hinaus stellt sich schon bei der Untersuchung von Sprache und Denken - und damit um so mehr bei einem individualpsychologisch vorgestellten Kausalverhältnis von Sprachgebrauch und Denken - das Problem der Tautologie; empirisch unabhängig voneinander geprüft wurden die beiden Bereiche bis jetzt nur auf dem - im Vergleich zu den weitreichenden Implikationen der Whorf-Hypothese - trivialen Gebiet der Kodierung von Farben (siehe Lennberg 1967).

Die Kritik zielt nun in zwei Richtungen zugleich an der Bernsteinischen Theorie vorbei, die weder eine rein sozialpsychologische noch eine rein psychologische Theorie ist, weder eine Theorie sozialer Selbstdarstellung und sozialer Wahrnehmung noch eine Theorie über den Zusammenhang sprachlicher Strukturen und psychischer Kompetenzen. Sie ist dagegen eine soziologisch angesetzte Sozialisations-theorie, die den Prozessen nachgeht, wie im Laufe der Lebensgeschichte objektive Sozialstrukturen sich im individuellen Lebenserfolg durchsetzen; sie ist m.a.W. eine Theorie der "kulturellen"

und "sozialen Reproduktion" der Gesellschaft (Bourdieu/Passeruo 1973). Ein solcher Ansatz schließt natürlich die Fruchtbarkeit des psychologischen und sozialpsychologischen keineswegs aus. Über den innerpsychischen Zusammenhang zwischen Sprache und Denken bestehen ausgedehnte Forschungen - mit allerdings meist negativen Ergebnissen (Lenneberg 1967, Furth 1966, Piaget 1969). Auch die Bedeutung der Sprache als Kennzeichen sozialer Identität soll mit der Aussage, daß sie nicht der Gegenstandsbereich der Theorie linguistischer Kodes ist, keineswegs geleugnet werden. In psychologischen Theorien der Interaktion geht es jedoch um die Verwendung und Wahrnehmung von Zeichen, darunter bestimmten Sprachformen; in der Theorie Bernsteins um Strategien des Symbolgebrauchs und des sozialen Handelns.¹⁾

Selbst dann jedoch, wenn wir die Kritik an der Theorie der linguistischen Kodes theoretisch zurückweisen können, muß sie bei der empirischen Prüfung ernstgenommen werden. Zum einen muß gesichert sein, daß in die Auswertung sprachlichen Materials nicht schichtspezifische Vorurteile eingehen, daß wissenschaftliche nicht nach den gleichen Kriterien wie soziale Bewertungsprozesse verlaufen. Auch wenn die Möglichkeit schichtspezifischer Vorurteile in der Auswertung sich nie endgültig ausschließen läßt, da jeder Maßstab der die Vorurteilsfreiheit garantieren soll, selber wieder als vorurteilsvoll

7)-----
Dieser Unterschied läßt sich an einem Beispiel, das gegen die Theorie Bernsteins eingewandt wurde, stellvertretend für alle sozialpsychologischen Uminterpretationen verdeutlichen. Lawton (1970) erläutert, daß ein Fremdenführer im Vatikan eine hoch elaborierte Sprache spreche, die jedoch hoch voraussagbar sei. Da hohe Voraussagbarkeit eines der Kriterien für einen restringierten Kode sei, fände man restringierten und elaborierten Kode zugleich. Auch unabhängig vom Kriterium der Voraussagbarkeit läßt sich der Einwand so formulieren, daß eine ritualisierte, aber hoch gestochene Sprache der Trennung der beiden Kodes widerspräche. Dieser Widerspruch entsteht aber nur dadurch, daß die Ebene konkreter Sprachformen und -normen und die Ebene des Sprachgebrauchs vermengt werden. Ein Fremdenführer muß gebildet reden; selbst Formen der Rollen-Distanz, wie scherzhafte Einlagen, demonstratives Desinteresse etc. dienen dieser Norm. Sobald man aber sich auf die Ebene des Sprachgebrauchs beschränkt, wird deutlich, daß der Fremdenführer einen restringierten Kode gebraucht und keinen anderen. - Auf andere Mängel der Kritik Lawtons, wie die Verwechslung von situationsspezifischen und gruppenspezifischen Sprachnormen, Register und Kodes, soll hier nicht weiter eingegangen werden.

bezweifelt werden kann, so haben wir doch die Schichtneutralität des Testmaterials und der Auswertung hinreichend belegen können (siehe 1.2. und 5.4.3.). Zum anderen ergibt sich die Möglichkeit, sprachliches Material unter dem Gesichtspunkt seiner Funktion im sozialen Handeln und zugleich unter dem Gesichtspunkt möglicher sozialer Bewertungen seiner äußeren Form zu analysieren. Anders gesagt: Die Kodes Bernsteins in ihren genuinen Sinn als Arten des Sprachgebrauchs und in der Umdeutung als schichtspezifischer Sprachnormen an sich zu prüfen. Wir sehen die bisher betrachteten Ergebnisse der aktiven Worterklärung als eine solche genuine Überprüfung der Bernsteinschen Theorie an und haben in der folgenden "Linguistischen Auswertung" versucht, schichtspezifische Sprachformen für die Aufgabe der Worterklärung zu erfassen vor dem Hintergrund mutmaßlicher schichtspezifischer Bewertungen. Beide Analysen beziehen sich auf das gleiche Material und sind nur in den beschriebenen Aspekten unterschieden.

6.2. Stellung der Analyse sprachlicher Mittel der Erklärung im Aufbau des Experiments.

Die Analyse der sprachlichen Mittel der Erklärung war nicht von Beginn der Untersuchung geplant. Sie wurde begonnen, als die negativen Ergebnisse zur zweiten Messung deutlich wurden. Wir vermuteten, daß trotz fehlender Schichtunterschiede in der Bewertung objektiver Erklärungsleistungen Unterschiede im sprachlichen Stil der Erklärung und in der Anzahl von grammatischen und orthografischen Fehlern zwischen den beiden Schichten bestünden, die anstelle gültiger Vermutungen über die kategoriale Kenntnis des Erklärenden vom Zuhörer wahrgenommen würden und die sprachlichen Reaktionen des Zuhörers steuern würden. Wenn eine solche Hypothese nach dem negativen Hauptergebnis sich bestätigen ließe, so wäre nicht nur Bernsteins Theorie widerlegt, sondern auch die Ideologiekritik an ihr gerechtfertigt. Um diese Stoßrichtung der "linguistischen" Analyse zu erläutern, muß ihr Zusammenhang mit der Theorie hinter dem ursprünglichen Aufbau des Experiments dargestellt werden.

In der Theorie des Experiments wurden Nachteile der Unterschichtkinder in der kognitiven Entwicklung als kumulatives Ergebnis sozial typischer Arten des aktiven Sprachgebrauchs verstanden. Die Form sprachlicher Äußerungen des Senders führt unabhängig von seinen tatsächlichen Ausdrucksintentionen zu kognitiven Kategorisierungen beim Zuhörer, die wiederum die Reaktionen des Zuhörers und damit die Informationseingabe des Senders bestimmen. Selbst bei gleich gutem Verständnis kann in diesem sozialen Kommunikationsprozeß eine Strategie reduzierter Verbalisierung bei Kindern aus der Unterschicht zu entsprechend eingeschränkten Reaktionen der Zuhörer und, kumulativ im Laufe der Entwicklung, zu Benachteiligungen im Lernen führen. Unabhängig von möglichen innerpsychischen, vom passiven Wortschatz abhängigen Zusammenhängen zwischen sprachlichen und kognitiven Kategorisierungen, kann auf diese Weise über die Vermittlung aktiven Sprachgebrauchs eine Verbindung zwischen Sprache und Denken bestehen. Wir müssen nun die Bedeutung des aktiven Sprachgebrauchs im Prozeß

sozialer Kommunikation genauer analysieren, um die Auswertung der aktiven Worterklärung und die "linguistische" Auswertung miteinander vergleichen zu können.

Der Zuhörer hat zweierlei Arten von Kriterien zur Hand, wenn er aus der Aussage seines Gegenübers eine Botschaft herauslesen will. Zum einen kann er sich auf die objektiv im gesendeten Sprachkorpus vorliegenden Kategorien stützen. Zum anderen kann er seine Aufmerksamkeit, unabhängig vom objektiven Gehalt der Aussage, auf Merkmale der grammatischen Struktur, des Stils und der Korrektheit richten. Die Auswertung der aktiven Worterklärungsleistung versucht den ersten, die "linguistische" Auswertung den zweiten Prozeß nachzuvollziehen. In beiden Fällen liegt zwar das Kommunikationsmodell, das Fortschritte in der kognitiven Entwicklung mit der kumulativen Wirkung typischer Arten des Sprachgebrauchs zusammenbringt, zugrunde, aber die Motive des Sprechers und die Kriterien des Senders werden im ersten anders als im zweiten Falle gesehen: Im ersten Fall wird angenommen, daß die Sender durch ihrem dem Sozialmilieu angepaßte Strategie des Symbolgebrauchs veranlaßt werden, selbst wenn sie über das gleiche Verständnis der Sache verfügen, reduziert oder extensiv zu verbalisieren, die Empfänger aber das Verständnis des Senders nach dem objektiven Gehalt der explizierten Botschaft einschätzen; im zweiten Fall, daß die Sprecher je nach ihrer Schichtzugehörigkeit bestimmte sprachliche Merkmale seltener oder häufiger verwenden, die vom Sender als ein Indiz für bestimmte Fähigkeiten bewertet werden, ohne daß ein solcher Zusammenhang erwiesen wäre. In beiden Fällen sind die Reaktionen des Zuhörers unabhängig von den tatsächlichen Fähigkeiten des Sprechers; im ersten Fall werden sie jedoch von den durch die Arten des Sprachgebrauchs gesteuerten sozial manifestierten Erklärungsleistungen des Sprechers, im zweiten Fall aber von sozialen Vorurteilen des Sprechers über den "Wert" bestimmter Sprachformen bestimmt. In beiden Fällen wird angenommen, daß die Reaktionen des Zuhörers wiederum Lernprozesse für den Sprecher einleiten und die Entwicklung des Sprechers auf Dauer fördern oder hemmen. Kurz: Beide Auswertungen beziehen sich auf vermutliche Wirkungen kumulativer Lernprozesse in sozialer Kommunikation, die

jedoch einmal durch schichtspezifische Arten des Sprachgebrauchs, das andere Mal über die Bewertung schichtspezifisch verteilter Sprachformen vermittelt wird. Im ersten Fall bewegen wir uns im Rahmen der Bernsteinschen Theorie, im zweiten Fall im Rahmen einer Theorie sozialspezifischer Bewertung von Sprache, die - nach den negativen Ergebnissen zum Einfluß schichtspezifischer Strategien des Sprachgebrauchs - als eine Ideologiekritik analoger Bewertungen in der als Theorie von Sprachnormen uminterpretierten Theorie Bernsteins angesehen werden kann.

In alltäglichen Kommunikationssituationen wird es natürlich kaum möglich sein, die beiden analysierten Ebenen zu trennen; wir behaupten nur, daß beide wirksam sind. Ebensowenig muß der Zuhörer sich der kalkulierten Reaktion auf den objektiven Gehalt oder der Bewertung des Stils der Aussage bewußt sein. Im stilisierten Nachvollzug einer alltäglichen Kommunikationssituation durch die beiden Auswertungen der zweiten Messung des Experiments sind die beiden Ebenen jedoch trennbar. Als "aktive Worterklärungsleistung" haben wir den objektiven Gehalt der Erklärungen wie ein alltäglicher Zuhörer zu beurteilen versucht. Natürlich kann der Zuhörer im Alltag seinerseits sozialen Strategien der Dekodierung folgen, die den Strategien der Enkodierung mehr oder minder entgegenkommen. Auf diese Flexibilität des Zuhörers, soweit sie Defizite des Senders kompensieren oder potenzieren kann, kommt es aber in der Experimentalmessung, die allein den Effekt der sozialen Strategien des Senders erfassen soll, nicht an, wenn auch umso mehr in der Praxis des sozialisatorischen Alltags. Andererseits haben wir versucht die Flexibilität des Zuhörers in der Differenzierung zwischen Verständnis und Explikation bei allen Erklärungen nachzuvollziehen. Mit der Messung des objektiven Gehalts der Aussage ist die Möglichkeit flexibler Reaktionen des Zuhörers nicht geleugnet, sie beruht im Gegenteil auf der Konstanthaltung verschiedener Einstellungen des Zuhörers für alle Erklärungen. Die Auswertung der Erklärungsleistungen bezieht den Gehalt der Aussagen auf Strategien des Symbolgebrauchs auf der Seite des Senders.

Als "sprachliche Mittel der Erklärung" haben wir versucht, drei sinn-neutrale Merkmale der Erklärungen zu erfassen: (1) Die sprachliche Oberfläche und (2) typische "Stile" der Erklärung (Nominalphrase vs. Satzform) und (3) grammatische und orthografische Fehler. Solche Merkmale können je nach der Schichtzugehörigkeit des Senders mit unterschiedlicher Häufigkeit geäußert werden; ihnen wird vom Empfänger ein sozial standardisierter "Wert" beigemessen, der - objektiv natürlich ungesicherte - Rückschlüsse auf die Fähigkeiten des Senders zuläßt. Die Auswertung bezieht die verwandten sprachlichen Mittel nicht auf situationsübergreifende Strategien des Symbolgebrauchs, die schichtspezifischen Interaktionsformen entsprechen, sondern sie versucht eine bei der genau bestimmten Aufgabe der aktiven Worterklärung entstandenen Sprachkorpus nach Kriterien zu messen, die beim Zuhörer anstelle oder neben einer korrekten Dekodierung des objektiven Gehalts der Aussage die Reaktion bestimmen und durch die schichtspezifische Verteilung der nach diesen Kriterien gemessenen Formen schichtdiskriminierend wirken können. Selbst dort, wo sie sich auf Variable bezieht, die in früheren Untersuchungen nicht aufgabenbezogener Sprachkorpus als Operationalisierung von Arten des Sprachgebrauchs angesehen wurden, wie etwa Anzahl der Subordinationen verschiedener Art, Anzahl der Pronomina etc., soll die "linguistische" Auswertung nicht Arten des Sprachgebrauchs, sondern sozial bewertete Sprachmittel erfassen.

Vor allem zwei Gründe lassen sich für diese theoretische Interpretation der Auswertung sprachlicher Mittel anführen. Der erste Grund bezieht sich auf die strategische Aussagekraft dieser Auswertung innerhalb des gesamten Experiments. Der Einfluß unterschiedlicher Arten des Sprachgebrauchs wurde ja schon an der aktiven Erklärungsleistung, ihres Umfangs wie ihrer Expliziertheit im Einzelfall, gemessen. Nachdem Umfang und Qualität der Erklärungen offensichtlich nur sehr wenig von der sozialen Schichtzugehörigkeit abhängen, würde der Versuch, die linguistischen Variablen wiederum als Indikatoren der Strategien des Sprachgebrauchs anzusehen, wie eine Suche nach Bestätigung der Theorie um jeden Preis anmuten. Aus diesem ersten Grunde verstehen wir - wie zu Beginn dieses Abschnittes festgestellt - die gesamten sprachlichen Mittel der Erklärung als unab-

hängig vom objektiven Gehalt der Aussage bewertete Zeichen der Konformität mit sozialen Normen - ganz abgesehen von den Fehlermaßen und den Maßen des Erklärungsstils, die in jedem Fall nur auf diese Weise verstanden werden können.

Der zweite Grund liegt im Charakter der Testaufgabe. Wenn sprachliche Merkmale als Indikatoren für Strategien des Sprachgebrauchs gewertet wurden, so stammten sie aus Sprachmaterial, das mit sehr offenen Methoden erhoben wurde, wie freie Aufsätze, Nacherzählungen, Rekonstruktionen von Bildergeschichten. In solchen Aufgaben gibt es nicht nur ein breites Spektrum von Ausdrucksintentionen, sondern auch einen großen Spielraum für die Wahl sprachlicher Mittel, so daß von der Aufgabe her ein relativ großer Ereignisraum denkbar ist, in dem Strategien des Sprachgebrauchs selektiv wirken. Anders bei der Bedeutungserklärung eines Stimuluswortes: Hier gibt es eine Reihe von Sprachhülsen, die man für jedes Stimuluswort einsetzen kann und die unabhängig vom Inhalt als besser oder schlechter angesehen werden; wir haben versucht, diese standardisierten Sprachformen in der Variablen-Gruppe (2) "Erklärungsstile" zu erfassen. Hier gibt es weiterhin für sprachliche Formen, die sonst als Indikatoren von Strategien des Sprachgebrauchs gewertet werden, keinen breiten Raum unterschiedlicher Wahlmotive: Sie sind entweder in bezug auf die Aufgabe relativ standardisierte, gut oder schlecht bewertete Mittel, wie präpositionale Verknüpfungen und Relativsätze als Nominal-Erweiterungen, oder in bezug auf die Aufgabe recht ausgefallene, als Neuschöpfung oder Fehlleistung interpretierbare Formen, wie metasprachlich nicht gekennzeichnete Hauptsätze oder Nebensätze ohne übergeordneten Hauptsatz. Der Charakter der Aufgabe legt es also nahe, die sprachlichen Mittel nicht als Indikatoren für Codes anzusehen, sondern die mutmaßlichen Bewertungen dieser Mittel in bezug auf die Aufgabe in Hypothesen über ihr schichtspezifisch häufigeres Auftreten umzuformen. Wir können die Generalhypothese der Auswertung der sprachlichen Mittel so umschreiben, daß in der Mittelschicht diejenigen sprachlichen Mittel häufiger sind, die in der Regel als guter sprachlicher Ausdruck bewertet und als Zeichen für besondere Fähigkeiten angesehen werden. Wir können diese Generalhypothese für die einzelnen Variablen spezifizieren, wenn wir die Systematik der Variablen vorgestellt haben.

6.3. Hypothesen und Auswertung der sprachlichen Mittel der Erklärung.

Die Auswertung der sprachlichen Mittel ist keine im strengen Sinne linguistische Analyse. Sie verwendet nicht einmal wie die meisten sozio-linguistischen Untersuchungen Kategorien der normativen Grammatik, etwa des Dudens, die unter streng linguistischen Kriterien schon als inadäquate Beschreibungsmodelle angesehen werden. Die Auswertung der sprachlichen Mittel bezieht sich vielmehr auf eine Reihe von Merkmalen, die bei der Durchsicht der Antworten auffällig waren und die für die Erklärung von Konzepten typisch erschienen. Selbst wenn also die Auswertung Variablen enthält, die Kategorien der normativen bzw. sprachwissenschaftlichen Grammatik ähneln, so wurden sie hier unter dem besonderen Gesichtspunkt ihrer Brauchbarkeit für Worterklärungen ausgewählt. Das kann am Beispiel der Pronomina verdeutlicht werden: Sie wurden nicht nach grammatischen Klassen (wie etwa lokal, temporal etc. oder Demonstrativ-, Reflexiv-, Personal- etc. Pronomina) unterklassifiziert, sondern nach ihrer Brauchbarkeit für Definitionen (etwa "man" als atypisches Pronomen in Erklärungen (S 14)¹⁾ vs. alle übrigen Indefinitpronomen (S 13), oder falscher (d.h. exaphorischer) Pronominalgebrauch (S 15) vs. verweisendem (anaphorischem) Pronominalgebrauch (S 16); entsprechend sind die Definitionsregeln der Kategorien nicht unbedingt die der grammatischen Kategorien gleichen Namens.

Kategorien der Auswertung wurden zuerst gesammelt, indem drei Auswerter alle im Vortest zur aktiven Worterklärung erhobenen Erklärungen durchgingen. Nachdem auf diese Weise ein Schema praktikabler und nicht zu seltener Kategorien entwickelt war, wurde es an allen Erklärungen des Haupttests angewandt und nur noch unwesentlich modifiziert. Die Hauptauswertung wurde für alle

¹⁾-----
Die Zahlen beziehen sich auf den Anhang D.

Erklärungen von zwei Auswertern unabhängig voneinander durchgeführt, die anschließend in einem dritten Arbeitsgang ihre Auswertung miteinander verglichen und die endgültige Auswertung festlegten. Ganz wie bei der Bewertung der Erklärungen nach "Verständnis" und "Explikation" scheint uns auch hier die gemeinsame, über objektive Mehrdeutigkeiten entscheidende und Mißverständnisse ausräumende Diskussion zwischen den Auswertern wichtiger als eine formelle Reliabilitätsprüfung, auf die wir nicht zuletzt wegen der enttäuschenden Ergebnisse bei der Bewertung der Erklärungsleistungen (Tab. 1) verzichtet haben. Wegen der systematischen Mehrdeutigkeit sprachlicher und sprachbezogener Kategorien scheint eine endgültig entscheidende Diskussion zwischen mehreren Auswertern die Reliabilität der Auswertung eher zu garantieren als formelle Verfahren der Inter-Rater-Reliabilität (wie Prozent der Übereinstimmung oder Korrelationen), die ja nur den Anteil kontingenter Mißverständnisse erklären. Ebensowenig wie von einer regulären Reliabilitätsprüfung der Auswertung kann von einer Prüfung unseres Kategorienschemas die Rede sein.

Die Auswertung benutzte für die mündliche und schriftliche Messung die gleichen Kategorien. Bis auf die Fehlermaße, bei denen sich aus dem Charakter der Messung Unterschiede ergaben, bereitete dies keine Schwierigkeiten. Für beide Messungen wurden 30 Summenmaße ausgezählt; für die mündliche Messung zusätzlich spezifische (S 48, 49, 50, 51; noise, Wiederholung, semantische Korrektur, Anzahl der Worte in "mazes" und in semantischer Korrektur) und für die schriftliche Messung ein (S 3, Schreibfehler) Fehlermaß. Diese Summenmaße wurden auf die zugehörigen Obereinheiten: Anzahl der Antworten, Anzahl der Worte, Anzahl der Verbphrasen, relativiert; auf diese Weise ergaben sich 74 Quotienten für beide Messungen und zusätzlich 17 Quotienten für die mündliche, zusätzlich 6 Quotienten für die schriftliche Messung. Die Bildung der Quotienten war notwendig, um durchgehende Effekte der Textlänge aus dem Schichtvergleich herauszuhalten; sie sind die im Schicht- und Geschlechtsvergleich endgültig benutzten "linguistischen" Maße.

Die "linguistischen" Quotienten sind - ebenso wie die Summenmaße - in Anhang D ausführlich beschrieben; an dieser Stelle soll nur die Logik ihrer Klassifikation, die auch in den tabellarischen Darstellungen verwandt wurde, dargestellt werden. Die Quotienten zerfallen in drei größere Gruppen: (1) Maße der sprachlichen Oberflächenstruktur, (2) Maße des Erklärungsstils und (3) Fehlermaße. Die Maße der Oberflächenstruktur kommen Kategorien der üblichen Grammatiken am nächsten. Sie umfassen (1.1) "erklärungsweiternde Satzteile": Anzahl der Worte und Verbphrasen pro Antwort, präpositionale und konjunktionale sowie falsche präpositionale und falsche konjunktionale Verknüpfungen, jeweils relativiert auf die Anzahl der Worte und Verbphrasen; ferner (1.2.) den Gebrauch der verschiedenen Pronomina; und schließlich (1.3.) Arten der "Satzverknüpfung": Koordination von Hauptsätzen, Nebensätzen und Relativsätzen pro Verbphrase. Die zweite Gruppe von "linguistischen Maßen" sind die des Erklärungsstiles, d.h. der Sprachform der Erklärung auf der über der Ebene von Satzteilen ("sprachliche Oberflächenstruktur") liegenden Ebene von Satzformen und Satzfragmenten. Hierzu gehören (2.1.) ein nominaler oder ein verbaler "Erklärungsansatz", sowie (2.2.) verschiedene Möglichkeiten, Erklärungen in längere sprachliche Formen zu bringen ("Erklärungsbreite"): der Gebrauch von mehr als einer Verbphrase pro Erklärung, mehrere Erklärungsanläufe für ein Stimuluswort, die Wiederholung des Stimuluswortes, die explizite Differenzierung von Objekt- und Metasprache, der Gebrauch von Beispielen und Vergleichen. Als dritte Gruppe der "linguistischen" Maße folgen die Fehler. Sie sind z.T. an der normativen Grammatik orientiert: als (3.1.) "Grammatikfehler" haben wir in beiden Messungen kategorisiert, was ein Lehrer beanstanden würde, lediglich mit einer genaueren Unterklassifikation. Dasselbe gilt für (3.2.) die Schreibfehler der schriftlichen Messung. Schließlich wurden (3.3.) als Noises, Wiederholung und "semantische Korrektur" Planungsschwierigkeiten des "mündlichen Ausdrucks" kategorisiert. Alle Fehlermaße wurden auf alle drei möglichen Gesamtheiten, Antworten, Worte, Verbphrasen, relativiert; daher erklärt sich die Vielzahl der Fehlermaße. Einige Maße der sprach-

lichen Oberfläche hätten ebenso gut als Fehlermaße klassifiziert werden können: Falsche propositionale und konjunktionale Verknüpfungen (Q 28-33), und falscher Pronominalgebrauch (Q 37, 41, 45). Aus zwei Gründen haben wir sie jedoch an der systematischen Stelle gelassen: Erstens gibt es eine Reihe von "richtige" und "falsche" Oberflächenmerkmale verknüpfenden Maßen; zweitens stellen diese Maße weniger als die Schreib- und Sprechfehler eindeutig konventionell bewertete Verstöße dar.

Wir können nun die Generalhypothese für die einzelnen Variablengruppen spezifizieren: (1.1.) "Erklärungserweiternde Satzteile": Häufiger in der Mittelschicht. (1.2.) "Promominalgebrauch": Häufigeres Auftreten von unbestimmten Pronomina und "man" in der Mittelschicht; häufigeres Auftreten von rückbezüglichem ("anaphorischem") Pronominalgebrauch in der Mittelschicht, häufigeres Auftreten von falsch verweisendem ("exophorischem") Pronominalgebrauch in der Unterschicht. (1.2.) "Subordination": Häufiger Gebrauch von Unterordnungen in der Mittelschicht, von Relativsätzen und Beiordnungen in der Unterschicht. (2.1.) "Erklärungsansatz": Häufigerer Gebrauch von ersetzenden bzw. nominalen Erklärungen in der Mittelschicht, von umschreibenden bzw. verbbezogenen Erklärungen in der Unterschicht. (2.2.) "Erklärungsbreite": Häufigeres Auftreten von Beispielen, Vergleichen und Objekt-Meta-Sprache-Differenzierungen in der Mittelschicht, von Wiederholungen des Stimuluswortes in der Unterschicht. (3.1.) Grammatikfehler und (3.2.) Schreibfehler: Häufiger in der Unterschicht. (3.3.) Planungsfehler im Mündlichen: Häufiger in der Mittelschicht.

6.4. Ergebnisse des Schichtvergleichs der sprachlichen Mittel.

Die Ergebnisse der Analyse der sprachlichen Mittel der Erklärung liegen bis jetzt nur als Mittelwerte jedes Maßes der mündlichen und schriftlichen Messung für alle Gruppen der Experimentalstichprobe I (N=66) vor: Für die beiden Geschlechts- und Schicht- sowie die vier Experimentalgruppen, für jede der kleinsten Gruppen (N=4) und jede mögliche Zusammenfassung dieser Gruppen. Statistische Prüfungen der Effekte wurden noch nicht durchgeführt. Eine Normalitätsprüfung der Maße ergab, daß so gut wie kein Maß normal verteilt ist, so daß - ganz unabhängig von Fragen des Meßniveaus - nur nicht-parametrische Prüfverfahren (wahrscheinlich Wilcoxon-Tests) in Frage kommen. Weiterhin sind Rangkorrelationen innerhalb der sprachlichen Maße (sowie evtl. eine Faktorenanalyse) zur Klärung ihrer Beziehungsstruktur, und Rangkorrelationen zwischen den Maßen der Bedeutungserklärung und den Maßen der sprachlichen Mittel der Erklärung vorgesehen.

Die Gruppenmittelwerte werden in einem gesonderten Materialband veröffentlicht. Sie liegen für alle Maße in Tabellen vor, die Geschlechts- und Schichtvergleiche ermöglichen und die erst zusammen mit der vollständigen Darstellung der Ergebnisse der sprachlichen Analyse veröffentlicht werden sollen. Im folgenden wollen wir die Ergebnisse des Schichtvergleichs für die Quotienten der "linguistischen" Auswertung zusammenfassend darstellen. Als vorläufigen Test für die Bedeutung der Schichtdifferenzen haben wir zwei Kriterien genommen: Die Übereinstimmung zwischen mündlich und schriftlich in der Richtung der Differenz und die Bedingung, daß die Schichtdifferenz mindestens 5% des jeweiligen Gesamtmittelwertes betragen mußte. Das erste Kriterium ist problematisch, weil der Wechsel zwischen mündlicher und schriftlicher Kommunikationsform die sprachlichen Mittel wahrscheinlich stärker als die Bedeutungserklärung beeinflusst hat. Das zweite Kriterium soll natürlich kein Ersatz für die Prüfung von Mittelwertdifferenzen anhand der Varianz der Differenz sein. Wir gehen nicht mehr auf jedes einzelne, sondern nur Gruppen von Maßen ein.

(1) Maße der linguistischen Oberflächenstruktur:

1. Die Mittelschichtkinder gaben mehr Worte und Verbphrasen in jeder Antwort als die Unterschichtkinder.
2. Die Mittelschichtkinder benutzten mehr präpositionale und mehr konjunktionale Verknüpfungen pro Wort wie pro Verbphrase, wie vorausgesagt.
3. Entgegen unserer Annahme gebrauchten die Unterschichtkinder mehr unbestimmte und mehr richtig verweisende (anaphorische) Pronomina, die Mittelschichtkinder mehr falsch verweisende (exphorische) Pronomina. Daß die Unterschichtkinder Pronomina häufiger anaphorisch, die Mittelschichtkinder häufiger exophorisch gebrauchen, widerspricht sowohl der Theorie Bernsteins wie auch anderen Untersuchungen (Hawkins 1973).
4. Ebenfalls entgegen unserer Annahme gebrauchten die Unterschichtkinder häufiger Subordinationen als die Mittelschichtkinder. Das gilt für Relativsätze, wo wir einen häufigeren Gebrauch der Unterschicht vorausgesagt haben, und vor allem für alle übrigen untergeordneten Nebensätze. Die Mittelschichtkinder brauchen hingegen mehr "Koordinationen außer und". Für Erklärungen von Wortbedeutungen verwenden also die Unterschichtkinder die Hypotaxe, die Mittelschichtkinder die Parataxe.

Das zuletzt referierte Ergebnis legt die Vermutung nahe, daß die Mittelschichtkinder stärker als die Unterschichtkinder dazu tendieren, Erklärungen, die prinzipiell knapp gehalten werden könnten, zu erweitern. Als Gruppe erreichen sie dadurch, obwohl im Einzelfall das Gegenteil möglich ist, keine differenzierteren oder abstrakteren Erklärungen; denn bezüglich der Bedeutungserklärung bestehen keine Unterschiede zwischen den Schichten. In die gleiche Richtung weist, daß die Unterschichtkinder - und nicht die Mittelschichtkinder - Pronomina, die als abstrakte Platzhalter relevanter Dimensionen noch die geeignetsten sprachlichen Mittel zur Differenzierung von Erklärungen darstellen,

und Subordinationen, die eine sprachliche Explikation logischer Beziehungen verlangen, öfter gebrauchen. Auf der anderen Seite gebrauchten die Mittelschichtkinder mehr Koordinationen (wobei "und" nicht mitgezählt wurde), die auf eine ähnliche, aber sprachlich breitere Weise logische Beziehungen sprachlich ausführen wie die Subordinationen. Allgemein scheinen die Mittelschichtkinder mehr durch in die Breite gehende sprachliche Mittel Erklärungen gleicher Güte wie die Unterschichtkinder zu erreichen, die eher die traditionellen Mittel sprachlicher Differenzierung von Erklärungen benutzen. Wenn es überhaupt Gültigkeit hatte, so kehrt sich das übliche Bild schichtspezifischer Sprachstile hier am Beispiel einer spezifischen Leistung um.

(2) Maße des "Erklärungsstiles":

5. "Erklärungsansatz": Wiederum entgegen unseren Erwartungen tendiert die Unterschicht eher zu nomenbezogenen, die Mittelschicht eher zu verbbezogenen Erklärungen. Dies widerspricht auch den Ergebnissen von Henderson (1970), die bei Bildbeschreibungen fand, daß die Mittelschichtkinder eher mit Hilfe von Nomina und Nominalerweiterungen, die Unterschichtkinder eher mit Verben und Verberweiterungen arbeiten.

6. "Erklärungsbreite": Die Mittelschichtkinder setzen stärker als die Unterschichtkinder Mittel ein, die die Erklärungen als solche expliziter, gefälliger und kommunikabler machen: Sie benutzen öfter mehr als eine Verbphrase und öfter unabhängige Erklärungseinheiten für eine Erklärung; sie trennen öfter zwischen Objekt- und Metasprache und drücken, wenn sie ein Beispiel geben, dies öfter explizit aus; schließlich wiederholen sie seltener das S-Wort, geben aber gleich viel Vergleiche wie die Unterschichtkinder. Die Tendenz, daß die Unterschicht gleichsam mehr in die Tiefe, die Mittelschicht mehr in die Breite geht, um zu etwa gleich brauchbaren Erklärungen zu kommen, bestätigt sich auch an dem Ergebnis 5. und 6.

(3) Fehlermaße:

7. Die Unterschichtkinder machen allgemein mehr Grammatikfehler verschiedener Arten als die Mittelschichtkinder und im Schriftlichen mehr Schreibfehler, während im Mündlichen die Mittelschichtkinder öfter Planungsfehler (mazes, noises, Wiederholungen). Daß die Unterschichtkinder mehr Fehler gegen die hochsprachliche Norm, die Mittelschichtkinder mehr Planungsfehler aufweisen, entspricht durchaus den ursprünglichen Annahmen der Theorie der linguistischen Kodes (Bernstein 1962).

Was tragen diese Ergebnisse zur Überprüfung oder zur ideologiekritischen Durchleuchtung der Theorie der linguistischen Kode bei? Nur zwei der Ergebnisse können ideologiekritische Einwände gegen die Theorie Bernsteins direkt unterstützen: Die Unterschichtkinder brauchen weniger Worte und Verbphrasen für ihre Erklärungen, und sie verstoßen häufiger gegen kodifizierte Normen der Hochsprache (Schreibfehler, Grammatikfehler). Beide Ergebnisse können auf Kriterien hinweisen, die in Bewertungsprozessen wirksam sind und dem Urteil des Zuhörers über das Wortverständnis des Kindes anstelle der tatsächlichen Erklärungsleistungen des Kindes zugrundeliegen. Es kann sein, daß die wortreicheren und grammatisch korrekteren Erklärungen der Kinder der Mittelschicht bei Lehrern und Eltern den Eindruck besseren Verständnisses oder größerer "Intelligenz" erwecken und, wie eine sich selbst erfüllende Prophezie (Rosenthal/Jacobson 1971), durch die entsprechend anspruchsvolleren Reaktionen der Eltern und Lehrer tatsächlich zu Vorteilen der Mittelschichtkinder in der kognitiven Entwicklung führen; eine solche Entwicklung wäre unabhängig vom Ausgangspotential der Kinder.

In dieser Interpretation paßt auch noch der allgemeine Eindruck, den wir aus den Ergebnissen gewonnen hatten, daß die Mittelschichtkinder eher breitere und gefälligere Erklärungen bieten (Ergebnis 2. und 6.). Eine Vielzahl von konkreten Ergebnissen widerspricht jedoch dieser Interpretation: Die Kinder der Unter-

schicht gebrauchen mehr unbestimmte Pronomina, mehr Subordinationen, sie verwenden Pronomina häufiger anaphorisch, und sie geben häufiger nominale als verbbezogene Erklärungen. All dies sind Merkmale, die eher dem Sprachstil der Mittelschicht zugewiesen werden müßten - ganz unabhängig von der Frage, was sie als Indikator für einen elaborierten Sprachgebrauch bedeuten würden. Wenn wir also den gleichen Mechanismus schichtdiskriminierend wirkender Bewertungen, die sich im Laufe der Zeit selber bestätigen, annehmen wie bei den Fehlermaßen, so müßten eher die Unterschichtkinder gefördert werden. Wir können also schließen, daß an der Mehrzahl der Ergebnisse die Vermutung nicht bestätigt werden kann, die Mittelschichtkinder benutzten häufiger sprachliche Merkmale, die im Alltagsverständnis und in der wissenschaftlichen Überhöhung durch eine Theorie der linguistischen Kodes als Indizien für besonderes Verständnis oder für besondere "Intelligenz" angesehen werden.

6.3.5. Zusammenfassung der Analyse der sprachlichen Mittel der Erklärung.

Wir wollen den Eindruck vermeiden, als sollte durch die Untersuchung der sprachlichen Mittel der Erklärung die Theorie der linguistischen Kodes unter allen Umständen gerettet werden. Die Analyse der sprachlichen Mittel wurde begonnen, als die wichtigsten Hypothesen dieses Experiments, die aus der Theorie der linguistischen Kodes abgeleitet sind, erkennbar widerlegt waren. Unter diesen Umständen sollte die Analyse der sprachlichen Mittel prüfen, ob die Kritik an der Theorie der linguistischen Kodes nicht an den gleichen Ergebnissen bestätigt werden könne, d.h. wenn Voraussagen nicht über durch schichtspezifische Arten des Sprachgebrauchs bewirkte Unterschiede, sondern über die Häufigkeit manifester sprachlicher Merkmale getroffen werden, die im Alltag und durch die ideologie-kritisch umgedeutete Theorie der linguistischen Kodes als Zeichen besonderer Leistungen bewertet werden. Die Differenzierung zwischen der Theorie der linguistischen Kodes, die eine Theorie des Sprachgebrauchs ist und für die manifeste sprachliche Merkmale nur Indikatoren des Sprachgebrauchs sind, und der ideologie-kritischen Umdeutung der Theorie der linguistischen Kodes, die die schichtspezifische Verteilung und die Bewertung von sprachlichen Merkmalen direkt untersucht, mag rein analytisch erscheinen. Wir glauben, daß die meisten der hier untersuchten Merkmale als nicht kode-indizierende, sondern sozial bewertete sprachliche Merkmale angesehen werden müssen und nur bei einigen - z.B. den anaphorischen und exaphorischen Pronominalgebrauch - beide Deutungen möglich sind. Aber auch selbst dann, wenn die theoretische Bestimmung der untersuchten sprachlichen Mittel der Erklärung reine Dezision wäre, wäre es richtig gewesen, sie nach den negativen, die Theorie der linguistischen Kodes widerlegenden Hauptergebnissen als sozial bewertete sprachliche Merkmale anzusehen: Nur in diesem Fall können wir neue Informationen gewinnen und nicht nur die weitere Überprüfung einer Theorie, deren Hauptvoraussagen sich nicht bestätigt haben, auf einem Nebenschauplatz.

Unterschiede der sprachlichen Form, die für die aktive Erklärungsleistung irrelevant sind, hätten bei den Zuhörern nun ebenso wie die tatsächliche Erklärungsleistung, nur aufgrund gleichsam illegitimer Kriterien, zu einer zwischen beiden Schichtgruppen diskriminierenden Reaktion führen können. In beiden Fällen hätten wir Reaktionen des Zuhörers, die von der tatsächlichen Fähigkeit des Sprechers unabhängig sind - der passive Wortschatz der Vpn. war ja kontrolliert. Im ersten Fall jedoch wäre die Reaktion des Zuhörers den Erklärungen des Sprechers angemessen und die Benachteiligung der Unterschichtkinder durch ihre schichtspezifischen Strategien des Sprachgebrauchs bedingt, im zweiten Fall würde die Reaktion des Zuhörers selber auf sozialen Bewertungen über den "intelligenten" Charakter bestimmter sprachlicher Formen beruhen, die über die schichtspezifische Verteilung dieser Merkmale zu einer schichtdiskriminierenden Bewertung wird. Im ersten Falle werden Annahmen der Theorie der linguistischen Codes geprüft, im zweiten Falle eine ideologie-kritische Umdeutung dieser Theorie.

Die Ergebnisse der Analyse der sprachlichen Mittel der Erklärung können nun auch die Voraussage über die schichtspezifische Verteilung einzelner sprachlicher Merkmale nicht bestätigen. Zwar wird eine solche ideologie-kritische Umdeutung der Theorie der linguistischen Codes an zwei Ergebnissen bestätigt: Die Unterschicht gebraucht weniger Worte und macht mehr Fehler. Aber dies sind bekannte und vergleichsweise triviale Ergebnisse. Auf der anderen Seite gebrauchen die Unterschichtkinder eine Reihe von sprachlichen Merkmalen, die man eher als Domäne der Mittelschichtkinder ansehen würde: Sie verwenden mehr unbestimmte Fürwörter, gebrauchen Fürwörter häufiger anaphorisch, arbeiten häufiger mit Subordinationen (auch ohne Einbezug der Relativsätze) und geben mehr nominale als verbbezogene Erklärungen. Wenn überhaupt solche sprachlichen Merkmale den Zuhörer zu Mutmaßungen über die Fähigkeiten des Sprechers und entsprechenden Reaktionen veranlassen, so werden eher die Unterschichtkinder überschätzt und durch eine Selbsterfüllung der Prophezeiung im Laufe der Entwicklung gefördert. Kurz: Sowenig die Theorie Bernsteins bestätigt werden konnte, sowenig konnten Hypothesen, die man aus der Kritik an Bernsteins Theorie extrapolieren kann, schlüssig bestätigt werden.

Viertes Kapitel: Ergebnisse der dritten Messung.

=====

1. Operationalisierung der Prüfvariablen.

1.1. Konstruktion der Testaufgaben.

In der zweiten Messung hatten wir bessere aktive Worterklärungen der Mittelschichtkinder erwartet, auch wenn sie den Unterschichtkindern in passiver Wortkenntnis nicht überlegen sind. Die dritte Messung sollte belegen, daß die Überlegenheit der Mittelschichtkinder im passiven Wortschatz und langfristig in der kognitiven Entwicklung sich aus schichtspezifischen Arten des Sprachgebrauchs ergeben haben kann. Die theoretische Begründung der dritten Messung soll kurz rekapituliert werden.

Der für die Unterschicht typische restringierte Gebrauch von Sprache führt Kinder der Unterschicht dazu, Konzepte sprachlich weniger klar auszudrücken als Kinder der Mittelschicht, selbst wenn kein Unterschied zwischen Kindern beider Schichten im Verständnis des Konzepts besteht. Der Zuhörer aber wird nach den im sprachlichen Ausdruck des Kindes erkennbaren Kategorisierungen auf das Verständnis des Kindes schließen und seine eigene sprachliche Antwort entsprechend wählen. Für das Kind ergibt sich dadurch eine relativ ärmere und weniger stimulierende kognitive Eingabe, die ihrerseits wieder den möglichen sprachlichen Ausdruck kognitiver Kategorisierungen auf Seiten des Kindes bestimmt. Auf diese Weise kann durch die für die Unterschicht typische Art des Sprachgebrauchs ein zirkulärer Kommunikationsprozeß in Gang kommen, der auf Dauer die Lernchancen des Unterschichtkindes mindert, unabhängig von ihren kognitiven Fähigkeiten. Umgekehrt aber müßten Kinder beider Schichten, wenn sie die gleiche sprachliche Eingabe empfangen, aus ihr das gleiche kognitive Verständnis gewinnen. Sobald die sprachliche Stimulierung gleich ist, sollten Kinder beider Schichten aus ihr die gleichen Konzepte ablesen.

Da wir diese der dritten Messung zugrundeliegende Hypothese an 10jährigen Kindern überprüfen wollten, an Kindern also, die durch den vermuteten zirkulären Prozeß sozialer Anregung schon schichtspezifisch beeinflußt sein müßten, wurden die Kinder beider Schichten auf der Leistung im standardisierten passiven Wortschatztest parallelisiert; auf diese Weise sollten die akkumulierten Wirkungen des sozialen Anregungsprozesses neutralisiert und sein punktueller Einfluß freigelegt werden. Als Form der dritten Messung wählten wir wiederum den multiple-choice Test. Der Stimulus hatte, anders als in der ersten Messung, notwendig sprachlichen Charakter, da die Evokation eines Konzepts als Reaktion gerade auf sprachliche Eingaben getestet werden sollte; die Qualität des Stimulus sollte deshalb auch von Testaufgabe zu Testaufgabe wechseln. Die Antwortvorgaben, die die kategoriale Differenzierung messen, sollten nicht mehr "richtig" oder "falsch" sein, aber eine eindeutig treffendste Antwort enthalten. Im Unterschied zur Verständnis-Bewertung der zweiten Messung, wo die Beherrschung des Konzepts durch den Auswerter rekonstruiert wurde, wird sie hier durch eng verwandte Antwortvorgaben explizit abgefragt. Die in der dritten Messung geforderte Leistung wollen wir "kategoriale Differenzierung" nennen, im Unterschied zum in der ersten Messung geforderten passiven Wortverständnis.

Die dritte Messung zerfällt in zwei Untertests. Die Trennung ergab sich erst im Laufe der Arbeit aus zwei Motiven. Erstens war es schwierig, aus den von den Kindern in der zweiten Messung gegebenen Erklärungen neue Antwortvorgaben zu finden, die in der Qualität der Kategorisierung bei gleichem Bedeutungsgehalt erkennbar und in jeder Testaufgabe auf die gleiche Weise variierten. Zu diesem Zweck wurde ein System differenzierter, untereinander theoretisch verbundener Kategorien von Erklärungen entwickelt, das zugleich der Bewertung der Antworten in der zweiten Messung wie der Konstruktion des Tests der dritten Messung dienen sollte (siehe III.1.). Dies Kategoriensystem erwies sich als zu umfassend und konnte, vor allem wegen der ungelösten Frage der Inter-

Rater-Reliabilität, für eine zwischen den Testaufgaben vergleichbare Auswahl von Vorgaben nicht angewendet werden. Zweitens wurde es schwer, die Beziehungen zwischen dem Stimulus und den Vorgaben zu kontrollieren. Da der Stimulus auch in der sprachlichen Form und der logischen Qualität variieren sollte, standen in den Testaufgaben, wo der Stimulus nicht mehr mit dem des standardisierten Tests übereinstimmte, sondern selber aus den Erklärungen der Kinder für diese Stichworte ausgewählt war, Stichwort und Antwortvorgabe nicht mehr in der gleichen Beziehung wie bei der aktiven Worterklärung; Analoges gilt, bei Aufnahme des Stimulus unter die Vorgaben, für den Vergleich der Vorgaben untereinander.¹⁾

Um den beiden genannten Schwierigkeiten zu entgehen, haben wir, unabhängig von den Stimulusworten des Frankfurter WST und von den Erklärungen der Kinder in der zweiten Messung, zusätzliche Testaufgaben konstruiert, bei denen erstens die sprachliche Form der Vorgaben überhaupt nicht und die enthaltenen Kategorisierungen nur in einzelnen abstrakten Merkmalen variierten und bei denen zweitens die Beziehung zwischen Stimulus und bester Vorgabe sehr klar die einer Umschreibung zum umschriebenen Konzept ist und in jeweils zwei Testaufgaben zwei gleiche Vorgaben so voneinander zu trennen sind, daß einmal die eine, das andere Mal die andere die beste Vorgabe ist, - so daß also die Beziehungen zwischen den Vorgaben in beiden Richtungen überprüft werden können. Diese zusätzlich konstruierten Testaufgaben haben wir "kategorialen Präzisionstest" genannt; die übrigen Aufgaben der dritten Messung, die nach wie vor aus den Erklärungen der Kinder in der zweiten Messung ohne deren genaue Kategorisierung gebildet wurden und in dem die Beziehungen zwischen Stimulus und Vorgaben und zwischen Vorgaben untereinander

1)-----

Als Beispiel Testaufgabe 42 (siehe Anhang E): "Wenn man etwas nicht ganz genau weiß und es irgendwie sagen muß" wurde in der zweiten Messung als genaue Umschreibung von "Schätzung" klassifiziert. Wenn dieselbe Umschreibung als neuer Stimulus verwandt wird, so ist das beste Konzept zu diesem Stimulus nicht mehr "Schätzung", sondern "raten", obwohl zum Stimulus "Schätzung" "raten" als eine schlechtere Erklärung als die genaue Umschreibung bewertet wurde.

nicht immer die gleiche Form hatten, haben wir den "differenzierten Verständnistest" genannt. Diese beiden Untertests wollen wir genauer beschreiben.

Die 42 Testaufgaben des "differenzierten Verständnistests" und die 16 Testaufgaben des "kategorialen Präzisionstests" waren in der Abfolge auf dem Testbogen miteinander vermischt, die Aufgabenpaare weit voneinander getrennt. Im "differenzierten Verständnistest" trat fast immer das Stimuluswort des Frankfurter Wortschatztests wieder auf, als Stimulus oder unter den fünf Vorgaben; in allen Fällen charakterisierte es das abgerufene Bedeutungsfeld.¹⁾ Sprachliche Form und semantische Qualität der Vorgaben variieren unabhängig voneinander, so daß die semantisch treffendste Alternative in einer konventionell schlechten sprachlichen Form ("wenn...") auftreten kann. Nachträglich haben wir versucht, die beiden Dimensionen als "sprachlogische Form" und "semantische Qualität" in der qualitativen Analyse (siehe 3.4.1.) zu klassifizieren und dem ursprünglichen Plan einer theoretischen Kontrolle der Testaufgaben gerecht zu werden.

Für den "kategorialen Präzisionstest" wählten wir - unabhängig von den Stimuli des Frankfurter Wortschatztests - 8 Paare von Begriffen aus, die sich durch nur ein Merkmal, aber eindeutig voneinander unterscheiden. So etwa bezeichnen "Begegnung" wie "Verabredung" ein Zusammentreffen einander bekannter Menschen, unterscheiden sich aber in der Dimension "beabsichtigt-zufällig". Als Stimulus formulierten wir eine möglichst treffende Umschreibung, unter die Vorgaben nahmen wir bei beiden Aufgaben eines Paares beide Ausgangsbegriffe sowie drei aufgabenspezifische Begriffe auf, die - ähnlich wie die beiden Ausgangsbegriffe untereinander -

1)-----
Bei 12 Aufgaben wurde ein Stimuluswort des Frankfurter WST wieder als Stimuluswort benutzt; zur Auswahl standen Umschreibungen und Einzelworte. Bei 22 Aufgaben wurde eine Umschreibung als Stimulus genommen, das ursprüngliche Stimuluswort tauchte aber neben anderen Umschreibungen unter den Vorgaben wieder auf. Bei 6 Aufgaben wurde ein neues Einzelwort als Stimulus gewählt und das alte Stimuluswort erschien unter den Vorgaben. Bei 2 Aufgaben erschien das ursprüngliche Stimuluswort weder als Stimulus noch als Vorgabe wieder (siehe auch 3.4.1.).

vom jeweils verlangten Begriff nur auf einer kritischen Dimension unterscheidbar waren; die Unterschiede zwischen allen fünf Vorgaben versuchten wir so klein wie möglich zu halten (siehe die vollständige Darstellung der Aufgabenpaare in Anhang G). Innerhalb jeder einzelnen der 18 Testaufgaben wurden also minimale Differenzierungen verlangt und in jedem Testpaar wurde die gleiche Differenzierung zwei Mal verlangt, in leicht gewandelter Umgebung und mit jeweils anderer richtiger Lösung. Bei der Suche nach geeigneten Stimulus- und Vorgabeworten orientierten wir uns an Dornseiff (1959). Bezeichnenderweise ist es nur gelungen, für abstrakte oder intentionale Worte Aufgabenpaare mit einer ausreichenden Anzahl von Antwortvorgaben zu finden; Versuche, ähnliche Aufgabenpaare auch mit extraverbal leichter identifizierbaren Worten zu konstruieren (z.B. "Obst-Gemüse", "Geschirr-Besteck"), scheiterten vor allem an mangelnden Antwortalternativen.¹⁾ Von den 8 Aufgabenpaaren verlangen fünf eine Differenzierung zwischen Worten, die eigene oder fremde Intentionen beschreiben - wir werden sie "intentionale Aufgabenpaare" nennen -, eines die Differenzierung zwischen Worten für affektive oder neutrale Einstellungen zur Zeit (Paar 2), zwei die Differenzierung abstrakter Beziehungen (Paar 3 und 8) - wir werden diese drei "abstrakt-relationale Aufgabenpaare" nennen (siehe 3.3.).

Zusammenfassung: In den beiden Test-Teilen der dritten Messung wird eine genaue Differenzierung innerhalb eines Bedeutungsfelds, eine "kategoriale Differenzierung" verlangt. Im "differenzierten Verständnistest", der aus den Erklärungen der Kinder in der zweiten Messung konstruiert wurde, variieren die sprachliche Form des Stimulus und der Vorgaben, die semantisch treffendste Vorgabe kann in einer konventionell schlechten sprachlichen Form ausgedrückt sein, die Unterschiede zwischen den Vorgaben sind nicht bewußt kalkuliert. Im "kategorialen Präzisionstest", der unter bewußter Beschränkung auf bestimmte semantische Bereiche konstruiert wurde,

¹⁾-----
Dies liegt offensichtlich an einer Eigenheit von Sprache: Je weniger die Referenten von Worten extraverbal identifizierbar sind, desto mehr müssen kategoriale Differenzierungen, die in der Sprache festgehalten sind, die Identifizierung erleichtern.

ist die sprachliche Form des Stimulus und der Vorgaben konstant, die Unterschiede zwischen den Konzepten eines Aufgabenpaares und den Vorgaben jeder Aufgabe beziehen sich auf bestimmte Bedeutungselemente - von der Art etwa wie sie in der strukturalistischen Semantik beschrieben werden (Katz/Fodor 1963). Durch den Verzicht auf Variation der sprachlichen Form und die Konzentration auf nur noch abstrakt faßliche kategoriale Differenzen stellt der "kategoriale Präzisionstest" eine genauere Messung "kategorialer Differenzierung" dar als der "differenzierte Verständnistest".¹⁾

1.2. Auswertung und Bildung von Maßen

Die dritte Messung unseres Experiments kann nicht als Test im strengen Sinne der psychologischen Testtheorie angesehen werden; die Schwierigkeit weder der einzelnen Testaufgaben noch der Antwortvorgaben wurde an einer Eichstichprobe geprüft und standardisiert. Da wir die Häufung der Antworten auf einer für beide Schichten gleichen Alternative vorausgesagt hatten, war von den theoretischen Zielvorstellungen unseres Experiments eine Standardisierung nicht so dringend erforderlich, um den zeitlichen und finanziellen Aufwand zu rechtfertigen. Wir haben die Antwortalternativen lediglich in eine Rangfolge gebracht und auf der Grundlage dieser Rangfolgen Maße entwickelt. Genau besehen, ist die Bewertung der Präzision der Antwortalternativen Aufgabe einer qualitativen Analyse mit Hilfe theoretischer Kategorien. Da wir den Versuch, ein umfassendes theoretisches Beschreibungsinstrumentarium zu entwickeln, aufgeben mußten (siehe 1.1.), kann die quantitative Bewertung in Form einer Rangfolge nur als Kompromiß angesehen werden.

1)-----

Trotz längerer Versuche ist mir keine einfachere und anspruchslosere Terminologie eingefallen. Der Leser wird anerkennen, daß irgendeine Sprachregelung zur Identifizierung und Verständigung notwendig ist. Einen weitergehenden Anspruch enthalten die vorgestellten Begriffe nicht.

Drei Bewerter beurteilten unabhängig voneinander die Rangfolge der konzeptuellen Angemessenheit der Antwortalternativen an das Stimulus-Konzept ohne Berücksichtigung der sprachlichen Form. Bis auf 5 Aufgaben (d.h. in 94,8% der Aufgaben) herrschte Übereinstimmung aller drei Bewerter über die beste Antwortalternative, in 55,2% der Aufgaben über den ersten und zweiten Rang und in 67,4% der Aufgaben über den ersten Rang und den zweiten oder dritten und den vierten oder fünften Rang; eine Produkt-Momentkorrelation aller 290 Bewertungen von 2 Bewertern ergab ,80. Die endgültige Reihenfolge wurde von allen drei Bewertern nach einer Klärung der Fälle von Nichtübereinstimmung festgestellt. (siehe Anhang E: Testbogen der dritten Messung). In einigen Fällen schien es den Bewertern auch bei der Diskussion unmöglich oder sinnlos, zwischen zwei Antwortalternativen über die bessere Angemessenheit an das Stimulus-Konzept zu entscheiden. In diesen Fällen wurden "ties", d.h. Mittelwerte kombinierter Ränge, auch für die endgültige Bewertung festgelegt: insgesamt 13 mal, aber nur in drei Fällen für den ersten und zweiten Rangplatz (bei den Aufgaben 10, 35, 52). So sehr eine vollständige Einigung über alle Rangplätze angestrebt wurde, so sehr sind die Schwierigkeiten aus der Bewertungsaufgabe selbst verständlich. Wenn man bedenkt, daß nahezu gleichwertige Umschreibungen eines Stimulus-Konzepts in eine Rangfolge konzeptueller Angemessenheit gebracht werden mußten, so ist die erzielte Übereinstimmung der Bewerter befriedigend. Zudem haben wir die möglichen Wirkungen der Nichtfestlegbarkeit von Rangfolgen kontrolliert, indem wir auch mit Vereinfachungen dieser Rangfolge, bis zur Dichotomie, Maße gebildet haben. Die Maße sollten eine Reihe von Problemen der dritten Messung ohne großen Informationsverlust berücksichtigen. Deshalb haben wir zunächst eine Vielzahl teilweise redundanter Maße gebildet, um dann nach einer empirischen Prüfung der Zusammenhänge zwischen den Maßen uns auf die wesentlichen Maße konzentrieren zu können. Im folgenden sollen zunächst die Aufgaben, denen die Maße gerecht werden sollen, analysiert, dann die Maße beschrieben und schließlich die Zusammenhänge zwischen ihnen erläutert werden. Folgende Probleme sind in die Konstruktion der Maße eingegangen:

--Bewertung der Antwortalternativen. Da wir ein im Grunde qualitatives Problem quantitativ behandeln, sollten unsinnige Differenzierungen des Rangs der Alternativen vermieden werden. Andererseits sollte die treffendste Antwort eindeutig von den übrigen Alternativen unterscheidbar sein. Wir haben die ursprüngliche Bewertung der Alternativen auf fünf Rangplätzen deshalb zu Bewertungen in drei und zwei Kategorien vereinfacht und mit jeder dieser Bewertungen Maße gebildet, zugleich aber die Bewertung der treffendsten Alternative in allen diesen Bewertungen beibehalten.¹⁾

In alle gebildeten bewertenden Maße geht daher die beste Alternative mit dem gleichen Gewicht ein, während zwischen den übrigen Alternativen unterschiedlich stark differenziert wird. Dieser Zusammenhang läßt sich schematisch folgendermaßen darstellen:

Bewertung der Antwortalternativen einer Testaufgabe in Variablengruppe (siehe Tab. 150).

(2)	(3)	(4)	(5)
Rangfolge	Drei.-Kat.	Dichotomie	Anteil
1	1	1	1
2	2	2)
3	2	2) nicht berücksichtigt
4	3	2)
5	3	2)

--Behandlung der Antwortverweigerungen. Da es sich nicht um einen Leistungstest handelte und kein Zeitdruck herrschte, erwarteten wir, daß alle Kinder alle Aufgaben bearbeiteten (siehe II.4.). Dennoch haben einige Kinder Aufgaben überflogen, vergessen oder nicht verstanden. Einerseits deutet, bei den entspannten Testbedingungen, eine Antwortverweigerung auf völlige Unkenntnis des abgefragten Konzepts, so daß sie schlechter als jede gegebene Antwort

1) -----
Dabei konnten "ties" der ursprünglichen Rangfolge genau auf der Grenze der weniger differenzierenden Kategorisierungen liegen. Das war jedoch nur bei den 3 Aufgaben (10, 35, 52) mit "ties" auf dem ersten und zweiten Rangplatz der fall. Hier gaben wir beiden Antworten in allen weniger differenzierenden Kategorisierungen die beste Kategorie "1".

bewertet werden müßte. Andererseits sollte nicht Wortkenntnis, sondern die Präzision des Verständnisses bekannter Worte getestet werden, so daß eine Antwortverweigerung der Vp. nicht angelastet werden dürfte. Wir haben uns entschlossen, bei allen bewertenden Maßen beide Überlegungen zu beachten und jeweils eine Berechnung mit Berücksichtigung der Antwortverweigerungen ("M") und ohne ("O") durchzuführen. Im ersten Fall erhielt eine verweigerte Antwort einen um einen Rang schlechteren Platz als die schlechteste vorgegebene Alternative, die Summe der Rangplätze wurde bezogen auf die für alle Individuen gleiche Anzahl möglicher Antworten (58). Im zweiten Fall erhielt eine verweigerte Antwort den Rangplatz 0, die Summe der Rangplätze wurde bezogen auf die vom jeweiligen Individuum gegebene Anzahl der Antworten. Außerdem wurden die Antwortverweigerungen gesondert ausgezählt (Variablengruppe (1)).

--Getrennte Auswertung der beiden Subtests. Alle Maße wurden für den Gesamttest, den "kategorialen Präzisionstest" und den "differenzierten Verständnistest" errechnet.

Die Maße der dritten Messung lassen sich also nach drei Dimensionen klassifizieren: Feinheit der Bewertung der Antwortalternativen, Berücksichtigung der Antwortverweigerungen, Gesamt- oder Subtest (siehe Tab. 150; im Text werden die einzelnen Variablen nur noch mit ihrer Ordnungsnummer erwähnt und Zusammenhänge zwischen ihnen erläutert). Die vorangestellten Zahlen in Klammern beziehen sich auf die Gruppierung der Maße auf der ersten Dimension.

- (1) Antwortverweigerung: Anzahlen (1, 2, 3) und Anteil (28). Außer der Gesamtzahl der Antwortverweigerungen wurde auch die Gesamtzahl der Antworten errechnet; beide müssen jedoch voll negativ miteinander korrelieren. Da der Anteil der Antworten gleich der Anzahl, bezogen auf die konstante Zahl 58, ist, muß 28 ebenso mit 1 voll negativ korrelieren. (Tab. 152) 28 wurde deshalb, wie alle anderen durch eine andere Variable voll determinierten Variablen, in Tab. 150 in Klammern gesetzt.

- (2) Mittlerer Rangplatz: Bewertung der Antwortalternativen auf 5 Rangplätzen; Antwortverweigerungen erhalten den Rangplatz 6 (4, 6, 8) oder 0 (5, 7, 9), die Summe der Rangplätze wird durch 58 (4, 6, 8) oder die von der Vp. gegebene Anzahl von Antworten geteilt (5, 7, 9). Diese Maße stellen also, wie alle folgenden, Quotienten aus der Summe der Bewertungen und der Anzahl bewerteter Aufgaben dar. Ebenso wie die beiden folgenden Variablengruppen sind diese Maße negativ: je höher der numerische Wert, desto schlechter ist die tatsächliche Leistung.
- (3) Mittelwert der Dreier-Kategorisierung: Bewertung der Antwortalternativen in drei Gütekategorien; Antwortverweigerungen erhalten die Bewertung 4 (10, 12, 14) oder 0 (11, 13, 15), die Summe der Kategorisierungen wird durch 58 (10, 12, 14) oder die von der Vp. gegebene Anzahl von Antworten geteilt (11, 13, 15).
- (4) Mittelwert der Dichotomisierung: Bewertung der Antwortalternativen in zwei Gütekategorien; Antwortverweigerungen erhalten die Bewertung 3 (16, 18, 20) oder 0 (17, 19, 21), die Summe der Kategorisierungen wird durch 58 (16, 18, 20) oder die von der Vp. gegebene Anzahl von Antworten geteilt (17, 19, 21).
- (5) Anteil (Proportion) der in die beste Kategorie fallenden Antworten: Die Anzahl der mit "1" bewerteten Antworten wird durch 58 (22, 24, 26) oder durch die von der Vp. gegebene Anzahl von Antworten geteilt (23, 25, 27), da die Bewertung der besten Antworten in allen Kategorisierungen gleich ist, hätte dieses Maß, ebenso wie in der tatsächlich benutzten Dreier-Kategorisierung, auch in der Rangplatz-Bewertung und der Dichotomisierung errechnet werden können. Im Gegensatz zu allen Maßen der Gruppen (2) bis (4) handelt es sich in der Gruppe (5) um positive Maße: je höher der Wert einer Vp., desto besser ihre Leistung; sie müssen also mit den Maßen (2) bis (4) negativ korrelieren. Anders auch als bei den Maßen (2) bis (4) ist bei (5) der Zähler bei entsprechenden "M"- und "0"-Maßen gleich, während sich der Nenner wie bei den anderen Maßen natürlich ändert. Die Maße der Gruppe (5) hängen sehr eng mit den Maßen der Gruppe (4) zusammen. Im Zähler der Quotienten der Maße der Gruppe (4) und (5) werden die besten Antworten mit "1" bewertet; der Unterschied besteht lediglich in der Bewertung der weniger treffenden Antwortalternativen und der Verweigerungen: Während sie in (4) mit "2" - bei den "0"-Maßen - oder mit "2" oder "3" - bei den "M"-Maßen - berechnet werden, werden sie in (5) überhaupt nicht berücksichtigt. Wenn die Zahl der Antworten, die jedes Individuum tatsächlich gibt, konstant ist, dann müßten die Maße der Gruppe (4) und die entsprechenden der Gruppe (5) einander voll determinieren. Das Verhältnis von "richtigen" und "falschen" Antworten innerhalb der überhaupt gegebenen Antworten ist für jede Vp. konstant, gleich ob die "falschen" Antworten noch bewertet werden oder nicht. Das ist bei den "0"-Maßen der Fall, da "richtige" und "falsche" Antworten auf den für jede Vp. konstanten Rahmen von ihr gegebener

Antworten bezogen und Antwortverweigerungen nicht mit "3" bewertet werden. Das ist bei den "M"-Maßen jedoch nicht der Fall, da innerhalb der "falschen" Antworten noch einmal nach einem vom Gesamtnenner (58) unabhängigen Merkmal, gegebene Antworten, differenziert wird in "2" oder "3"-Bewertungen. Die Maße 16, 18, 20 können daher die Maße 22, 24, 26 nicht voll determinieren, obwohl sie jeweils hoch korrelieren. Die Maße 17, 19, 21 dagegen determinieren die Maße 23, 25, 27, die letzteren sind daher in Klammern gesetzt. In jedem Fall aber bringen die Maße (5) wenig über (4) hinausgehende Information. Wir haben sie hier der Vollständigkeit halber aufgeführt, werden sie aber bei den Interpretationen nur untergeordnet berücksichtigen.

Neben die Anteile der besten Antworten haben wir die Anzahlen gesetzt (29, 30, 31). Es ist offensichtlich, daß Anzahlen und Anteile ohne Berücksichtigung der Antwortverweigerungen voll miteinander korrelieren müssen. Wir benutzen die Anzahlen bei der Darstellung der Ergebnisse gelegentlich an Stelle der Anteile, weil sie leichter vergleichbar und interpretierbar sind.

- (6) Anzahlen der in die zwei oder drei Kategorien der Zweier- bzw. Dreier-Kategorisierung fallenden Antworten. Die Maße 29, 30, 31 gehören formal auch in diese Gruppe; da sie jedoch voll mit 22, 24, 26, korrelieren, haben wir sie in die Gruppe (5) gesetzt. Die Maße A223, A233 sowie A212 und A222 sind zwar gebildet, in den Analysen aber nicht mehr verwandt worden. Nur ein Maß wird gelegentlich genannt und interpretiert: GSAZ33, 16AZ33, 42AZ33 d.h. die Anzahl der Antworten in der schlechtesten Kategorie der Dreier-Kategorisierung. Allein dieses Maß haben wir zu der gesonderten Gruppe (6) gezählt, da es eine spezifische Information gibt. In der Systematik von Tab. 150 ist es nicht genannt. Es ist natürlich unabhängig von der Behandlung der Antwortverweigerungen.

Die Maße der Gruppe (1) und (6) können wir als "Häufigkeitsmaße", die der Gruppen (2) bis (6) als "bewertende" und die der Gruppen (2) bis (5) als "richtige Antworten bewertende Maße" bezeichnen. Wenn wir die fünf vorgestellten Variablengruppen überblicken und nach der Logik ihrer Konstruktion analysieren, so lassen sich ihnen nur drei Arten von Informationen entnehmen: die Gruppe (1) gibt Auskunft über gegebene Antworten bzw. Antwortverweigerung; die Gruppen (2) bis (5) geben Auskunft sowohl über die besten Antworten wie über die Gewichtigkeit der "falschen" Antworten. Über die besten Antworten geben (2) bis (5) in der gleichen Weise Auskunft; nach dem Gewicht, das sie den "falschen" Antworten geben, aber unterscheiden sie sich. Wenn wir nun die Vielzahl der Variablen

(2) bis (5), die ja alle zu den zwei genannten Punkten gleichermaßen Informationen geben, auf die wesentlichen reduzieren wollen, so müssen drei Fragen untersucht werden: (1) Wieweit unterscheiden sich die Maße mit und ohne Berücksichtigung der Antwortverweigerung innerhalb jeder Variablen-Gruppe? (2) Wieweit hängen die Maße der vier Variablengruppen untereinander in jedem Subtest zusammen? (3) Wieweit hängen die Maße der Subtests untereinander in jeder Variablengruppe zusammen? Diese Fragen sollen zuerst mit Hilfe von Häufigkeitszählungen und Korrelationen, dann mit Hilfe einer Faktorenanalyse untersucht werden.

(Zu 1) Die Häufigkeit von Antwortverweigerungen (Tab. 151) ist in der Gesamtgruppe sehr niedrig, in der Mittelschicht etwas höher als in der Unterschicht. In jeder Schichtgruppe aber haben 13 Vp. die 58 Testaufgaben vollständig bearbeitet, die größere Anzahl der Antwortverweigerungen in der Mittelschicht ergibt sich in der Hauptsache aus dem Verhalten zweier Vp. (722, 233) im "differenzierten Verständnistest". Entsprechend verbessern sich die Mittelwerte aller Maße in der Gesamt- und den beiden Schichtgruppen durch die Nichtbewertung der Antwortverweigerungen nur geringfügig (Tab. 151). Wenn Antwortverweigerungen selten sind und nur bei wenigen Vpn. auftreten, so wird ihre Berücksichtigung die Stellung der einzelnen Vpn. zueinander kaum verändern. Diese Vermutung läßt sich an den Interkorrelationen der entsprechenden "M"- und "O"-Maße überprüfen (Tab. 154): Sie sind tatsächlich sehr hoch und erwartungsgemäß am höchsten bei den Maßen der Gruppe (5), die die Antwortverweigerungen nur einmal, in Nenner des Quotienten, berücksichtigen. In der Unterschicht gilt dies für beide Subtests, in der Mittelschicht jedoch für den 42er-Test etwas weniger: hier macht sich das häufige Schweigen der beiden erwähnten Vpn. im "differenzierten Verständnistest" bemerkbar.

Betrachtet man die Struktur der Interkorrelationen aller Maße in der Gesamtgruppe, so zeigt sich für die "bewertenden" Maße der Gruppen (2) bis (5) kaum ein Unterschied zwischen der Struktur der "M"- und "O"-Maße; dies gilt für die Gesamtgruppe (Tab. 155)

ebenso wie für beide Schichtgruppen (Tab. 157 und 158). Für die Beziehung der "Häufigkeitsmaße" der Gruppe (1) zu den Maßen (2) bis (5) ergibt sich jedoch ein differenziertes, aus der Konstruktion der Maße aber verständliches Bild: In der Gesamtgruppe korrelieren Antwortverweigerungen mäßig positiv mit schlechten Gesamtleistungen in den "M"-Maßen, aber nur geringfügig negativ in den "O"-Maßen; häufige Antworten schlagen bei den "M"-Maßen zu Buche, verschlechtern die "O"-Leistungen aber nur wenig. Da im "differenzierten Verständnistest" mehr Testaufgaben enthalten sind und auch mehr verweigert wurden, gilt dieser Zusammenhang hier stärker als im "kategorialen Präzisionstest". Die schichtspezifischen Interkorrelationen der Maße (1) mit den Maßen (2) bis (5) wiederholen bei Berücksichtigung der Antwortverweigerungen ("M") das Bild der Gesamtgruppe; ohne Berücksichtigung der Antwortverweigerungen ("O") erhält man jedoch einen Gegensatz zwischen den beiden Schichtgruppen. Während sich in der Mittelschicht der negative Zusammenhang zwischen Antworthäufigkeit und "O"-Leistungen verstärkt, ergibt sich für die Unterschicht ein ganz schwacher positiver Zusammenhang zwischen Antworthäufigkeit und "O"-Leistung.¹⁾ Wichtig ist weniger die an sich recht geringe absolute Höhe der positiven bzw. negativen Korrelationen als die Gegensätzlichkeit der Beziehungen zwischen beiden Variablen in der Unter- und Mittelschicht. Wir können vermuten, daß die Mittelschichtkinder eher bewußt auf eine Beantwortung einer Aufgabe verzichtet haben.

¹⁾-----
Außerdem finden wir einen diesem Strukturunterschied entsprechenden Niveauunterschied, der um so interessanter ist, als Unterschiede der Beziehungen von Variablen zwischen den beiden Schichtgruppen sich nicht notwendig in entsprechenden Unterschieden der Größe von Variablen zwischen den beiden Schichtgruppen niederschlagen müssen. Die Mittelschicht beantwortet weniger Testaufgaben, aber mit einer besseren Durchschnittsbewertung (siehe Tab. 151), sowohl der "M"- wie der "O"-Leistungen. Siehe dazu "Darstellung der Ergebnisse", Abschnitt 3.

(Zu 2) Wieweit müssen wir die verschiedenen "bewertenden" Maße berücksichtigen? Sie korrelieren, sobald sie sich auf den gleichen Untertest beziehen, extrem hoch untereinander, sowohl in der Gesamt- wie in den beiden Schichtgruppen. Wir können sie also als alternative Maße ansehen. Die hohe Austauschbarkeit der "bewertenden" Maße stellt für sich schon ein Ergebnis dar. Sie deutet auf einen hohen Anteil an "richtigen" Alternativen - denn hierin stimmen alle "bewertenden" Maße überein - und eine Häufung der Antworten bei bestimmten "falschen" Alternativen - denn durch eine differenziertere Bewertung werden alle Vpn. gleichmäßig besser oder schlechter beurteilt. Anders gesagt: Die Art der Fehler ist für die Unterscheidung der Vp keine wichtige Information, wenn die meisten Vp die "richtige" Antwortvorgabe oder die gleichen "falschen" wählen.

(Zu 3) Wieweit gleichen oder unterscheiden sich die Subtests, gemessen an den entsprechenden Maßen (Tab. 159)? Zunächst korreliert der Gesamttest in allen Experimentalgruppen höher mit dem 42er- als mit dem 16er-Test. Wieweit ist dies aber nur die Folge des größeren Anteils am Gesamttest oder Ergebnis von Unterschieden zwischen beiden Subtests? Die beiden Subtests korrelieren nur mäßig in der Gesamtgruppe, etwas höher in der Mittelschicht und so gut wie nicht in der Unterschicht. Während aber die Beziehung des 42er-Test zum Gesamttest in beiden Schichtgruppen konstant ist, schwankt die Korrelation zwischen Gesamttest und 16er-Test je nach den Beziehungen der Subtests zueinander. Wir glauben deshalb, daß die Maße des "kategorialen Präzisionstests" (16er Test) von den Gesamt- und 42er-Maßen in einem stärkeren Maße unabhängig sind, als man es lediglich von der geringeren Anzahl von Testaufgaben her erwarten würde.

Wir können die Ergebnisse unseres Versuchs, die Vielzahl unserer Maße mit Hilfe von Häufigkeitszählungen und Korrelationen zu reduzieren, in sechs Sätzen zusammenfassen: Erstens sind die "bewertenden" Maße mit und ohne Berücksichtigung der Antwortver-

weigerung einander gleichwertig; sie korrelieren hoch miteinander und die Struktur ihrer jeweiligen Interkorrelationen ist gleich. Zweitens ist die Beziehung der "Häufigkeits"- zu den "bewertenden" Maßen von der Berücksichtigung der Antwortverweigerung abhängig; hier hat die Antworthäufigkeit einen der Logik der Maßkonstruktion entsprechenden Einfluß. Drittens unterscheiden sich die beiden Schichtgruppen bezüglich des Zusammenhangs der Maße ohne Berücksichtigung der Antwortverweigerung mit den Antwortverweigerungen. Viertens können wir die verschiedenen "bewertenden" Maße als gleichwertige Alternativen, müssen sie aber nicht alle zusammen betrachten; zwischen den "falschen" Antworten braucht nicht mehr differenziert zu werden. Fünftens müssen die Maße für den "kategorialen Präzisionstest" gesondert, die für den "differenzierten Verständnistest" und den Gesamttest können zusammen betrachtet werden. Sechstens hängen die beiden Subtests in der Unterschicht überhaupt nicht, in der Mittelschicht aber recht stark zusammen.

Wieweit können wir ähnliche Ergebnisse in der Faktorenanalyse der Interkorrelationen der nicht-redundanten Maße wiederfinden? Von den insgesamt 31 vorgestellten Maßen sind 16 in dem Sinne nicht redundant, daß sie nicht von der Gesamtheit der übrigen Maße voll vorausgesagt werden können; "Nicht-Redundanz" ist hier also in einem statistischen Sinne, nicht aber in einem theoretischen Sinne zu verstehen. Technisch gesprochen, sind in diesem Sinne alle Variablen nicht redundant, die in einer multiplen Varianzanalyse einen standardisierten Diskriminanzfunktionskoeffizienten ungleich 0 erhalten. Die rotierte Faktorenmatrix einer Faktorenanalyse mit diesen 16 Maßen (Tab. 160) ergibt eine klare Faktorenstruktur von drei Faktoren. Auf dem ersten Faktor laden hoch positiv alle "bewertenden" Maße des Gesamt- und 42er-Test und noch geringfügig positiv alle Maße des "kategorialen Präzisionstests".¹⁾

¹⁾-----
Es sei daran erinnert, daß "16A13M" das einzige positive der "bewertenden" Maße unter den hier aufgeführten 16 Maßen ist. Es muß daher notwendig mit den anderen Maßen negativ korrelieren und entgegengesetzte Faktorenladungen aufweisen.

Auf dem zweiten Faktor laden hoch positiv die Antwortzahlen bzw. negativ die Antwortverweigerungen und geringer die Gesamtmaße mit Berücksichtigung der Antwortverweigerung. Auf dem dritten Faktor laden hoch negativ alle Maße des "kategorialen Präzisionstests", geringer negativ alle Gesamtmaße und überhaupt nicht die Maße des "differenzierten Verständnistests". Der zweite Faktor repräsentiert eindeutig unsere "Häufigkeitsmaße", der erste und dritte Faktor unsere "bewertenden" Maße. Der erste und dritte Faktor decken in unseren Maßen zwei entgegengesetzte "Fertigkeiten" auf, von denen die eine in beiden Subtests, und vorwiegend im "differenzierten Verständnistest", die andere nur im "kategorialen Präzisionstest" sowie in der Gesamtleistung bewertet wird.

Diese Faktorenstruktur reproduziert die Ergebnisse der Analyse der Korrelationen genau. Erstens differenziert sie innerhalb der "bewertenden" Maße nicht nach Antwortverweigerung. Zweitens streicht sie die Häufigkeit von Antwort und Antwortverweigerung als unabhängigen Faktor heraus. Viertens unterscheidet sie nicht zwischen den verschiedenen "bewertenden" Maßen. Fünftens belegt sie die Besonderheit des "kategorialen Präzisionstests". Zur Bestätigung unseres dritten und sechsten Ergebnisses hätten wir zwei weitere Faktorenanalysen mit den schichtspezifischen Interkorrelationen durchführen müssen.

2. Darstellung der Untersuchungsgruppe der dritten Messung.

Die Stichprobe der dritten Messung des Experiments stellt eine durch die Schichtparallelisierung bedingte Auswahl von 42 Individuen aus der Gesamtstichprobe von 392 Kindern dar.¹⁾ Es wurden Paare innerhalb des gleichen Geschlechts von Unter- und Mittelschichtkindern gebildet. Als erste Kontrollvariable verwendeten wir die Leistung auf dem ersten passiven Wortschatztest; als zweite die Gesamtleistung auf den nichtverbalen Subtests des PSB. Die nichtverbale Intelligenz wurde als zweiter Kontrollfaktor gewählt, weil sie am meisten zusätzliche und von der ersten Kontrollvariable unabhängige Information über vermutlich für die geforderte Testleistung notwendige Fähigkeiten, präzise Kategorisierungen an sprachlichem Material, liefern konnte (siehe II.3.). Eine Darstellung des "Matching-Design" gibt Tab. 161. Zwischen den Mittelwerten der 4 Gruppen für WST liegt die größte Differenz bei 0,13 z-Werten; bezüglich der Mittelwerte für NONVER ist die Gruppe der Mittelschicht-Jungen den übrigen 3 Gruppen um Werte zwischen 0.42 und 0.6 z-Werten überlegen. Die Kontrolle von NONVER ist also weniger scharf als die Kontrolle des WST.

Wie in der zweiten Messung (Kapitel III.2.1.), soll auch bei der Darstellung der Untersuchungsgruppe der 3. Messung der Experimentalplan überprüft und die Verteilung einiger weiterer Variablen in den Untersuchungsgruppen dargestellt werden. Die Überprüfung des Experimentalplans teilt sich wiederum in zwei Schritte auf. Zuerst wollen wir - anhand der Varianz- und Korrelationsanalyse - prüfen, wieweit der Schichteinfluß auf für die Experimentalleistungen wichtigen Variable ausgeschaltet werden konnte. Zweitens werden wir darstellen, wieweit durch die Auswahl auf dem Wortschatz-

¹⁾-----
Ursprünglich sollten nur 40 Individuen getestet werden. Da die dritte Messung als Gruppentest angelegt war, bildeten wir mehr als die erforderlichen 20 Match-Paare und ließen auch Vpn., die einander bezüglich der Kontrollvariablen ersetzen konnten, zum Test zu. Auf diese Weise wurden insgesamt 49 Vpn. getestet, aus denen wir ein Paar zusätzlich, d.h. insgesamt 42 Vpn., auswählten. Die 7 überzähligen Vpn. wurden in keiner statistischen Analyse berücksichtigt

test gleicher Paare die Korrelationen des Wortschatztests sich in der Experimentalstichprobe gegenüber der Gesamtstichprobe verändert haben. Die Beschreibung weiterer, nicht in den Experimentalplan eingegangener Variablen werden wir auf die differenzierte Schichteinteilung und die Schulklassen beschränken.

2.1. Kontrolle des Schichteinflusses.

Die Varianzanalyse der Kontrollvariablen WST (Tab. 162) zeigt für keinen der drei Effekte einen F-Wert über 1; die Korrelation zwischen sozialer Schicht (IFILT) und Wortschatzleistung (Tab. 165) ist von .30 ($p < .05$) in der Gesamtstichprobe auf .05 in der Experimentalstichprobe gesunken. Die Varianzanalyse der zweiten Kontrollvariablen NONVER (Tab. 163) zeigt ebenso keinen signifikanten Effekt; auch hier ist die Korrelation von .26 ($p < .05$) auf $-.01$ gesunken (Tab. 165). Durch die Parallelisierung ist der Schichteinfluß auf den Wortschatztest - wie in der zweiten Messung - und auch auf die zusätzliche Kontrollvariable NONVER völlig ausgeschaltet worden.

Wie in der zweiten Messung haben wir auch eine Reihe weiterer verbaler und nicht-verbaler Leistungen innerhalb des Experimentaldesigns überprüft. Keine der verbalen Testleistungen und Schulnoten (Tab. 162) zeigt einen signifikanten Schicht- oder Geschlechtereffekt in unserer Experimentalgruppe; für keine der verbalen Testleistungen ergibt sich eine Korrelation mit Sozial-schicht (IFILT) über .06 (Tab. 165), für die beiden Schulnoten in Deutsch nicht-signifikante Korrelationen von .14. Ebenso zeigt keine der nicht-verbaler Testleistungen einen signifikanten Schicht- oder Geschlechtereffekt in unserer Experimentalgruppe und von den übrigen Schulnoten sind nur in Kunst und Musik die Mädchen den Jungen signifikant überlegen (Tab. 163) - entsprechend den Ergebnissen in der Gesamtstichprobe (siehe Bosse u.a. 1971);

für keine der nicht-verbale Testleistungen und übrigen Schulnoten ergeben sich signifikante oder auch nur nennenswerte Korrelationen (Tab. 165).

Ein weiteres Ergebnis der Varianzanalysen muß jedoch noch kurz gestreift werden. Bei den verbalen Tests VERBIT und TEST 5 ("Wort-einfall") zeigt sich ein signifikanter Schicht-Geschlecht-Interaktionseffekt (Tab. 162): in der Unterschicht sind die Mädchen, in der Mittelschicht die Jungen überlegen. Dieses Muster läßt sich bei den TEST 1/2 ("Allgemeine Wortkenntnis"), TEST 6 ("Verstümmelte Worte") und den Deutsch-Schulnoten noch wiederentdecken, sowie bei TEST 9 ("Rechenfähigkeit") und der Note für Heimatkunde. Der Tendenz nach (allerdings auch unter .20 nicht mehr signifikant) findet es sich noch bei GESINT und NONVER. Wenn auch auf den beiden Kontrollvariablen keinerlei signifikante Effekte bei der Prüfung des Designs auftauchten, so finden wir jedoch bei Variablen, die mit der einen oder anderen Kontrollvariablen zusammenhängen, durchgängig den gleichen Schicht-Geschlecht-Interaktionseffekt. Bei der Prüfung der Hypothese, daß in der dritten Messung keine Schichtdifferenzen mehr auftauchen, werden wir deshalb den Geschlechtsfaktor mitberücksichtigen.

2.2. Beziehungen der Kontrollvariablen WST sowie der zweiten Kontrollvariablen NONVER zu anderen Variablen.

Die Korrelationen des WST mit den Intelligenzsubtests sind in der Experimentalstichprobe im Vergleich zur Gesamtstichprobe (Tab. 166) stark gesunken, und, mit Ausnahme von VERBIT und TEST 6 ("Verstümmelte Wörter"), nicht mehr signifikant. Die Korrelationen des WST mit den beiden Deutsch-Noten sind in der Experimentalstichprobe auffällig gestiegen; dies gilt für beide Schichtgruppen des Experiments. Von den vier Schichtindikatoren zeigen der Beruf des Vaters und die Ausbildung der Mutter leicht negative, die Ausbildung des Vaters und das Einkommen positive Korrelationen

mit dem Wortschatztest. Obwohl also durch die Parallelisierung der Einfluß der Sozialschicht (IFILT) auf die Wortschatzleistung ausgeschaltet wurde, blieben in der Experimentalstichprobe geringe, aber gegenläufige Korrelationen des Wortschatztests mit den einzelnen Indikatoren bestehen, aus denen der Filterindex der sozialen Schicht (IFILT) gebildet wurde. Dies wird beim Vergleich der Prädiktorengruppen Schichtindikatoren und Intelligenzsubtests in der multiplen Regressionsanalyse (3.6.3.) wichtig werden.

Die Korrelationen der zweiten Kontrollvariablen NONVER mit den Intelligenzsubtests bleiben in der Experimentalstichprobe etwa gleich hoch wie in der Gesamtstichprobe, die Korrelationen der Schulnoten sinken nur geringfügig. Bei den vier Schichtindikatoren zeigen sich zwar entgegengesetzte, aber sehr geringe Korrelationen, so daß der Zusammenhang der zweiten Kontrollvariablen NONVER mit dem Filterindex der sozialen Schicht wie seine einzelnen Dimensionen ausgeschaltet werden konnte.

Die Interkorrelationen der beiden Kontrollvariablen geben Auskunft über die Effektivität der doppelten Parallelisierung; sie lassen sich an Tab. 161 anschaulich überprüfen. Sie mußten in der Experimental- gegenüber der Gesamtstichprobe fast zwangsläufig (von .38 auf .14; Tab. 166) sinken, da bei der Parallelisierung zu einer Vp., wenn sie - was oft vorkam - ein inkonsistentes Profil beider Variablen hatte, eine zweite Vp. mit einem ähnlich inkonsistenten Profil gesucht werden mußte - was nicht immer gelang. Innerhalb der Unterschichtgruppe (.06) ist die Interkorrelation der beiden Kontrollvariablen noch stärker als innerhalb der Mittelschichtgruppe des Experiments (.22) gesunken. Für die Parallelisierung kann das nur heißen, daß durch die erlaubten Differenzen zwischen den Partnern auf der zweiten Kontrollvariablen¹⁾ der Abstand der zweiten Kontrollvariablen von der ersten sich bei den Vpn.

¹⁾-----
Maximal 1.0 Z-Werte, was jedoch bei 7 von den 21 Paaren bis zur größten Differenz von 3.0 Z-Werten überschritten werden mußte (siehe Tab. 161).

der Unterschicht häufiger vergrößert hat als bei den Vpn. der Mittelschicht. Dennoch sind die Interkorrelationen der beiden Kontrollvariablen höher als man sie erwartet hätte, wenn die zweite Variable nicht mehr kontrolliert worden wäre: Das kann ein Vergleich mit der zweiten Messung zeigen, in der NONVER nicht kontrolliert wurde und mit WST in der Gesamtgruppe .04, der Unterschichtgruppe -.01 und der Mittelschichtgruppe .07 korreliert (siehe Tab. 21).

Die Effektivität der zweifachen Kontrolle bei der Auswahl des Experimentalsamples haben wir also an drei Kriterien beurteilt. Erstens sind die Differenzen zwischen den Mittelwerten der Experimentalgruppen für NONVER größer als für WST (Tab. 161) - in dieser Hinsicht ist die Parallelisierung von NONVER nicht sehr gut gelungen. Zweitens sind die Interkorrelationen beider Kontrollvariablen nicht sehr hoch und in den beiden Schichtgruppen unterschiedlich groß (Tab. 166) - auch in dieser Hinsicht ist die Parallelisierung von NONVER nicht sehr gut gelungen. Drittens hat die Parallelisierung von NONVER trotz dieser Mängel bewirkt, daß eine Verselbständigung der verbalen Intelligenzbereiche wie in der zweiten Messung im Experimentalsample der dritten Messung nicht zu beobachten ist (Tab. 167). Insgesamt aber werden wir die zusätzliche Kontrolle wegen der genannten Mängel bei der Interpretation der Ergebnisse weitgehend ignorieren.

Die Interkorrelationen der verschiedenen Tests und Schulnoten (Tab. 167) sind wichtig für die Interpretation der Korrelationen und multiplen Regressionen der Tests und Schulnoten mit den Experimentalleistungen (Abschnitt 3.6.). Wie in der zweiten Messung sind auch hier die Interkorrelationen der Schulnoten merklich höher als die Interkorrelationen der Intelligenzsubtests.

2.3. Verteilung der Vpn. auf eine differenzierte Schichterteilung und auf die Schulen und Schulklassen.

Die Verteilung der Vpn. auf differenzierte Schichtgruppen (KFIL3) (Tab. 169) gibt Aufschluß über den Ausgleich der Unterschiede auf der ersten Kontrollvariablen durch die Parallelisierung und die soziale Distanz zwischen den Vpn. Die soziale Distanz zwischen den Untersuchungsgruppen ist, eher zugunsten der Hypothese fehlender Schichtunterschiede in der dritten Messung, nicht vergrößert worden: bei der Auswahl der Vpn. wurden die Kinder von mittleren Angestellten und Beamten (o8/12) zur "Mittleren Mittelschicht" gerechnet; die meisten Vpn. gehören Schichten des mittleren Bereichs OUS und MMS) an. Die Wortschatzleistungen wurden durch die Experimentalauswahl in den beiden unteren Schichten gleichmäßig gehoben, in den beiden oberen Schichten gleichmäßig gesenkt. In den beiden unteren Schichtgruppen entstammen die besseren Vpn. bei den Jungen der Unteren, bei den Mädchen der Oberen Unterschicht, während innerhalb der beiden oberen Schichtgruppen die besseren Vpn. sowohl bei den Jungen wie bei den Mädchen aus der Mittleren Mittelschicht kommen.

Die Verteilung der Vpn. auf Schulen und Schulklassen (Tab. 170; siehe auch IV.2.1.) zeigt eine leichte Überrepräsentation von Kindern aus Frankfurt-Goldstein im Experimentalsample, die auch die besten Leistungen auf der Kontrollvariablen WST aufweisen. Hauptsächlich wurden Kinder aus der Unterschicht und Jungen für unser Experimentalsample aus der Schule in Goldstein ausgewählt. Wenn wir die Schichtgruppen des Samples innerhalb der Schulen auf der Kontrollvariablen vergleichen, so zeigt sich, daß innerhalb der Schulen in Oberursel-Mitte und Oberursel-Süd die besseren Kinder eher aus der Mittelschicht, innerhalb der Schulen in Oberursel-Nord und Frankfurt-Goldstein aber eher aus der Unterschicht für das Experiment ausgesucht wurden; das Einzugsgebiet der Schulen Oberursel-Mitte und Oberursel-Süd ist vorwiegend mittelständisch, der Schwerpunkt des Einzugsgebiets der Schule Oberursel-Nord und Goldstein liegt in der Oberen Unterschicht (Bosse u.a. 1971:207, Tab. 5.2.).

3. Ergebnisse der dritten Messung.

3.1. Kurze Wiederholung der Hypothese der dritten Messung und ihres Zusammenhangs mit der zweiten Messung.

In der zweiten Messung wird die aktive Worterklärung geprüft, in der dritten Messung das passive Wortverständnis. Die Theorie der linguistischen Kodes bezieht sich auf Unterschiede im aktiven Gebrauch der Sprache, die unabhängig sind vom passiven Verständnis. Sie sagt daher voraus, daß die Kinder der Mittelsicht aufgrund ihrer schichtspezifischen Art des Sprachgebrauchs bessere aktive Erklärungen von Worten geben, auch wenn sie den gleichen passiven Wortschatz haben wie Kinder der Unterschicht. Sie sagt daher ebenso voraus, daß Kinder beider Schichten Sprache gleich gut verstehen, wenn sie nur die gleiche Sprache empfangen (Evokations-Hypothese). Für unsere Experimentalanordnung heißt das: Unter- und Mittelschichtkinder wählen aus einer Reihe ähnlicher Erklärungen für ein Konzept dieselbe als beste aus und die Leistung von Kindern beider Schichten hängt nur noch von ihrer verbalen und nicht-verbalen Intelligenz ab.

In der dritten Messung wird also der Einfluß sozialer Strategien des Sprachgebrauchs indirekt geprüft: durch den Fortfall der Wirkung beim Fortfall der Ursache. Solange wir also die Hypothesen des Experiments prüfen, müssen wir die in der dritten Messung erhobenen "kategorialen Differenzierung" als eine individuelle Leistung betrachten und können sie nicht wiederum als Produkt sozialer Strategien des Sprachgebrauchs interpretieren. Sobald wir jedoch die Ergebnisse unabhängig von den Ausgangshypothesen beschreiben, können wir auch die "kategoriale Differenzierung" als das Produkt sozialer Strategien des Sprachgebrauchs ansehen.

Entsprechend wollen wir bei der Darstellung der Hypothesenprüfung der dritten Messung die Experimentalleistung als individuelle Fähigkeit und erst bei den zusätzlichen Analysen (Abschnitt 3.4. bis 3.7.) auch als soziales Produkt interpretieren.

3.2. Quantitative Analyse der Ergebnisse der dritten Messung.

Bevor die Ergebnisse dargelegt werden können, müssen zwei einander widersprechende Einwände gegen die Aussagekraft des neu-konstruierten Tests erörtert werden. Auf der einen Seite kann man vermuten, daß die Wahl einer Alternative nur noch zufällig bedingt ist, da nur geringe kategoriale Unterschiede zwischen den Vorgaben, die von den Kindern hinter den auffälligeren Unterschieden der sprachlichen Form vielleicht nicht einmal wahrgenommen würden, eine Entscheidung begründen könnten. Auf der anderen Seite läßt sich einwenden, daß die Alternativen der Testaufgaben nicht vorgetestet sind, so daß alle Kinder die optimale Vorgabe wählen, weil die anderen zu ausgefallen sind. Im ersten Fall könnte der neukonstruierte Test die kategoriale Differenzierungsleistung nicht erfassen, weil er Unterscheidungen verlangt, die von den Kindern noch nicht beherrscht werden; im zweiten Fall, weil er Unterscheidungen verlangt, die längst allen Kindern selbstverständlich sind. Zu einem Teil haben wir uns gegen beide Einwände abgesichert, indem wir den Test aus den von den Kindern selbst gegebenen Erklärungen konstruiert haben. Zum anderen Teil müssen wir an den Häufigkeitsverteilungen der Antworten überprüfen, ob der eine oder der andere Einwand zutrifft, oder ob die Antworten sich bei einigen Alternativvorgaben häufen, ohne daß andere völlig ignoriert werden; wir müssen eine Vorstellung von der "Streuung" der Antworten auf die Alternativen gewinnen.

Daß beim Paarvergleich für die einzelnen Testaufgaben mit Hilfe des Wilcoxon-Tests viele Paare wegen gleicher Antworten nicht geprüft werden, zeigt, daß Kinder mit ähnlichen Wortschatzkenntnissen die gleiche Vorgabe gewählt haben. Die hohen Korrelationen der Testleistungen mit dem Wortschatztest (erste Messung) deuten in die gleiche Richtung (siehe Tab. 177, 196 und 200). Zufällig wurden die Alternativen also nicht gewählt, aber gaben die Aufgaben auch genug Anreiz, nicht optimale Alternativen zu wählen? Ein kurzer Blick auf die Verteilung der Wahlen beider Schichten für alle Aufgaben (siehe Anhang F) zeigt, daß bei fast allen Aufgaben eine Vorgabe besonders häufig gewählt wurde und dennoch die übrigen Vorgaben nicht übersehen wurden, der Test also trotz seiner Mängel für die Prüfung kategorialer Differenzierungen bei 10jährigen angemessen war. Wir haben versucht, diesen Eindruck auch quantitativ festzuhalten (Tab. 171).

Einerseits wurde bei etwa 60% der Testaufgaben die häufigste (und meist ist das die beste) Alternative von mehr als der Hälfte der Vpn. gewählt, aber nur bei noch etwa 20% der Testaufgaben von mehr als zwei Dritteln der Vpn. (Tab. 171, 1. Teil). Anders gesagt: Zwar häufen sich die Antworten bei bestimmten Alternativen, jedoch selten so massiv, daß andere Alternativen nur noch mit sehr wenigen Vpn. besetzt sein können. Auf der anderen Seite sind bei etwa 70% der Testaufgaben eine oder mehr Alternativen nicht besetzt, aber nur noch bei etwa 25% der Testaufgaben zwei oder mehr Alternativen (Tab. 171, 2. Teil). Anders gesagt: Zwar werden durchaus einzelne Alternativen übergangen, aber bei den wenigsten Aufgaben ist eine einzelne Vorgabe so attraktiv, daß zwei oder mehr andere Vorgaben nicht mehr berücksichtigt würden. Nur in 4 Aufgaben war eine (oder zwei) Alternativen so dominierend, daß die anderen praktisch nicht in Betracht kamen (13: "Streit", 24: "Liebe", 37: "Veränderung", 11: "Proviant" und "Lebensmittel"; bezeichnenderweise sind alle Aufgaben mit nur einer vorherrschenden Alternative aus dem "kategorialen Präzisionstest"). Inwiefern hängen nun die Bevorzugung einer Antwortvorgabe und die Vernachlässigung aller übrigen zusammen?

Eine solche Beziehung ergibt sich zwar in gewissem Maß aus der Form des multiple-choice Tests, aber die Abweichungen von der vollständigen Übereinstimmung kann uns eine Vorstellung von der "Streuung" der Antworten auf die vorgegebenen Alternative geben (Tab. 171, 3. Teil). Von 39 Aufgaben, bei denen eine oder mehr Alternativen unbeantwortet blieben, häufen sich bei 28 Aufgaben mehr als die Hälfte der Antworten bei einer Alternative; oder: von den 34 Aufgaben, bei denen mehr als die Hälfte der Vpn. dieselbe Alternative wählten, sind bei 28 Aufgaben eine oder mehr Vorgaben überhaupt nicht beantwortet worden. Die gleiche Überlegung kann man mit Hilfe des Vergleichs des 3. Teils der Tab. 171 mit dem 1. oder dem 2. Teil auch für das verschärfte Kriterium "zwei oder mehr nicht beantwortete Alternativen" durchführen. Aus der Nichtwahl einer bzw. zweier Alternativen folgt also nicht in jedem Fall die Überhäufung einer anderen Alternative und die häufige Beantwortung einer einzigen Alternative zieht nicht unbedingt die Nichtwahl anderer Alternativen nach sich. Auch wenn bestimmte Alternativen des Tests viele Antworten und andere gar keine hervorlocken, sind die Antworten insgesamt recht gut auf die gesamten Alternativen verteilt.

Die Unterschiede in der Verteilung der Antworten auf die Alternativen zwischen den Testformen und den Schichtgruppen (Tab. 171) leiten schon über von der Rechtfertigung unseres Tests zur Darstellung seiner Ergebnisse. Bei allen untersuchten "Streuungs"-Maßen zeigt sich eine geringere "Streuung" im "kategorialen Präzisionstest" und in der Mittelschichtgruppe, die geringste beim "kategorialen Präzisionstest" der Mittelschichtgruppe, wo 13 von 16 Aufgaben von mehr als der Hälfte der Vpn. mit derselben Alternative und zugleich mit mindestens einer anderen Alternative überhaupt nicht beantwortet wurden. Dies ist ein wichtiges Ergebnis, weil es auf die unterschiedliche Behandlung des Tests durch die beiden Schichtgruppen und die Notwendigkeit, beide Untertests getrennt darzustellen, hinweist.

Die Ergebnisse der dritten Messung nun bestätigen im wesentlichen die Hypothese fehlender Schichtunterschiede. Wir werden dies an drei verschiedenen statistischen Prüfverfahren darstellen: dem Wilcoxon-Test - als dem unseren Daten angemessensten Verfahren¹⁾ -, der Varianzanalyse und der Korrelationsrechnung. Die Verwendung des Wilcoxon-Tests bedarf einer Erläuterung bezüglich der Richtung der Hypothese und der Kontrollvariablen. Da wir die Annahme fehlender Schichtunterschiede statistisch prüfen wollen, hätte die Hypothese eigentlich zweiseitig und mit einem sehr hohen Wahrscheinlichkeitsniveau für die Zurückweisung geprüft werden müssen. Das hätte programmtechnisch zur Folge gehabt, daß die T-Werte und Signifikanzniveaus für den Fall nicht ausgedruckt worden wären, daß in einer der beiden Richtungen signifikante Unterschiede auftauchen. Aus diesem Grund haben wir allen statistischen Prüfungen die wahrscheinlichere der beiden alternativen Annahmen zugrundegelegt, daß nämlich die Mittelschicht bessere Leistungen als die Unterschicht zeige. Da sich von einem solchen Test auf zweiseitige Tests auch mit höheren p-Werten zurückschließen läßt, ist die Logik der Hypothesenprüfung durch dieses Verfahren nicht betroffen. - Da von den beiden ursprünglichen Kontrollvariablen der Wortschatztest recht hoch, die nicht-verbale Intelligenz aber nur sehr schwach mit den Leistungen der dritten Messungen korrelierte (siehe Tab. 177), haben wir Tests mit verschiedenen verbalen und nicht-verbalen Kontrollvariablen durchgeführt (Tab. 173, 174). Die Paare sind dabei innerhalb des einmal ausgewählten Samples je nach dem Wert der Kontrollvariablen neu zusammengestellt worden; nur in den Tests mit "WST-z" als Kontrollvariable sind die Paare, wie sie in Tab. 161 dargestellt sind, geprüft worden.

¹⁾-----
Zur Charakterisierung der statistischen Prüfverfahren und ihrer Angemessenheit gelten auch hier die Überlegungen innerhalb der Darstellung der Ergebnisse der zweiten Messung (siehe Kap. III. 5.1.)

Die Ergebnisse der Wilcoxon-Tests können wir kurz zusammenfassen. Erstens ergeben sich nur für einige Maße des "differenzierten Verständnistests", aber für kein Maß des "kategorialen Präzisionstests" signifikante Schichtunterschiede. Beim "differenzierten Verständnistest" - und, da der "kategoriale Präzisionstest" ja keinen für die Schichtgruppen unterschiedlichen zusätzlichen Einfluß darstellt, auch beim Gesamttest - zeigen zweitens die Maße der Zweier-Kategorisierung, gleich ob mit oder ohne Berücksichtigung der Antwortverweigerungen, keine signifikanten, aber die Signifikanz knapp verfehlende Schichtunterschiede; drittens die Maße der Rangplätze und der Dreier-Kategorisierung mit Berücksichtigung der Antwortverweigerungen keine, ohne Berücksichtigung der Antwortverweigerungen signifikante Schichtunterschiede zugunsten der Mittelschicht; viertens die Maße, die nur die schlecht bewerteten Antwortvorgaben erfassen, durchweg signifikante Schichtunterschiede zugunsten der Mittelschicht. Dieses Muster der Ergebnisse findet sich bei allen untersuchten Kontrollvariablen, sogar denen der nicht-verbale Intelligenz, allerdings mit der erstaunlichen Ausnahme der verbalen Intelligenz (VERBIT), bei der einige der genannten Unterschiede ganz verwischt sind (siehe 3.6.2. und 3.6.3.).

Entgegen der Logik der statistischen Hypothesenprüfung haben wir in der dritten Messung nicht eine der inhaltlichen Hypothese entgegengesetzte statistische Nullhypothese gebildet, die zur Beibehaltung der inhaltlichen Hypothese falsifiziert werden muß, sondern - wenn man so will - eine inhaltliche "Nullhypothese" unmittelbar geprüft. Wie oben schon erläutert, müssen wir in einem solchen Fall relativ hohe beidseitige Zufallswahrscheinlichkeiten angeben, um die inhaltliche "Nullhypothese" bestätigen zu können. Da in den Tab. 173 und 174 - aus den erwähnten programmtechnischen und sachlichen Gründen - die einseitigen Wahrscheinlichkeiten für eine Überlegenheit der Mittelschichtkinder angegeben sind, haben wir sie für die insgesamt 22 "bewertenden" Maße (d.h. alle Maße der

Tab. 173 und 174, mit Ausnahme der 3 "k.A."-Maße und von GSANT) in zweiseitige Zufallswahrscheinlichkeiten umgerechnet. Unter dem 20%-Niveau sind dann 7 der 8 Gesamtmaße, keines der 7 16er-Test-Maße und alle der 7 42er-Test-Maße signifikant; unter dem 30%-Niveau sind auch alle Gesamtmaße, aber immer noch keines der 16er-Test-Maße signifikant; das niedrigste zweiseitige Signifikanzniveau, das ein Maß des 16er-Tests erreicht, liegt bei 54% (16MS30) bzw. bei 27% für die Zufälligkeit der Überlegenheit der Mittelschicht. Wenn wir also die Testbedingungen gegen die Bestätigung der inhaltlichen "Nullhypothese" verschärfen, zeigen alle Maße des "differenzierten Verständnistests", aber kein Maß des "kategorialen Präzisionstests" Schichtunterschiede. Wir gewinnen die Anweisung, die beiden Untertests getrennt zu behandeln, verlieren aber die Einsicht in die Struktur der Ergebnisse verschiedener Variablen. Die Entscheidung über die Hypothese wird durch die unterschiedlichen Ergebnisse in beiden Untertests nicht leichter. Wir wollen in der qualitativen Analyse untersuchen, ob die Unterschiede im "differenzierten Verständnistest" auf Mängel seiner Konstruktion, Schichtvorurteile bei der Bewertung der Alternativen oder tatsächliche Unterschiede zurückgeführt werden müssen, zuvor aber die Ergebnisse der beiden anderen Prüfverfahren darstellen.

Die Varianzanalyse mit Geschlecht und dichotomer Schicht als Faktoren (Tab. 175) bestätigt genau das durch die Wilcoxon-Tests gewonnene Bild. Signifikante Schichtunterschiede finden sich nur bei der Anzahl der Antworten in der schlechtesten Kategorie der Dreier-Kategorisierung (42AZ33), fast signifikante bei GSAZ33 und bei 42MRA0. Es gibt keine signifikanten Geschlechtsunterschiede und keine signifikanten Interaktionseffekte. Auch zeigen die Schichtunterschiede innerhalb der beiden Geschlechtsgruppen (Spalte 7 und 8) und die Geschlechtsunterschiede innerhalb der beiden Schichtgruppen (Spalte 5 und 6) keine Unterschiede der Richtungen. Die Interaktion zwischen Schicht und Geschlecht auf einigen vorwiegend verbalen Testleistungen innerhalb der Untersuchungs-

gruppe (siehe Abschnitt 2., Tab. 162 und 163) ist also ohne Auswirkungen auf die Ergebnisse der dritten Messung geblieben. - Die Einwegvarianzanalyse mit der differenzierten Schichtenteilung KFIL3 (Tab. 176) ergibt keinerlei signifikante Differenzen. Immerhin wird deutlich, daß innerhalb der oberen Straten die OMS auch hier durchweg schlechter abschneidet als die MMS, während innerhalb der unteren Straten die OUS im 16er-Test, die UUS im 42er-Test bessere Leistungen zeigt. Insgesamt schneidet die MMS am besten ab. Bei den Maßen für die gewählte Anzahl schlechter Kategorien der Dreier-Kategorisierung für den 42er-Test (GSAZ42 und 42AZ33) schneiden beide unteren Straten schlechter ab als beide höheren.

Als drittes Prüfverfahren wurden die Produkt-Moment-Korrelationen der einzelnen (Tab. 177) sowie die multiple Regressionsanalyse aller Schichtindikatoren (Tab. 178) mit den Maßen der dritten Messung verwandt. Bei den Schichtindikatoren handelt es sich um mehrstufige Skalen des Berufsprestiges, der Schulabschlüsse und der Einkommen, bei der Variable IFILT - die in der multiplen Regressionsanalyse natürlich nicht verwandt wurde - um die Integration der übrigen Schichtindikatoren. Wir werden hier die Korrelationen der Schichtzugehörigkeit (IFILT) und der Kontrollvariablen WST mit den Leistungen der dritten Messung gegeneinander abwägen; auf die Bedeutung der einzelnen Schichtindikatoren werden wir in 3.6.1., auf die Bedeutung der einzelnen Intelligenzleistungen in 3.6.2. und 3.6.3. genauer eingehen. Für die Überprüfung des Schichteinflusses können wir aus den Korrelations- und Regressionsanalysen dreierlei festhalten. Erstens sind in der Gesamtgruppe die Korrelationen zwischen IFILT und den Leistungen der dritten Messung sehr gering, zwischen WST und den Leistungen der dritten Messung jedoch sehr hoch (Tab. 177); es tauchen keine signifikanten multiplen Regressionen der vier Schichtindikatoren auf die richtige Antworten bewertenden Maße der dritten Messung auf, die durch sie erklärte Varianz (R^2) ist sehr gering (Tab. 178). Daß die verbale Testleistung und nicht die Sozial-

schicht auf die Leistung der dritten Messung Einfluß hat, wird vollauf bestätigt. Zweitens werden in der Gesamtgruppe beim Gesamt- und 42er-Test die Korrelationen für jeden Schichtindikator und IFILT um so geringer, je weniger bei den Maßen der dritten Messung zwischen den falschen Antworten differenziert wird; das kann nur zu einem Teil die Folge der Notwendigkeit abnehmender Varianzen bei den weniger differenzierenden Maßen (siehe Tab. 151) sein, die ja für die Maße des 16er-Tests ebenso wie für die des 42er und des gesamten Tests gilt, ohne dort ebenso stark in abnehmenden Korrelationen zu resultieren. Drittens ergeben sich bei den Maßen für die gewählte Anzahl schlechter Kategorien (GSAZ33 und 16AZ33) nahezu signifikante multiple Regression (Tab. 178). Wenn also irgendwo Schichteinflüsse wirksam wären, dann nur bei der Wahl der schlechten Vorgaben.

Varianz- wie Korrelationsanalyse bestätigen die Ergebnisse des Wilcoxon-Tests; die Korrelationen konnten neben der Zurückweisung eines Schichteinflusses auch den tatsächlichen Einfluß des Wortschatzes nachweisen. Der Wilcoxon-Test ist aber das für Anlage und Meßniveau unseres Experiments angemessenste Prüfverfahren; deshalb wollen wir bei der folgenden Beurteilung der Ergebnisse vom Wilcoxon-Test ausgehen. Inwiefern können wir das Muster der Ergebnisse des Wilcoxon-Tests als Beleg für die Hypothese fehlender Schichtunterschiede interpretieren? Zuallererst: Es tritt nur im 42er Test, nicht aber im 16er Test auf, der ja die Fähigkeit kategorialer Differenzierung schärfer erfaßt. Statt Schichtunterschiede in dieser von beiden Untertests geprüften Fähigkeit zu akzeptieren, müssen wir also besondere, für die Unterschicht erschwerende Bedingungen im 42er Test suchen. Das wird eine Aufgabe der qualitativen Analyse sein. - Zweitens ergeben sich auch innerhalb des 42er Tests nur da Schichtunterschiede zugunsten der Mittelschicht, wo die Antwortvorgaben über die Alternative "richtig - falsch" hinaus bewertet werden; bei den Rangplätzen und der Dreier-Kategorisierung, nicht aber der Zweier-Kategorisierung. Außerdem wählen die Kinder der Unterschicht signifikant häufiger als die Kinder der Mittelschicht die schlechtesten Vorgaben nach der Dreier-

Kategorisierung. Da alle Kategorisierungen sich in der Bewertung der besten Vorgaben nicht unterscheiden, zeigen die beiden genannten Ergebnisse im Zusammenhang - ebenso wie die mit den weniger differenzierenden Maßen abnehmenden Korrelationen der Schichtindikatoren mit den Maßen des 42er und der Gesamttests -, daß das schlechtere Abschneiden der Unterschichtkinder mehr durch die häufige Wahl der am schlechtesten bewerteten Vorgaben als durch eine seltene Wahl der "richtigen" Antworten zustandekommt. Dann aber könnte es sein, daß die Bewerter - aufgrund impliziter Mittelschichtvorurteile - innerhalb der "falschen" Antwortvorgaben gerade diejenigen am schlechtesten bewertet haben, die die Unterschichtkinder am ehesten wählen,¹⁾ die auftretenden Schichtunterschiede also auf methodische Schwierigkeiten ebensogut wie auf objektive Gegebenheiten zurückzuführen sind. Kurz: Weil signifikante Schichtunterschiede nur vereinzelt und nur da auftreten, wo sie zugleich methodisch in Frage gestellt werden können, werden wir die Hypothese fehlender Schichtdifferenzen beibehalten. Wiederrum kann erst eine qualitative Analyse der Antwortvorgaben über die Leistungen beider Schichten im "differenzierten Verständnistest" entscheiden. - Drittens ergeben sich auch innerhalb des 42er Test nur dann Schichtunterschiede zugunsten der Mittelschicht, wenn die Antwortverweigerungen nicht bewertet werden.²⁾ Da die Vpn. der Mittelschicht öfter als die der Unterschicht Aufgaben nicht lösen, besonders des 42er Tests, (siehe Tab. 175, 179, 180, Variable GSANTW, Spalte 14), arbeitet die Nichtberücksichtigung der Ant-

1) -----
Dieser methodische Verdacht gegen unsere Auswertungen gilt allgemein für sozio-linguistische Untersuchungen. Solange er aber nicht - wie im genannten Fall - an spezifischen Ergebnissen konkret erhärtet werden kann, bleibt er folgenlos. Da jede "Auswertung sprachlichen Materials" nur eine Formalisierung alltäglicher Verstehens- und Bewertungsprozesses ist, wäre die einzige Konsequenz, auf sozio-linguistische Untersuchungen zu verzichten.

2) Dies Ergebnis rechtfertigt nachträglich die Verdoppelung der Maße als Lösung für das Problem der Behandlung von Antwortverweigerungen. Wie immer wir uns vorher entschieden hätten, die Bedeutung der Antwortverweigerungen gerade für die Annahme oder Ablehnung der Hypothese wäre uns entgangen und wir hätten uns mit unberechtigter Gewißheit für das eine oder andere entschieden.

wortverweigerung also zu ihren Gunsten. Es kann sein, daß die Kinder der Mittelschicht dort die Antwort verweigern, wo die Kinder der Unterschicht die schlechtesten Vorgaben wählen. Weil die Behandlung der Antwortverweigerungen auf diese Art den Bewertungen für beide Schichten einen unterschiedlichen Sinn gäbe, werden wir die Hypothese fehlender Schichtdifferenzen beibehalten.

Die Instabilität der Schichtunterschiede innerhalb der Untertests, der Kategorisierungsarten und der Behandlungsarten der Antwortverweigerungen veranlassen uns also, insgesamt keine Schichtunterschiede für die dritte Messung zu akzeptieren. Wir wollen versuchen, die aus diesen drei Argumenten gewonnenen Vermutungen noch quantitativ zu belegen. Zur Illustration der Unterschiede zwischen den Untertests und zur Prüfung des beim Vergleich der Kategorisierungsarten gewonnenen Eindrucks, die besseren Leistungen der Kinder aus der Mittelschicht ergäben sich aus der Wahl gerade der schlechtesten Alternativen bei den Kindern aus der Unterschicht, haben wir die durchschnittliche Anzahl der von jeder Gruppe gewählten Alternativen nach der Dreier- (und der Zweier-)Kategorisierung ausgerechnet (Tab. 179). Am auffälligsten ist, daß die Unterschiede zwischen den beiden Untertests für beide Schichtgruppen erheblich größer sind als die Schichtdifferenzen innerhalb des 42er Tests (siehe Tab. 179 (2)). Es wird sichtbar, daß durch die gegenüber dem "kategorialen Präzisionstest" gelockerte Form des "differenzierten Verständnistests" die Aufgabe für beide Schichten erschwert wurde, für die Unterschicht jedoch in stärkerem Maße; Ursachen hierfür müssen in der qualitativen Analyse gesucht werden. - Während sich im 16er Test die Anzahlen der gewählten Alternativen auf keinem der drei Bewertungsniveaus zwischen den Schichten unterscheiden, treten beim 42er Test - und damit beim Gesamttest - bei der

Bewertung "2" geringfügige Schichtunterschiede auf und entspricht einem Überwiegen der Mittelschicht bei der Bewertung "1" recht genau ein Überwiegen der Unterschicht auf Bewertung "3". Auch häufen sich die Antwortverweigerungen der Mittelschichtkinder im 42er Test. Das Ergebnis, daß die Schichtunterschiede des 42er Tests nur für die Maße ohne Berücksichtigung der Antwortverweigerungen nach dem Wilcoxon-Test signifikant sind, und zu einem Teil auch die Vermutung, daß die Unterschichtkinder dort die "schlechtesten" Antworten geben, wo die Mittelschichtkinder keine Antworten geben, finden hier eine Bestätigung.

Zur Illustration des Einflusses der Antwortverweigerungen auf die Signifikanz der Schichtunterschiede haben wir die Anzahl der Antwortverweigerungen in den verschiedenen Experimentalgruppen (Tab. 180) und ihre Verteilung auf die Vp und Testaufgaben (Tab. 181) ausgerechnet. Wiederum zeigt sich, daß die Kinder beider Schichtgruppen im 42er Test häufiger als im 16er Test, und die Mittelschichtkinder häufiger als die Unterschichtkinder Antworten verweigert haben (Tab. 180). Die Verteilung der Antwortverweigerungen auf die Vpn. (Tab. 181 (1)) ist in beiden Schichten ziemlich ähnlich; in der Mittelschichtgruppe finden sich jedoch mehr Vpn. mit extrem vielen Antwortverweigerungen (Unterschicht: Vp. 904 mit 5, Vp. 928 mit 5; Mittelschicht: Vp. 707 mit 5, Vp. 233 mit 7, Vp. 722 mit 12 Antwortverweigerungen), so daß die häufigeren Antwortverweigerungen in der ganzen Mittelschichtgruppe auf einzelne Vpn. zurückgeführt werden kann.¹⁾ Keine Vp. verweigert jedoch mehr als vier Aufgaben hintereinander, so daß die häufigen Verweigerungen nicht auf Überlesen, Überspringen einer Seite des Testheftes etc. zurückgeführt werden kann. Die Verteilung der Antwortverweigerungen auf die Testaufgaben (Tab. 181 (2)) zeigt in den Extremen keine Unterschiede zwischen den

¹⁾-----
Wenn nur die 12 Antwortverweigerungen der Vp. 722 ignoriert werden, so treten in beiden Schichtgruppen insgesamt 20 Antwortverweigerungen auf (siehe Tab. 180).

Schichten; die insgesamt zahlreicheren Antwortverweigerungen der Mittelschichtkinder erscheinen in der größeren Anzahl von Versuchsaufgaben wieder, auf die einzelne Mittelschichtkinder keine Antwort gegeben haben. In der Gesamtgruppe tauchten nur bei drei Testaufgaben mehr als drei Antwortverweigerungen auf (34 "Projekt": 4; 41 "Prinzip": 5; 48 "Zuwachs": 3 Antwortverweigerungen).

Was folgt aus diesen Tatsachen für die Wahl der Maße und damit auch für die Entscheidung über die Signifikanz der Schichtunterschiede? Daß nur einzelne Kinder die häufigeren Antwortverweigerungen in der Mittelschicht bewirken, legt nahe, die Maße ohne Berücksichtigung der Antwortverweigerungen in den Vordergrund zu stellen und die dort auftretenden signifikanten Schichtunterschiede zu akzeptieren; andererseits ist die größere Anzahl von Kindern, die viele Aufgaben verweigern, in der Mittelschicht ein eigenes Ergebnis, das nicht ignoriert werden kann. Daß aber keine Aufgabe von einer Schichtgruppe besonders gemieden wurde, legt nahe, die Maße mit Berücksichtigung der Antwortverweigerungen zu wählen und keine signifikanten Schichtunterschiede zu akzeptieren. Die beiden letzten Überlegungen scheinen uns mehr Gewicht zu haben als die erste, so daß wir nach wie vor die Hypothese fehlender Schichtunterschiede beibehalten wollen.

Die Ergebnisse der quantitativen Analyse sollen kurz zusammengefaßt werden; zugleich wollen wir Konsequenzen für die Aufgaben der qualitativen Analyse ziehen. (1) Rechtfertigung des Tests: Die Antworten der Kinder verteilen sich innerhalb des ganzen Tests hinreichend auf alle Vorgaben und konzentrieren sich zugleich auf einer bestimmten Vorgabe stark genug, so daß der Test als eine Prüfung von Formen kategorialer Differenzierung angesehen werden kann. Innerhalb des "kategorialen Präzisionstests" und in der Mittelschichtgruppe war die Konzentration der Antworten auf bestimmte Vorgaben, innerhalb des "differenzierten Verständnistests" und der Unterschichtgruppe die "Streuung" auf alle Vorgaben größer. (2) Prüfung der Hypothese: Da auftauchende signifikante Schichtunterschiede nicht über die Untertests, die

Kategorisierungsarten und die Arten der Berücksichtigung von Antwortverweigerungen stabil waren, halten wir die Hypothese fehlender Schichtunterschiede für bestätigt. - Aus beiden Ergebnissen folgt, daß in der qualitativen Analyse beide Untertests getrennt behandelt und Anhaltspunkte für die schlechtere Leistung der Unterschichtkinder im "differenzierten Verständnistest" gesucht werden müssen. Außerdem sollte die qualitative Analyse die Bewertung der Alternativen, die den Ergebnissen der quantitativen Analyse zugrundeliegt, überprüfen und endgültig über die Annahme oder Ablehnung der Hypothese entscheiden.

3.3. Qualitative Analyse des "kategorialen Präzisionstests".

Da im "kategorialen Präzisionstest" insgesamt sich keine Schichtunterschiede ergeben haben, kann die qualitative Analyse nur noch einander ausgleichende Unterschiede in der schichtspezifischen Schwierigkeit der Testpaare und die Stimmigkeit der Schichtdifferenzen innerhalb jedes Testpaars untersuchen. Wir haben die Ergebnisse für jedes Testpaar und jede Schichtgruppe anhand von zwei Maßen dargestellt: den mittleren Rangplätzen (Tab. 182) und der Anzahl der Vpn., die die jeweils andere Vorgabe gewählt haben (Tab. 183). Da die Ergebnisse für beide Maße recht gut übereinstimmen, werden wir sie zusammen darstellen.

Nennenswert bessere Ergebnisse zeigt die Unterschicht bei dem Testpaar "Veränderung - Wechsel" und "sich erinnern - entsinnen", die Mittelschicht bei dem Testpaar "Wunsch - Absicht" und "langweilig - langwierig". Die Schichtunterschiede innerhalb jeden Paares sind entweder für beide Aufgaben minimal oder sie weisen in die gleiche Richtung; allein bei dem Paar "Ehrlichkeit - Zuverlässigkeit" gibt es eine Umkehrung. Für die Gesamtgruppe wie für jede Schichtgruppe einzeln sind die zwei Aufgaben eines Paares von sehr unterschiedlicher Schwierigkeit (Tab. 182);

die Diskriminierbarkeit zweier gleicher Konzepte hängt also stark davon ab, ob sie im Kontext mit verwandten Worten des einen oder des anderen Konzepts stehen und ob das eine oder das andere Konzept getroffen werden soll. Angesichts der Tatsache, daß es sich in beiden Testaufgaben um das gleiche Bedeutungsfeld handelt, ist dies ein erstaunliches Ergebnis, das aber auch darauf zurückgeführt werden kann, daß der "kategoriale Präzisionstest" nicht vorgetestet ist. Sowohl in den "intentionalen" wie den "abstrakt-relationalen" Aufgabenpaaren zeigen sich keine Schichtunterschiede. Wenn überhaupt, so sind die "abstrakt-relationalen" Aufgaben für beide Schichten die schwierigeren, sowohl nach den mittleren Rängen (Tab. 182) wie der Anzahl der gewählten alternativen Vorgaben (Tab. 183). Insgesamt kann die qualitative Analyse des "kategorialen Präzisionstests" keine Anhaltspunkte für unterschiedliche Kategorisierungsprozesse des gleichen sprachlichen Materials zwischen Kindern beider Schichten geben.

3.4. Qualitative Analyse des "differenzierten Verständnistests".

3.4.1. Ziele und Instrumente der qualitativen Analyse.

Im Unterschied zum "kategorialen Präzisionstest" ergeben sich bei der quantitativen Analyse des "differenzierten Verständnistests" in beiden Schichten unterschiedliche Verteilungen der Antworten auf die Vorgaben, die, seien sie nun signifikant oder nicht, eine Erklärung verlangen. Wir wollen die Erklärung in dem Prozeß suchen, wie unterschiedliche Formen des Test-Stimulus unterschiedliche Kategorisierungen, die durch die Antwortvorgaben der Testaufgaben repräsentiert werden, hervorgerufen haben. Nach entsprechenden theoretischen Gesichtspunkten werden wir das sprachliche Material unseres Tests klassifizieren, um erstens zu versuchen, quantitativ festgestellte Schichtunterschiede durch qualitative Merkmale zu erklären und zweitens unter völligem Verzicht auf die quantitative Bewertung Unterschiede in den Kategori-

sierungsprozessen von Kindern aus beiden Schichten zu beschreiben. Die Abtrennung einer "qualitativen" von der "quantitativen" Analyse hat also sowohl das Ziel, die bisherigen Ergebnisse anhand nachträglich getroffener Unterscheidungen weiterzuverfolgen, wie auch die bisherige Bewertung der Antworten als "richtig" oder "mehr oder minder richtig" durch theoretisch besser ausgewiesene Klassifizierungen zu ersetzen und den Vorwurf versteckter Vorurteile der Bewertung zu prüfen. - Natürlich hat die "qualitative" Analyse Beschränkungen. Die neuen Klassifikationen sind ex post; die Ergebnisse meistens statistisch nicht geprüft und inhaltlich sehr begrenzt. Aber wenn die folgenden Analysen auch kaum Hypothesen belegen, so können sie doch Schichtunterschiede mit einer für die strikte Hypothesenprüfung schwer zugänglichen Anschaulichkeit illustrieren.

Drei Faktoren, die auf die Kategorisierungsprozesse der Kinder nach unseren theoretischen Überlegungen Einfluß haben, konnten durch eine qualitative Analyse des sprachlichen Materials unseres Tests erfaßt werden. Erstens wollten wir wissen, ob die sprachlogische Form der Stimuli die Leistung der Kinder und die Wahl bestimmter Vorgaben beeinflußt; deshalb haben wir die 42 Testaufgaben nach der sprachlogischen Form des Stimulus klassifiziert. Ferner interessierte uns die Frage, ob Kinder verschiedener sozialer Herkunft aus dem gleichen sprachlichen Reiz gleiche oder unterschiedliche Konzepte - und durch den beschriebenen Zirkelprozeß auch: Lernchancen - gewinnen; deshalb haben wir die 210 Antwortvorgaben - zweitens - nach der sprachlogischen Form und - drittens - der semantischen Qualität klassifiziert. Diese drei Kategorisierungen wurden von einem Auswerter, dem Verfasser des Berichts, durchgeführt. Sie behandeln, da nach der sprachlogischen Form sowohl Stimuli wie Vorgaben kategorisiert wurden, zwei theoretische Dimensionen, deren Unterscheidung kurz begründet werden soll, so daß jede vorgestellt werden kann.¹⁾

1) -----
Eine erschöpfende Darstellung der Klassifikation der Stimuli findet sich in Tab. 184 sowie im Anhang E: Testformular.

Unter der sprachlogischen Form verstehen wir grammatische Merkmale der Sprache, die die Definition von Konzepten präformieren. Unter der semantischen Qualität verstehen wir die Nähe der Vorgaben zum Bedeutungsgehalt des Stimulus; sie wird hier - anders als in der intuitiven Bewertung, durch eine Rangfolge - durch eine Analyse der logischen Beziehung zwischen Stimulus und Vorgabe erfaßt. Auch wenn sich empirisch Beziehungen zwischen bestimmten semantischen Qualitäten und sprachlogischen Formen herstellen lassen, sind beide Dimensionen analytisch voneinander unabhängig. Ein bestimmtes Bedeutungsverständnis kann prinzipiell in allen sprachlogischen Formen ausgedrückt werden und eine bestimmte sprachlogische Form sagt nichts über die semantische Genauigkeit einer Definition. Der Unterschied zwischen beiden Dimensionen liegt darin, daß die sprachlogische Form eines Stimulus oder einer Vorgabe ohne die Berücksichtigung des angesprochenen Bedeutungsfelds beurteilt werden kann, während als semantische Qualität die logische Form der semantischen Beziehung zwischen Stimulus und Vorgabe kategorisiert wird. Da das angesprochene Bedeutungsfeld durch den Stimulus definiert wird, können nach der semantischen Qualität nur die Vorgaben, nach der sprachlogischen Qualität aber sowohl die Stimuli wie die Vorgaben kategorisiert werden.

Die Klassifikation der Antwortvorgaben nach ihrer semantischen Qualität sollte die Rang-Bewertung ersetzen, indem die semantische Nähe zwischen Stimulus und Vorgabe auf einem Kategorienschema logischer Beziehungen dargestellt wird. Die Zuordnung einer Antwortvorgabe zu einer bestimmten semantischen Qualität geschieht mit Hilfe der Konstruktion einer Begriffshierarchie des durch den Stimulus angesprochenen Bedeutungsfelds. Der mengenlogische Umfang der Vorgabe wird - das ist die Modellvorstellung bei dieser Auswertung - am Umfang des Stimulus gemessen. Als Auswerter war ich mir darüber klar, daß eine Begriffshierarchie, ganz nach dem Charakter des Bedeutungsfelds, leichter oder schwerer vorstellbar ist und die Zuordnung einzelner Vorgaben, je nach dem Bedeutungs-

verständnis verschiedener Sprecher, anders ausfallen könnte, so daß eine Reliabilitätsprüfung notwendig gewesen wäre. Weil sie aus zeitlichen und personellen Gründen nicht durchgeführt werden konnte, habe ich die gleiche Kategorisierung im Abstand von zwei Wochen unabhängig voneinander zweimal durchgeführt und die kritischen Fälle verglichen. Wenigstens garantiert die Auswertung durch nur eine Person die Einheitlichkeit des semantischen Verständnisses der im Test abgefragten Bedeutungsfelder. Der Leser, der die Kategorisierung der Vorgaben nach der semantischen Qualität kritisch nachvollziehen will, wird deshalb gebeten, das ursprüngliche Testformular (Anhang E) durchzulesen, um die Vorgaben zu einem Stimulus im Zusammenhang betrachten zu können.

Klassifikationsschema der semantischen Qualität

=====

"Begriff" (B)¹⁾. "Synonym": 20C "schnaufen / ächzen". 1A "Du kriegst die Tracht Prügel zurück / heimzahlen". 43A "Stärker sein / man besiegt seinen Gegner" (Unabhängig von der sprachlichen Form des Stimulus oder der Vorgabe ist die semantische Beziehung zwischen den sinntragenden Worten für die Klassifikation entscheidend. Die Übergänge zur Unterklasse "explizite Umschreibung" sind allerdings fließend: siehe z.B. 43A und 44E.) - "Oberbegriff mit relevanter Differenz": 12D "Wolkenkratzer / ein Haus, das bis in die Wolken steigt". 30A "Kapsel / geschlossener runder Behälter". (Auch für Adjektive; 16C "Logisch / klar gedacht". 33C "Allmählich / langsam aber sicher". - "Disjunktive erschöpfende Aufzählung": 9E "Lebewesen / ein Mensch oder ein Tier". - "Explizite Umschreibung": 6A "Bedächtig / Wenn man erst überlegt, was man tut". 18D "Wenn man etwas spendet / Wenn man etwas hat, das man lieb hat, und es dem anderen gibt". 25D "Wenn man etwas hergibt oder geben muß / Wenn man den armen Leuten etwas abgibt".

¹⁾-----
Die in Klammern gesetzten Abkürzungen werden in einigen Tabellen und im Anhang verwandt.

44E "Röcheln / Wenn man leise hustet". 52D "Wenn man sehr mutig ist / daß man etwas macht, das gefährlich ist". - Die Gruppe der "Begriffe" umfaßt alle 45 in der quantitativen Bewertung auf dem ersten Rang eingestuft und 10 weitere Vorgaben, die den Stimulus sehr gut treffen, in der quantitativen Bewertung aber nur den zweiten Rangplatz erhielten, weil eine bessere Vorgabe vorlag; in diesen Fällen glaubten wir also, mit der quantitativen Bewertung besser als mit der qualitativen Kategorisierung differenzieren zu können. - Als Kriterium für die Klassifikation einer Antwortvorgabe als "Begriff" diente die Überlegung, ob sie in einer für das angesprochene Bedeutungsfeld konstruierten Begriffshierarchie den gleichen Punkt wie der Stimulus beschreibt oder nicht.

"Allgemein" (A). "Zu allgemeiner Oberbegriff": 1E "Du kriegst die Tracht Prügel zurück / Strafe". 11E "Essen, was man auf die Reise mitnimmt / Lebensmittel". 20C "Schnaufen / Wenn man ein Geräusch macht". 29B "wenn man nicht alles haben will / anständig". 42C "Wenn man etwas nicht ganz genau weiß und es irgendwie sagen muß / Wenn man meint". 49C "Beschäftigter bei einem Amt / Beruf". 50C "Wenn man etwas kauft und es unterschreibt / bedeutet, daß man etwas abgemacht hat". - Die Gruppe der "allgemeinen" Vorgaben umfaßt 36, in der quantitativen Bewertung mit Rangplätzen zwischen 2 und 5 versehene Antwortvorgaben. - Als Kriterium für die Klassifikation einer Antwortvorgabe als "Allgemein" galten zwei Bedingungen: Stimulus und Antwortvorgabe können nicht als bedeutungsgleich angesehen werden; die im Stimulus beschriebene Klasse kann als Untermenge der in der Vorgabe beschriebenen Klasse angesehen werden. Die Antwortvorgabe steht also in einer für das angesprochene Bedeutungsfeld konstruierten Begriffshierarchie auf einer höheren Ebene.

"Konkret" (K). "Zu spezifischer Unterbegriff": 23D "Musikschreiber / ein Mann der Lieder ausdenkt". 30B "Kapsel / Raketenober-
teil". 36B "Export / eine Sorte Bier". 49A "Beschäftigter bei einem Amt / ein Mann, der Geld in einer Bank ausgibt". -

"Beispiel": 8B "Wenn viele Fahrzeuge oder Menschen hintereinander stehen / soviel wie eine Armee". 15C "Verschwenden / Wenn jemand Taschengeld bekommen hat und verprascht es schnell". 19B "Tier, das bei Menschen lebt / Hund, Katze, Pferd, Kuh, Schwein". 21A "gddig oder süß / eine Katze könnte das sein. - "Nicht verallgemeinerte Umschreibung": 11B "Essen, was man auf die Reise mitnimmt / Wenn man wandern geht und Brot und Saft dabei hat". 20A "Schnaufen / Wenn man rennt und dann außer Atem ist". 31E "Wenn man auf den Rummelplatz geht und sich amüsiert / Wenn man keine Sorgen hat und tanzt und fröhlich ist". - Die Gruppe der "konkreten" Vorgaben umfaßt 54, in der quantitativen Bewertung mit Rangplätzen zwischen 2 und 5 versehenen Antwortvorgaben. - Als Kriterium für die Klassifikationen einer Antwortvorgabe als "Konkret" galten zwei Bedingungen: Stimulus und Antwortvorgabe können nicht als bedeutungsgleich angesehen werden; die in der Antwortvorgabe beschriebene Klasse kann als Untermenge der im Stimulus beschriebenen Klasse angesehen werden. Der Stimulus steht also in einer für das angesprochene Bedeutungsfeld konstruierten Begriffshierarchie auf einer höheren Ebene. - Sowohl "allgemeine" wie "konkrete" Beziehungen zwischen Stimulus und Antwortvorgaben bezeichnen ein Einschlußverhältnis und unterscheiden sich nur durch die Position des allgemeineren Begriffs in der Testaufgabe. Für die beiden letzten Gruppen von Vorgaben, die "parallelen" und "partikularen" ist ein solches Einschlußverhältnis nicht mehr konstruierbar. Die Beziehung zwischen zwei in "allgemeinem" oder "konkretem" Verhältnis zueinander stehenden sprachlichen Kategorisierungen wird "allgemein" oder "konkret", je nach dem, welche Kategorisierung in der Testaufgabe an der Stelle des Stimulus steht. Die Beziehung zwischen zwei "parallelen" oder "partikularen" sprachlichen Kategorisierungen bleibt "Parallel" oder "partikular", gleich welche Kategorisierung die Stelle des Stimulus einnimmt.

Parallel (PRL). "Begriff auf gleicher Ebene". 4C "Wenn man eine Sache zu etwas nimmt / verbrauchen". 4D "Wenn man eine Sache zu etwas nimmt / wenn man etwas ausnützt". 6B "Bedächtig / fleißig". 14B "einer, der sich großtut / einer der etwas sagt, was gar nicht stimmt". 14D "einer, der sich großtut / wenn man einem etwas aufbindet". 23E "Musikschreiber / dasselbe wie ein Dirigent". 29A "Wenn man nicht alles haben will / Wenn man nachgibt". 40E "Wenn man etwas unbedingt haben will / geizig". 47C "Alte kaputte Sachen / Dreck". 58A "Grausam / Gehässig". - Die Gruppe der "parallelen" Vorgaben umfaßt 27, in der quantitativen Bewertung mit Rangplätzen zwischen 2 und 5 versehene Antwortvorgaben. - Als Kriterium für die Klassifikation einer Antwortvorgabe als "parallel" galten drei Bedingungen: Stimulus und Antwortvorgabe können nicht als bedeutungsgleich angesehen werden; die in der Antwortvorgabe beschriebene Klasse kann als Untermenge der im Stimulus beschriebenen Klasse angesehen werden; die im Stimulus beschriebene Klasse kann als Untermenge der in der Antwortvorgabe beschriebenen Klasse angesehen werden. Stimulus und Antwortvorgabe stehen als in einer für das angesprochene Bedeutungsfeld konstruierten Begriffshierarchie auf der gleichen Ebene, so daß sie einander nicht untergeordnet werden können und als Unterbegriff eines gemeinsamen, nicht explizierten Oberbegriffs angesehen werden müssen. Da sie nicht wie "allgemeine" oder "konkrete" Vorgabe in eindeutiger Unter- oder Überordnung zum Stimulus stehen, können "parallele" Vorgaben am ehesten aus einem semantisch anderen, nicht allgemeineren oder konkreteren Verständnis des Stimulus gewählt werden.

Partikular (PTK). "singuläres nicht-kriteriales Bedeutungselement": 23A "Musikschreiber / ein Mann oder eine Frau, die ins Studio kommen". 25E "Wenn man etwas hergibt oder geben muß / Gott schenken". 33B "allmählich / Zeit". - "Exemplarischer Gebrauch": 15A "Verschwenden / Man kann seine Zeit verschwenden, indem man faul herumliegt". 16D "Logisch / Wenn ein Fall logisch ist". 20A "Schnaufen / Wenn ein Bub von einem anderen Bub gejagt wird und er dann am Schluß schnaufend bei seinen Eltern ankommt".

33C "Allmählich / Mir wird es allmählich klar". 41C "Prinzip / Er hat es im Prinzip richtig erfaßt". - "Situation": Nennung einer Situation, in der das Stimuluswort gebraucht werden kann. 16A "Logisch / wenn einer nichts kapieren will". 41A "Prinzip / Wenn man etwas erklären tut und sich nicht richtig ausdrücken kann". (Die beiden letzten Unterklassen "partikularer" Antwortvorgaben fallen durchweg mit der "inexpliziten" sprachlichen Form zusammen; das scheint uns jedoch kein Mangel der Klassifikation, sondern eine sprachliche Notwendigkeit zu sein.) - "Neudeutung von Bedeutungspartikeln": 12A "Wolkenkratzer / der Wolkenkratzer kommt vom Flugzeug". 12C "Wolkenkratzer / Regenbogen". - Die Gruppe der "partikularen" Antwortvorgaben umfaßt 38, in der quantitativen Bewertung mit Rangplätzen zwischen 3 und 5 versehene Antwortvorgaben; ähnlich wie "Begriff" nur den besseren, so decken die "partikularen" Antwortvorgaben nur den schlechteren Bereich der quantitativen Bewertung ab. - Als Kriterium für die Bewertung einer Antwortvorgabe als "Partikular" galten zwei Bedingungen: Stimulus und Antwortvorgabe können nicht als bedeutungsgleich angesehen werden; zwischen Stimulus und Antwortvorgabe kann nicht mehr, wie bei allen bisherigen Klassifikationen der "Logischen Qualität", eine kategorial hierarchische Beziehung hergestellt, sondern nur noch eine Gemeinsamkeit nicht-kriterialer Bedeutungselemente gefunden werden. Die Klasse "partikularer" Antwortvorgabe kann als eine Restgruppe bezeichnet werden, in der die Beziehungen zwischen zwei Kategorisierungen sich nicht mehr mengenlogisch, als Einschluß oder Überschneidung, konstruieren lassen.

Die Folge der Darstellung der semantischen Qualitäten ist natürlich nicht willkürlich; sie enthält verschiedene Möglichkeiten der Zusammenfassung der einzelnen semantischen Qualitäten und latent auch Bewertungen. Zum Abschluß dieser Klassifikation sollen die möglichen Zusammenfassungen skizziert und das Problem der Bewertung erläutert werden. Zuerst können die "allgemeinen" und "konkreten" den "parallelen" und "partikularen" als Vorgaben mit einer ein- bzw. zweiseitigen Beziehung zum Stimulus gegenübergestellt werden; dies Prinzip wurde bei der Darstellung benutzt. Weiterhin aber

können "begriffliche" und "allgemeine" Vorgaben von den "konkreten", "parallelen" und "partikularen" abgesetzt werden, weil die ersten allgemeinere semantische Merkmale als der Stimulus, die letzten aber nur konkretere enthalten. Schließlich können die "Begriffe" als besonders gute, die "partikularen" Vorgaben als besonders schlechte, d.h. in einem Begriffssystem optimal bzw. überhaupt nicht mehr lokalisierbare Vorgaben von den drei übrigen Gruppen unterschieden werden. Die letzte Zusammenfassung zeigt, daß dem Klassifikationsschema als solchem ein Kontinuum der Bewertung zugrundeliegt. Im Vergleich mit den Rangreihen jedoch kann bei einzelnen Aufgaben eine der semantischen Qualität nach bessere Vorgabe einen schlechteren Rangplatz als eine der semantischen Qualität nach schlechtere Vorgabe erhalten, und umgekehrt. Im Unterschied zu den Rangreihen ist die Bewertung testaufgabenübergreifend und theoretisch interpretierbar.

Die Klassifikation der Antwortvorgaben nach der sprachlogischen Form - unabhängig von der semantischen Qualität - sollte darstellen, wieweit die grammatische Form der Aufgabe gerecht wird, ein Konzept verständlich und brauchbar zu definieren. "Definitionsangemessenheit der sprachlichen Form" wäre vielleicht eine genauere Bezeichnung gewesen, wenn ihre Umständlichkeit sie nicht verboten hätte. Die Klassifikation sollte die sprachlichen Merkmale einer Definition erfassen, die sie von vornherein unabhängig von ihrer semantischen Genauigkeit als eine Definition explizit machen. Die Zuordnung einer Antwortvorgabe zu einer bestimmten sprachlogischen Form richtet sich danach, ob sie eine präzise sprachliche Kategorisierung, eine Wortklasse, oder eine Umschreibung darstellt, die bei zukünftigen eigenständigen Gebrauch nicht ohne Umformungen verwendbar ist. Die sprachlogische Qualität einer Vorgabe wird - das ist die Modellvorstellung bei dieser Auswertung - daran gemessen, wieweit sie als Definition entzifferbar, speicherbar und im aktiven Sprachgebrauch wieder verwendbar ist. Eine Wortklassen-Definition kann in einen beliebigen Kontext des aktiven Sprachgebrauchs sofort eingehen; durch den Verzicht auf weitere grammatische Mittel ist sie als

Definition erkennbar und im Lexikon speicherbar. Eine umschreibende Definition benutzt - im günstigen Fall - die Mittel eines Satzes, d.h. zusätzlich zur Nominalklasse eine Verbphrase, zur Kennzeichnung ihres Definitionscharakters. In einer solchen Form ist sie nur schwer speicherbar und nicht wieder verwendbar. Die Mittel, die für eine aktive Nutzung des Konzepts bereit liegen sollten, sind schon zu seiner Definition verbraucht. Auch das Klassifikationsschema der sprachlogischen Form enthält also eine Bewertung, deren Kriterien: Erkennbarkeit, Speicherbarkeit, Verwendbarkeit - sich jedoch aus dem Zweck der gestellten Aufgabe: Suche der besten sprachlichen Realisierung eines Konzepts - rechtfertigen lassen. Anders als bei der quantitativen Bewertung beider Untertests und der qualitativen Kategorisierung nach semantischer Qualität wird als sprachlogische Form eine spezifische Art des Gebrauchs der Sprache bewertet, die nur im "differenzierten Verständnistest" erfaßt werden konnte.¹⁾

Klassifikationsschema der sprachlogischen Form
=====

"Kurz_kategorisierend" (kk)²⁾. "Stichwort": Einzelwort, gleich welcher Wortklasse und gleich welchen semantischen und logischen Bezug zum Stimuluswort. 1A "Du kriegst deine Tracht Prügel zurück / heimzahlen". 1E "du kriegst die Tracht Prügel zurück / Strafe". 6B "bedächtig / fleißig". - "Zusammengesetztes Stichwort": Zur Konzeptbeschreibung notwendige Verbindung von Einzelworten, gleich welcher Wortklasse und gleich welchen semantischen und logischen Bezugs zum Stimuluswort. 30B "Kapsel / Raketenober- teil". 35C "Wenn eine Aufgabe schwierig ist / sehr schwer" (da eine Steigerung sich nur mit zwei Worten darstellen läßt). - Insgesamt wurden 68 Vorgaben als "kurz kategorisierend" klassifiziert.

¹⁾-----
Diese Unterscheidung wurde schon beim "vorläufigen Kategoriensystem" der zweiten Messung unter den Bezeichnungen "Ersatz"- und "Umschreibungsstrategie" benutzt (siehe III.1.)

²⁾ Die in Klammern gesetzten Abkürzungen werden in einigen Tabellen und im Anhang verwandt.

"Erweitert kategorisierend" (ek). "Durch Apposition, Adverbien oder Objekte erweitertes Stichwort": 12C "Wolkenkratzer / amerikanisches Hochhaus". 16C "Logisch / klar gedacht". 25E "Wenn man etwas hergibt oder geben muß / Gott schenken". 36D "Export / eine Ladung, die ins Ausland muß". - "Getilgtes oder erhaltenes Indefinitpronomen mit Relativsatz- oder Gerundiverweiterung": 9D "Lebewesen / was Gott geschaffen hat". 14B "einer, der sich großtut / einer, der etwas sagt, was gar nicht stimmt". 18B "Wenn man etwas spendet / etwas zum Hergeben". - "Expliziter Vergleich": 8B "wenn viele Fahrzeuge oder Menschen hintereinander stehen / soviel wie eine Armee". 23E "Musikschreiber / dasselbe wie ein Dirigent". 58C "grausam / das heißt soviel wie ein Schläger". - "Zwei nah verwandte Stichworte": 9E "Lebewesen / ein Mensch oder ein Tier". 14E "einer, der sich großtut / ein Betrüger oder ein Gauner". 55C "Wenn man unordentlich oder mit schmutzigen Kleidern rumläuft oder man ist nicht gekämmt / frech oder unartig". - Insgesamt wurden 52 Vorgaben als "erweitert kategorisierend" klassifiziert. - Die "erweitert kategorisierende" sprachlogische Form entspricht der klassischen Definitionsform durch den "nächsten Oberbegriff" und die "spezifische Differenz", da wir ja vom semantischen Gehalt absehen, d.h. von der Prüfung, ob der verwandte Oberbegriff tatsächlich der nächste und die verwandte Differenz tatsächlich die spezifische ist. Im Unterschied zu allen anderen sprachlogischen Formen verwendet die "erweitert kategorisierende" zwei getrennte Elemente: Während in der "kurz kategorisierenden" Form nur eine Wortklasse benutzt wird, enthalten "umschreibende" und "inexplizite" Formen komplexe Sätze, die eine Bedeutung ohne die Differenzierung bestimmter Elemente kennzeichnen. Insofern lassen sich "erweitert kategorisierende" Formen als "zweidimensionale" allen übrigen sprachlogischen Formen als "ein-dimensionalen" gegenüberstellen. - Die beiden bisher beschriebenen sprachlogischen Formen können an jeder ihrer Wortklasse gemäßen Satzstelle in einen Satz eingesetzt werden. Für die beiden folgenden sprachlogischen Formen gilt dies nicht mehr, da sie selber die Satzform verwenden. Insofern können wir die sprachlogischen Formen insgesamt als "kategorisierende" bzw. "nicht kategorisierende" zusammenfassen.

Umschreibend (um). Wenn-Definitionen: 1A "Bedächtig / wenn man erst überlegt, was man tut". - Objektsatz-Definition: 50C "Wenn man etwas kauft und es unterschreibt / bedeutet, daß man etwas abgemacht hat". 52C "Wenn man sehr mutig ist / daß man etwas macht, das gefährlich ist". - Insgesamt wurden 57 Vorgaben als "umschreibend" klassifiziert. - Beiden Arten "umschreibender" sprachlogischer Formen ist gemeinsam, daß sie ein Konzept nicht durch eine Wortklassen-Kategorisierung, sondern durch einen Satz beschreiben. Beide Arten sind im Lexikon schwer speicherbar und im aktiven Sprachgebrauch nicht ohne Umformung einsetzbar. In beiden Arten aber ist der Definitionszweck der Satz-Umschreibungen expliziert, durch die grammatisch korrekte Form einer Satz-Erklärung ("Objektsatz-Definition") oder durch die umgangssprachliche Anerkennung einer hochsprachlich nicht korrekten Erklärungsform ("Wenn-Definition"). Dies unterscheidet sie von der letzten der von uns klassifizierten "sprachlichen Formen":

Inexplizit (in). Satz: Satz-Erklärung ohne Hinweis auf Definitionscharakter. 6C "Bedächtig / man denkt nach". 12A "Wolkenkratzer / Der Wolkenkratzer kommt vom Flugzeug". 39B "Wenn man sehr traurig ist und sich nicht beherrschen kann / man weiß nicht, was man tun soll". - Beispiel: Aufzählung ohne Erläuterung des Beispielcharakters. 9A "Lebewesen / Fisch, Vögel, Vieh". 19B "Tier, das bei Menschen lebt / Hund, Katze, Pferd, Kuh, Schwein". - Satzfragment: Satzerweiterung ohne zugehörige Nominal- oder Verbalphrase. 18A "Wenn man etwas spendet / in der Kirche". 34A "Projekt / das ist wo man kauft". - Gebrauch: Nennung einer für das Stimulus-Wort angemessenen Situation ohne Erklärung seiner Bedeutung. 16A "Logisch / Wenn einer nichts kapieren will". 41A "Prinzip / Wenn man etwas erklären tut und sich nicht richtig ausdrücken kann". - Wiederholung: Zur Kategorisierung notwendige Wiederholung des Stimulus. 16D "Logisch / Wenn ein Fall logisch ist". 33D "Allmählich / mir wird es allmählich klar". 41C "Prinzip / er hat es im Prinzip richtig erfaßt". - Manche der genannten Beispiele können unter mehreren der genannten Unterklassen als

"inexplizite sprachliche Formen" klassifiziert werden (z.B. 33D als "Satz" und "Wiederholung"). Alle Beispiele, die sowohl als "Umschreibend" ("Wenn-Definition") wie "inexplizit" charakterisiert werden können, haben wir als "inexplizit" klassifiziert (16A, 41A, 16D), da in allen diesen Fällen die "Umschreibung" nicht ein Konzept, sondern einen Sprachgebrauch anzeigen soll. Natürlich mußten wir, um diese Unterscheidung zu treffen, semantische Überlegungen nutzen und das Prinzip durchbrechen, keine Informationen, die bei der Klassifizierung der "Logischen Qualität" verwertet wurden, in die der "sprachlichen Form" einfließen zu lassen. Aber - ganz abgesehen von der geringen Anzahl der Fälle - hätte eine rigorose Anwendung dieses Prinzips gegen den Sinn der Trennung zwischen umschreibenden, aber definierenden und den Definitionszweck nicht mehr explizierenden sprachlichen Formen verstoßen. (s.a. Anmerkungen zu "Partikular" bei "Logischer Qualität"). - Insgesamt wurden 33 Vorgaben als "inexplizit" klassifiziert.

Nach der sprachlogischen Form haben wir nicht nur die Antwortvorgaben, sondern auch die Stimuli der 42 Testaufgaben klassifiziert. Da dieselben Klassifikationsregeln angewandt wurden, kann auf eine neuerliche Vorstellung von Beispielen verzichtet werden; die Zuordnung der sprachlogischen Formen zu jedem Stimulus findet sich in Tab. 184 und im Testformular des Anhangs. Die Übertragung der Klassifikation bereitete keine Schwierigkeit, da die Stimulus ja wie die Vorgaben aus den aktiven Erklärungen der Kinder in der zweiten Messung gewonnen waren (siehe 1.1.). Die Stimuli von 16 Testaufgaben wurden als "kurz kategorisierend", von 9 als "erweitert kategorisierend", von 16 als "umschreibend" und von einer Testaufgabe als "inexplizit" kategorisiert; "umschreibende" und "inexplizite" Stimuli haben wir als "nicht kategorisierend" (nk) zusammengefaßt.

Auch die sprachlogische Form der Stimuli und Antwortvorgaben wurde nur von einem Auswerter klassifiziert. Da diese Klassifikation aber auf Merkmalen, die am gegebenen Sprechmaterial identifizierbar sind, und - bis auf die erwähnte Ausnahme der Unterscheidung zwischen "Gebrauch" und "Wiederholung" - nicht auf einem individuellen Bedeutungsverständnis beruht, kann das Fehlen der üblichen Reliabilitätsprüfungen leichter verschmerzt werden.

3.4.2. Ergebnisse der qualitativen Analyse.

Wir können die für die qualitative Analyse entwickelten Instrumente zur weiteren Untersuchung der Testaufgaben und zur genaueren Betrachtung der bevorzugt gewählten Antwortvorgaben, unabhängig von ihrer Zugehörigkeit zu bestimmten Testaufgaben, benutzen. Auf der ersten Linie werden wir (1) den Einfluß der sprachlogischen Form der Stimuli wie der richtigen Antwortvorgabe auf die quantitative Leistung jeder Sozialschichtgruppe und (2) die von jeder Schicht besonders häufig richtig beantworteten Aufgaben sowie von der anderen Schicht als Alternative zur richtigen bevorzugten Vorgaben untersuchen. Auf der zweiten Linie werden wir (3) den Einfluß der sprachlogischen Form und der (4) semantischen Qualität der Antwortvorgaben auf ihre Bevorzugung durch eine Schichtgruppe betrachten und dabei die sprachlogische Form der zu jeder Antwortvorgabe zugehörigen Stimulusvorgabe mitberücksichtigen; schließlich soll (5) der gemeinsame Einfluß von sprachlogischer Form und semantischer Qualität des Stimulus und (6) der kombinierte Einfluß aller drei kategorisierten Faktoren geprüft werden. Beide Linien sind in einer letzten Analyse (7) vertreten, in der wir Übereinstimmungen beider Schichten bei Testaufgaben und Antwortvorgaben aufdecken und mit den Unterschieden vergleichen werden. Die zweite Gruppe von Analysen verzerrt in gewissem Maße die ursprüngliche Testsituation: während dort die Vorgaben unter einen

bestimmten Stimulus gruppiert waren, werden sie hier nachträglich nach qualitativen Merkmalen unabhängig von ihrer ursprünglichen Testaufgaben-Zugehörigkeit zusammengefaßt. Da sie aber nicht mehr - wie die erste Gruppe von Analysen - die Schichtunterschiede des Testergebnisses weiter aufklären sollen, sondern die Bevorzugung bestimmter qualitativ beschreibbarer Kategorisierungen durch die eine oder andere Sozialschicht, die über den ganzen Test beiden Schichten im gleichen Maße zur Auswahl standen, haben wir diesen Mangel in Kauf genommen. - Ein weiterer Mangel ist allen folgenden Analysen gemein: sie ignorieren das spezifische in einer Testaufgabe erfragte Bedeutungsfeld. Trotz Versuchen ist es uns nicht gelungen, eine sinnvolle Zusammenfassung der 42 Testaufgaben nach rein semantischen Gesichtspunkten zu finden, insbesondere nicht die Klassifikation nach "extraverbaler Explizierbarkeit" aus der zweiten Messung auf die dritte Messung zu übertragen. Im übrigen sprechen aber auch die Ergebnisse der vergleichbaren Analysen zur ersten und zweiten Messung nicht für die Wahrscheinlichkeit des Erfolgs einer solchen Analyse (siehe Kapitel III.4.2. und 5.4.1.-3.).

Die Ausgangshypothese des Experiments besagte, daß Kinder beider Schichten aus den gleichen sprachlichen Eingaben die gleichen kognitiven Kategorisierungen entnehmen würden; die kognitiven Kategorisierungen ließen sich nur wiederum sprachlich in einem Test erfassen (siehe 1.1.). Für eine weitere Qualifizierung der Hypothese sind daher von den drei in 3.4.1. dargestellten Dimensionen die sprachliche Form des Stimulus und die semantische Qualität wichtiger als die sprachliche Form der Vorgaben. Entsprechend sind die Analysen (1), (2), und besonders (4) wichtiger als (3), (4) und (6); die Analyse (7) versucht, im Gegensatz zu allen anderen von schichtspezifischen Strategien des Sprachgebrauchs unabhängige Einflüsse auf die Testleistungen herauszuarbeiten.

(1) Einfluß der sprachlogischen Form des Stimulus und der richtigen Antwortvorgabe auf Schichtunterschiede bei einzelnen Testaufgaben. Die Unterschiede der mittleren Leistung (MRAM) wie der durchschnittlichen Anzahl richtig antwortender Vpn. schwanken beträchtlich zwischen den Testaufgaben nach der sprachlogischen Form ihres Stimulus (Tab. 184). Während die Mittelschichtkinder auf "erweitert kategorisierende" und "nicht kategorisierende" Stimuli viel leichter als Kinder aus der Unterschicht die richtige Antwortvorgabe finden, schneiden Kinder beider Schichten auf "kurz kategorisierende" Stimuli gleich gut ab. Bei "erweitert kategorisierenden" Stimuli haben Mittelschichtkinder den größten Vorsprung. Dieser Unterschied kann für das differenziertere der beiden Maße sogar statistisch gesichert werden.¹⁾ Unabhängig von der sprachlogischen Form des Stimulus sind die Mittelschichtkinder auch bei den 32 Testaufgaben (ohne Aufgabe 52) mit "kurz kategorisierender" richtiger Antwortvorgabe insgesamt überlegen: MRAM + .14 ; N (Vp.) + 1.9 (diese Zusammenfassung nicht in Tab. 184). Innerhalb der Aufgaben mit "kurz kategorisierendem" Stimulus ist die sprachlogische Form der richtigen Antwortvorgabe als zweiter Faktor wichtig: Während "kurz kategorisierende" richtige Antwortvorgaben auf "kurz kategorisierende" Stimuli von Mittelschichtkindern eher gewählt werden, schneiden Kinder aus der Unterschicht bei Aufgaben mit "kurz kategorisierenden" Stimuli, aber nicht "kurz kategorisierenden" richtigen Antwortvorgaben besser ab. Da sowohl bei den sprachlogischen Formen der Stimuli wie der richtigen Antwortvorgaben die für den Wechsel des Vorzeichens der Schichtdifferenz entscheidende Grenze zwischen "kurz kategorisierendem" und "erweitert kategorisierendem" sprachlichem Material liegt, scheint zunächst die Länge der sprachlichen

1) -----
Eine Varianzanalyse mit transformierten Schichtdifferenzen der beiden Maße (MRAM: + 1.00; N (Vp.): +10) und den drei Arten sprachlicher Formen des Stimulus als Faktor ergab für MRAM einen F-Wert von: $F(2,39) = 3.21$; $p < .05$ und für N (Vp.) einen F-Wert von: $F(2,39) = 1.73$; $p = n.s.$ Eine solche Varianzanalyse prüft, genau besehen, einen Interaktionseffekt zwischen Schicht und Form des Stimulus; korrekter wäre eine 2 x 3 Zwei-Weg-Varianzanalyse.

Äußerung das entscheidende Merkmal. Allerdings sprechen zwei andere Tatsachen gegen diese Interpretation: Zum ersten hat die "Länge" widersprechende Einflüsse; während bei den Stimuli die kürzere Form die Leistung der Unterschicht verbessert, zeigt bei den "kurz kategorisierenden" richtigen Antwortvorgaben die Mittelschicht das bessere Ergebnis. Zum zweiten steigt bei den sprachlogischen Formen der Stimuli die Leistung der Mittelschicht nicht kontinuierlich mit der "Länge" an, sondern ist am besten bei den "erweitert kategorisierenden" Stimuli und nicht, wie eine Verursachung durch die Länge nahelegen würde, bei den "nicht-kategorisierenden" Stimuli. Wir wollen diese Frage in Analyse (3) weiter verfolgen.

Was tragen die Ergebnisse der Analyse (1) zur Bewertung der quantitativen Ergebnisse und zur qualitativen Beschreibung von Unterschieden oder Übereinstimmung der Kategorisierungsprozesse in den beiden Schichten bei? Zunächst läßt sich eine Konsequenz ziehen, die für die Hypothese fehlender Schichtunterschiede spricht: Hätten wir alle Testaufgaben mit "kurz kategorisierenden" Stimuli konstruiert, so ist nicht sicher, ob wir auch die wenigen signifikanten Schichtunterschiede der quantitativen Analyse erhalten hätten. Hätten wir dagegen noch mehr "erweitert kategorisierende" Stimuli konstruiert, so hätten sich vielleicht noch mehr quantitative Ergebnisse zugunsten der Mittelschicht ergeben. Diese Überlegungen sind insofern überzogen, als sie den Einfluß der besonderen Bedeutungen der Stichworte und den Einfluß der Proportionen der verschiedenen Formen des Stimulus ignorieren. Sie decken jedoch eine mögliche Ursache und die Brüchigkeit der quantitativ-bewertenden Schichtunterschiede auf. - Weiterhin aber weisen die Ergebnisse der Analyse (1) qualitativ-bewertungsneutral auf unterschiedliche Reaktionen von Kindern beider Schichten auf sprachliche Eingaben und sprechen damit, wie wir im folgenden ausführen wollen, gegen die Ausgangshypothese.

Die gängige, kurze Beschreibung eines Konzepts in einem Wort können Kinder beider Schichten gleich gut entziffern. Mit längeren sprachlichen Erklärungen aber verbinden Kinder aus der Mittelschicht leichter die richtigen Konzepte. Der Unterschied liegt nicht in der "Länge" an sich, sondern in zwei Faktoren: Zuerst verlangen längere Äußerungen eine grammatische Entschlüsselung der Beziehung ihrer Elemente; das gilt sowohl für die "erweitert kategorisierenden" wie die "umschreibenden" Stimuli. Zweitens aber verlangen speziell die "erweitert kategorisierenden" Vorgaben eine simultane Analyse zumindest zweier Ebenen; sie sind das sprachliche Äquivalent für die klassische Form der logischen Definition durch den "nächsten Oberbegriff" und die "spezifischen Differenz". Die Überlegenheit der Mittelschichtkinder könnte mit einer besseren sprachlichen Performanz,¹⁾ und - bei den "erweitert kategorisierenden" Stimuli, bei denen Kinder aus der Mittelschicht den größten Vorsprung haben - zusätzlich mit einer größeren Vertrautheit mit der klassischen Definitionsform zusammenhängen. Beide Annahmen sind nicht unwahrscheinlich: Schichtunterschiede in grammatischen Leistungen von Kindern konnten nur für die Performanz, nicht für die Kompetenz festgestellt werden (Lenneberg 1967; Menyuk 1963, 1971). Eine größere Gewandtheit der Mittelschichtkinder bei der klassischen Definitionsform entspreche dem Vorwiegen "analytischer" kognitiver Stile bei Erwachsenen der Mittelschicht (Miller/Swanson 1961, 1963) und der größeren "Bildungsnähe" der Mittelschichtkultur.

Was immer die Erklärung sein mag, in jedem Fall ist das Ergebnis, daß eine Interaktion zwischen der sprachlichen Form des Stimulus und den Schichtdifferenzen bei der Wahl der Antwortvorgaben auftritt, ein Argument gegen die Ausgangshypothese der dritten Messung

1)-----

Keineswegs - wie man glauben könnte - mit einer größeren sprachlichen Kompetenz. Kinder beider Schichten sind prinzipiell fähig, die grammatischen Strukturen "erweitert kategorisierender" und "umschreibender" Stimuli (wie auch Vorgaben) zu entschlüsseln - das beweisen ja gerade auch die Ergebnisse der Tab. 184. Die Speicherung einer grammatischen Analyse und ihre Reaktivierung beim Vergleich mit einer weiteren grammatischen Analyse aber - Aufgaben, die durch unseren Test gestellt waren - fallen in den Bereich sprachlicher Performanz.

und gegen die zugrundeliegende Theorie der linguistischen Kodes. Wenn die Kinder der Mittelschicht auf alle Formen des Stimulus eine etwa gleich große, aber nicht signifikante Überlegenheit gezeigt hätten, so hätten wir eine bessere Test-Adaption der Mittelschichtkinder o.ä. als Ursache ansehen und die Annahme beibehalten können, daß prinzipiell Kinder beider Schichten aus dem gleichen sprachlichen Material die gleichen Konzepte gewinnen. Die signifikante Abhängigkeit der Schichtdifferenz von der Form des Stimulus widerspricht direkt einer solchen Annahme.

Am Beispiel der sprachlogischen Form des Stimulus hat sich die Analyse qualitativer Faktoren recht erfolgreich erwiesen. Die weiteren qualitativen Analysen sollen klären, ob die anfänglich festgestellten großen Unterschiede zwischen den Testaufgaben auf weitere qualitative Faktoren oder verschwiegene Vorurteile der quantitativen Bewertung zurückgeführt werden müssen.

(2) Schichtdiskriminierende Testaufgaben und alternativ bevorzugte Antwortvorgaben. Wenn bestimmte Testaufgaben von der einen oder anderen Schicht bevorzugt wurden, wie schon Tab. 184 zeigte, was kennzeichnet sie, außer der sprachlogischen Form des Stimulus und der richtigen Vorgabe? Welche alternativen Vorgaben wurden von der jeweils anderen Schichtgruppe gewählt? Wie hätte eine andere Bewertung dieser Vorgaben das quantitative Ergebnis geändert? Um diese Fragen zu beantworten, haben wir alle Testaufgaben danach kategorisiert, ob die richtige Vorgabe von Kindern aus der Unterschicht oder der Mittelschicht häufiger oder von beiden gleich häufig gewählt wurden, und zu jeder Aufgabe die von einer Schichtgruppe besonders bevorzugten Vorgaben mit ihrer Klassifikation nach "semantischer Qualität" herausgeschrieben (Tab. 185).

Zunächst einmal zeigt diese Aufstellung wiederum, daß mehr Aufgaben von der Mittelschicht besser gelöst wurden, aber auch die große Anzahl von Aufgaben, bei denen sich keine Unterschiede bei der Wahl der richtigen (23 von 42 Aufgaben) wie aller Vorgaben (14 von 42 Aufgaben) zwischen den Schichtgruppen zeigten. Zumindest

drei weitere Aussagen lassen sich aus Tab. 185 lesen: Erstens werden Aufgaben, die intentionale Bedeutungsfelder abfragen, in der Hauptsache von Kindern aus der Mittelschicht besser bearbeitet; von den 9 richtigen Vorgaben, die sich am leichtesten als "intentionale" Worte kategorisieren lassen, wurden 5 (25A, 29D, 39D, 40C, 58E) von Mittelschichtkindern häufiger gewählt, nur eine von Unterschichtkindern (1A) und zwei (31B, 46C) von Kindern aus beiden Schichten etwa gleich häufig. Zweitens wählen die Unterschichtkinder bei Aufgaben, die von Mittelschichtkindern häufiger oder von Kindern beider Schichten gleich häufig richtig beantwortet werden, sehr häufig "konkrete" Vorgaben als Alternative zur richtigen; von den 12 Aufgaben, bei denen Kinder aus der Unterschicht an Stelle der von der Mittelschicht bevorzugten richtigen Vorgabe eine andere Vorgabe gewählt haben, bevorzugten sie in 8 Fällen (10, 15, 20, 26, 27, 29, 48, 58) und von den 3 nicht schichtdifferenzierenden Aufgaben, bei denen die Unterschichtkinder bestimmte Vorgaben häufig gewählt haben, in 2 Fällen (23, 36) "konkrete" Vorgaben. Umgekehrt bevorzugen die Mittelschichtkinder an Stelle der richtigen eher "allgemeine" oder "Begriffs"-Vorgaben - wenn auch in geringerem Maße (von 12 Aufgaben 6: 1, 8, 43, 44; 36, 42). Drittens - schon das zuletzt erwähnte Ergebnis weist darauf hin - kann der quantitativen Kategorisierung kein Vorurteil zugunsten der Mittelschicht nachgesagt werden; fünfmal wurde ein "Begriff", der nicht den besten Rangplatz hatte, von der Mittelschicht bevorzugt (8, 43, 44; 9, 12)¹⁾ und von den 3 Testaufgaben, die zwei "Begriffe" mit dem besten Rangplatz aufweisen (10, 35, 52), wird bei 2 Aufgaben die Bevorzugung der einen Vorgabe durch die Mittelschicht durch die Bevorzugung der anderen durch die Unterschicht ausgeglichen. Wenn das zweite Ergebnis zeigt, daß als Alternative zu der als richtig bewerteten Vorgabe Kinder aus der Unterschicht "konkrete" und Kinder aus der Mittelschicht "allgemeine" Vorgaben bevorzugen, so hätte eine quantitative Bewertung der Vorgaben nach der kontrollierten Einschätzung des Allgemein-

1) -----
Die Schichtdifferenz bei allen 10 "Begriffs"-Vorgaben, die nicht den ersten Rangplatz erhielten, beträgt 13 Vpn. zugunsten der Mittelschicht.

heitsgrads statt nach dem intuitiven Bedeutungsverständnis der Auswerter ein eher schlechteres Ergebnis für die Kinder der Unterschicht produziert - die Bevorzugung "allgemeinere" Erklärungen als Mittelschichts-Vorurteil einmal unterstellt.¹⁾

Was tragen diese Ergebnisse zur Interpretation der quantitativen Ergebnisse bei? Neben der sprachlichen Form des Stimulus scheint noch der "intentionale" Charakter bestimmter abgefragter Bedeutungsfelder zur besseren Leistung der Mittelschichtkinder beigetragen zu haben, nicht aber eine Bevorzugung allgemeinerer Vorgaben in der quantitativen Bewertung. Auf der anderen Seite können die dargelegten Ergebnisse auch unabhängig von ihrer Aufgabe der Kontrolle der quantitativen Bewertung beurteilt werden: sie zeigen eine gewisse Bevorzugung "allgemeiner" Antwortvorgaben durch die Kinder der Mittelschicht und eine komplementäre "konkreter" Vorgaben durch die Kinder der Unterschicht bei der Wahl der Antwortvorgaben trotz besser bewerteter und von der jeweils anderen Gruppe auch häufiger gewählter Alternativen; sie beziehen sich also direkt auf gruppenspezifische Neigungen der Vpn., nicht allein auf die Qualität des gewählten sprachlichen Materials.

In Bezug auf die Ausgangshypothese sind die Ergebnisse dieser Analyse ambivalent. Einerseits spricht die Tatsache, daß in 23 der 42 Testaufgaben keine Differenzen in der Wahl der richtigen (und in 14 in der Wahl aller) Vorgaben auftauchen, für gleiches Verständnis bei Kindern beider Schichten. Andererseits gibt es offensichtlich - wenn auch quantitativ gering - schichtspezifische Strategien bei der Wahl der Alternativ-Vorgaben. Wir wollen an dieser Stelle keine Entscheidung treffen.

¹⁾-----
Ich bin mir bewußt, daß diese ganze Argumentation gegen die "qualitative" Auswertung gewendet werden kann, so als zeige sie, daß die Bewertung nach semantischer Qualität noch mittelschichtbelasteter sei als die quantitative Rang-Bewertung, da sie zudem durch ein Mitglied der "Mittelschicht" vorgenommen wurde. Richtig an diesem Einwand ist, daß man immer, auch bei der x-ten Kontrolle, mit Vorurteilen rechnen muß, ihnen nicht absolut entgehen kann. Sophistisch ist dieser Einwand, weil er über dem abstrakten Nachweis der Unumgänglichkeit von "Vorurteilen" auch in Kontrollen vergißt, die spezifischen Vorurteile der vorliegenden Kontrollverfahren, d.h. hier der Bewertung nach semantischer Qualität, zu suchen.

(3) Schichtunterschiede bei den nach sprachlogischer Form klassifizierten Antwortvorgaben aller nach der sprachlogischen Form des Stimulus klassifizierten Testaufgaben. In dieser und allen folgenden Analysen werden wir nicht mehr einzelne Testaufgaben, sondern Gruppen von Antwortvorgaben untersuchen, unabhängig von ihrer Zugehörigkeit zu bestimmten Testaufgaben; alle, nicht nur die richtigen oder besonders stark schichtdiskriminierende Vorgaben werden dabei berücksichtigt. Bei jeder folgenden Analyse werden wir zuerst die Verteilung der Antwortvorgaben auf die Gruppen betrachten, die sich durch die jeweils betrachteten Klassifikationen ergeben; die Verteilungen können auch als ex post Charakterisierungen unseres Tests angesehen werden. Danach erst werden wir die Schichtdifferenzen beschreiben; wir haben ausgezählt, wieviel von den 865 Wahlen der Vpn. aus der Unter- bzw. 856 Wahlen der Vpn. aus der Mittelschicht auf jede Gruppe der Antwortvorgaben entfallen, und die Differenz zwischen den beiden Zahlen durch die Anzahl der Vorgaben in der jeweiligen Gruppe dividiert. Die resultierenden Zahlen zeigen an, wieviel Vpn. aus der einen mehr als aus der anderen Sozialschicht eine Vorgabe der betreffenden Gruppe im Durchschnitt gewählt haben, und sind in allen folgenden, Schichtdifferenzen analysierenden Tabellen dargestellt (Tab. 187, 189, 191, 193). Wenn diese Zahlen sich auf nicht allzu kleine Vorgabengruppen beziehen, lassen sie sich, isoliert betrachtet, als Präferenz der einen oder anderen Sozialschicht für die betreffende Gruppe von Vorgaben, und im Vergleich mit Zahlen des Spalten- bzw. Zeilen-Rands, als zusätzliche Differenzierung beider Schichten bei gegebener Präferenz einer Schicht verstehen. In beiden Fällen erhalten wir Aussagen über schichtspezifische Evokationsprozesse aus dem Muster der Schichtdifferenzen und nicht

aus der Betrachtung schichtspezifischer Verteilungsmuster.¹⁾

Die Klassifikation der Antwortvorgaben nach der sprachlogischen Form ergibt eine recht gleichmäßige Verteilung innerhalb jeder sprachlogischen Form des Stimulus (Tab. 186); allein auf "nicht kategorisierende" Stimuli tauchen "umschreibende" Antwortvorgaben häufiger als bei allen Stimuli, "erweitert kategorisierende" dafür weniger auf. Wir können die Einflüsse dieser Klassifikation auf die Schichtunterschiede daher ohne Einschränkung analysieren.

1) -----
Mängel dieser Berechnung sind: (1) die Vernachlässigung der ungleichen Anzahlen der Antwortverweigerungen in beiden Schichten; die Differenz ist jedoch sehr gering (Tab. 180), (2) die ungleiche Erwartung für die Wahlen der Vorgaben, die ja unabhängig von der Zugehörigkeit zu Testaufgaben klassifiziert wurden, so daß die Anzahl von Vorgaben gleicher Klasse zwischen den Aufgaben variiert; diese Differenzen sind jedoch ebenfalls sehr gering (siehe Testformular mit Klassifikationen im Anhang) und gelten für beide Schichtgruppen in der gleichen Weise. Als eine alternative Berechnung, die Nachteil (1) vermeidet, kann man innerhalb jeder Aufgabengruppe die schichtspezifische prozentuale Verteilung der Wahlen auf die Vorgaben und die Differenzen der Prozentsätze für jede Vorgabengruppe ausrechnen. Diese Prozentsatzdifferenzen variieren jedoch in Abhängigkeit von der Größe der Vorgabengruppe, oder genauer: mit der Häufigkeit, mit der bestimmte Vorgabengruppen von beiden Schichten häufiger als andere gewählt wurden. Wir haben diese zweite Rechenverfahren für die Analysen (3) und (4) durchgeführt und für die Analyse (4) in Abb. 18 illustriert; bis auf die Verwischung der Schichtdifferenzen bei kleinen Vorgabengruppen ergeben sich zwischen beiden Verfahren keine Unterschiede. Da wir die maximale Profilierung der Schichtdifferenzen darstellen wollten, haben wir uns für das erste Rechenverfahren entschieden. Die Ähnlichkeiten beider Schichtgruppen werden auch bei diesem Verfahren noch sichtbar, zumal wir nur die größeren Schichtunterschiede interpretieren wollen. Daß es sich in keinem Fall um eine statische, d.h. auf der Zufallswahrscheinlichkeiten der Wahlen beruhende Prüfung von Ähnlichkeiten oder Differenzen handelt, versteht sich von selbst.

Unabhängig von der sprachlogischen Form der Antwortvorgaben kann die sprachlogische Form der Stimuli, da alle Vorgaben jeder Aufgabengruppe berücksichtigt wurden, keinen Einfluß auf Schichtunterschiede ausgeübt haben; die negativen Vorzeichen ergeben sich aus der größeren Anzahl von Antworten überhaupt in der Unterschicht (Tab. 187, letzte Zeile). Die sprachlogische Form der Antwortvorgaben unabhängig von der sprachlogischen Form der Stimuli jedoch (Tab. 187, letzte Spalte) differenziert deutlich zwischen den beiden Sozialschichten. Die Mittelschichtkinder bevorzugen "kurz kategorisierende", die Unterschichtkinder "erweitert kategorisierende" und, weniger stark, "umschreibende" und "inexplizite" Vorgaben. Für alle Antwortvorgaben ohne Berücksichtigung der sprachlogischen Form des Stimulus wird damit bestätigt, was sich schon für die richtigen Vorgaben auf "kurz kategorisierende" Stimuli angedeutet hatte (siehe Tab. 184). Wie aber verändert sich das Muster der Schichtdifferenzen, wenn man die Testaufgaben nach der sprachlogischen Form des Stimulus gesondert betrachtet?

Innerhalb der Testaufgaben mit "kurz kategorisierenden" Stimuli zeigt sich das allgemeine Muster nur noch prononcierter. Innerhalb der Aufgaben mit "erweitert kategorisierenden" Stimuli wählen Kinder der Unterschicht besonders häufig "erweitert kategorisierende" und innerhalb der Aufgaben mit "nicht kategorisierenden" Stimuli "umschreibende" und "inexplizite" Antwortvorgaben, während die Kinder der Mittelschicht in beiden Gruppen von Testaufgaben "kurz kategorisierende" Antwortvorgaben bevorzugen. Offensichtlich also haben die Kinder der Mittelschicht eine Vorliebe für "kurz kategorisierende" Antworten unabhängig von der Vorgabe, während die Kinder der Unterschicht auf "kurz" und "erweiternd" kategorisierende Stimuli "erweiternd kategorisierende" Vorgaben und auf "nicht kategorisierende" Stimuli "nicht kategorisierende" Vorgaben wählen, d.h. die Antwort annähernd an der Vorgabe orientieren.

Wir haben schon zu Beginn dieses Abschnitts darauf hingewiesen, daß die Ergebnisse dieses Abschnittes nicht im Rahmen der Ausgangshypothese interpretiert werden können. Sie bezieht sich auf das Verständnis des Stimulus-Konzepts als abhängige Variable, das allein durch die Klassifikation der "semantischen Qualität", nicht der "sprachlogischen Form" erfaßt wird. Dennoch können wir, unabhängig von der ursprünglichen Hypothese, die Wahl bestimmter sprachlicher Formen als eine Operationalisierung schichtspezifischer Strategien des Sprachgebrauchs ansehen. Wir wollen versuchen, die Ergebnisse über den Einfluß der sprachlichen Form des Stimulus (Analyse 1) und der sprachlichen Form der Antwortvorgaben (Analyse 3) zusammenfassend mit Hilfe von Annahmen über schichttypische Einstellungen zur Sprache zu erklären.

Offensichtlich verstehen es die Kinder aus der Mittelschicht besser, von der sprachlogischen Form der Eingabe zu abstrahieren und eine allgemeine Tendenz zur Vereinfachung einzusetzen; darauf deuten die häufige Wahl der richtigen Antworten bei "erweitert kategorisierenden" und "nicht kategorisierenden" Stimuli und bei "kurz kategorisierenden" Stimuli mit "kurz kategorisierenden" richtigen Vorgaben (Tab. 184) sowie die Bevorzugung "kurz kategorisierender" Antworten unabhängig von der sprachlogischen Form der Stimulus (Tab. 187). Kurz: Die Kinder der Mittelschicht ersetzen den Ausgangsstimulus, von seiner Form wenig beeinflusst, durch ein knappes Alternativ-Stichwort, selbst wenn weniger "kategorisierende" sprachliche Formen die Ausgangsbedeutung besser treffen.¹⁾ Sie bezahlen

1) -----
Bei den 8 Aufgaben mit "kurz kategorisierendem" Stimulus und "nicht kategorisierender" richtiger Vorgabe, bei denen die Unterschicht eine erheblich bessere Leistung zeigte (siehe Tab. 184), wählten die Kinder der Mittelschicht vorzugsweise "kategorisierende" aber nicht richtige Vorgaben (siehe Tab. 187); nämlich 16B "Typisch" +3, 30C "Tabletten für den Schnupfen" +4, 36A "was exportiert, also befördert wird" +5, 43D "Überwältigen" +5. Alle diese Vorgaben sind "kategorisierend" oder "kurz kategorisierend".

die Unabhängigkeit von der Form des sprachlichen Angebots mit dem Risiko, Bedeutungen nicht wahrzunehmen, wenn sie einer ausgefalleneren Form gespeichert werden müssen. - Die Kinder der Unterschicht dagegen sind von der sprachlogischen Form der Eingabe nicht unbeeindruckt; darauf weist die Tatsache, daß sie bei den "kurz kategorisierenden" Testaufgaben den Mittelschichtkindern ebenbürtig sind (Tab. 184). Der Bedeutung anderer sprachlogischer Formen aber vergegenwärtigen sie sich in der jeweils gleichen sprachlogischen Form (Tab. 187). Da sie keine von der sprachlogischen Form der Eingabe unabhängige Tendenz zur Antwort aufgebaut haben, fällt es ihnen nicht schwer, auf "kurz kategorisierende" Stimuli "erweitert kategorisierende" und "nicht kategorisierende" Vorgaben zu wählen, wenn diese semantisch am besten treffen (Tab. 184). Kurz: Die Kinder der Unterschicht reagieren auf schwierige und ungewohnte Formen von Stichworten mit einer Nachbildung der sprachlogischen Form, auf die einfache und gewohnte Form des Stichworts jedoch frei von Vorlieben für bestimmte Formen mit der semantisch treffendsten Antwort. Sie wiegen die Abhängigkeit von der sprachlogischen Form durch semantische Genauigkeit auf, wenn die Aufgaben in einer einfachen Form gestellt werden. In jedem Fall scheint entscheidend, ob eine verknappende und stilisierende Einstellung zur Erklärung vorliegt oder nicht.

(4) Schichtunterschiede bei den nach semantischer Qualität klassifizierten Antwortvorgaben aller nach der sprachlogischen Form des Stimulus klassifizierten Testaufgaben. Die Klassifikation der Antwortvorgaben nach der semantischen Qualität ergibt eine nicht immer gleichmäßige Verteilung innerhalb jeder sprachlogischen Form des Stimulus (Tab. 188); bei den Aufgaben mit "erweitert kategorisierenden" Stimuli tauchen besonders häufig allgemeine Vorgaben auf und, während auf "kurz kategorisierende" Stimuli "partikulare" Vorgaben auf Kosten "paralleler" besonders stark vertreten sind, ist es bei den Aufgaben mit "nicht kategorisierenden" Stimuli umgekehrt. Wir werden die Ungleichmäßigkeiten der Zellenbesetzung bei der Interpretation berücksichtigen.

Die semantische Qualität der Antwortvorgaben für alle Testaufgaben, unabhängig von der sprachlichen Form des Stimulus (Tab. 189, letzte Spalte; Abb. 18) differenziert deutlich zwischen den Schichten.¹⁾ "Begriffe" werden weitaus häufiger von Kindern der Mittelschicht gewählt; nach den Ergebnissen der quantitativen Analyse und der qualitativen Analyse (2) ist das nicht verwunderlich. Wichtiger ist, welche von den übrigen, nicht "richtigen" semantischen Qualitäten von den Kindern der Unterschicht als Alternative bevorzugt werden: Während nämlich für "allgemeine" und "partikulare" Antwortvorgaben keine Schichtunterschiede nachweisbar sind, wählen die Unterschichtkinder "konkrete" und "parallele" Vorgaben häufiger als die Mittelschichtkinder. Läßt sich dies Muster der Schichtdifferenzen wiederfinden, wenn wir die Testaufgaben nach der sprachlichen Form des Stimulus gesondert betrachten?

¹⁾-----
Die Gleichheit der Profile in Abb. 18 rührt aus den zwischen den Vorgaben verschiedener semantischer Qualität wechselnden Anzahlen der Vorgaben (siehe oben S. 371 Anm. 1: Abb. 18 illustriert die dort ausgeführte alternative Darstellungsmethode).

Innerhalb der Testaufgaben mit "kurz kategorisierendem" Stimulus wird die Vorliebe der Kinder der Unterschicht für "konkrete" Vorgaben noch deutlicher, während keine weiteren klaren Schichtunterschiede mehr auftauchen; die für den ganzen Test beobachtete Bevorzugung der "Begriffe" durch die Mittelschichtkinder verteilt sich hier auf "begriffliche", "allgemeine" und "partikulare" Vorgaben, denn Aufgaben mit "kurz kategorisierendem" Stimulus bearbeiten die Kinder der Mittelschicht nicht besser, wie wir bei der Analyse der richtigen Vorgaben (Tab. 184) gesehen haben. Innerhalb der Testaufgaben mit "erweitert kategorisierendem" Stimulus zeigen die Mittelschichtkinder eine besonders starke Überlegenheit bei der Wahl der "Begriffe" - ebenfalls eine Wiederholung der bei der Analyse der richtigen Vorgaben (Tab. 184) gewonnenen Ergebnisse -, die Kinder der Unterschicht bevorzugen dagegen "allgemeine" und "parallele" Vorgaben. Innerhalb der Testaufgaben mit "nicht kategorisierenden" Stimuli bevorzugen die Unterschichtkinder - wie bei der Gesamtheit der Aufgaben - "konkrete" und "parallele", zusätzlich aber in besonders starkem Maße auch "partikulare" Vorgaben, während die Mittelschichtkinder neben den "begrifflichen" auch "allgemeine" Vorgaben häufiger wählen. Die Bevorzugung der "partikularen" Vorgaben auf "nicht kategorisierende" Stimuli spiegelt wahrscheinlich die in der quantitativen Analyse festgestellte besonders häufige Wahl gerade der am schlechtesten bewerteten Vorgaben durch die Unterschichtkinder wider.

Was läßt sich aus diesen Ergebnissen zur Charakterisierung schichtspezifischer Evokationsprozesse gewinnen? Auch hier ist - wie in der Analyse (3) - die Abhängigkeit der Vorgaben von der sprachlogischen Form des Stimulus selber eine schichtspezifische Erscheinung, - Wie die "kurz kategorisierende" Form so bevorzugen die Mittelschichtkinder die "begriffliche" semantische Qualität der Vorgaben unabhängig von der sprachlogischen Form des Stimulus. Aus zwei Gründen aber muß diese Tendenz hier geringer veranschlagt werden: erstens stellen die "partikularen" Vorgaben auf "kurz kategorisierende" Stimuli (23 Vorgaben) und die "allgemeinen"

Vorgaben auf "nicht kategorisierende" Stimuli (14 Vorgaben) größere Ausnahmen dar. Zweitens aber ergibt sich die durchgängige Bevorzugung der "Begriffe" durch die Mittelschichtkinder aus ihrer allgemeinen Überlegenheit im "differenzierten Verständnistest"; ob man diese aber als bessere semantische Differenzierungsfähigkeit oder als Vorliebe für eine bestimmte semantische Qualität verstehen soll, ist ein müßiger Streit. - Das recht undurchsichtige Muster von der Unterschicht häufiger gewählter Vorgaben kann beschrieben werden, wenn wir "kurz kategorisierende" und "nicht kategorisierende" (d.h. "umschreibende" und "inexplizite") Stimuli als "eindimensionale" zusammenfassen, da sie nicht wie die "erweitert kategorisierenden" Stimuli eine Kategorisierung mehrerer Ebenen und deren Synthese verlangen. Die Unterschichtkinder scheinen "eindimensionale" Kategorisierung, als Aufforderung zur Konkretisierung, "mehrdimensionale" Kategorisierung aber als Aufforderung zur Verallgemeinerung zu verstehen. Dafür spricht die Bevorzugung "konkreter" auf "kurz kategorisierende" und "allgemeiner" Vorgaben auf "erweitert kategorisierende" Stimuli; beide Vorgabengruppen gehören zu den am stärksten besetzten (Tab. 188). Wenn man weiterhin anerkennt, daß "parallele" Vorgaben bei einem - für die Kinder der Unterschicht am ehesten zu unterstellenden - abweichenden Grundverständnis des angesprochenen Bedeutungsfelds als "Begriffe" oder "allgemeine" Vorgaben verstanden werden können, so fügt sich auch die Bevorzugung der "parallelen" Vorgaben auf "erweitert kategorisierende" Stimuli (Tab. 189) dieser Annahme. Schließlich scheinen auch "nicht kategorisierende" Stimuli besonders häufig zur Wahl von Vorgaben geführt zu haben, die sich nur in mehr ("konkret") oder minder ("partikular") genau bestimmten Bedeutungsteilen mit dem Stimulus decken, d.h. eher konkretisieren als verallgemeinern.

Wie immer wir das Muster der bevorzugten Wahlen beschreiben, für die Ausgangshypothese des Experiments hat es, ganz wie die Ergebnisse der vergleichbaren Analyse (1), eher negative Bedeutung. In der Analyse (1) hatte sich gezeigt, daß die Wahl der richtigen

Vorgaben von der sprachlogischen Form des Stimulus abhängt. In dieser Analyse zeigt sich nun, daß die Wahl aller Vorgaben von der sprachlogischen Form des Stimulus und - das ist das entscheidend neue Ergebnis - auch von der semantischen Qualität der Vorgaben abhängen. Offensichtlich verbinden die Kinder verschiedener Schichten, auch wenn sie im passiven Wortschatz gleich gut sind, mit dem gleichen sprachlichen Stimulus doch unterschiedliche Konzepte. Einen Eindruck von der Stärke dieses Einflusses kann ein Vergleich der Tab. 189 mit der Tab. 187 geben: Durch die Unterschiede der semantischen Qualität bedingte Schichtdifferenzen sind auffallend profilierter als durch Unterschiede der sprachlogischen Form der Vorgaben bedingte Schichtdifferenzen.

(5) Schichtunterschiede bei den nach sprachlogischer Form und semantischer Qualität klassifizierten Vorgaben aller Testaufgaben. Wenn die sprachlogische Form des Stimulus die Wahl der sprachlogischen Formen und semantischen Qualitäten der Vorgaben dirigiert, lassen sich dann die von einer Schicht typisch gewählten Vorgaben durch eine Kombination beider Dimensionen charakterisieren? Und ist diese Bevorzugung dieser doppelt charakterisierten Vorgaben ihrerseits von der sprachlogischen Form des Stimulus abhängig? Die erste Frage wollen wir in diesem Abschnitt, die zweite im folgenden untersuchen. Beide Untersuchungen unterliegen zwei zusätzlichen Beschränkungen. Erstens wird eine Aufteilung der Antwortvorgaben in noch kleinere Gruppen als bisher notwendig sein. Zweitens gruppieren wir nicht mehr zwei Klassen, Stimuli und Vorgaben, sondern eine Klasse, Vorgaben, in zwei Dimensionen. Deshalb lassen sich die Ergebnisse nicht mehr als Einfluß der einen auf die andere Dimension interpretieren. Noch weniger als die vorangegangenen stellen die folgenden Klassifikationen kontrollierte "Experimentalbedingungen" dar. Sie sind ein Suchverfahren für die Aufdeckung von Gemeinsamkeiten besonders schichtdifferenzierender Antwortvorgaben.

Die Klassifikation der Antwortvorgaben nach ihrer sprachlogischen Form und semantischen Qualität (Tab. 190) ergibt eine recht un-

gleichmäßige Verteilung. Die starke Besetzung der in beiden Dimensionen extremen ("kurz kategorisierende" "Begriffe" und "inexplizite" "partikulare" Vorgaben) und mittleren Gruppen ("umschreibend" "konkrete" Vorgaben) legt einen linearen Zusammenhang nahe, der jedoch durch die Konzentration der "parallelen" Vorgaben auf "kurz kategorisierende" sprachliche Formen durchbrochen wird. Wir glauben, daß der sich abzeichnende lineare Zusammenhang zwischen den "besseren" semantischen Qualitäten und den "besseren" sprachlogischen Formen die Möglichkeit einer unabhängigen Klassifikation beider Dimensionen nicht widerlegt, sondern selber schon ein Ergebnis darstellt. Selbst bei dem nicht umfangreichen sprachlichen Material des Tests, das nicht nach Normalität zensuriert, sondern eher mit einem Vorurteil für Ausgefallenes ausgewählt wurde, treten die Erklärungen am häufigsten auf, bei denen Form und Inhalt übereinstimmen: kurze Synonyma, umschreibende Beispiele, sprachlich konfuse Nebensächlichkeiten. Deshalb werden wir die Klassifikation trotz der unterbesetzten Zellen in der negativen Diagonale weiter analysieren. Um dennoch auch eine Verteilung mit durchweg stärker besetzten Zellen zu haben, haben wir die Klassifikation mit einer Zusammenfassung der "kategorisierenden" und "nicht kategorisierenden" Gruppen noch einmal dargestellt.

Die Kinder aus der Mittelschicht bevorzugen "kurz kategorisierende" "Begriffe" - was aus den beiden bisher festgestellten entsprechenden Einzeltendenzen resultiert - und "umschreibende" "allgemeine" Vorgaben - was den bisherigen Tendenzen widerspricht (Tab. 191; die schon besprochenen Ergebnisse der isolierten Klassifikationen sind in der Randspalte und -zeile zum Vergleich noch einmal dargestellt). Die Kinder aus der Unterschicht bevorzugen "erweitert kategorisierende" "konkrete" und "parallele" sowie "umschreibend" "partikulare" Vorgaben - in allen drei Fällen scheinen sich die Einzeltendenzen gegenseitig auffällig zu verstärken - sowie "nicht kategorisierende" Begriffe. Der große Abstand zwischen den von den Kindern der Mittelschicht bevorzugten "kategorisierenden" und den von den Kindern der Unterschicht

bevorzugten "nicht kategorisierenden" "Begriffen" unabhängig von der sprachlichen Form des Stimulus ist das auffälligste Ergebnis dieser Analyse. Selbst wenn man die 38 "kurz kategorisierenden" "Begriffe" den 16 nicht "kurz kategorisierenden" "Begriffen" gegenüberstellt, bleibt der durch die sprachlogische Form innerhalb der gleichen semantischen Qualität bewirkte Abstand recht groß (von + 1.79 bis - .31). Und auch noch wenn man ohne Berücksichtigung des einen "inexpliziten" "Begriffs", den aus der Unterschicht 7 Vp. mehr als aus der Mittelschicht beantworten (43A "stärker sein / man besiegt seinen Gegner"), "kategorisierende" und "nicht kategorisierende" "Begriffe" gegenüberstellt, bleibt zwischen den Schichtdifferenzen ein Abstand von 1.43 (zwischen +1.43 und 0.00), der bei dieser Gegenüberstellung innerhalb keiner anderen semantischen Qualität erreicht wird. - Als letztes Ergebnis sei noch die große Präferenz der Unterschichtkinder für die "kategorisierenden" gegenüber den "nicht kategorisierenden" "konkreten" Vorgaben (sowie natürlich die entgegengesetzte Vorliebe bei den Mittelschichtkindern) erwähnt; die Differenz zwischen den Schichtdifferenzen der beiden sprachlogischen Formen erreicht bei dieser semantischen Qualität mit 1.38 fast die zuletzt erwähnte bei den "Begriffen".

Normen der Hochsprache wie der gängigen Spracherziehung legen eine positiv funktionale Beziehung zwischen semantischer Qualität und sprachlogischer Form einer Erklärung fest, etwa so, daß treffende Synonyme kategorisierend, Beispiele umschreibend und semantische ungenaue Erklärungen auch sprachlich konfus sind; eine ähnliche Beziehung fanden wir ja auch in der Klassifikation unseres sprachlichen Materials (Tab. 190). Offensichtlich bevorzugen nun Kinder der Unterschicht Antwortvorgaben, bei denen sprachliche Form und logische Qualität - an dieser Norm gemessen - nicht zueinander passen: "nicht kategorisierende" "Begriffe" und "kategorisierende" "Konkreta" d.h. treffende Umschreibungen und semantisch eingeschränkte Kategorisierungen, besonders in der "erweitert kategorisierenden" Form, wo die "Erweiterung" der Form

auch als mit einer Einschränkung des Inhalts konform verstanden werden kann. Umgekehrt bevorzugen Kinder der Mittelschicht nach der genannten Norm stimmige Vorgaben: "kurz kategorisierende" "Begriffe" und, wenn überhaupt: "nicht kategorisierende" "Konkreta"; vielleicht lassen sich auch die von den Mittelschichtkindern bevorzugten "umschreibenden" "allgemeinen" Vorgaben als dieser Norm entsprechend bewerten. Kurz: Die Kinder der Mittelschicht scheinen die Vorgaben nach Regeln des üblichen Sprachgebrauchs zu bewerten und entsprechend zu bevorzugen, während die Kinder der Unterschicht keine solche Bewertung vornehmen oder zumindest andere Regeln der Stimmigkeit zwischen Form und Inhalt anwenden. Für die zweite Möglichkeit spricht ihre Präferenz für "konkrete" Vorgaben besonders der "erweitert kategorisierenden" Form; natürlich kann die Erweiterung eines Nomens oder Verbs auch als das angemessene Mittel zur Spezifizierung eingeschränkter Bedeutungen angesehen werden. Zum Schluß dieses Abschnitts soll wiederholt werden, daß alle vorgeschlagenen Interpretationen sich nicht auf schichtspezifische Funktionszusammenhänge zwischen semantischer Qualität und sprachlogischer Form beziehen können, sondern nur auf die Beurteilung von Vorgaben nach Normen des Zusammenpassens von sprachlogischer Form und semantischer Qualität, die schichtspezifisch sein können.

(6) Schichtunterschiede bei den nach sprachlogischer Form und semantischer Qualität klassifizierten Vorgaben aller nach der sprachlogischen Form des Stimulus klassifizierten Testaufgaben. Diese Analyse hat zwei Aufgaben: Erstens soll sie die Ergebnisse der vorangegangenen an den nach der sprachlogischen Form des Stimulus getrennten Testteilen überprüfen, wo dies trotz der geringen Zellenbesetzungen noch möglich ist. Zweitens soll sie, als die detaillierteste Aufgliederung der Antwortvorgaben, mögliche weitere qualitative Ursachen von Schichtunterschieden bei immer noch hinlänglich großen Vorgabengruppen aufdecken. - Die Klassifikation der Antwortvorgaben nach sprachlogischer Form und semantischer Qualität zeigt innerhalb jeder sprachlogischer Form des

Stimulus andeutungsweise die in Analyse (5) besprochene lineare Beziehung (Tab. 192).

In allen drei Gruppen von Aufgaben bevorzugen die Kinder der Mittelschicht - wie in Analyse (5) - "kurz kategorisierende" "Begriffe" und "umschreibende" "allgemeine" Vorgaben (Tab. 193); eine Inspektion der insgesamt 11 Vorgaben der letzten Gruppe zeigt, daß sie geringfügig aber durchgängig von den Mittelschichtkindern, bei keiner Vorgabe aber von den Unterschichtkindern häufiger gewählt wurden. In allen drei Gruppen von Aufgaben bevorzugen die Kinder der Unterschicht - wie in Analyse (5) - "erweitert kategorisierende" "konkrete", "erweitert kategorisierende" "parallele" und "umschreibend" "partikulare" Vorgaben; da bei den Aufgaben mit "erweitert kategorisierendem" und "nicht kategorisierendem" Stimulus nur ganz wenige "nicht kategorisierende" "Begriffe" vorgegeben waren, können wir die Bevorzugung "nicht kategorisierender Begriffe" durch die Unterschicht, die wir in Analyse (5) dargestellt haben, nur bei den Aufgaben mit "kurz kategorisierendem" Stimulus wiederfinden. Soweit also von der Verteilung der Vorgaben her sich eine Möglichkeit ergibt, finden wir die Ergebnisse der Analyse (5) in dieser Analyse bestätigt.

Auf eine ähnliche Weise wie die Ergebnisse der Analyse (5) lassen sich natürlich auch die der Analyse (3) und (4) anhand der Tab. 193 auf ihre Konsistenz überprüfen. Für die Überprüfung der Ergebnisse der Analyse (3) und der Tab. 187 muß innerhalb jeder der drei sprachlogischen Formen des Stimulus das Gesamtergebnis für eine sprachlogische Form der Antwortvorgabe über alle Qualitäten der Vorgabe verglichen werden: für auf "kurz kategorisierende" Stimuli folgende "erweitert kategorisierende" Vorgaben ergeben sich starke Schwankungen, für auf "erweitert kate-

gorisierende" Stimuli folgende "erweitert kategorisierende" Vorgaben und auf "nicht kategorisierende" Stimuli folgende "umschreibende" und "inexplizite" Vorgaben ergeben sich fast durchgängige Vorzeichen. Für die Überprüfung der Ergebnisse der Analyse (4) und der Tab. 189 muß innerhalb jeder der drei sprachlogischen Formen des Stimulus das Gesamtergebnis für eine semantische Qualität der Antwortvorgabe über alle sprachlogischen Formen der Antwortvorgabe verglichen werden: Für alle besprochenen größeren Gesamtdifferenzen zeigt sich dabei weitgehend Konsistenz -- mehr als bei den Ergebnissen der Analyse (3).

Für die zweite Aufgabe dieser Analyse - die Aufdeckung weiterer qualitativer, mit Schichtunterschieden verbundener Merkmale der Vorgaben - haben wir innerhalb jeder Aufgabengruppe die von der Mittelschicht als Alternative für die "Begriffe" bevorzugten Klassen von Antwortvorgaben analysiert. Bei den Aufgaben mit "kurz kategorisierenden" Stimulus bevorzugen die Kinder der Mittelschicht "erweitert kategorisierende" "allgemeine" und "partikulare" sowie "umschreibend" "parallele" Aufgaben; bei den Aufgaben mit "erweitert kategorisierendem" Stimulus keine Vorgaben neben den "Begriffen"; und bei den Aufgaben mit "nicht kategorisierendem" Stimulus "allgemeine" Vorgaben "kategorisierender" oder "umschreibender" sprachlogischer Form. Auch eine genauere Inspektion der entsprechenden Antwortvorgaben führte jedoch zu keiner sinnvollen Hypothese über gemeinsame schichtdifferenzierende Faktoren. - Auffällig sind schließlich in Tab. 193 die häufigen Nulldifferenzen: bei 9 der insgesamt 50 Klassen von Antwortvorgaben. Zusammen mit den auch bei stärker besetzten Vorgabengruppen manchmal recht geringen Schichtdifferenzen erinnern sie an die wichtigste Beschränkung aller in den Abschnitten (3) bis (6) dargestellten Analysen: sie beschreiben Unterschiede zwischen den Schichtgruppen, ohne sie mit den Ähnlichkeiten beider Schichtgruppen messen zu können.

Die beiden Analysen (5) und (6) können wie die Analyse (3) nicht im Rahmen der Ausgangshypothese interpretiert werden, da die Vorgaben zwar nach der semantischen Qualität zusätzlich aber nach der sprachlogischen Form klassifiziert wurden; wohl aber geben die genannten Analysen Indizien für schichtspezifische Strategien des Sprachgebrauchs, da die Wahl bestimmter sprachlogischer Formen als eine überdauernde Einstellung zur Sprache angesehen werden kann.¹⁾ Wir können die schichtspezifischen Strategien bei der Wahl bestimmter sprachlogischer Formen etwa folgendermaßen charakterisieren: Die Mittelschichtkinder orientieren sich nicht an der sprachlogischen Form des Stimulus und haben eine allgemeine Neigung für "kurz kategorisierende" Vorgaben (Analyse 3) oder für Vorgaben, deren sprachlogische Form mit der semantischen Qualität nach mutmaßlichen hochsprachlichen Normen übereinstimmen (Analyse 6). Die Unterschichtkinder sind an der sprachlichen Form der Vorgabe orientiert und haben eine allgemeine Neigung für Vorgaben, deren sprachlogische Form der des Stimulus entspricht (Analyse 3), oder für Vorgaben, bei denen sprachlogische Form und semantische Qualität wenig übereinstimmen (Analyse 6). Beide Dimensionen - Stimulusunabhängigkeit und stilisierte Vorgabenwahl in der Mittelschicht, Stimulusabhängigkeit und nicht-stilisierte Vorgabenwahl in der Unterschicht - zeigen, daß die Mittelschichtkinder die Testaufgabe mit Hilfe der Analyse ihrer abstrakten Form lösen, die Unterschichtkinder dagegen sich an konkreten sprachlichen Formen oder einzelnen von der sprachlichen Form unabhängigen Bedeutungsgehalten orientieren. Man kann fast sagen, die Mittelschichtkinder wählten eher "konvergente", die Unterschichtkinder eher "divergente" Lösungen.

1) -----
Mit den "schichtspezifischen Strategien" in der Theorie der linguistischen Kodes teilen "schichtspezifische Strategien", wie sie in den hier durchgeführten Analysen untersucht werden, die Ebene der Einstellung zur Sprache (im Unterschied zu sprachlichen Merkmalen); sie unterscheiden sich jedoch von den erstgenannten dadurch, daß sie nicht mehr auf spezifische Formen von Sozialbeziehungen theoretisch bezogen sind. Dies ist jedoch, bei dem ex post-Charakter der Analysen, unvermeidlich.

Alle vorangehenden Analysen waren auf die Entdeckung von Schichtunterschieden angelegt, sie zielten also auf eine Widerlegung der Hypothese. Mit wenigen Ausnahmen - in Analyse (2) und (6) - haben sie nur der Ausgangshypothese widersprechende Ergebnisse zu Tage gebracht. Die folgende Analyse zielt auf die Herausarbeitung von Gleichförmigkeiten in beiden Schichtgruppen, auf eine Bestätigung der Hypothese. Erst wenn auch hier die Ausgangshypothese scheitert, können die qualitativen Analysen insgesamt und mit den quantitativen Ergebnissen verglichen werden.

(7) Ähnlichkeiten beider Schichtgruppen. Zwei Tatsachen verwiesen bis jetzt bei der Untersuchung der Unterschiede auf die Ähnlichkeiten beider Schichtgruppen: der relativ hohe Anteil nicht schichtdifferenzierender an der Gesamtzahl der Aufgaben (Tab. 185) und die zahlreichen Gruppen von Vorgaben mit nur geringen oder überhaupt keinen Schichtunterschieden (Tab. 193). Zum Abschluß der qualitativen Analysen, die vorrangig die Aufgabe hatten, Unterschiede zwischen den Schichten zu prüfen, soll daher geklärt werden, welche Merkmale Aufgaben und Antworten haben, die von beiden Schichtgruppen gleich beantwortet werden, und welches Gewicht Übereinstimmungen und Unterschieden zwischen den Schichten zukommt. Zwei Gründe sprechen dafür, diese Fragen an den für beide Gruppen besonders schwierigen Aufgaben zu untersuchen. Zwar können sich Übereinstimmungen prinzipiell daraus ergeben, daß die Aufgaben für Kinder der untersuchten Altersstufe zu leicht oder daß sie zu schwer sind. Bei einem multiple-choice-Test aber werden im ersten Fall von allen Kindern vorwiegend die richtigen und kaum andere Vorgaben gewählt, so daß mögliche Unterschiede sich kaum im Wechsel der Vorgaben ausdrücken; im zweiten Fall werden von allen Kindern die richtigen Vorgaben selten gewählt, so daß die Alternativen für beide Schichtgruppen gleichartig oder verschieden sein können. Zweitens sinkt die Wahrscheinlichkeit großer Schichtunterschiede mit der Anzahl der Vpn. aus beiden Schichten, die überhaupt eine richtige Vorgabe wählen; Übereinstimmungen haben die größte, Unterschiede die geringsten Chancen beobachtet zu werden. Kurz:

Unterschiede trotz Übereinstimmungen sind bei den schwierigen Aufgaben qualitativ besser beschreibbar und quantitativ schärfer prüfbar. Am Anteil der schichtdiskriminierenden und nicht diskriminierenden innerhalb der schwierigen Aufgaben und am Vergleich dieser Anteile zwischen allen und der Untermenge der schwierigen Aufgaben werden wir Übereinstimmungen und Differenzen messen.

Als "schwierige" haben wir Aufgaben, die von jeder Schicht extrem selten richtig und extrem häufig falsch beantwortet wurden, in Tab. 194 zusammengestellt. Als Kriterium für "Extremität" galt, daß mehr als ein Drittel der Vp. aus jeder Schicht (7) eine falsche oder weniger als ein Drittel der Vp. aus jeder Schicht eine richtige Vorgabe gewählt hatte; da solche Kriterien nur relativ willkürlich sein können, haben wir sie durch leichtere Kriterien ergänzt. Auf diese Weise kamen nur wenig weitere Vorgaben hinzu, so daß wir alle analysieren konnten. Bei Annahme des leichtesten Kriteriums finden wir 12 "extreme" richtige Vorgaben; an ihrer Stelle wurden 16 "extreme" falsche Vorgaben gewählt - bei 8 Testaufgaben je eine und bei vier Testaufgaben (6, 9, 16, 34) je zwei. Ferner finden wir bei 13 Aufgaben nur eine "extreme" falsche Vorgabe, so daß insgesamt 29 "extreme" falsche und 12 "extreme" richtige Vorgaben oder 25 "extreme" Testaufgaben in der gesamten Tabelle registriert sind.

Zumindest zwei Besonderheiten "extremer" Vorgaben als solcher, gleiche wie oft sie von den beiden Schichtgruppen im einzelnen gewählt wurden, lassen sich Tab. 194 entnehmen. Unübersehbar ist, daß die "extremen" richtigen Vorgaben nahezu ausschließlich abstrakte, außersprachlich schwer erklärbare Sachverhalte bezeichnen ("Export", "Geschöpf", "Prinzip", "Projekt", "bedächtig", "logisch", "allmählich", "raten", "vergeuden"; "Kapsel" meint zwar einen konkreten, doch sehr präzise zu spezifizierenden Gegenstand; "ächzen" ist die einzige Ausnahme). Und hinzu kommt, daß bei den "extremen" nicht richtigen Vorgaben solche überwiegen, die an Stelle des geforderten abstrakten Begriffs konkrete Vorfälle,

Beispiele oder einen bestimmten Sprachgebrauch setzen. Aus "Geschöpf" wird auf diese Weise "was Gott geschaffen hat", "Vergeuden" (Taschengeld vernaschen)", "ächzen" (vom Rennen außer Atem); aus einer "Kapsel" wird ein "Raketenoberteil" und aus einem "Projekt" eine "(Mappe)" oder ein "(Bauwerk)"; statt "langsam aber sicher" sagen die Kinder lieber "mir wird es allmählich klar", statt "Gier" "wild auf das und das", statt "Grundsatz" "er hat es im Prinzip richtig erfaßt".¹⁾ Sehr abstrakte Begriffe können sich also Kinder beider Schichten nur an konkreten Beschreibungen verdeutlichen. Eine ähnliche Tendenz zeigen die Kinder der Unterschicht bei den Aufgaben, bei denen ihnen die Kinder der Mittelschicht überlegen sind (Tab. 185): als Alternativen zur richtigen wählten sie konkretisierende Vorgaben.²⁾

1) -----
Dieser Eindruck läßt sich mit Hilfe der Kategorisierung nach semantischer Qualität quantitativ erhärten. Von den 16 als Alternative zu vernachlässigten richtigen Vorgaben gewählten Vorgaben sind 5 "konkret" und 5 "partikular". Von den gesamten 29 "extremen" nicht richtigen Vorgaben, die wir durch Vergleich mit Tab. 184 als Anteil der gesamten nicht richtigen Vorgaben darstellen können, sind 4 (von 10 = 40%) "Begriffe", 6 (von 37 = 16%) "allgemeine", 12 (von 53 = 23%) "konkrete", 2 (von 27 = 7%) "parallele", 5 (von 38 = 13%) "partikulare" Vorgaben. Die starke Repräsentanz nicht richtiger "Begriffe" ist nicht verwunderlich und unterstützt unsere Klassifikation. Die geringe Repräsentanz der "parallelen" Vorgaben zeigt, daß in den seltensten Fällen das gesamte Bedeutungsfeld von beiden Schichten anders verstanden wurde; die Ausnahmen selber sind hier aufschlußreich: "Lügner" für "Aufschneider" und "Quittung" für "Vertrag" (Aufgabe 14 und 50); in beiden Fällen wird außerdem die richtige Vorgabe nicht "extrem" selten gewählt.

2) Diese Tendenz findet sich auch bei den "extremen" Aufgaben wieder: bei allen drei Aufgaben, bei denen die Mittelschichtkinder die "extreme" richtige Vorgabe häufiger wählen, wählen die Unterschichtkinder "extreme" Konkreta: 15B, 20A, 40A.

Welche Schichtunterschiede lassen sich bei den "extremen" Testaufgaben feststellen? Für diese Frage müssen neben der "extremen" Qualität einer Vorgabe auch die Quantität der Wahlen berücksichtigt. Schon bei flüchtiger Betrachtung der Tab. 194 fallen die zahlreichen und starken Schichtdifferenzen selbst in den "extremen" Testaufgaben auf. Um diesen Eindruck zu überprüfen, haben wir ausgezählt, wieviele von den "extremen" Testaufgaben (Tab. 195, Zeile 1-4) und von den "extremen" falschen Vorgaben (Tab. 195, Zeile 7-9) von einer Schicht besser bearbeitet bzw. häufiger gewählt wurden und die gleichen Auszählungen für alle Aufgaben (Tab. 195, Zeile 5-6, 10) danebengestellt (siehe auch Tab. 185). Zunächst zeigen sich auch bei den "extremen" Aufgaben die gleichen Tendenzen wie bei der Gesamtheit der Aufgaben. Mittelschichtgünstige Aufgaben sind häufiger als unterschichtgünstige (Spalte 1-3) und die "extremen" falschen Vorgaben werden häufiger von der Unterschicht gewählt werden (Spalte 7-9). Wenn man alle Unterschiede von mehr als drei Vpn. berücksichtigt (Zeile 8), so sind von den 7 von der Unterschicht häufiger gewählten falschen Vorgaben 5 "konkret" und von den 6 von der Unterschicht gewählten Vorgaben 3 "Begriffe" - wie in Analyse (2) können wir keine Bevorzugung der Mittelschicht durch die Auswertung und wie in Analyse (4) eine Vorliebe der Unterschicht für "konkrete" Vorgaben feststellen.

Auch die Anteile schichtdiskriminierender Aufgaben zwischen den "extremen" und der Gesamtzahl der Aufgaben können verglichen werden. Wenn wir Schichtunterschiede nur bei den richtigen Vorgaben untersuchen (Zeile 3 und 5), so ist der Anteil nicht schichtdiskriminierender Aufgaben in beiden Fällen höher als der schichtdiskriminierender; wenn wir die richtige oder wenigstens eine andere Vorgabe über die Schichtunterschiede entscheiden lassen (Zeile 4 und 6), so verhält es sich umgekehrt. Beide Male liegt der Anteil nicht schichtdiskriminierender Aufgaben bei den "extremen" geringfügig höher als bei allen Aufgaben. Da in der Gruppe der "extremen" Aufgaben die richtigen Vorgaben in beiden Schichten von weniger als 9 Kindern gewählt wurden - das konstante Kriterium für die

Schichtunterschiede hier also wesentlich strenger ist -, ist die Konstanz des Anteils schichtdiskriminierender Aufgaben in beiden Testteilen und der hohe Anteil schichtdiskriminierender Aufgaben auch noch bei den "extremen" Aufgaben erstaunlich.¹⁾ Auch unabhängig von der Variation der allgemeinen Schwierigkeit der Aufgaben bleibt der Anteil schichtdiskriminierender Aufgaben ziemlich konstant; selbst bei den extrem schwierigen Aufgaben halten sich diskriminierende und nicht diskriminierende Aufgaben - je nach Maß - die Waage. Daher scheint uns die Betonung der Schichtunterschiede in den Analysen (1) bis (6) nicht ungerechtfertigt.

Wir können die Ergebnisse der auf Schichtunterschiede und der auf Übereinstimmung zielenden Analysen nun resümieren. Die Betrachtung der Schichtunterschiede ergab in den meisten Analysen ein recht profiliertes Muster von der einen oder der anderen Schicht bevorzugter Aufgaben bzw. Vorgaben. Unklar aber blieb, wie stark die Gemeinsamkeiten beider Schichtgruppen bei der Wahl von Vorgaben veranschlagt werden muß. Aus diesem Grunde haben wir die für beide Gruppen schwierigen Aufgaben, bei denen am ehesten Gemeinsamkeiten auftreten müßten, gesondert untersucht und mit den Ergebnissen im Gesamtest verglichen. Da aber in den extremen Testaufgaben ähnlich starke Schichtunterschiede auftreten wie in den Testaufgaben insgesamt, scheinen sich die Schichtunterschiede gegen gleichförmige Tendenzen in beiden Schichtgruppen durchzusetzen und die Konzentration auf die Analyse der Unterschiede gerechtfertigt zu sein.

¹⁾-----
Eine Rangkorrelation der allgemeinen Schwierigkeit einer Aufgabe, gemessen in durchschnittlichen Rangplatz der von allen Vp. gewählten Antwortvorgaben (MRAM), mit der Schichtdifferenz des gleichen Maßes zugunsten der Mittelschicht ergab für alle 42 Aufgaben nur $r(s) = +.05$. Auch bei einer sehr viel genaueren Variable und mit einem sehr viel feineren Maß ergibt sich kein Zusammenhang zwischen Schwierigkeit und Schichtunterschieden; auch dies Ergebnis rechtfertigt es, Unterschiede unter Vernachlässigung der Übereinstimmungen zu betrachten.

Für die Ausgangshypothese bedeutet das, daß endgültig die negativen Tendenzen überwiegen. Wir haben nicht nur dort, wo sie wahrscheinlich waren, Schichtdifferenzen entdeckt, sondern auch dort, wo Übereinstimmung wahrscheinlich war, keine Übereinstimmung zwischen den Schichten. Insgesamt also kann die qualitative Analyse des "differenzierten Verständnistests" die Ergebnisse der quantitativen Analyse des "differenzierten Verständnistests" bestätigen: Auch bei gleich großem passiven Wortschatz, verbinden Kinder verschiedener sozialer Schichten mit dem gleichen sprachlichen Eingabe-Material unterschiedliche Konzepte.

3.4.3. Zusammenfassung der qualitativen Analyse.

Die qualitative Analyse des "differenzierten Verständnistests" ging davon aus, daß sich in der quantitativen Analyse nur für diesen Test Schichtunterschiede ergeben hatten. Sie hatte nach diesem Ergebnis zwei Aufgaben zu erfüllen: erstens mögliche versteckte Vorurteile in der quantitativen Bewertung aufzudecken und zweitens die Schichtunterschiede weiter aufzuklären und über die Ausgangshypothese innerhalb des "differenzierten Verständnistests" zu entscheiden. Eine dritte Aufgabe kann man in der möglichst genauen Beschreibung von schichtspezifischen Kategorisierungsprozessen sehen, unabhängig von der Hypothesenprüfung. Während nämlich in der quantitativen Analyse die aufsummierten Schichtunterschiede sich aggregieren und gegenseitig durchkreuzen, können sie in der qualitativen Analyse bis auf Aufgaben und Vorgaben desaggregiert und nach der Richtung getrennt werden; insofern kann die qualitative Analyse maximale Unterschiede herausarbeiten. Wir wollen in der Zusammenfassung zuerst darstellen, wieweit Schichtunterschiede aus Vorurteilen der Bewertung resultieren können; zweitens eine Darstellung schichtspezifischer Aufgabenlösungen, unabhängig von den als Beleg dienenden Analysen und der angestrebten Hypothesenprüfung geben; und drittens die Analysen selber nach ihrem Beitrag zur Hypothesenprüfung resümieren.

Für die Bevorzugung der Mittelschicht durch die quantitative Bewertung ergab sich kein Anhaltspunkt. Im Gegenteil: Kinder der Mittelschicht bevorzugten Vorgaben, die in der quantitativen Analyse schlecht, in der qualitativen aber recht gut bewertet wurden. Eine Rangkorrelation aller Aufgaben zwischen der Schwierigkeit und Schichtunterschieden zugunsten der Mittelschicht betrug .05. Der Anteil der Aufgaben, bei denen die Unterschicht der Mittelschicht unterlegen ist, war innerhalb der extrem schwierigen Aufgaben nicht wesentlich höher als innerhalb aller Aufgaben.

Unabhängig von der Hypothesenprüfung fügen sich die einzelnen Analysen etwa zu folgendem Gesamtbild zusammen. Die Überlegenheit der Mittelschichtkinder bei der Wahl der "richtigen" Vorgabe hängt von der Form des Stimulus ab; sobald die Testaufgabe einen "kurz kategorisierenden" Stimulus hat, sind die Kinder der Unterschicht den Kindern der Mittelschicht ebenbürtig. Aber nicht nur die Wahl der richtigen, sondern auch aller übrigen Vorgaben scheint von der sprachlogischen Form des Stimulus abzuhängen. Die Unterschichtkinder zeigen eine Tendenz, die sprachlogische Form der Vorgaben nach dem Vorbild der sprachlogischen Form der Stimuli zu wählen und unter den semantischen Qualitäten der Vorgabe Konkretionen als Reaktion auf eindimensionale und Verallgemeinerungen als Reaktion auf mehrdimensionale Stimuli zu bevorzugen. Die Mittelschichtkinder bevorzugen - relativ unabhängig von der sprachlogischen Form der Stimuli - "kurz kategorisierende" Vorgaben und "begriffliche" Vorgaben; beide Tendenzen können aber auch als Widerspiegelung der generellen Überlegenheit der Mittelschichtkinder angesehen werden, da viele der richtigen Vorgaben "kurz kategorisierend" und alle richtigen Vorgaben "Begriffe" sind. Wenn man allerdings die Analyse auf den einzigen Testteil beschränkt, in dem die sprachliche Qualität der richtigen Vorgabe variiert, - auf die Aufgaben mit "kurz kategorisierendem" Stimulus -, so zeigen die Mittelschichtkinder sehr deutlich bei "kurz kategorisierende" richtige Vorgaben und die Unterschichtkinder ebenso deutlich bei nicht "kurz kategorisierende" richtige Vorgaben die besseren Leistungen.

Verzichtet man auf eine Differenzierung der Aufgaben nach der sprachlogischen Form des Stimulus, so ergeben sich einige weitere Ergebnisse. Die Mittelschichtkinder schneiden besonders bei solchen Aufgaben besser als die Unterschichtkinder ab, die ein intentionales Bedeutungsfeld behandeln. Als Alternative zur "richtigen" Vorgabe bevorzugen die Unterschichtkinder "konkrete", die Mittelschichtkinder "allgemeine" Vorgaben; unabhängig von der Zugehörigkeit zu bestimmten Testaufgaben aber werden "konkrete" und "parallele" Vorgaben, nicht aber "allgemeine" und partikulare" von den Unterschichtkindern bevorzugt, während die Mittelschichtkinder natürlich "Begriffe" bevorzugen. Schließlich wählen die Mittelschichtkinder häufiger als die Unterschichtkinder Vorgaben, bei denen sprachliche Form und logische Qualität nach Konventionen oder Notwendigkeiten des Sprachgebrauchs übereinstimmen. Kinder beider Schichten haben Schwierigkeiten bei abstrakten, extraverbal schwer erklärbaeren Sachverhalten und weichen auf konkretisierende Beschreibungen aus. Auch bei den extrem schwierigen Aufgaben zeigen sich fast ebenso starke Schichtunterschiede wie im gesamten "differenzierten Verständnistest".

Die Ausgangshypothese des Experiments besagt, daß die Konzepte von Kindern beider Schichten sich nicht unterscheiden würden, sondern allein von der Form des sprachlichen Eingabe abhängen. Für sie sind die Analysen (1), (2) und (4) relevant, weil sie entweder die sprachliche Form des Stimulus oder die semantische Qualität der Vorgaben erfassen, die für die sprachfreien Konzepte der Kinder steht. Die Analyse (2) wies zwar alternative Vorgabe-Wahlen beider Schichten innerhalb derselben Aufgabe, aber auch eine Reihe von schichtindifferenten Aufgaben nach; sie brachte keinen eindeutigen Schluß für die Hypothese. Die Analysen (1) und (4) konnten jedoch belegen, daß Schichtunterschiede von der Form des Stimulus abhängen und daß auch auf die gleiche Form des Stimulus die Kinder beider Schichten Vorgaben unterschiedlicher semantischer Qualität wählen. Insgesamt scheint uns daher durch die qualitative die quantitative Analyse bestätigt und für den "differenzierten Verständnistest" die Ausgangshypothese widerlegt. Wir

wollen im nächsten Abschnitt untersuchen, wieweit sich dies Ergebnis mit der Bestätigung der Ausgangshypothese im "kategorialen Präzisionstest" verträgt.

Die Analysen (3), (5) und (6) konnten nicht die Ausgangshypothese prüfen, aber Indizien über schichtspezifische Arten des Sprachgebrauchs bringen. Insgesamt scheint in der Unterschicht eher eine "divergente", in der Mittelschicht eher eine "konvergente" Art, die Aufgabe zu behandeln, vorzuliegen. Die Analyse (7) sollte prüfen, ob in den insgesamt schwierigeren Aufgaben nicht stärker die Gemeinsamkeiten beider Schichtgruppen durchschlagen würden. Da dies nicht bestätigt werden konnte, schien die Konzentration der vorangehenden Analysen auf Unterschiede nicht ungerechtfertigt.

3.5. Zusammenfassung und Bewertung der quantitativen und qualitativen Analyse sowie der beiden Testteile.

Der "kategoriale Präzisionstest" sollte zwei Mängel des "differenzierten Verständnistests" umgehen (siehe 1.1.). Die sprachliche Form von Stimulus und Vorgaben sollte konstant und die Bedeutungsvariation der Vorgaben sollte deutlich erkennbar sein. Die Beziehungen zwischen den Vorgabe-Konzepten sollten selber untersucht werden; aus diesem Grund wurde die zweimalige Differenzierung der gleichen Konzepte vom jeweils anderen Ausgangspunkt eingeführt. Entsprechend konnten im "differenzierten Verständnistest" die Variation der sprachlichen Form von Stimulus und Vorgaben und der semantische Gehalt der Vorgaben nur im Nachhinein kontrolliert werden. Die Beziehungen zwischen zwei Konzepten werden in ihm nicht untersucht, allein die aus einer isolierten sprachlichen Form abgelesenen Konzepte.

Für den "kategorialen Präzisionstest" konnte weder in der quantitativen noch in der qualitativen Analyse ein Schichtunterschied belegt werden. Die qualitative Analyse konnte von der Form des Tests her kaum mehr Ergebnisse erbringen. Zwischen den Gesamtleistungen für die Worte eines Paares tauchten erstaunlich große Unterschiede auf, so daß der Ausgangspunkt für die Diskriminierung zweier Konzepte nicht unerheblich ist. Für die quantitative Auswertung ergeben sich daraus keine Konsequenzen, da in ihr die beiden Aufgaben jedes Paares voneinander unabhängig verrechnet wurden. Selbst unter höheren Signifikanzniveaus ergeben sich im Wilcoxon-Test keine Schichtunterschiede; die niedrigste Zufallswahrscheinlichkeit für einen einseitigen Test ist 27%. Insgesamt findet sich weder in der Konstruktion des Tests noch in den Ergebnissen ein Hinweis auf die Hinfälligkeit der Bestätigung unserer Ausgangshypothese.

Für den "differenzierten Verständnistest" zeigte sich in der quantitativen Analyse bei der Prüfung der Hypothese der Überlegenheit der Mittelschicht auf den stärker differenzierten Maßen ohne Berücksichtigung der Antwortverweigerungen tatsächlich eine Überlegenheit der Mittelschicht und bei der Prüfung der Hypothese fehlender Schichtunterschiede auf allen Maßen Schichtunterschiede. In der qualitativen Analyse fanden wir, daß die Kinder der Unterschicht bei komplexeren Stimuli schlechtere Vorgaben wählen und allgemein ausgefallene Vorgaben bevorzugen. Die qualitative Analyse ergab jedoch keine Anhaltspunkte für Schichtvorurteile in der quantitativen Bewertung, ihre Ergebnisse boten vielmehr eine Reihe von plausiblen Interpretationsmöglichkeiten. Insgesamt ergeben sich aus der Konstruktion des Tests und den Ergebnissen der quantitativen Analyse zwar Vorbehalte gegen die Zuverlässigkeit des Tests, die in der qualitativen aber Analyse nicht erhärtet werden. Das weniger ausgearbeitete Instrument hat gerade wegen seiner geringeren Bestimmtheit Möglich-

keiten für qualitative Analysen gelassen, die einer nachträglichen Analyse des präziseren Instruments, des "kategorialen Präzisionstests", nicht mehr offenstanden. Wir müssen also die widersprüchlichen Ergebnisse in den beiden Testteilen anerkennen und wollen versuchen, die Unterschiede zwischen den Untertests sowie das schlechtere Abschneiden der Unterschicht im "differenzierten Verständnistest" erklären.

Für die Erklärung der unterschiedlichen Ergebnisse in den beiden Untertests scheint uns ihre unterschiedliche Schwierigkeit der Schlüssel (siehe Tab. 179). Der allgemein schwierigere Untertest ist zugleich der Test, in dem die Unterschicht schwächer ist. Einen weiteren Hinweis geben die schichtspezifischen Interkorrelationen analoger Maße der beiden Untertests, die in der Mittelschicht recht hoch, in der Unterschicht gleich Null sind (siehe Tab. 159). Offensichtlich sind bei den Mittelschichtkindern durch die beiden Untertests gleiche, bei den Unterschichtkindern aber ganz verschiedenartige Fähigkeiten angesprochen. Hier können uns die auf die Hypothese bezogenen wie auch die nicht auf die Hypothese bezogenen qualitativen Analysen des "differenzierten Verständnistests" weiterhelfen. Wir hatten die erste Gruppe von Analysen so resümiert, daß die Dekodierung "erweitert kategorisierender" Stimuli, den Unterschichtkindern schwerer fällt. Wir hatten die zweite Gruppe so resümiert, daß die Kinder der Mittelschicht unabhängig von der Stimulusform stilistisch als Erklärung besonders taugliche Vorgaben bevorzugen, die Kinder der Unterschicht dagegen sich an der Form des Stimulus orientieren und eher ausgefallene Vorgaben wählen oder kurz: daß die Kinder der Mittelschicht die Aufgaben eher in einer "konvergenten", die Kinder der Unterschicht eher in einer "divergenten" Weise lösen. Es ist nun sehr wahrscheinlich, daß die Vorgaben, die aus einer "konvergenten" Einstellung gewählt wurden, besser bewertet wurden, als die Vorgaben, die aus einer "divergenten" Einstellung gewählt wurden - nicht allein deshalb, weil auch bei den Bewertern "konvergente" Normen vorgeherrscht haben mögen, sondern auch, weil die entsprechenden Vorgaben objektiv treffender gewesen sein können.

Genau derartige Differenzierungen zwischen Stimuli und Vorgaben waren im "kategorialen Präzisionstest" von der Art des sprachlichen Materials her nicht möglich. Die Unterschichtkinder werden dort nicht durch semantisch und sprachlich ausgefallene von einfacheren Vorgaben abgelenkt und finden an den sprachlichen Formen des Stimulus keine Anhaltspunkte für semantische Abschweifungen. Sie sind durch die Gleichheit der äußeren Form und die minimalen semantischen Differenzen der Vorgaben gezwungen, kategorial zu differenzieren. Die Mittelschichtkinder auf der anderen Seite verfügen schon zu einem größeren Teil über die Einstellung zu einer von Äußerlichkeiten unabhängigen kategorialen Differenzierung, die bei den Unterschichtkindern noch in den meisten Fällen durch die objektiven Bedingungen der Aufgabe erzwungen werden muß. Aus diesem Grunde können die Mittelschichtkinder im "kategorialen Präzisionstest" nicht besser sein als die Unterschichtkinder, während sie unter den Bedingungen des "differenzierten Verständnistests" sofort besser werden. Aus diesem Grunde auch erreichen beide Schichten im "kategorialen Präzisionstest" bessere Leistungen und fallen beide Schichten im "differenzierten Verständnistest" zurück - wenn auch die Unterschicht stärker als die Mittelschicht.

Wenn diese Interpretation sinnvoll ist, dann können wir auch endgültig zwischen der Bestätigung der Hypothese im "kategorialen Präzisionstest" und der Widerlegung im "differenzierten Verständnistest" entscheiden. Obwohl wir die Ergebnisse des "differenzierten Verständnistests", gerade nach der qualitativen Analyse für methodisch hinlänglich gesichert halten, scheinen uns aus theoretischen Gründen die Ergebnisse des "kategorialen Präzisionstests" entscheidend zu sein. Der "differenzierte Verständnistest" prüft die kategoriale Differenzierung der Kinder unter den genannten Randbedingungen, die für Kinder beider Schichten erschwerend, aber stärker erschwerend für die Unterschichtgruppe, sind; diese Randbedingungen können als Auslöser schichtspezifischer Strategien angesehen werden, die sich jedoch stärker auf manifeste sprachliche Merkmale, die im Test nur die Konzepte indizieren sollten, beziehen, als auf die von sprachlichen Stimulus evozierten Konzepte. Der "kategoriale Präzisionstest" dagegen ruft

unmittelbar die evozierten Konzepte oder die Fähigkeiten der Individuen zu kategorialer Differenzierung ab. Sofern wir uns streng im Rahmen der Ausgangshypothesen des Experiments bewegen, müssen wir also allein den "kategorialen Präzisionstest" als maßgeblich und die Hypothese der dritten Messung als bestätigt betrachten. Sobald wir jedoch Schichtunterschiede des Sprachverständnisses untersuchen wollen, müssen wir uns auf den "differenzierten Verständnistest" konzentrieren und können eine Reihe von unterschiedlichen Verständnisweisen bei Kindern aus der Unter- und Mittelschicht feststellen, die sich auch im Rahmen der Theorie linguistischer Kodes interpretieren lassen (siehe Abschnitt 3.1.).

3.6. Zusammenhänge der Leistung in der dritten Messung mit anderen Faktoren als der dichotomen Sozialschicht.

Die Ausgangshypothese der dritten Messung war, daß sich zwischen Kindern beider Schichten mit gleich gutem passivem Wortschatz keine Unterschiede in der "kategorialen Differenzierung" ergeben würden. Die vorangegangenen Analysen hatten erbracht, daß im "kategorialen Präzisionstest" überhaupt keine Schichtunterschiede auftreten und die minimalen Schichtunterschiede im "differenzierten Verständnistest" sich aus Schwierigkeiten der Unterschichtkinder bei der Dekodierung von "erweitert kategorisierenden" Stimuli und aus ihrer Vorliebe für ausgefallene Formen von Vorgaben ergeben würden. Wir hatten argumentiert, daß die gleichen Probleme sich offenbar auch für die Mittelschichtkinder ergeben würden, lediglich in geringerem Maße, und daß sie mit der gestellten Aufgabe kategorialer Differenzierung im Alltag zwar durchaus als eine Erschwernis verquickt seien, sie aber selbst nicht wesentlich beträfen; oder m.a.W.: daß der "kategoriale Präzisionstest" als der eigentliche Test "kategorialer Differenzierung" angesehen werden muß, die im "differenzierten Verständnistest" unter zusätzlichen Bedingungen abgefragt wird. Von diesen Überlegungen her haben wir dann die Ausgangshypothese als bestätigt angesehen.

Die folgenden Abschnitte bringen nun in gewissem Sinn eine Umkehr der Argumentation. Wenn keine Schichtunterschiede vorliegen, so suchen wir jetzt die Faktoren, die tatsächlich die "kategoriale Differenzierung" beeinflussen; wir hatten vorausgesagt, daß die Leistungen der dritten Messung nur noch von der Intelligenz und dem passiven Wortschatz abhängen. In diesem Abschnitt prüfen wir also, wie in der zweiten Messung (siehe Kapitel III.5.6.) alternative Einflüsse zur Sozialschicht. Da wir jedoch in der dritten Messung den Einfluß alternativer Faktoren vorausgesagt haben, sind die folgenden Analysen nicht mehr ganz mit den entsprechenden Analysen zur zweiten Messung vergleichbar.

Wie in der zweiten Messung handelt es sich um eine Untersuchungsgruppe, in der die Schichtunterschiede im passiven Wortschatz schon ausgeschaltet sind. Genau so wenig wie in der zweiten Messung ließe sich hier aus dem Vorherrschen individuell-psychologischer über soziale Einflußfaktoren auf die Bedeutungslosigkeit der sozialen Einflüsse schließen. Hinzukommt jedoch eine für die dritte Messung besondere Überlegung. Anders als in der zweiten Messung, war das Experiment hier ja auf die Bestätigung der Hypothese fehlender Schichtunterschiede angelegt: Das Nichtauftreten von Schichtunterschieden bei Kontrolle des passiven Wortschatzes sollte gerade, in Zusammenhang mit den Voraussagen zur ersten und zweiten Messung, ein indirekter Beleg für die Wirksamkeit sozialer Strategien des Sprachgebrauchs unter alltäglichen Bedingungen, d.h. ohne Kontrolle des passiven Wortschatzes, sein: Aus den gleichen sprachlichen Eingaben - eine Bedingung die alltäglich nicht erfüllt ist - sollten Kinder beider Schichten mit gleich gutem passiven Wortverständnis - eine Bedingung die alltäglich im Durchschnitt ebenfalls nicht erfüllt ist - gleichartige Konzepte gewinnen und Unterschiede zwischen den Kindern nur noch von ihrer allgemeinen Intelligenz oder von ihrem passiven Wortschatz abhängen. Kurz: Insofern als wir in den folgenden Abschnitten gleichsam die Kehrseite der Hypothese betrachten, haben sie - anders als die inhaltlich vergleichbaren Abschnitte zur zweiten Messung - noch die Funktion der Hypothesenprüfung.

Unabhängig aber von der Bestätigung des Einflusses von Intelligenz, Schulleistung und Wortschatz auf die "kategoriale Differenzierung" können wir in den folgenden Analysen jedoch auch die schichtspezifischen Einflußstrukturen dieser Faktoren untersuchen. Dabei handelt es sich nicht um Hypothesenprüfung, sondern Deskription und evtl. Entwicklung von Hypothesen. Nach den unterschiedlichen Ergebnissen für beide Untertests der dritten Messung werden wir die Deskription auf den Vergleich spezifischer Einflußstrukturen in beiden Subtests - sei es für die gesamte Gruppe oder die beiden Schichtgruppen getrennt - ausdehnen. Insbesondere hoffen wir durch die für die Untertests und die Experimentalgruppen spezifischen Einflußstrukturen Aufschluß zu erhalten, warum der "differenzierte Verständnistest" für die Unterschicht relativ schwieriger war.

Für beide Funktionen der folgenden Analysen verwenden wir die gleichen statistischen Verfahren: Korrelations- und multiple Regressionsanalyse. Die Ergebnisse der statistischen Verfahren werden wir jedoch - je nach der Zielsetzung - unterschiedlich betrachten müssen. Für die Hypothesenprüfung können wir nur hohe und signifikante Korrelationen bzw. multiple Regressionen beachten. Für die Deskription werden wir auch geringere Korrelationen und insbesondere die Struktur der Korrelationen und die Struktur der Beta-Gewichte auch für nicht-signifikante R^2 berücksichtigen.

3.6.1. Dimensionen der Soziallage und "kategoriale Differenzierung".

Wir haben die Korrelationen des Filterindex mit der "kategorialen Differenzierung" in 3.2. dargestellt: sie sind unerheblich. Dennoch könnten die einzelnen Dimensionen des Filterindex der sozialen Schicht und der (mit diesen Dimensionen und dem Filterindex um etwa .50 korrelierende) Buchbesitz als eines Indikatoren für kulturelle Gewohnheiten spezifische und einander entgegengesetzte Korrelationen mit den Maßen der dritten Messung aufweisen. Auch für diese Variablen gelten die Überlegungen der Ausgangshypothese: wir

sagen für alle vier Dimensionen und - mit Vorsicht - auch für den Buchbesitz unbedeutende Korrelationen voraus. Genau genommen, versuchen wir also in dieser Analyse noch - wie in Abschnitt 3.2. -, den fehlenden Einfluß sozialer Faktoren nachzuweisen. Wir haben sie aus zwei Gründen trotzdem in den Abschnitt 3.6. aufgenommen, der der Prüfung der vorausgesagten tatsächlich wirksamen Einflüsse gewidmet ist. Erstens sind die hier dargestellten Korrelationen, genauso wie die Korrelationen der Intelligenz- und Schulleistungen mit den Maßen der dritten Messung (3.6.2.), Grundlage der beiden Seiten der Hypothesenprüfung verbindenden multiplen Regressionsanalyse (3.6.3.). Zweitens kann diese Analyse, anders als die Analysen in 3.2. für die beiden Schichtgruppen und die beiden Untertests spezifische Einflußstrukturen beschreiben.¹⁾

In der Gesamtgruppe finden wir - wie erwartet - keine Korrelationen der vier Dimensionen des Filterindex mit den Untertests der dritten Messung (Tab. 199). Wenn wir jedoch beachten, daß wir in der dritten Messung nicht eine statistische Nullhypothese falsifizieren, sondern eine "inhaltliche Nullhypothese" verifizieren wollen, so müssen wir auch schon recht geringe und nicht signifikante Korrelationen berücksichtigen. Korrelationen über .20 finden sich jedoch beim Einkommen für alle Untertests, und bei der Ausbildung des Vaters für den 42er und den Gesamttest.²⁾ Wir erinnern jedoch daran, daß Ausbildung des Vaters und

1) -----
Für die genaue Darstellung der untersuchten unabhängigen Variablen siehe die entsprechenden Abschnitte der zweiten Messung: Kapitel III.5.6.2. und 5.6.5.

2) Daß die Korrelationen absinken, je weniger differenzierte Maße wir betrachten (d.h. in der Tab. 199 innerhalb jedes Untertests beim Lesen von links nach rechts), ist eine Folge der Tatsache, daß sich bei den weniger differenzierten Maßen die Varianzen verringern (siehe Tab. 151). Bei den meisten unabhängigen Variablen fallen die Korrelationen zwischen der Dreier- und der Zweier-Kategorisierung abrupt ab. Die einzige Ausnahme ist der Beruf des Vaters in der Unterschicht, dessen Korrelationen mit den weniger differenzierten Maßen des gesamten und des 42er Tests zunehmen (s.u.)

das Einkommen schon mit der Kontrollvariablen WST recht hoch positiv, der Beruf des Vaters und die Ausbildung der Mutter jedoch geringfügig negativ korrelierten (siehe Abschnitt 2., Tab. 166); der gleichen Inkonsistenz bei der Voraussage der Experimentalleistung, die mit der Kontrollvariablen hoch korreliert (siehe Tab. 200), brauchen wir daher keine große Bedeutung beimessen. Aus dem gleichen Grunde sind die unwesentlichen Korrelationen der Ausbildung der Mutter nicht verwunderlich, für die wir nach den Ergebnissen der zweiten Messung noch am ehesten höhere Korrelationen erwartet hätten. Insgesamt kann die Hypothese fehlender Schichteinflüsse also auch für die vier Einzelindikatoren des Filterindex bestätigt werden.

Versuchen wir deskriptiv unterschiedliche Gewichte der einzelnen Dimensionen innerhalb der beiden Schichtgruppen herauszuarbeiten, so wiegt die Übereinstimmung vor. In beiden Schichten findet man positive Zusammenhänge der Testleistungen mit dem Einkommen wieder. In beiden Schichtgruppen tauchen positive Zusammenhänge zwischen dem Beruf des Vaters und der Leistung im 16er Test auf, die sich in der Gesamtgruppe nur sehr schwach beobachten ließen. In beiden Schichtgruppen erscheinen schließlich (nicht signifikante) Mittelwertunterschiede der Varianzanalyse mit 4 Schichten (Tab. 176; Abschnitt 3.2.) als relativ hohe Korrelationen mit dem Filterindex wieder: die OMS ist der UMS im 42er Test unterlegen (negative Korrelationen mit IFILT in der Mittelschichtgruppe), die OUS der UUS im 16er Test überlegen (positive Korrelationen mit IFILT in der Unterschichtgruppe).

In der Unterschichtgruppe finden sich jedoch zusätzliche Besonderheiten. Vergleichen wir die Unterschichtgruppe mit der Gesamt- und Mittelschichtgruppe, so fallen innerhalb des 16er Test keinerlei Unterschiede zwischen den Gruppen auf; innerhalb des 42er Tests stechen aber die starken negativen Korrelationen des Berufs des Vaters und die relativ starken positiven Korrelationen des Einkommens in der Unterschicht von den unbedeutenden entsprechenden

Korrelationen in der Mittelschichtgruppe ab. Vergleichen wir die beiden Untertests innerhalb der Unterschichtgruppe, so stehen die negativen Korrelationen des Berufs des Vaters mit den 42er positiven Korrelationen des Berufs des Vaters mit dem 16er Test gegenüber, aber es finden sich recht starke Korrelationen des Einkommens mit beiden Untertests. Kann sich ein solches Muster aus den unterschichtspezifischen Interkorrelationen oder Varianzen der Schichtindikatoren ergeben haben?

Die Korrelationen zwischen Einkommen und Beruf des Vaters in der Unterschicht ist positiv und mit .37 höher als 2 der insgesamt 6 Interkorrelationen der Schichtindikatoren (Tab. 168). Zudem ist korreliert der Beruf des Vaters mit der Ausbildung des Vaters .62, und dennoch hat nur der Beruf des Vaters einen Einfluß auf die Experimentleistung. Weiterhin korreliert innerhalb der Unterschichtgruppe mit den Kontrollvariablen WST das Einkommen extrem hoch (.62) der Beruf des Vaters jedoch niedriger als die Ausbildung des Vaters. Die Interkorrelationen der Schichtindikatoren führen uns zwar nicht weiter, aber nach den Korrelationen mit der Kontrollvariablen WST können wir die Korrelationen des Einkommens mit den Experimentalleistungen in der Unterschicht getrost als Scheinkorrelation betrachten. Für die Korrelationen des Berufs des Vaters mit den Experimentalleistungen in der Unterschicht aber kann nicht einmal eine zu geringe Varianz verantwortlich gemacht werden: sie ist im Vergleich zur Mittelschicht nicht wesentlich kleiner¹⁾ (Tab. 199). Weder Korrelationen mit der Kontrollvariablen WST noch Varianzen schließlich könnten Ursache für den gegensätzlichen Einfluß des Berufs des Vaters auf die beiden Untertests in der Unterschicht sein.

Wir wissen, daß sich im 16er Test keine Schichtunterschiede ergeben; wir sehen hier, daß sich auch die Einflußstrukturen der Schicht-

¹⁾ Für eine Übersicht über die Rangfolge der Berufe des Vaters, sowie über die in den Korrelationen verwandten z-Werte vgl. die entsprechende Tab. zur Stichprobe der zweiten Messung: Tab. 26.

indikatoren zwischen den beiden Schichtgruppen nicht unterschieden. Wir wissen ferner, daß im 42er Test die Unterschicht der Mittelschicht unterlegen war; wir sehen hier, daß sich auch die Einflußstruktur der Schichtindikatoren auf die Maße des 42er Test zwischen den beiden Schichten und innerhalb der Unterschicht vom 16er Test unterscheidet. Wir wissen schließlich, daß in der Mittelschicht die Leistungen in den beiden Untertests mit etwa .50, in der Unterschicht aber um .00 korrelieren (Tab. 159). Ganz offensichtlich muß das unterschiedliche Abschneiden der Unterschichtkinder in den beiden Untertests mit dem Beruf des Vaters zusammenhängen. Je höher der Beruf des Vaters innerhalb des (sowieso schon eingeschränkten) Bereichs der Berufe der Unterschicht, desto besser sind die Leistungen der Kinder im "kategorialen Präzisionstest", desto schlechter aber im "differenzierten Präzisionstest". Daß der erste Zusammenhang umso stärker gilt, je differenzierter die betrachteten Maße der dritten Messung sind, muß aus den abnehmenden Varianzen bei weniger differenzierten Maßen verstanden werden (siehe Tab. 151); dieselbe Tendenz findet sich für fast alle unabhängigen Variablen in Tab. 199. Daß der zweite Zusammenhang jedoch um so stärker gilt, je weniger differenziert die betrachteten Variablen der dritten Messung sind, widerspricht direkt der aus der Maßkonstruktion sich ergebenden Tendenz und findet sich nur bei den Korrelationen des Berufs des Vaters mit den Maßen der 42er Messung innerhalb der Unterschicht. Diese Tatsache gibt einen Hinweis für eine Interpretation des Gegensatzes zwischen den beiden Untertests. Wir wissen aus der qualitativen Analyse, daß die Kinder der Unterschicht im "differenzierten Verständnistest" formal eher abweichende, semantisch aber ganz gut treffende Vorgaben wählen - eine Möglichkeit, die ihnen im "kategorialen Präzisionstest" nicht gegeben ist. Es könnte nun sein, daß der Beruf des Vaters innerhalb der Unterschicht in Wirklichkeit durchgängig positiv mit der "kategorialen Differenzierungsfähigkeit" zusammenhängt, daß sich dies auf die Leistungen der Kinder im 16er Test direkt auswirkt, im 42er Test, der von der Sprache und der Semantik her mehr Möglichkeiten der Ablenkung bietet, die Kinder aber zur Wahl von treffenden, aber ungewöhnlichen statt einfacher Vorgaben verleitet d.h. zur Wahl von im übertragenen

und buchstäblichen Sinne zweitbesten Alternativen. Ein solcher Zusammenhang müßte immer stärker werden, je undifferenzierter die Maße sind. Endgültig verifizieren ließe er sich nur, wenn man die Leistungen derselben Vpn. in beiden Untertests einzeln analysiert.

Der Buchbesitz ist ein Indikator für das "kulturelle Kapital" oder die soziokulturelle Lebenswelt der Familie ; er korreliert hoch mit der Schichtzugehörigkeit (IFILT) in unserer gesamten (.50) und der Mittelschichtgruppe (.43), nicht aber in der Unterschichtgruppe (.06). Von der Ausgangshypothese her können wir auch hier das Fehlen eines Einflusses auf die Experimentalleistungen voraussagen; da aber mit dem Buchbesitz gerade auch nicht mit der vertikalen Dimension sozialer Schicht übereinstimmende Faktoren gemessen werden sollen, ist dies eine riskante Voraussage. In der Gesamtgruppe korreliert der Buchbesitz etwa so hoch wie das Einkommen positiv mit den Leistungen in beiden Untertests (Tab. 199). Anders als das Einkommen korreliert er jedoch nur sehr gering positiv mit der Kontrollvariablen innerhalb der Untersuchungsgruppe (.14). Trotz hoher Korrelationen mit dem Filterindex und den einzelnen Schichtindikatoren scheint der Buchbesitz also einen, wenn auch schwachen Einfluß der soziokulturellen Lebenswelt auf die Experimentalleistung zu indizieren, der nicht mit "sozialer Schicht" voll erfaßt ist.

Innerhalb der Unterschichtgruppe hat der Buchbesitz keinerlei Bedeutung für die Experimentalleistung. Da der Buchbesitz in der Gesamtstichprobe (N=329) sich kaum zwischen den beiden unteren Straten unterscheidet (Oevermann / Krappmann / Krepner 1972) und auch in unserer Experimentalstichprobe ein im Vergleich zur Mittelschicht geringe Varianz aufweist, ist dies nicht verwunderlich. Innerhalb der Mittelschichtgruppe hängt der Buchbesitz auffällig stark mit den Maßen des 16er Tests zusammen.

Zusammenfassung: In der Gesamtgruppe kann auch anhand der einzelnen Schichtindikatoren die Hypothese fehlender Schichtunterschiede bestätigt werden. Die vier Schichtindikatoren korrelieren mit den Maßen des "kategorialen Präzisionstests" in beiden Schichten gleich. Bei den Maßen des "differenzierten Verständnistests" kontrastieren jedoch negative Korrelationen des Berufs des Vaters in der Unterschicht mit positiven oder nicht vorhandenen Korrelationen des Berufs des Vaters in der Mittelschicht und der Gesamtgruppe. Innerhalb der Unterschichtgruppe kontrastiert die positive Korrelation des Berufs des Vaters mit dem "kategorialen Präzisionstest" zur negativen Korrelation des Vaters mit dem "differenzierten Verständnistest". Die Gleichartigkeit der Einflußstruktur auf den "kategorialen Präzisionstest" in beiden Schichten entspricht dem Fehlen von Schichtunterschieden in diesem Test, die Verschiedenartigkeit der Einflußstruktur auf den "differenzierten Verständnistest" in beiden Schichten dem Auftreten von Schichtunterschieden in diesem Test. - Der Buchbesitz korreliert schwach mit den Experimentalleistungen in der Gesamtgruppe, in der Mittelschicht auffällig stark mit den Leistungen im "kategorialen Präzisionstest".

3.6.2. Intelligenz, Schulleistung und kategoriale Differenzierung.

Die Vpn. unseres Experiments bearbeiteten den PSB ("Prüfsystem für Schul- und Bildungsberatung") von Horn und den Frankfurter Wortschatztest (WST); außerdem wurden die Schulnoten erhoben und innerhalb der Schulklassen z-Transformiert. Die Korrelationen dieser Variablen mit dem Median der 7 Maße jeder Testform für die Gesamt-, Unterschicht- und Mittelschichtgruppe sind in Tab. 200, die Einzelkorrelationen jedes Maßes in den Tab. 196, 197 und 198 dargestellt.

Der Wortschatztest korreliert in allen drei Experimentalgruppen mit allen Maßen der dritten Messung hochsignifikant. Insofern ist die Ausgangshypothese der dritten Messung auch in ihrem zweiten Aspekt bestätigt. Kein weiteres Intelligenzgesamtmaß und erst recht kein Intelligenz-Subtest korreliert mit beiden Untertests der dritten Messung in allen Gruppen. Insofern bleibt die Bestätigung des zweiten Aspekts der Hypothese auf den Wortschatztest beschränkt; wir müssen nach den für die beiden Untertests und die beiden Schichtgruppen besonderen Einflüsse der Intelligenz suchen.

In der Gesamtgruppe fällt sofort auf, daß die Korrelationen der Intelligenz und Schulnoten mit den 16er Test auffällig höher (Tab. 200) und zahlreicher (Tab. 196) sind als mit dem 42er Test.¹⁾ Allein der 16er Test korreliert mit Maßen der verbalen Intelligenz (VERBIT, TEST 1/2, TEST 5) und den beiden Schulnoten in Deutsch, zugleich aber auch mit den nicht-verbalen Intelligenz-Subtests 6 und 7; der 42er Test korreliert nur mit nicht-verbalen Intelligenz-Subtests (TEST 3: logisches Schließen anhand von Figuren), aber auch mit den Schulnoten in Deutsch-Schriftlich. Wegen ihrer Verstreutheit und ihrer geringen Systematik lassen sich die Einflüsse auf die Leistungen in den beiden Untertests der dritten Messung in der Gesamtgruppe kaum, und nur grob mit der Gegenüberstellung verbaler (16er Test) und nicht-verbaler (42er Test) Faktoren beschreiben. - In der Mittelschichtgruppe korrelieren alle Intelligenz-Gesamtmaße, alle Intelligenz-Subtests (mit Ausnahme von TEST 8: embedded figures, und TEST 9: Rechenfähigkeit) und alle Noten positiv mit den Maßen beider Untertests. Größere Unterschiede in der Höhe der Korrelationen zwischen den beiden Untertests finden sich nur bei der Note für Deutsch-Mündlich, wo der 42er Test überhaupt keine, und bei TEST 6 (Verstümmelte Wörter), wo der

¹⁾-----
Allein die Religions-Note korreliert durchgängig in beiden Untertests, und - nebenbei - auch in allen Experimentalgruppen.

42er Test wesentlich niedrigere Korrelationen zeigt. Auch hier also lassen sich die Einflüsse auf die beiden Untertests kaum beschreiben, nicht jedoch wie in der Gesamtgruppe wegen ihrer Verstreutheit, sondern wegen ihrer durchgängigen Gleichartigkeit. In der Unterschichtgruppe finden sich weniger und geringere Korrelationen als in der Mittelschicht. Um so auffälliger sind die Kontraste zur Mittelschicht in der Einflußrichtung. Übereinstimmend mit der Mittelschicht sind lediglich die positiven Korrelationen der verbalen TEST 1/2 (allgemeine Wortkenntnis), des nicht-verbalen TEST 7 (Faltaufgaben) und der Schulnote in den vier Leistungsfächern mit dem 16er Test sowie der Religionsnote mit beiden Untertests der dritten Messung. Wir wollen im folgenden die Gegensätze zur Mittelschichtgruppe genauer darstellen und versuchen, unter Hinzunahme der schichtspezifischen Interkorrelationen zu einer Interpretation der schichtspezifischen Einflußfaktoren zu kommen.

Nehmen wir die Höhe und die Konsistenz der Korrelationen über die beiden Untertests der dritten Messung als Kriterium, so ist nur ein in den Schichten gegensätzlicher Einfluß bemerkenswert: TEST 10 (Wahrnehmungstempo) hängt in der Mittelschicht nahezu signifikant positiv, in der Unterschicht nahezu signifikant negativ mit den Maßen der dritten Messung zusammen. Die schnelle und korrekte Wahrnehmung ist eine Vorbedingung für die Bearbeitung jedes multiple-choice Tests; insofern ist kaum erklärlich, warum sie für einen verbalen multiple-choice Test besonders relevant sein sollte, und noch weniger, warum sie die Leistung der Mittelschichtkinder gefördert, die Leistung der Unterschichtkinder aber beeinträchtigt haben sollte. Eine Erklärungsmöglichkeit bietet sich jedoch, wenn wir den nicht-verbalen Subtest 10 nur als Beispiel für nicht-verbale Intelligenz überhaupt sehen und den Gegensatz der Einflußrichtung auf die gleiche Weise erklären, wie der Gegensatz positiver und negativer Korrelationen zwischen nicht-verbaler Intelligenz und sprachlichen Leistungen in der Unter- bzw. Mittelschicht allgemein erklärt worden ist (Overmann 1970). Wir müssen dazu die weitem, oft nur schwachen schichtspezifischen Korrelationen zwischen

Intelligenz-Subtests und den Maßen der dritten Messung heranziehen und zwischen den beiden Untertests der dritten Messung unterscheiden; in der Regel betrachten wir nur noch die Maße für die Untertests, da die Gesamtmaße einen "bias" zugunsten des 42er Tests allein aufgrund seiner höheren Anzahl von Testaufgaben enthalten.

Bei den drei Intelligenzgesamtmaßen (GESINT, NONVER, VERBIT) finden sich in der Unterschichtgruppe positive Korrelationen von GESINT und VERBIT und negative von NONVER mit dem 16er und negative Korrelationen aller drei Intelligenzgesamtmaße mit dem 42er Test (Tab. 197). Obwohl sich das Muster für beide Untertests negativer Einflüsse von NONVER, und auf den 16er Test positiver, den 42er Test negativer Einflüsse von VERBIT an keinem der entsprechenden einzelnen Intelligenztests wiederfindet, mit Ausnahme des TEST 10, von dem wir ja ausgegangen waren,¹⁾ scheint es uns bemerkenswert wegen des Kontrasts zu den Ergebnissen in der Mittelschichtgruppe. Wir können aber, auf der nächst allgemeineren Stufe, den Gegensatz zwischen der Einflußstruktur in den beiden Schichtgruppen beschreiben. Erstens finden sich - im krassen Gegensatz zu den durchgängigen Einflüssen aller Intelligenzmaße und Schulnoten auf beide Untertests in der Mittelschicht - in der Unterschicht weder in beiden Untertests Einflüsse einzelner Intelligenzmaße noch einzelner Schulnoten (mit den schon erwähnten Ausnahmen der Schulnote in Religion und TEST 10). Zweitens tauchen allein in der Unterschichtgruppe negative Korrelationen mit vorwiegend nicht-verbale Intelligenzmaßen auf, und zwar in beiden Untertests der dritten Messung. Es sieht so aus, als hätte in der Unterschichtgruppe eine

1)-----

Suchen wir die ursächlichen Einzeltests, so stehen hinter den positiven Korrelationen von VERBIT mit dem 16er Test vor allem TEST 6 (Verstümmelte Wörter) und - weniger stark TEST 1/2 (allgemeine Wortkenntnis) und hinter den negativen Korrelationen von NONVER mit dem 42er Test natürlich TEST 10 (Wahrnehmungstempo) und - weniger stark - TEST 3 und 4 (logisches Schließen mit Figuren bzw. Buchstaben und Zahlen). Hinter den positiven Korrelationen beider Intelligenzarten mit dem 42er Test stehen TEST 5 (Worteinfall), TEST 8 (embedded figures) und TEST 10. Zwei Maße korrelieren zudem positiv mit dem 16er und negativ mit dem 42er Test: TEST 8 und die Note in Rechnen.

unstabile und ungleichartige Beziehung der einzelnen Untertests der dritten Messung¹⁾ zu verschiedenen Einzelbereichen der Intelligenz die zugrundeliegende negative Beziehung zwischen nicht-verbaler Intelligenz und Experimentalleistung in beiden Untertests verschleiert.

Wenn diese Vermutung zutrifft - sie stützt sich ja nur auf ein sehr undurchsichtiges Muster von Einflüssen -, dann ließe sich in der dritten Messung, anders als in der zweiten Messung, ein Muster schichtspezifischer Zusammenhänge zwischen Intelligenz und Sprachverhalten wiederfinden, das Oevermann (1970:230-234) beschrieben hat. Oevermann fand bei einer Untersuchung von Aufsätzen der 7. Realschulklasse innerhalb der Mittelschichtgruppe positive Korrelationen der verbalen und der nicht-verbalen Intelligenz mit Variablen der grammatischen Struktur, in der Unterschichtgruppe positive Korrelationen der verbalen, aber negative Korrelationen der nicht-verbalen Intelligenz. Er vermutet, daß gerade die intelligenteren Kinder der Unterschicht sich besser an die der Mittelschicht- und Schulkultur widersprechenden Bedingungen ihres Herkunftsmilieus anpassen und der Anpassung an mittelschicht- und schulspezifische "elaborierte" Sprachformen länger Widerstand leisten. Im Rahmen einer solchen Erklärung bringt die Tatsache, daß die hier berichteten Korrelationen sich auf Kinder des 4. Grundschuljahrgangs beziehen, zwei neue Aspekte. Erstens kann aus dem Auftauchen der gleichen Struktur von Korrelationen auch bei jüngeren Kindern auf die Langwierigkeit und Konflikträchtigkeit des "Umpolens" der

1)-----
Ein Vergleich der 6 Korrelationen des mittleren Rangs mit NONVER zwischen beiden Schichten (Variable 71, 72, 73, 74, 75, 76 in Tab. 197 und 198) mit Hilfe der Fisher-Transformation von r nach z (siehe Hays 1963:532) ergab für die Variable 71 (MRAM) eine signifikante Differenz bei einem zweiseitigen Test: $Z = 1.883, p < .06$. Für die fünf übrigen Variablen ergaben sich Z-Werte zwischen 1.08 für 42MRA0 und 1.602 für 42MRAM. Insgesamt kann die Signifikanz der Differenz zwischen beiden Schichtgruppen nur schwach gesichert werden. Es ist nach diesen Ergebnissen jedoch sicher, daß die Schichtdifferenz der Korrelationen von TEST 10 mit den Maßen der dritten Messung, von der wir ja ausgegangen waren, signifikant ist.

Unterschichtkinder von Herkunfts- auf das Schulumilieu geschlossen werden. Zweitens ist durch die Tatsache, daß die untersuchten Kinder noch nicht für weiterführende Schulen ausgelesen sind, eine alternative Erklärung der Korrelationsstruktur ausgeschlossen. In der Realschulgruppe konnte sie auch so erklärt werden, daß durch die Vorherrschaft sprachlicher Kriterien bei der Auslese aus der Unterschicht viele Kinder aufgrund ihrer guten "angepaßten" verbalen Leistungen trotz unterdurchschnittlicher nicht-verbaler Leistungen und ebenso viele Kinder nur aufgrund der Kompensation mangelnder verbaler Leistungen durch überdurchschnittliche nicht-verbale für den weiterführenden Schulbesuch ausgewählt wurden.

3.6.3. Sozialschicht oder Wortkenntnis: Ergebnisse der multiplen Regressionsanalyse.

Wir haben gezeigt, daß die Schichtindikatoren keinen oder nur einen aus ihrem Zusammenhang mit der Kontrollvariablen WST erklärbaren Einfluß auf die Experimentalleistungen haben (3.6.1.) und daß der Wortschatztest in beiden Schichtgruppen mit allen Experimentalleistungen hoch korreliert (3.6.2.). Wenn wir den Determinationskoeffizienten errechnen, so erklärt der Wortschatztest 24% der Varianz der Gesamt-, 37% der Varianz der Unterschicht- und 26% der Varianz der Mittelschichtgruppe als einzelner Faktor (Tab. 200). Aus dem Vergleich der Korrelationen haben wir Hinweise, daß der Einfluß des Wortschatztests sowohl von den Schichtindikatoren wie den übrigen Testleistungen unabhängig ist. Wir können diese Vermutungen statistisch mit Hilfe der multiplen Regressionsanalyse prüfen; dazu wollen wir kurz an die Hypothesen des Experiments erinnern und die Besonderheiten des statistischen Verfahrens erläutern.

Die Hypothese der dritten Messung war, daß Kinder mit dem gleichen Wortschatz die geforderten kategorialen Differenzierungen gleich gut bewältigen würden - unabhängig von der Schichtzugehörigkeit. Die beiden Einzelelemente dieser Hypothese, Unabhängigkeit von der Sozialschicht und Abhängigkeit vom Wortschatz, sind bestätigt worden. Beide Seiten der Hypothese können zugleich in der multiplen Regressionsanalyse geprüft werden, die die spezifischen, von allen anderen unabhängigen Beiträge jedes einzelnen Faktors für die dritte Messung herausarbeitet. Die multiple Regressionsanalyse gibt an, wieviel von der Varianz der abhängigen Variablen (Kriterium) durch eine Reihe von unabhängigen Variablen (Prädiktoren) erklärt wird: R^2 , und wieviel jede einzelne unabhängige Variable für sich zur Voraussage der abhängigen Variablen beiträgt: das sog. Beta-Gewicht; die Signifikanz von R^2 kann mit einem F-Test überprüft werden. Multiple Regressionsanalysen können mit unterschiedlichen Anzahlen von Prädiktoren beliebig oft wiederholt werden. Wir haben eine Analyse mit den (1) Schulnoten und den Schichtindikatoren als Prädiktoren, die (2) mit den Schulnoten als Prädiktoren wiederholt wurde, sowie eine Analyse (3) mit den zusammen mit dem Wortschatztest insgesamt 10 Intelligenzsubtests und den Schichtindikatoren als Prädiktoren durchgeführt, die (4) mit den Intelligenzsubtests als Prädiktoren und (5) den Schichtindikatoren als Prädiktoren wiederholt wurden. Als Kriterien nahmen wir bei den insgesamt 5 Analysen die negativen Gesamtmaße. In die multiple Regressionsanalyse können nur Vpn. eingehen, die für alle Prädiktoren Werte haben; eine Vp., die auch nur auf einer der Prädiktoren keinen Wert hat, fällt aus der ganzen Analyse. Daher sind 7 von 42 Vpn. nicht berücksichtigt worden.¹⁾

¹⁾-----
Es handelt sich um die Vpn. 205, 233, 618, 625, 904, 919, 1120; 4 Jungen aus der Unterschicht, 1 Junge aus der Mittelschicht, 2 Mädchen aus der Mittelschicht. Die Mittelwerte auf den Kontrollvariablen für die 4 Kinder aus der Unterschicht sind 4.1/4.6, für die 3 Kinder aus der Mittelschicht 5.8/5.7. (siehe Tab. 161).

Die multiplen Regressionen der Schulnoten auf die bewertenden Maße der dritten Messung sind nicht signifikant, die Beta-Gewichte der einzelnen Schulnoten sehr niedrig. Dies mag nach den hohen Einzelkorrelationen der Schulnoten mit den Experimentalmaßen in der Gesamtgruppe (Tab. 196 und 200) überraschend sein, ergibt sich jedoch wahrscheinlich aus den hohen Interkorrelationen der Schulnoten (Tab. 167); sobald eine "Multikollinearität" der Prädiktoren vorliegt, ist eine der Anwendungsbedingungen der multiplen Regressionsanalyse zur Abschätzung des relativen Gewichts der Prädiktoren verletzt (siehe Kapitel III.5.6.2.). Die gleichen Überlegungen und Ergebnisse gelten für die multiple Regression mit den Schulnoten und den Schichtindikatoren als Prädiktoren. Zwar sind die Korrelationen der Noten mit den Schichtindikatoren gering, aber die Interkorrelationen der Schichtindikatoren den Noten vergleichbar hoch; wir haben jedoch schon in Kapitel III.5.6.2. Gründe angegeben, warum uns im Falle der Schichtindikatoren die Multikollinearität der Prädiktoren kein Hindernis für die Anwendung der multiplen Regressionsanalyse zu sein schien.

Die drei multiplen Regressionsanalysen mit den Intelligenz-Subtests und den Schichtindikatoren als Prädiktoren betreffen unmittelbar die Prüfung der Hypothese; sie sollen zudem über unterschiedliche Voraussetzungen für die beiden Untertests der dritten Messung Auskunft geben. Die Interkorrelationen der Intelligenzsubtests sind minimal, die hohen Interkorrelationen der Schichtindikatoren u.E. unumgänglich, die Korrelationen zwischen Schichtindikatoren und Intelligenzsubtests unbedeutend. In der Analyse mit den 4 Schichtindikatoren als Prädiktoren ergaben sich nur für die Wahlhäufigkeit der sehr schlechten Vorgaben signifikante R^2 ; diese Maße könnten auf durch die Bewertung künstlich produzierte Schichtunterschiede hinweisen, wir haben sie bei der Darstellung der Schichtunterschiede besprochen (siehe 3.2., Tab. 178). In der Analyse mit den 9 Intelligenzsubtests und dem Wortschatztest als Prädiktoren ergaben sich für die Maße des Gesamt- und des 42er Tests keine signifikanten R^2 , wohl aber für die Maße des 16er Tests.

In der Analyse mit den Intelligenz- und Wortschatztests sowie den Schichtindikatoren als Prädiktoren ergaben sich ebenfalls für den gesamten und den 42er Test keine signifikanten R^2 , und für den 16er Test fast signifikante R^2 . Nach diesen Ergebnissen werden wir die Betrachtung der Beta-Gewichte zur Prüfung der Ausgangshypothese auf den 16er Test einschränken (Tab. 201), zum Vergleich der Einflußstruktur beider Untertests aber dennoch die Beta-Gewichte der nicht signifikanten multiplen Regressionen auf die Maße des 42er Tests heranziehen (Tab. 202). Wir wollen zuerst auf die Prüfung der Ausgangshypothese anhand der Ergebnisse des 16er Tests eingehen.

Betrachtet man zunächst die Intelligenzsubtests, so zeigt sich, daß der Wortschatztest bei weitem die höchste spezifische Voraussagekraft hat, mit etwas Abstand gefolgt von TEST 5, dessen Sonderstellung schon bei den Korrelationen aufgefallen war. Die von den Korrelationen gewonnene Vermutung, daß der Wortschatztest den größten Einfluß auf die kategoriale Differenzierung ausübt, wird damit bestätigt. Mehr noch: Alle 10 Intelligenzsubtests können die Experimentalleistung weniger sicher und nur geringfügig besser voraussagen als der Wortschatztest allein, wie ein Vergleich mit der Signifikanz und der erklärten Varianz der einfachen Korrelation zwischen Wortschatztest und 16er Test der dritten Messung in der Gesamtgruppe zeigt (Tab. 200). Allerdings schwankt die Signifikanz auch noch der multiplen Regression der Intelligenzsubtests auf die Maße des 16er Tests um das 5%-Niveau. Die Multiple Regression der Schichtindikatoren auf die Maße des 16er Tests hingegen ist für kein Maß signifikant. Den stärksten spezifischen Einfluß hat zweifellos die Ausbildung des Vaters und das Einkommen, während der gegenwärtige Beruf des Vaters eher negativ auf die Leistung im 16er Test wirkt. Auch diese Ergebnisse stimmen mit dem überein, was von den Korrelationen her zu erwarten war (Tab. 199). Wie die Korrelationen können wir auch die Beta-Gewichte der Schichtindikatoren auf ihre Korrelation mit der Kontrollvariablen WST zurückführen (Tab. 166). Anders als in der zweiten Messung ergibt sich also keine Differenzierung zwischen den einzelnen Schichtindikatoren bezüglich der Voraussagekraft der Experimentalleistungen.

Betrachtet man schließlich die multiple Regression der Intelligenzsubtests und Schichtindikatoren, so hat die Signifikanz gegenüber der der Schichtindikatoren zu-, gegenüber der der Intelligenzsubtests aber abgenommen; dadurch daß die Schichtindikatoren zu den Intelligenzsubtests als Prädiktoren hinzukommen, ist die erklärte Varianz nur minimal gestiegen. Die Struktur der Beta-Gewichte der Intelligenzsubtests bleibt erhalten, die der Schichtindikatoren jedoch nicht. Nach wie vor hat der Wortschatztest die größte spezifische Voraussagekraft.

Als Ergebnis der drei multiplen und der einfachen Regressionsanalysen kann festgehalten werden, daß der Wortschatztest alleine die Leistung im 16er Test fast genau so gut, aber sicherer voraussagt als in Verbindung mit den Intelligenzsubtests und den Intelligenzsubtests sowie den Schichtindikatoren, und erst recht besser als die Schichtindikatoren alleine. Durch die Tatsache, daß er als Kontrollvariable diente, ist dies Ergebnis noch keineswegs verbürgt. Die beiden Seiten der Ausgangshypothese sind damit auch in einer simultanen Analyse bestätigt.

Wir wollen als zweites die Prädiktorenstruktur der beiden Untertests der dritten Messung vergleichen (Tab. 2o1 und 2o2). Daß die R^2 allein für den 16er Test signifikant sind, kann nach den ungleich häufigeren und höheren Einzelkorrelationen der Intelligenzmaße mit dem 16er Test nicht mehr überraschen (Tab. 2oo). Daß von den Prädiktoren nur der Wortschatztest in beiden Untertests positive Beta-Gewichte über .4oo aufweist, war von den hohen Korrelationen des WST mit beiden Untertests ebenfalls zu erwarten. Weiterhin finden sich in beiden Untertests bei TEST 1o (Wahrnehmungstempo) negative Beta-Gewichte über .2oo. TEST 1o korrelierte innerhalb der Unterschicht negativ, innerhalb der Mittelschicht positiv mit beiden Untertests der dritten Messung (Tab. 2oo). Daß er in der multiplen Regressionsanalyse innerhalb der Gesamtgruppe ein negatives Beta-Gewicht erhält, deutet darauf, daß die positiven Einflüsse des TEST 1o in der Mittelschichtgruppe schon durch die zahlreichen anderen positiven Einflüsse von Intelligenz-Subtests innerhalb der

Mittelschichtgruppe abgedeckt sind, so daß sich nur noch die negativen Einflüsse innerhalb der Unterschicht als eigenständige Voraussage in der Gesamtgruppe bemerkbar machen können. Beim 16er Test zeigen noch die beiden verbalen Subtests TEST 5 (Worteinfall) und TEST 6 (Verstümmelte Wörter) höhere (meist über .400 liegende) Beta-Gewichte, beim 42er Test dagegen der nicht-verbale Subtest TEST 3 (logisches Schließen anhand von Figuren). Von der multiplen Regressionsanalyse her deutet sich also etwas besser als von den einfachen Korrelationen her wiederum an, daß der "kategoriale Präzisionstest" stärker von der verbalen Intelligenz, der "differenzierte Verständnistest" stärker von der nicht-verbalen Intelligenz beeinflusst ist. Wir müssen aber wiederholen, daß es schwierig war, dieses Muster bei Trennung der beiden Schichtgruppen wiederzufinden.

3.6.4. Zusammenfassende Beurteilung des Einflusses von Schicht, Intelligenz und Wortschatz auf die kategoriale Differenzierung.

Zu Beginn hatten wir dem Abschnitt 3.6. zwei Aufgaben zugewiesen: Er sollte erstens die Voraussagen über die tatsächlich in der dritten Messung relevanten Variablen prüfen und zweitens schichtspezifische Einflußstrukturen der Schichtindikatoren, der Intelligenztests, des Wortschatztests und der Schulnoten auf die beiden Untertests, sowie spezifische Einflußvariablen für jeden der beiden Untertests der dritten Messung herausarbeiten. Die Ausgangshypothese konnte sowohl durch die unwesentlichen Korrelationen der einzelnen Schichtindikatoren wie die hohen Korrelationen des Wortschatztests mit den Maßen beider Messungen bestätigt werden. Sie konnte ferner bestätigt werden an der Struktur der Beta-Gewichte der Intelligenzmaße und der Schichtindikatoren in der multiplen Regressionsanalyse.

Überraschend war jedoch, daß in der Gesamtgruppe kein Intelligenzmaß mit den Maßen der beiden Untertests stärker korreliert. Genauso wie in der zweiten Messung finden wir auch in der dritten Messung, daß nur der Wortschatztest, aber kein Intelligenzmaß mit den Experimentalleistungen zusammenhängt (siehe III.5.6.1.).

In der Mittelschichtgruppe fanden sich mit beiden Testteilen durchgängig hohe positive Korrelationen der Schulnoten und Intelligenzsubtests, in der Unterschichtgruppe jedoch eine undurchsichtige Struktur von Zusammenhängen, in der allein die negativen Korrelationen nicht-verbaler Testleistungen bemerkenswert waren. Der "kategoriale Präzisionstest" zeigt in allen Gruppen mehr und stärkere Zusammenhänge mit den Intelligenztests und Schulnoten als der "differenzierte Verständnistest". Vorwiegend ergeben sich positive Korrelationen und Beta-Gewichte des "kategorialen Präzisionstests" mit verbalen, des "differenzierten Verständnistests" mit nicht-verbalen Testleistungen.

3.7. Einflüsse der Kommunikations- und Kontrollweisen der Mutter auf die kategoriale Differenzierung des Kindes.

Mit der Darstellung der Einflüsse der Kommunikations- und Kontrollweisen der Mutter verlassen wir die Hypothesenprüfung - sei es der Widerlegung von Einflüssen der sozialen Schicht (3.2.), sei es der Bestätigung des Einflusses von Wortschatz und allgemeiner Intelligenz in beiden Schichten (3.6.) - und gehen zur Darstellung im Experimentalplan nicht berücksichtigter Faktoren über. Eine Reihe von Variablen, die wir in der zweiten Messung geprüft haben, werden in der dritten Messung jedoch nicht mehr berücksichtigt, weil die Untersuchungsgruppe der dritten Messung zu klein ist und weil die Ergebnisse schon von der zweiten Messung her wenig ermutigend waren. Allein die Kommunikations- und Kontrollweisen der Mutter, die in der zweiten Messung ohne Bedeutung waren (siehe Kapitel III.5.7.), zeigen auf die dritte Messung nennenswerte Einflüsse; da sie theoretisch mit der sozialen Schichtzugehörigkeit verbunden sind, werden wir sie zusammen mit ihren Korrelationen mit dem Filterindex in der Gesamt- und der Experimentalstichprobe betrachten.

Die Instrumente zur Erhebung der mütterlichen Kontroll- und Kommunikationsweisen wurden schon bei der Darstellung der Ergebnisse der zweiten Messung genauer beschrieben. Es handelte sich um Fragen über die Bereitschaft der Mutter, auf vom Kind initiierte verbale Interaktion einzugehen ("Kommunikationsfrage"), über die Bedeutung der Sprache für die Mutter-Kind-Beziehung ("Sprachfrage") und über die Beteiligung des Kindes an einer Auseinandersetzung zwischen den Ehepartnern ("Nachbarnfrage"). Alle diese Fragen sollten nur von der Mutter beantwortet werden. Daß diese Fragen - in ihrer deutschen Übersetzung - kein einheitliches Instrument mit statistisch gesicherten Dimensionen darstellen, wurde ebenfalls schon bei der Darstellung der Ergebnisse der zweiten Messung ausgeführt. Als Hypothese liegt nahe, daß größere Bereitschaft zu verbaler Interaktion seitens der Mutter - oder die erklärte Bereitschaft, denn wir haben nur die Norm, nicht das Verhalten erfragt -

die kategoriale Differenzierungsfähigkeit des Kindes fördert. Dies müßte in jeder Schichtgruppe gelten. Hypothesen über den Zusammenhang der Bewertung der Sprache durch die Eltern und der kategorialen Differenzierungsfähigkeit des Kindes sind nicht einfach zu formulieren. Einerseits kann eine hohe Einschätzung der Sprache mit einem bewußten, sogar lehrenden Gebrauch der Sprache seitens der Eltern verbunden sein, der zu guter Differenzierungsfähigkeit des Kindes führt. Andererseits kann eine geringe Betonung der Sprache durch die Eltern selber schon eine Reaktion auf gute sprachliche Fertigkeiten des Kindes sein. Ob die Eltern ihrer Bewertung entsprechend die Sprache gebrauchen oder ob sie mit ihrer Bewertung nur auf eine Einschätzung des Kindes reagieren, kann zudem selber schichtspezifisch sein. Daher kann man unterschiedliche Zusammenhänge des gleichen Erziehungsverhaltens mit den Experimentalleistungen in beiden Schichtgruppen vermuten. Bei der Darstellung der Ergebnisse wollen wir von der plausibelsten und einfachsten Hypothese ausgehen: daß hohe Bewertung der Sprache in allen Schichtgruppen positiv mit der kategorialen Differenzierungsfähigkeit der Kinder zusammenhängt.

Die Korrelationen der Einstellungen der Eltern zum Sprachgebrauch mit den Leistungen des Kindes in der dritten Messung sind in Tab. 203 so dargestellt, daß das mit den guten Leistungen des Kindes korrelierende Einstellungsextrem sichtbar wird. Für jeden Untertest ist der Median der Korrelation aller Maße aufgeführt, sofern er über .20 liegt. Bei den Kommunikations- und Sprachfragen sind die Korrelationen mit dem 42er Test größer und häufiger als mit dem 16er Test - umgekehrt also wie bei den Intelligenztests und Schulnoten. Der "kategoriale Präzisionstest" hängt stärker mit den Intelligenz- und Schulleistungen, der "differenzierte Verständnistest" mit den Einstellungen der Eltern zu Sprache und Erziehung zusammen. Aus dem Charakter der beiden Testleistungen ist das verständlich: im "kategorialen Präzisionstest" wurden minimale kritische Unterschiede einer semantischen Dimension, im "differenzierten Verständnistest" unterschiedliche Weisen des Gebrauchs der Sprache (wie Umschreibung oder kategoriale Definition) abgefragt.

Zwischen den beiden Untertests gibt es keine Unterschiede in der Richtung der Korrelationen - mit einer Ausnahme bei der Nachbarnfrage in der Unterschichtgruppe, die uns unverständlich ist. Fast alle Items mit Korrelationen beim 42er Test zeigen auch Korrelationen beim Gesamttest - woran der größere Anteil des 42er Tests am Gesamttest natürlich mitwirkt. Eine Reihe von Items schließlich haben Korrelationen in beiden Untertests. Aus diesen Gründen werden wir in der folgenden anders als in den vorausgegangenen Analysen nicht mehr zwischen den beiden Untertests unterscheiden, stützen die Darstellung aber hauptsächlich auf den 42er Test.

Auffällig ist in der Gesamtgruppe die geringe Anzahl und Höhe der Korrelationen. Treten bei der Kommunikationsfrage Korrelationen auf, so hängt die Unterhaltungsbereitschaft der Mutter mit guten Leistungen in der dritten Messung zusammen; treten bei der Sprachfrage Korrelationen auf, so hängt die geringe Betonung der Sprache mit guten Leistungen in der dritten Messung zusammen. Die positiv korrelierenden Situationen der Kommunikationsfrage sind dadurch gekennzeichnet, daß sie von der Sache her die Teilung der Aufmerksamkeit auf zwei Aktivitäten erlauben, gleichsam durch eine "Offenheit" oder "Zweiwertigkeit" (C1: Hausarbeiten, C2: Straße, C6: Bus oder Straßenbahn); von den nicht korrelierenden Situationen hat nur C7 (Mahlzeit) diesen Charakter. Trennt man die beiden Schichtgruppen, so fällt sofort auf, daß Anzahl und Stärke der Korrelationen zunehmen. Teilweise verstärken sich dabei schon in der Gesamtgruppe beobachtete Korrelationen, teilweise brechen vorher verwischte Gegensätze auf.

Verstärkt hat sich in der Unterschichtgruppe und in Kommunikationsfrage der Zusammenhang zwischen Unterhaltungsbereitschaft der Mutter und Differenzierungsfähigkeit des Kindes. Zusätzlich taucht er jetzt auch in den Situationen C7 (Mahlzeit) und C3 (Entspannung) auf. Der ersten, nicht aber der zweiten Situation hatten wir von Anfang eine "Zweiwertigkeit" zugestanden. Nachträglich aber läßt sich "Entspannung" doppelt interpretieren: als Anspruch auf abgeschirmte, individuell auszulebende Erholung oder als Bereitschaft zu äußeren Anregungen und Zerstreuungen. Wenn wir das zweite Verständnis für in der Unterschicht wahrscheinlicher halten, dann fügt sich auch diese Situation dem schon bei der Gesamtgruppe verwandten Interpre-

tationsschema. Den gleichen für beide Schichtgruppen gegensätzlichen Einfluß des mütterlichen Kontrollverhaltens hatten wir schon für die zweite Messung feststellen können (siehe III.5.7.).

Verstärkt hat sich in der Mittelschichtgruppe und in der Sprachfrage der Zusammenhang zwischen geringer Betonung der Sprache und guten Leistungen in der dritten Messung, am deutlichsten bei den Items 9 ("Sich mit ihnen beschäftigen, wenn Kinder sich unglücklich fühlen"), 11 ("Den Kindern seine Zufriedenheit zeigen, wie gut sie sich gemacht haben."), 8 ("Den Kindern zeigen, was richtig und falsch ist") und 3 ("Die Kinder wissen lassen, was man selbst fühlt und empfindet"). Der gemeinsame Nenner der Items 9, 11 und 3 ist ein diffuser affektiver Austausch zwischen Eltern und Kind, das Item 8 bleibt isoliert. Die negativen Korrelationen des Items 8 mit dem gesamten und dem 42er Test stehen ferner in gewissem Widerspruch zu den positiven Korrelationen des Items 13 mit dem 16er Test; beide behandeln die Übermittlung moralischer Prinzipien. Eine weitere Gruppe mit negativen Korrelationen wird durch die beiden "kognitiven" Items 10 und 12 gebildet; aber auch hier stehen die negativen Korrelationen des Items 10 mit dem 42er Test in gewissem Widerspruch zu den positiven Korrelationen des eng verwandten Items 4 mit dem 16er Test. Die Zusammenhänge der kategorialen Differenzierung mit dem Erziehungsverhalten im evaluativen und kognitiven Bereich sind also nicht eindeutig. Eindeutig hängt in der Mittelschichtgruppe eine geringe Betonung der Sprache durch die Eltern nur im affektiv diffusen Bereich mit besseren Leistungen der Kinder in der kategorialen Differenzierung zusammen.¹⁾ - Das häufigere Auftreten von Korrelationen mit der Sprachfrage in der Mittelschicht wird durch das häufigere Auftreten von Korrelationen mit der Kommunikationsfrage in der Unterschicht nicht ganz kompensiert. Ähnlich wie

¹⁾-----
Die schichtspezifischen Interkorrelationen der affektiv diffusen Items in der Mittelschicht (N=21) sind: S 3 mit S 9 +.31;
S 3 mit S 11 +.33; S 9 mit S 11 +.81 ($p < .001$).

Intelligenz und Schulnoten scheint auch das Erziehungsverhalten der Eltern mit dem sprachlichen Lernen der Kinder der Unterschicht schwächer als der Mittelschicht integriert zu sein.

Vergleicht man die Einflüsse des Erziehungsverhaltens auf die kategoriale Differenzierung in beiden Schichtgruppen, so ist der Gegensatz unübersehbar. Während in der Unterschicht Bereitschaft zu verbaler Interaktion und hohe Bewertung der Sprache für die Eltern-Kind-Beziehung mit besseren Leistungen des Kindes in der dritten Messung zusammenhängen, so sind es in der Mittelschicht Abweisung verbaler Interaktion und geringe Bewertung der Sprache. Man kann jedoch annehmen, daß die Bereitschaft zu verbaler Interaktion generell bei Müttern aus der Mittelschicht größer ist - was mit dem gleichen Instrument von Brandis und Henderson (1970) wie von Bosse u.a. (1971) nachgewiesen werden konnte - und daß Eltern der Mittelschicht die Bedeutung der Sprache im personalen bzw. affektiv diffusen Bereich¹⁾ höher einschätzen - was mit dem gleichen Instrument von Bernstein und Henderson (1969), nicht aber von Bosse u.a. (1971)²⁾ nachgewiesen werden konnte. M.a.W.: In jeder

- 1)-----
- 1) Die ursprünglichen Hypothesen Bernsteins (Bernstein/Henderson 1969) lauteten
- Beide Schichten bewerten Sprache im "person" Bereich höher als im "skill" Bereich.
 - Die Zunahme der Bedeutung der Sprache im personalen Bereich ist in der Mittelschicht größer.
 - Innerhalb des sachbezogenen Bereichs bewerten die Mittelschichteltern die Sprache für die Übertragung von Prinzipien (Item 10) am höchsten.

Alle drei Hypothesen konnten von Bernstein/Henderson bestätigt werden. Für unseren Zweck genügt, als eine vereinfachte Darstellung dieser Hypothesenstruktur, die Aufzählung der Bereiche, für die eine höhere Bewertung der Sprache durch die Mittelschicht vorausgesagt wird. Auch wenn wir unter die "affektiv-diffusen" Items ein "dummy"-Item gezählt haben, deckt sich diese Gruppe von Items mit dem wichtigsten Teil der Bernsteinschen "person" Gruppe.

- 2) In der Stichprobe unseres Experiments (N=42) finden sich die gleichen Ergebnisse wie in der von Bosse u.a. (N=392) -, mit verstärkter Tendenz (siehe Tab. 79). Das ist nicht verwunderlich, da unsere Stichprobe ja eine Untergruppe der Stichprobe von Bosse u.a. darstellt.

Schichtgruppe hängen die untypischen Beziehungsweisen mit guten Leistungen der Kinder im passiven Wortverständnis zusammen. Um diesen etwas paradoxen allgemeinen Zusammenhang aufzuklären, soll er in den einzelnen im Fragebogen angesprochenen Lebensbereichen überprüft werden; dazu müssen sie nach ihren Beziehungen zu den Leistungen des Kindes und zu der Schichtzugehörigkeit der Eltern klassifiziert werden. Bei der Suche schichtspezifischer Einflußfaktoren müssen wir ihre Gleichwertigkeit für die Experimentalleistung im Auge behalten: da keine Schichtunterschiede in der kategorialen Differenzierung festgestellt wurden, müssen sie auf verschiedene Weise zum gleichen Ergebnis geführt haben.

Für die Kommunikationsfrage gilt die Hypothese, daß in allen Bereichen die Interaktionsbereitschaft bei den Müttern der Mittelschicht größer ist und daß größere Bereitschaft zu verbaler Interaktion zu besserer kategorialer Differenzierungsfähigkeit des Kindes führe. Zwei Bereiche zeigen übereinstimmend in Mittel- und Unterschicht einen positiven Zusammenhang zwischen Gesprächsbereitschaft der Mutter und Differenzierungsfähigkeit des Kindes: C 1, Hausarbeiten; und C 6, Bus oder Straßenbahn. Es handelt sich um einen "privat offenen" und einen "öffentlich offenen" Bereich. Sowohl in der Stichprobe unseres Experiments wie in der Gesamtstichprobe, aus der sie gezogen wurde, finden sich relativ hohe Korrelationen des Items 7, aber geringe des Items 6 mit dem Schichtindex (Tab. 204). Während Gesprächsbereitschaft der Mutter in öffentlichen Verkehrsmitteln (Item 6) also unabhängig von der Schichtzugehörigkeit auftritt und in jeder Schichtgruppe zu dem gleichen, positiven Erfolg beim Kind führt, scheinen bei Item 1 gegenläufige soziale Einflüsse zu den Mittelschichtkindern vergleichbaren Leistungen der Unterschichtkinder geführt zu haben: Mütter der Unterschicht sanktionieren zwar öfter Schwatzhaftigkeit der Kinder während der Hausarbeit, aber auch in der Unterschicht wird durch Unterhaltungsbereitschaft in dieser Situation die Differenzierungsfähigkeit des Kindes gefördert. Der gleiche gegenläufige Einfluß innerhalb der Unterschicht findet sich bei einem weiteren

"offenen" Bereich, für den sich in der Mittelschicht keine Zusammenhänge mit der Testleistung des Kindes zeigen: C 2, "Straße". Der Rückblick auf die Beziehungen zwischen Sozialschicht und Erziehungsverhalten ergab also zwei Typen von Bereichen des Erziehungsverhaltens: im ersten fördert sprachliche Interaktion das kategoriale Differenzierungsvermögen des Kindes unabhängig von der Schichtzugehörigkeit der Eltern (C 6); im zweiten fördert sprachliche Interaktion das kategoriale Differenzierungsvermögen des Kindes in beiden (C 1) oder allein der Unterschicht (C 2), auch wenn die Eltern der Unterschicht in beiden Bereichen zu sprachlicher Interaktion seltener bereit sind. Alle Bereiche zeigen keine Unterschiede im Zusammenhang zwischen Verhalten der Mutter und Leistung des Kindes; alle sind "offene" Bereiche.

Schichtunterschiede des Zusammenhangs zwischen mütterlichem Verhalten und Leistung des Kindes zeigen sich in zwei Bereichen der Kommunikationsfrage: In C 3, "Entspannung", und C 7, "Mahlzeit", führt die Gesprächsbereitschaft der Mutter in der Unterschicht zu guten, in der Mittelschicht zu schlechten Leistungen des Kindes. Beides sind "private" Bereiche, während ihre Charakterisierung als "offen" oder "geschlossen" nicht eindeutig ist. Beide korrelieren unterschiedlich innerhalb der Experimentalgruppe mit dem Filterindex: Während die Mutter der Mittelschicht bei Mahlzeiten auf das Schwätzen des Kindes eingeht ($r = -.54$; $p < .001$), bei der Entspannung ihm aber den Mund verbietet ($r = +.34$; $p < .05$; Tab. 204), ist es bei den Müttern der Unterschicht umgekehrt. Offenbar stellen sich Mahlzeiten den Müttern der Mittelschicht als "offene" und den Müttern der Unterschicht als "geschlossene" Situationen dar, während Entspannung von den Müttern der Mittelschicht als "geschlossene" und den Müttern der Unterschicht als "offene" Situationen verstanden werden.¹⁾ Damit sind zwei weitere Typen von Bereichen des Erziehungsverhaltens gefunden. Für den Bereich der Entspannung -als Beispiel

7)-----
Die schichtspezifischen Interkorrelationen zwischen C 3, "Entspannung", und C 7, "Mahlzeit" sind $-.24$ in der Mittelschicht (trotz des gleichartigen Einflusses auf die Leistungen der dritten Messung) und $+.58$ in der Unterschicht.

des dritten Typs - hängt das jeweils schichtspezifische Verhalten der Mutter mit besseren Leistungen des Kindes zusammen. Für den Bereich der Mahlzeiten - als Beispiel des vierten Typs, der dem anfangs vorgestellten allgemeinen Fall entspricht - hängt das jeweils schichtuntypische Verhalten der Mutter mit besseren Leistungen des Kindes zusammen. Zum vierten Typ gehört auch der Bereich C 5, "Einkauf", für den sich in der Unterschicht keine Zusammenhänge mit der Testleistung des Kindes zeigen. Auch hier geht die Mutter der Mittelschicht auf das Schwätzen des Kindes ein ($r = -.36$; $p < .05$; Tab. 204), aber ihre Gesprächsbereitschaft korreliert mit schwachen Leistungen des Kindes; "Einkauf" scheint wie "Mahlzeiten" für die Mutter der Mittelschicht eine "offene" Situation zu sein.

Für die Sprachfrage müssen die Hypothesen innerhalb der Bereiche differenziert werden. Eine größere Betonung der Sprache in der Mittelschicht läßt sich mit Bernstein voraussagen für die personalen, die (teilweise mit den personalen identischen) affektiv-diffusen und die auf Funktionsprinzipien bezogenen sachlichen Items (S 10), insgesamt also für die Items 3 und 7 bis 13. Wenn wir von den vereinzelt recht niedrigen, Korrelationen der Items 2, 4, 5 und 6 absehen, so decken sich die Bereiche, für die eine größere Betonung der Sprache in der Mittelschicht vorausgesagt wurde, mit den Bereichen, bei denen nennenswerte Zusammenhänge der Erziehungsweisen mit kategorialer Differenzierung auftauchen; zumindest ist für alle Bereiche mit einem signifikanten Einfluß auf kategoriale Differenzierung (S 8 - S 11) eine größere Betonung der Sprache in der Mittelschicht wahrscheinlich. Diese Hypothese kann jedoch - anders als in Bernsteins Untersuchung - für keine der zuletzt genannten Bereiche bestätigt werden (Tab. 204) - weder in der Experimental- noch in der Gesamtgruppe. Hinweise auf die Schichttypik der Bereiche lassen sich daher nur noch aus einem gegensätzlichen Zusammenhang mit kategorialer Differenzierung in beiden Schichtgruppen gewinnen.

Ein scharfer Gegensatz des Einflusses ergibt sich in einem affektiv diffusen Bereich (S 11). Daß Eltern der Mittelschicht, die auf Sprache im affektiv diffusen Bereich großen Wert legen, das Differenzierungsvermögen des Kindes negativ beeinflussen, ist unwahrscheinlich. Vielmehr scheint die Richtung des Einflusses umgekehrt zu sein: Gerade bei den verbal gewandteren Kindern ist eine affektiv diffuse Kommunikation so unproblematisch, daß eine Hervorhebung der Leistungen der Sprache für die Eltern unnötig erscheint. Für die Unterschichtgruppe - so kann man annehmen - tritt eine derartige nachträgliche Reaktion der Eltern auf erkannte Fähigkeiten des Kindes nicht ein und die ursprünglich angenommene Richtung des Einflusses bleibt gewahrt: Die Eltern, die auf Sprache in affektiv-diffuser Kommunikation Wert legen, fördern die kategoriale Differenzierungsfähigkeit ihrer Kinder. - Die gleiche Umkehrung der Einflußrichtung - nun allerdings in der Unterschicht - scheint sich bei einem kognitiven Item (S 10) wiederzufinden: die Korrelationen geringer Betonung der Sprache mit guter Differenzierungsfähigkeit sind hier in der Unterschicht häufiger und größer als in der Mittelschicht. Offenbar schätzen in der Unterschicht die Eltern gerade der verbal geschickten Kinder die Übermittlung von Funktionsprinzipien so unproblematisch ein, daß sie die Notwendigkeit von Sprache nicht besonders betonen. Für die beiden letzten Items (S 10 und S 11) haben wir die Interpretation der Ergebnisse von der Ebene der Betonung der Sprache auf die Ebene der Notwendigkeit der Betonung von Sprache oder der Selbstverständlichkeit von Kommunikation überhaupt verlagert: ¹⁾ Eltern beider

1)-----

Diese Verlagerung wird dem Sinn der Ausgangsfrage gerecht, um wieviel schwerer es ohne Sprache wäre "die folgenden Dinge mit den Kindern zu tun", will sagen: "über folgende Probleme mit den Kindern zu kommunizieren". Wir sehen die Problematisierung der Sprache als eine Folge der schichtspezifischen Selbstverständlichkeit der Kommunikation an: auch ohne Sprache können die Eltern der Unterschicht über Funktionsprinzipien, die Eltern der Mittelschicht über affektiv-diffuse Probleme relativ leicht verständigen. Ob eine solche Leichtigkeit der Kommunikation sich der außerkommunikativen Offensichtlichkeit der Objektwelt verdankt - wie für die Funktionsprinzipien wahrscheinlich -, oder Produkt redundant gewordener sprachlicher Leistungen ist - wie für die affektiv diffusen Probleme wahrscheinlich - kann hier außer Betracht bleiben. In jedem Fall nehmen wir an, daß Normen des Sprachgebrauchs sich aus schichtspezifischen Notwendigkeiten der Kommunikation ergeben. Ein solches Verfahren stimmt überdies mit dem bekannter ex post Erklärungen von Normdifferenzen zwischen den Schichten aus - angenommen - objektiven Zwängen überein. "Sauberkeit" als Ziel der Erziehung muß in der Mittelschicht z.B. weniger betont werden, weil sie von den Wohnungsverhältnissen her selbstverständlich sein kann (Kohn 1969).

Schichten halten Sprache für jeweils schichttypische Bereiche der Kommunikation erst dann für wichtig, wenn ihre Kinder schwächere verbale Leistungen zeigen. Auf dieser Ebene scheint die Übermittlung von Funktionsprinzipien nicht, wie Bernstein annehmen würde, eine Domäne der Mittel-, sondern der Unterschicht. Daß in der Mittelschicht affektiv diffuse, in der Unterschicht Kommunikation über die Funktionsprinzipien von Dingen größere Selbstverständlichkeit besitzt, kann mit der "Substanz der Arbeit" zusammengebracht werden, die in der Unterschicht typischerweise "Umgang mit Sachen", in der Mittelschicht "Umgang mit Personen oder Daten und Symbolen" ist (Kohn 1969).

Zusammenfassend wollen wir versuchen, die gegebenen Erklärungen schichtspezifischer Einflüsse der Kommunikationsbereitschaft und Sprachbewertung der Eltern auf das passive Wortverständnis zu systematisieren. Wir sind von dem allgemeinen Eindruck ausgegangen, daß das schichtuntypische Verhalten der Eltern die Leistungen der Kinder fördert, und haben untersucht, ob er sich in den einzelnen Bereich des Erziehungsverhaltens wiederfindet. Dabei haben wir vier Typen gefunden. (1) In zwei Bereichen ergaben sich keine Schichtunterschiede des Elternverhaltens und ein vergleichbarer Zusammenhang mit der Differenzierungsfähigkeit des Kindes in beiden Schichten: C 6, "Bus oder Straßenbahn"; S 13, "Gut und Böse". Ohne daß diese Einstellungen der Eltern sich inhaltlich zusammengruppierten, müssen wir sie als wirksame schichtunabhängige Einflüsse verbuchen. (2) In zwei weiteren Bereichen: C 1, "Hausarbeiten" und C 2, "Straße" hängt zwar die Interaktionsbereitschaft der Mutter in jeder Schicht positiv mit der Differenzierungsfähigkeit des Kindes zusammen, tritt aber selber stärker in der Mittelschicht auf. Obwohl sie sich als "offene" Bereiche zusammenfassen lassen, können wir nicht entscheiden, ob es sich um soziale Einflüsse innerhalb oder außerhalb von "Schicht" handelt. Fassen wir die beiden ersten Typen zusammen, so haben wir immerhin drei "offene" Bereiche, in denen die Interaktionsbereitschaft der Mütter in beiden Schichten, besonders aber in der Unterschicht, in der sie seltener vorkommt, mit gutem kategorialen Differenzierungsfähigkeit des Kindes zusammenhängt. Zwei Ergebnisse verdienen also festgehalten zu werden:

Erstens führt die Bereitschaft der Mutter, mit dem Kinde bei Beschäftigungen, die durch eine Unterhaltung nicht gestört werden, zu sprechen, in jeder Schicht zu guter Differenzierung des Kindes. Zweitens gilt dies besonders in der Unterschicht, wo die positiven Folgen der Gesprächsbereitschaft teilweise ihre geringere Häufigkeit kompensieren.

Während in den beiden ersten Typen das Erziehungsverhalten der Eltern in beiden Schichten gleichartig wirkt, zeigt es in den beiden folgenden Typen gegensätzliche Wirkungen in beiden Schichten. (3) In fünf Bereichen - weitaus dem größten Typus - fanden wir den allgemeinen Sachverhalt wieder, daß das schichtuntypische Verhalten der Eltern mit guten Leistungen des Kindes zusammenhängt: C 5 "Einkauf" und C 7, "Mahlzeiten"; sowie den affektiv diffusen Bereichen der Sprachfrage: S 3, S 9, S 11. (4) In zwei Bereichen hing das schichttypische Verhalten der Eltern mit guten Leistungen des Kindes zusammen: C 3 "Entspannung" - hier hatten wir uns allerdings über die Schichttypik des Verhaltens getäuscht und nicht erwartet, daß die Mütter der Mittelschicht bei der Entspannung Schwätzversuche des Kindes zurückweisen würden -; sowie S 10, "Funktionsprinzipien von Dingen" - hier ist eine geringe Betonung der Sprache in der Unterschicht wahrscheinlicher. Wir wollen einige Erklärungen diskutieren, warum - als eine am größten Typus (3) belegbare Regel - die schichtuntypische Einstellung der Eltern die Leistungen der Kinder fördert, und warum diese Regel in zwei Bereichen des Typus (4) durchbrochen wird.

In der Unterschicht hängt Unterhaltungsbereitschaft und Betonung der Sprache, in der Mittelschicht Abweisung von Unterhaltung und geringe Betonung der Sprache durch die Eltern mit gutem Differenzierungsvermögen des Kindes zusammen. Zwei Erklärungen scheinen uns möglich. Da wir - auch aus dieser Untersuchung - wissen, daß die Mütter der Mittelschicht allgemein - mit der Ausnahme der "Entspannung" (C 3) - eher als die Mütter der Unterschicht zu verbalen Interaktion bereit sind, vermuten wir, daß in der Mittelschichtgruppe der Kausalzusammenhang sich umgekehrt hat. Nicht die Gesprächsbereitschaft der Mütter führt zu schlechtem Wort-

verständnis der Kinder, sondern die geringere Gesprächsbereitschaft ist eine Reaktion auf die verbale Aufdringlichkeit gerade der verbal gewandteren Kinder. Allgemein ist wahrscheinlich, die Eltern der Mittelschicht Fähigkeiten ihrer Kinder erkennen und ihr Verhalten wie ihre Berichte über ihr Verhalten und ihre Einstellungen nach dieser Erkenntnis richten, die Eltern der Unterschicht jedoch nicht in dieser Weise reagieren, so daß die hypothetischen "natürlichen" Zusammenhänge bestehen bleiben. Auf die gleiche Weise kann die geringe Bedeutung, die die Eltern der Mittelschicht der Sprache für die Interaktion mit den verbal besseren Kindern beimessen, eine Reaktion auf die korrekte Wahrnehmung der Fähigkeiten der Kinder sein. Kurz: Eine erste Erklärung kann sein, daß die Eltern der Mittelschicht - im Gegensatz zu den Eltern der Unterschicht - ihre sozialisatorische Arbeit bewußt kontrollieren und deshalb auf schichttypische und für förderlich gehaltene Einstellungen und Verhaltensweisen verzichten, sobald sie offenbar Erfolg gezeigt haben. Die zweite Erklärung geht insofern weiter, als das Verhalten der Eltern beider Schichten als Reaktion auf die Wahrnehmung des Kindes gesehen wird; sie schließt aus dem schichtspezifischen Charakter des Elternverhaltens auf Normen, die die Notwendigkeit des Eingreifens der Eltern und die Problematisierung des Gebrauchs von Sprache für beide Schichten unterschiedlich definieren. Eltern der Mittelschicht erwarten im allgemeinen einen sicheren kommunikativen Umgang mit ihrem Kind; erst wenn das Kind verbal größere Schwierigkeiten hat als erwartet, erklären die Eltern, daß sie sich besonders oft mit ihm unterhalten, und erkennen die Wichtigkeit der Sprache für die Interaktion mit dem Kind. Eltern der Unterschicht aber unterhalten sich im allgemeinen weniger mit ihrem Kind; erst wenn das Kind verbal gewandter ist als erwartet, sehen die Eltern Veranlassung, sich mit ihm zu unterhalten und erkennen Schwierigkeiten in der verbalen Interaktion mit dem Kind. Die Gründe können verschieden sein - vielleicht fühlen die Eltern eine Überlegenheit des Kindes, vielleicht können sie es weniger leicht zurückweisen, vielleicht wollen sie es fördern, nachdem seine Fähigkeiten aufgefallen sind -, führen aber zu den gleichen Reaktionen.

Warum nun hängt in zwei Situationen das schichttypische Verhalten der Eltern mit guten Leistungen des Kindes zusammen? Zweifellos muß die Erklärung in der mehr oder minder großen schichtspezifischen Vertrautheit der verschiedenen Situationen gesucht werden. Wir vermuten (im Sinne der ersten Erklärung), daß die Eltern der Unterschicht nur solange nicht in bewußter Reaktion auf die Einschätzung des Kindes erziehen, als sie selber für die Beurteilung von Situationen keine bewußten Maßstäbe besitzen. Sobald ihnen die Situation vertraut ist, sehen sie (im Sinne der zweiten Erklärung) nicht mehr die verbal gewandteren Kinder als Herausforderung, sondern glauben wie die Eltern der Mittelschicht, sich den verbal weniger gewandten bewußt widmen zu müssen. Wir glauben, daß die Übermittlung von Funktionsprinzipien (S 10) den Unterschichteltern von den Aufgaben der Arbeitswelt besonders vertraut ist; deshalb halten sie hier Sprache erst dann für wichtig, wenn ihre Kinder schlechte sprachliche Leistungen zeigen. Wir wissen, daß während der Entspannung (C 3) Unterschichtmütter sich mit ihrem Kind unterhalten, während Mittelschichtmütter von ihm Ruhe verlangen, und das jeweils schichttypische Verhalten mit gutem kategorialen Differenzierungsvermögen zusammenhängt. Die Voraussetzung unserer allgemeinen Erklärung ist hier also nicht mehr gegeben. Dennoch scheinen uns die beschriebenen schichtspezifischen Normen der Problematisierung von Sprache hier durch das schichtspezifische Verständnis der Situation verstärkt zu werden. Wenn die Mütter der Mittelschicht Entspannung als ausschließlich individuelle Angelegenheit ansehen, werden sie sie erst recht nur für verbal schwächere Kinder opfern. Wenn die Mütter der Unterschicht Entspannung als Gelegenheit zur Unterhaltung verstehen, werden sie sie erst recht für verbal außergewöhnlich gewandte Kinder nutzen.

Fünftes Kapitel: Diskussion und Bewertung der Ergebnisse.

=====

Das in den vorangegangenen Kapiteln dargestellte Experiment bezog sich auf die aktive und passive Verfügung über den Wortschatz der Sprache. Über diese beiden Aspekte des Sprachgebrauchs wurden zwei Voraussagen getroffen: Kinder der Mittelschicht erklären aktiv Wortbedeutungen besser als Kinder der Unterschicht, obwohl sie über keine besseren passiven Wortkenntnisse verfügen als Kinder der Unterschicht, allein aufgrund der für die Mittelschicht typischen Strategie eines expliziten Sprachgebrauchs. Und: Kinder beider Schichten, die über gleich gute passive Wortkenntnisse verfügen, gewinnen aus der gleichen sprachlichen Eingabe die gleiche kognitive Kategorisierung; Inhalt und Form der Kategorisierung hängen bei Kindern beider Schichten in der gleichen Weise von Inhalt und Form der sprachlichen Eingabe und von der passiven Wortkenntnis (bzw. der gemessenen Intelligenz) der Kinder ab. Beide Hypothesen beziehen sich auf Schritte eines sozialen Kommunikationsprozesses, in dem schichtspezifische Strategien des Sprachgebrauchs, auf die Dauer gesehen, die kognitive Entwicklung der Unterschichtkinder nachteilig beeinflussen, ohne daß ihre tatsächlichen Fähigkeiten und Kenntnisse schlechter wären.

Auf der einen Seite - der Seite des Senders - explizieren Kinder der Unterschicht ihre Kenntnis sprachlich weniger genau als die Mittelschichtkinder, obwohl sie über ein gleich gutes Verständnis der Wortbedeutung und der Sachzusammenhänge verfügen. Unabhängig jedoch vom Verständnis des Kindes evoziert beim Zuhörer die objektiv vorliegende sprachliche Form die Kategorisierungen, aufgrund derer er auf das Verständnis des Sprechers schließt, und die Antwort auf den Sprecher kalkuliert. Er wird dem Unterschichtkind, aufgrund des objektiven Gehalts seiner sprachlichen Äußerungen, weniger zutrauen und es durch seine Reaktionen weniger herausfordern. Langfristig können in dieser Weise tatsächlich die soziale Anregung und die intellektuellen Lernprozesse des Unterschichtkindes eingeschränkt werden. Auf der anderen Seite - der Seite des Empfängers - evoziert das

gleiche sprachliche Material bei Kindern beider Schichten die gleichen Kategorisierungen.

Beide Hypothesen zusammen besagen, daß die Kinder der Unterschicht in durch soziale Kommunikation vermittelten kognitiven Lernprozessen benachteiligt sind, ohne daß sie über ein schlechteres Verständnis verfügen und ohne daß sie in aktuellen Lernprozessen die Informationen schlechter verstehen. Die Verbindung beider Hypothesen wird durch die Annahme hergestellt, daß sprachliche Formen kognitive Kategorisierungen hervorrufen. Sie gibt den Rahmen, in dem die beiden Hypothesen der Theorie der linguistischen Kodes sich bewegen: daß Kinder der Unterschicht aufgrund der kode-abhängigen objektiven Form ihrer Verbalisierung beim Zuhörer schlechtere Kategorisierungen evozieren als die Mittelschichtkinder, dennoch aber für sich aus den gleichen sprachlichen Formen keine schlechteren Kategorisierungen evozieren als die Mittelschichtkinder. Das Evokationsmodell ist neben der Theorie der linguistischen Kodes das zweite theoretische Element, das unserem Experiment zugrundeliegt.

Keine der beiden Hypothesen konnte bestätigt werden. Die aktiven Worterklärungen der Kinder bei kontrolliertem passivem Wortschatz zeigte in keinem der vielen untersuchten Aspekte irgendeinen bemerkenswerten Unterschied zwischen den beiden Schichtgruppen. Das kategoriale Verständnis von Bedeutungserklärungen unterschied sich nicht zwischen den beiden Schichtgruppen, solange es in reiner Form d.h. nur mit minimalen kategorialen Differenzen, aber doch signifikant, sobald es ausgehend von verschiedenen sprachlichen Formen der Bedeutungserklärung mit semantisch und sprachlich stärker variierenden Vorgaben getestet wurde. Pointiert läßt sich daraus die genaue Umkehrung der ursprünglichen beiden Hypothesen entnehmen: die Kinder beider Schichten unterscheiden sich nicht in der aktiven Erklärung, aber sehr wohl im passiven Verständnis; weniger zugespitzt müssen wir sagen, daß weder im aktiven Wortgebrauch noch im passiven Wortverständnis Schichtunterschiede auftreten.

Diese Ergebnisse widersprechen nicht nur den Hypothesen, sondern auch den Ergebnissen anderer, gleich angelegter Untersuchungen (Carson/Rabin 1960). Wir wollen in diesem Schlußkapitel zunächst einige mögliche Ursachen für die negativen Ergebnisse innerhalb der Experimentalanlage und dann ihre Konsequenzen für die beiden theoretischen Elemente des Experiments: die Evokationsannahme und die Theorie der linguistischen Kodes diskutieren; abschließend werden wir die Bedeutung der Ergebnisse für die sozialpolitische und pädagogische Praxis erörtern.

1. Mögliche Ursachen für die den Hypothesen und den Ergebnissen ähnlicher Untersuchungen widersprechenden Ergebnisse in der Anlage des Wortschatzexperimentes.

1.1. Wahl der Schicht- und Altersgruppen.

Eine Erklärung für die deutschen wie angelsächsischen Untersuchungsergebnisse widersprechenden Ergebnisse kann die unterschiedliche soziale Distanz zwischen den verglichenen Gruppen und - beim Vergleich zwischen Untersuchungen aus verschiedenen Nationen - die Unterschiede zwischen den Sozialstrukturen und Kulturen in der Akzentuierung der Schichtgrenzen und in den Kriterien der Schichtung gewesen sein. Hinzu kommt, daß der Einfluß schichtspezifischer Sprachen oder Arten des Sprachgebrauchs sich in den höheren Altersgruppen verstärkt oder vermindert haben kann. Wir wählen den Vergleich mit anderen Untersuchungen gleichsam als Methode, um diese Ursachen für das Versagen unserer Hypothesen bestätigen oder ausschließen zu können.

Carson/Rabin (1960), deren Experiment in der zweiten Messung dieses Experiments repliziert wurde, verglichen drei Gruppen, die durch Rassen- und geographische Schranken voneinander getrennt waren: "Southern Negro" (SN), "Northern Negro" (NN), "Northern White" (NW). Obwohl alle drei Gruppen den gleichen Mittelwert auf einem Bild-Vokabel-Test hatten, war in der aktiven Erklärung die Gruppe NW

der Gruppe NN und SN und die Gruppe NN der Gruppe SN signifikant überlegen. Carson/Rabin betonen, daß der Berufsstatus der Väter der beiden "nördlichen" Gruppen vergleichbar war, so daß der normalerweise mit dem Rassenunterschied gegebene Schichtunterschied für die unterschiedlichen Leistungen der beiden "nördlichen Gruppen" nicht wirksam gewesen sein kann. Die Gruppenunterschiede im Experiment von Carson/Rabin müssen als soziale Unterschiede in den Dimensionen der Rasse und der geographischen Region zurückgeführt werden. Die Diskrepanz zu unserer Untersuchung kann sich also aus dem Wechsel der Dimensionen sozialer Ungleichheit zwischen beiden Untersuchungsgruppen ergeben haben.

Vor allem beim Vergleich mit amerikanischen Studien muß bedacht werden, daß unter "Schicht" die Faktoren des sozio-ökonomischen Status und der Rassenzugehörigkeit zusammenfallen können. Whiteman/Brown/Deutsch (1967) zeigten an einer Gruppe von Kinder des 1. und 5. Schuljahres aus New York, daß Rassenzugehörigkeit, Status und sozio-ökologische Deprivation unabhängig voneinander sprachliche und nichtsprachliche Intelligenzleistungen negativ beeinflussen. Aber auch beim Vergleich mit den Studien Bernsteins in England muß beachtet werden, daß die meisten der dort durchgeführten Schichtvergleiche zugleich auch Vergleiche zwischen Wohngebieten sind, deren typische Bevölkerung aus der Unter- bzw. Mittelschicht stammt (Bernstein 1962a, 1970).

Auch in unserer Untersuchung ist, jedoch in geringerem Maße, Schichtzugehörigkeit mit schichttypischem Wohngebiet (Oberursel, Frankfurt-Goldstein) verbunden; dies gilt zudem nur für die dritte, nicht aber die zweite Messung (siehe III.2 und IV.2, jeweils die Tabellen für die Verteilung der Vpn. auf Schulen: Tab. 31a und Tab. 170). Insgesamt jedoch ist die durch den Wechsel der Schichtdimensionen zwischen Kulturen bedingte Veränderung des Schichteinflusses schwer abzuschätzen.

Wichtiger sind die zwischen den Kulturen wechselnde Kontinuität des Schichtgefälles und die zwischen den Untersuchungen wechselnde Strategien von kontinuierlichen oder extremen Schichtvergleichen. Williams/Naremore (1969:80) vergleichen in ihrer Untersuchung der Sprachfunktionen 10 bis 12jähriger Kinder das obere mit dem unteren Ende der Statusskala. Zusammen mit der Tatsache, daß in den USA (Detroit) Straten unterhalb der Arbeiterschaft relativ größer sind als z.B. in der BRD, kann eine experimentelle Vergrößerung der Distanz zwischen den Schichtgruppen, für einen Teil der festgestellten Schichtunterschiede verantwortlich sein. Auch in Bernsteins alten (1962a) und neuen (1970:9) Untersuchungen wurde die Strategie verfolgt, die Distanz zwischen beiden Schichtgruppen zu maximieren. In unserer Untersuchung jedoch wurde nur die untere Mittelschicht, der weniger als ein Zehntel der in der Gesamtstichprobe erhobenen Vpn. zugerechnet wurden, nicht für die Wahl der parallelisierten Schichtgruppen berücksichtigt. Vergleicht man die Ergebnisse der verbleibenden 4 Schichtgruppen, so zeigen sich in der zweiten Messung zwischen den beiden mittleren Gruppen überhaupt keine, zwischen den beiden extremen Gruppen jedoch beachtliche Differenzen (siehe III.5.6.1.); in der dritten Messung, bei der allerdings das Fehlen von Schichtunterschieden auch vorausgesagt wurde, ist der Unterschied zwischen den 4 Schichtgruppen nicht nennenswert (siehe IV.3.2.). Von den Ergebnissen der zweiten Messung her kann vermutet werden, daß bei einer Extremisierung des Schichtvergleichs die Hypothesen bestätigt worden wären. Gegenüber den frühen Untersuchungen Bernsteins (1962a) sind die Testbedingungen für Schichtunterschiede in unserem Experiment also verschärft worden. Jedoch ergaben sich auch in späteren Untersuchungen Bernsteins, bei denen Kinder aus dem gesamten Kontinuum sozialer Schichtung als Unter- und Mittelschicht miteinander verglichen wurden, die Theorie der linguistischen Kodes bestätigende Ergebnisse (siehe Henderson 1970:20). Zudem scheint uns die größere Kontinuität des Schichtvergleichs allein keine hinreichende Erklärung für das Versagen der Hypothesen der Bernsteinschen Theorie in einer deutschen Untersuchung.

Eine weitere Ursache für das Versagen unserer Hypothesen könnte sein, daß ursprüngliche Unterschiede des schichtspezifischen Sprachgebrauchs durch vier Jahre Grundschule bereits kompensiert sein könnten. Eine Reihe von neueren deutschen Untersuchungen beschäftigen sich mit Schichtunterschieden der Sprache von Schulanfängern (Bühler 1972, Sommer 1972, Wiederhold 1971). Wenn auch gegen die Untersuchungen Bühlers und Wiederholds verschiedene methodische und theoretische Einwände vorgebracht werden müssen (Gloy 1973), so kann die Übereinstimmung aller drei Untersuchungen, daß Schichtunterschiede in der Sprache von Schulanfängern nicht oder nur in geringem Maße gegeben sind, kein Zufall sein. Sommer (1972) verglich den Wortschatz, die grammatischen Fähigkeiten, die Artikulation und die Wortflüssigkeit von Kindern der Unter- und der Mittelschicht. Solange die Intelligenz nicht beim Schichtvergleich kontrolliert war, ergaben sich verschiedene, aber keineswegs durchgängige Schichtunterschiede. Sobald die verbale, nicht-verbale oder gesamte Intelligenz jedoch kontrolliert war, traten insgesamt keine Schichtunterschiede mehr auf. Obwohl Sommers Untersuchung sich auf manifeste Sprachmerkmale und nicht auf Arten des Sprachgebrauchs bezieht, kann sie durch den referierten doppelten Schichtvergleich in der Theorie der linguistischen Kodes interpretiert werden: sie sagt auch dann noch Schichtunterschiede voraus, wenn die gemessene Intelligenz kontrolliert ist.

Ein Mangel der Untersuchung Sommers (1972:85) schränkt die letzte Folgerung jedoch wieder ein. In den entscheidenden Wilcoxon-Tests mit Kontrolle der verschiedenen Intelligenzwerte verglich Sommer als Unter- und Mittelschicht zwei Statusgruppen, die einander entschieden näher stehen als die von uns verglichenen Gruppen. Zur Unterschicht zählt sie un- und angelernte Arbeiter, zur Mittelschicht gelernte Arbeiter, Büroangestellte und Akademiker. M.a.W.: die Trennung zwischen ihren beiden Schichtgruppen liegt dort, wo wir die Trennung zwischen UUS und OUSgezogen haben, die Mittelschicht enthält in sich sehr inhomogene Gruppen und vor allem die Grenze zwischen manueller und nicht-manueller Arbeit. In der ersten Messung unserer Untersuchung ergaben sich die stärksten Schicht-

unterschiede zwischen der OUS und UMS, an genau jener Grenze, die Sommer in die Mittelschicht verlegt (siehe III.4.1.). In der zweiten Messung unserer Untersuchung ergaben sich jedoch die stärksten Differenzen bei einer Einteilung in 4 Schichten, in der die UUS allerdings sehr schwach besetzt war, zwischen der UUS und OUS, an jener Grenze also, die Sommer mit der Grenze zwischen Unter- und Mittelschicht identifiziert (siehe III.5.6.1.). Wie dem auch sei: In jedem Fall besteht die Möglichkeit, daß das Fehlen von Schichtunterschieden in Sommers Untersuchung sich daraus ergibt, daß sie Kinder zu eng verwandter Straten vergleicht, und besonders Kinder von Vätern mit manuellen Berufen sowohl zur Unter- wie zur Mittelschicht rechnet. Dann aber könnte das Fehlen signifikanter Schichtunterschiede ein zufälliges Ergebnis sein, wofür auch das recht verwirrende Muster der Ergebnisse mit verschiedenen Kontrollen hinweist.¹⁾

Die Gefahr, daß - wie bei Sommer - wegen der zu engen Verwandtschaft der verglichenen Schichtgruppen keine Schichtunterschiede mehr auftreten, ist das Gegenstück zu der Gefahr, daß sie - wie in den zuerst referierten Untersuchungen - durch eine Maximierung der Distanz in jedem Fall gesichert sind. Beide Mängel gelten nicht mehr für die Untersuchung Wiederholds (1971), der seinen Vpn. mit Hilfe des Schichtmodells von Janowitz nach dem Beruf der Väter und dem Familieneinkommen in 4 Schichten einteilte und zudem den kompensierenden Effekt der Schule - zwischen dem Schulbeginn und dem Beginn des 2. Schuljahres - direkt überprüfte. Wiederhold ließ Schulanfänger ein Bild beschreiben, ein Märchen nacherzählen und forderte von ihnen einen Erlebnisbericht. Er untersuchte sprachliche Merkmale und eine Reihe von Lösungsstrategien der Aufgaben, die sich direkt im Rahmen linguistischer Kodes interpretieren lassen und bei denen eine Reihe von Schichtdifferenzen auftauchen, die als Bestätigung linguistischer Kodes angesehen werden können. Für den

¹⁾-----
In vergleichbaren Abschnitten unserer Untersuchung haben wir ebenfalls die gleiche Gruppe bei Kontrolle verschiedener Variablen auf Schichtunterschiede hin überprüft (siehe III.5.2.2. und IV.3.2.). Es ergab sich, anders als bei Sommer, beim Wechsel der Kontrolle, kein Neuauftreten von Schichtunterschieden oder gar ein Wechsel der Richtung der Schichtunterschiede.

Wortschatz ergeben sich zu Beginn des 1. Schuljahres geringfügige Unterschiede zwischen den Schichten in den verschiedenen Wortklassen, die aber nach einem Jahr weitgehend nivelliert sind. Für die gesamte Untersuchung resümiert der Autor, daß zu Beginn der Schule nur geringe Unterschiede zwischen den Kindern der 4 Schichten bestehen, die sich nach einem Jahr fast vollständig ausgeglichen haben, daß die meisten Unterschiede zwischen der OMS und den 3 übrigen Schichten auftauchen und daß die allgemein fördernden Effekte der Schule die kompensierenden überwiegen (Wiederhold 1971:202). Durch das Fehlen statistischer Prüfungen sind die Schichtunterschiede bestätigenden und widerlegenden Ergebnisse schwer abzuschätzen und die Effekte der Schule und der Sozialschicht schwer zu vergleichen. Da Wiederhold aber beim Schichtvergleich keinerlei Intelligenzmaße kontrolliert, d.h. da er Gruppen vergleicht, bei denen die unteren Straten nach den meisten bekannten Ergebnissen und auch den Ergebnissen unserer ersten Messung (siehe III.4.1.) wahrscheinlich in einer schlechteren Ausgangslage waren, ist die Tatsache, daß bei vielen Variablen keine Schichtunterschiede gefunden werden konnten, ein Argument gegen die Hypothese schichtspezifischer Arten des Sprachgebrauchs bei Schulanfängern.

Bühler (1972) kommt - allerdings mit Verfahren, die für den Leser des Untersuchungsberichts kaum nachvollziehbar sind - bei einem Vergleich der Sprache von 260 Erstkläßlern zu dem Ergebnis, daß die verschiedenen Sprechsituationen, nicht aber die soziale Schichtzugehörigkeit der Sprecher die grammatischen Formen und den Informationsgehalt der Sprache bestimmen. Auch er kontrolliert keinerlei Intelligenzwerte beim Schichtvergleich. Ein Mangel der Experimentalanlage, der die Bestätigung von Schichtunterschieden in der Sprache sofort als bekanntes und theoretisch wenig aussagekräftiges Ergebnis trivialisieren würde, legt im Falle der Nichtbestätigung um so gravierendere Schlüsse für die Theorie der linguistischen Codes nahe.

Allein Schlee (1973), der genau wie unsere Untersuchung mit 5 Schichten (nach Moore/Kleining) arbeitete und aus den beiden oberen die "Mittelschicht", den beiden unteren die "Unterschicht" für den

dichotomen Schichtvergleich auswählte, stellte bei Vorschulkindern und Kindern des 4. Schuljahres, deren Intelligenz kontrolliert war, eine signifikante Überlegenheit der Mittelschichtkinder im passiven Sprachverständnis fest. Diese Untersuchung gibt also, obwohl die Testbedingungen für Schichtunterschiede mit unserer Untersuchung vergleichbar scharf sind, Argumente für die Bedeutung schichtspezifischer Strategien des Sprachgebrauchs, nicht jedoch für ihren altersspezifisch variierenden Einfluß auf das Sprachverständnis.

Insgesamt also weist die im Vergleich zu anderen Untersuchungen geringe soziale Distanz zwischen den Schichtgruppen zusammen mit dem Ergebnis minimaler Differenzen zwischen den beiden "mittleren", maximaler zwischen den beiden "äußeren" Schichtgruppen, auf die Verschärfung des Schichtvergleichs als einer Ursache für die Diskrepanz unserer zu anderen Ergebnissen. Eine solche Erklärung ist jedoch weder theoretisch hinreichend, da zwischen den beiden "mittleren" Schichtgruppen die für alle sozio-kulturellen Untersuchungen wichtige Grenze zwischen manueller und nicht-manueller Arbeit liegt, noch empirisch hinreichend, da vergleichbare Untersuchungen trotz ähnlich scharfer Testbedingungen (Bernstein 1973, Brandis/Henderson 1970, Schlee 1973) der Sozialschicht die Theorie der linguistischen Codes bestätigen können. Wenig sinnvoll scheint es hingegen zu sein, das Versagen unserer Hypothesen auf die kompensierende Wirkung der 4 Jahre Grundschule, die unsere Vpn. bereits durchlaufen haben, zurückzuführen. Theoretisch hat auch die entgegengesetzte Annahme Plausibilität, daß der Bruch zwischen Herkunfts- und Schulumilieu die Lage des Kindes aus den unteren Straten zunehmend erschwert; sie ist für die USA durch das von Whiteman/Brown/Deutsch (1967) gemessene "kumulative Defizit" verbaler Intelligenzleistungen bei Kindern der Unterschicht sogar bestätigt worden. Empirisch aber zeigen in der BRD eine Reihe von Untersuchungen geringe oder keine Unterschiede in der Sprache von Schulanfängern beider Schichten, obwohl sie die Intelligenz als einen schichtabhängigen Einflußfaktor auf Sprachleistungen nicht kontrollieren (Bühler 1972, Sommer 1972, Wiederhold 1971). Untersuchungen hingegen, die die Intelligenz kontrollieren, zeigen ent-

weder keine Schichtunterschiede in der Sprache von Schulanfängern (Sommer 1972), oder die gleichen wie bei Viertkläblern (Schlee 1973).

1.2. Eliminierung möglicher Schichtunterschiede durch die Kontrolle des passiven Wortschatzes.

Die Theorie der linguistischen Kodes sagt Unterschiede zwischen der Unter- und Mittelschicht unabhängig von der gemessenen Intelligenz voraus. Zudem können nur bei Kontrolle der Intelligenz Unterschiede zwischen den Schichten in sprachlichen Leistungen auf schichtspezifische Arten des Gebrauchs von Sprache zurückgeführt und eine Erklärung durch die schichtspezifische "natürliche" Verteilung der für die verschiedenen manifesten Sprachformen wichtigen Intelligenz ausgeschlossen werden. Da die Intelligenz von der Schichtzugehörigkeit (in unserer Gesamtstichprobe (N=357): $r = .28$) und die Sprachleistungen von der Intelligenz (in unserer Gesamtstichprobe: $r = .38$, GESINT mit WST, Tab. 22) abhängen, sind mit sozialer Schicht ansteigende sprachliche Leistungen an sich kein erstaunliches Ergebnis.¹⁾ Theoretische Überlegungen der Theorie der linguistischen Kodes wie methodische Überlegungen zur Aussagekraft der Experimentalanlage sprechen also für die Kontrolle selber schichtabhängiger Variabler für den Schichtvergleich. Wir haben dies an verschiedenen Stellen unseres Berichts betont (III.3.) und in unserer Untersuchung auf vier verschiedene Weisen versucht, den intelligenzunabhängigen Einfluß schichtspezifischer Arten des Sprachgebrauchs herauszustreichen.

Als erstes haben wir die Leistung im Frankfurter WST (erste Messung) nicht nur in ihrer Abhängigkeit von Schicht, sondern auch in ihrer Abhängigkeit von Schicht bei statistischer Kontrolle der verschiedenen verbalen und nicht-verbalen Intelligenzleistungen des PSB untersucht

1)-----

Um so verwirrender sind Untersuchungen, die trotz fehlender Kontrolle der Intelligenz bei Schulanfängern keine Unterschiede in sprachlichen Leistungen feststellen konnten (Wiederhold 1971, Bühler 1972); siehe dazu weiter oben.

(siehe III.4.1., Tab. 41). Beide Male ergab sich ein signifikanter Effekt der Sozialschicht (5 Straten) auf die Leistung im WST. Als zweites haben wir die aktive Worterklärung (zweite Messung) zwischen den Schichten bei experimenteller Kontrolle des passiven Wortschatzes untersucht. Hier ergaben sich keine Schichtunterschiede (siehe III.5.2.). Als drittes haben wir die Explikations- und die Verständnisleistung in der aktiven Worterklärung versucht gesondert zu erfassen, mit der Hypothese, daß trotz gleich guten Verständnisses die Mittelschichtkinder besser explizieren würden; auch diese Trennung sollte vom Sprachgebrauch abhängige Unterschiede zwischen den Schichten bei gleicher Kenntnis herausarbeiten. Auch in der Differenz zwischen Verständnis und Explikation ergaben sich jedoch keine Unterschiede zwischen den Schichten (siehe III.5.4.4.). Als viertes haben wir untersucht, ob die Wahl bestimmter Formen der Erklärung trotz gleich guter Gesamtleistung zwischen den beiden Schichtgruppen differenzieren; auch dies würde auf den Einfluß von Strategien des Gebrauchs von Sprache hinweisen, die unabhängig sind von den durchschnittlichen Erklärungsleistungen. Auch hier ergab sich jedoch nicht nur kein Unterschied zwischen den beiden Schichtgruppen, sondern eine geradezu frappante Ähnlichkeit der beiden schichtspezifischen Profile (siehe III.5.4.5.). Wir können also zusammenfassen, daß die vier Mal wiederholte Strategie der Kontrolle nicht auf schichtspezifische Strategien des Sprachgebrauchs zurückführbarer Einflüsse der Schichtzugehörigkeit auf Leistungen, die theoretisch von schichtspezifischen Strategien des Sprachgebrauchs abhängen sollen, dann zum Erfolg geführt hat, wenn die Kontrolle sich auf Variablen der verbalen und nicht-verbalen Intelligenz bezog, nämlich in der ersten Messung, immer dann aber nicht zum Erfolg geführt hat, wenn die Kontrolle sich auf i.e.S. sprachliche Leistungen bezog, nämlich auf die passive Wortkenntnis, das in der Erklärung offenbarte Wortverständnis und die aktive Erklärungsleistung insgesamt, d.h. in allen drei in der zweiten Messung angewandten Kontrollen.

In dieser Lage haben wir zwei Möglichkeiten, das Versagen unserer Hypothesen zu begründen. Als erstes können wir auf die nach wie vor nicht widerlegten naturalistischen Unterschiede der Sprache zwischen den Schichtgruppen hinweisen, d.h. auf jene Unterschiede, die sich bei der Erhebung nicht schon unter Gesichtspunkten des Experimentalplans ausgesuchter Gruppen unmittelbar oder mit statistischen Kontrollverfahren ergeben. Die Ergebnisse der ersten Messung gewinnen so gegenüber der zweiten Messung sehr stark an Bedeutung. Insbesondere die Schichtunterschiede im WST bei Kontrolle der Subtests des PSB können direkt als eine Bestätigung der Theorie der linguistischen Kodes gesehen werden, jedoch auf einem Gebiet, auf dem nach der Gesamtstrategie des Experiments keine Schichtunterschiede auftauchen sollten, dem passiven Wortverständnis¹⁾. Eine solche Interpretation der ersten Messung gewinnt durch unsere Ergebnisse der dritten Messung und die Ergebnisse einer anderen Untersuchung zusätzliches Gewicht. Wir konnten, im "differenzierten Verständnistest" Unterschiede im Verständnis des Stichworts zwischen den Schichtgruppen trotz kontrolliertem passivem Wortschatz feststellen; Schlee (1973) konnte ebenfalls im passiven Sprachverständnis eine Überlegenheit der Mittelschichtkinder feststellen, trotz kontrollierter nicht-verbaler Intelligenz.

Mit dieser Reaktion auf die negativen Ergebnisse unseres Experiments sind jedoch zwei Risiken verbunden. Sofern wir uns auf naturalistische Unterschiede der Sprache ohne auch nur die statistische Kontrolle anderer schichtabhängiger Einflüsse zurückziehen, verlieren die Er-

¹⁾-----
Wir können so weit gehen, die ursprüngliche Hypothese 2 des Experiments (Schichtunterschiede in der ersten Messung, auch bei Kontrolle der gemessenen Intelligenz) als inkonsistent mit der Hypothesen-Gesamtstrategie des Experiments anzusehen, die ja Unterschiede in der aktiven Erklärung und Gleichheit der Schichten im passiven Verständnis nachweisen sollte, jeweils bei Kontrolle von Wortschatz bzw. Intelligenzleistungen.

gebnisse an theoretischer Aussagekraft und damit zugleich an praktisch pädagogischer Bedeutung.¹⁾ Die von der Intelligenz abhängigen Unterschiede sprachlicher Fertigkeiten zwischen den Schichten und die schichtspezifische Verteilung bestimmter sprachlicher Merkmale können den Prozeß, wie sich in der Sozialstruktur verankerte kulturelle und ökonomische Lebensbedingungen des Elternhauses in Lebenschancen des Kindes umsetzen, nicht besser aufschlüsseln als irgendein anderes zwischen den Schichten variierendes psychosoziales Merkmal. Solche Ergebnisse können auch der pädagogischen Praxis keine begründeten Empfehlungen geben, außer dem neuerlichen Hinweis auf weitere Dimensionen mit schichtvariiender Merkmale. Sofern also naturalistische Unterschiede betrachtet werden, ist es konsequent, auf sozialisationstheoretische Ansprüche und damit auf pädagogische und sozialpolitische Bedeutung der Ergebnisse zu verzichten und Sprache als einen sozialstrukturellen Indikator u.a. und Gegenstand der makrosozialen Grundlagenforschung zu betrachten (Labov 1966, 1970).

Aber auch sofern wir naturalistische Unterschiede des Wortverständnisses mit statistischen Kontrollen betrachten, ergibt sich für die Theorie der linguistischen Codes ein zweites Risiko. Sie konnte die Wirksamkeit sozio-kultureller, prinzipiell nicht-sprachlicher, Regeln des Sprachgebrauchs innerhalb universeller für die jeweilige Kultur verbindlicher und ⁱⁿihrem für alle Sprachen gemeinsamen Kern wahrscheinlich biologisch begründeter (Lenneberg 1967) Regeln einer Hochsprache anhand der Gegenüberstellung von sprachlicher Kompetenz und sprachlicher Performanz verdeutlichen (Oevermann 1970:39, Bernstein 1971). Mit dieser Trennung ließ sich unmittelbar einsichtig machen, daß eine "Kompensation" allein

¹⁾-----
Dies gilt, wie gezeigt, für die Anlage der Untersuchungen von Bühler (1972) und Wiederhold (1971), aber ironischerweise nicht für ihre Ergebnisse. Dieser entscheidende Punkt wird in der Kritik beider Untersuchungen von Gloy (1973) nicht berücksichtigt.

auf Gebieten der Performanz sinnvoll ansetzen könne und gerade durch der Gemeinsamkeiten der Kompetenz ihre Wirkungschancen habe. Der aktive und der passive Wortschatz wurden dabei als Analogon zur ursprünglich auf den Bereich der Grammatik beschränkten Trennung zwischen Kompetenz und Performanz im Bereich des Mexikons betrachtet. Je mehr nun Einflüsse schichtspezifischer Strategien des Sprachgebrauchs auf kompetenzanaloge Bereiche, auf performanzanaloge jedoch nicht nachgewiesen werden können, desto mehr schwimmt für die Theorie der linguistischen Kodes der Bezugsrahmen, in dem der Einfluß von Strategien des Sprachgebrauchs sich fassen und die Möglichkeit pädagogischer Korrekturen sich darstellen ließ. Ein Ausweg aus diesem Dilemma auf theoretischer Ebene, der den Ergebnissen unseres Experiments gerecht wird, wäre es, das passive Wortverständnis und den aktiven Wortgebrauch gleichermaßen als Modi der sprachlichen Performanz anzusehen und auf ein Analogon zur grammatischen Kompetenz im Bereich des Lexikons zu verzichten; wir werden auf diesen Ausweg bei der Diskussion der Theorie der linguistischen Kodes (siehe 3.2.) zurückkommen.

Die zweite Möglichkeit, das Versagen der Hypothesen unseres Experiments zu erklären, greift an der Trennung zwischen kontrollierten mit Schicht korrelierenden Einflüssen psychischer Fähigkeiten und unter dieser Kontrolle zu untersuchenden sozial-determinierten Arten des Sprachgebrauchs an. Wir hatten den passiven Wortschatz als kompetenzanaloge Fähigkeit mit dem Frankfurter WST kontrolliert und die aktiven Worterklärungen als von sozialen Strategien des Sprachgebrauchs determinierte performanzanaloge Leistung gemessen. Wenn wir die Konsequenz aus den Ergebnissen der dritten Messung und den Ergebnissen Schlees (1973) ziehen, daß auch das passive Sprachverständnis von schichtspezifischen Arten des Sprachgebrauchs abhängig sein kann, wenn wir ferner die theoretische Überlegung berücksichtigen, daß passives Wortverständnis und aktiver Wortgebrauch gleichermaßen als Modi der Performanz angesehen werden können, dann scheint der passive Wortschatz nicht die angemessene Kontrollvariable gewesen zu sein. Durch die Kontrolle des WST sind nicht nur von der sozialen Schicht abhängige psychische Fähigkeiten, sondern auch schon selber Einflüsse schichtspezifischer Arten des Sprachgebrauchs ausgeschlossen worden, so daß in der Experimentalleistung ein Teil dessen, was ge-

messen werden sollte, nicht mehr gemessen werden konnte. Die Trennung zwischen mit Schicht korrelierenden psychischen Leistungen und Einflüssen schichtspezifischer Arten des Sprachgebrauchs ist zugegebenermaßen analytisch; allein aus diesem Grund können wir ja unsere Interpretation der Kontrolle des WST nachträglich ändern. Es gibt jedoch durchaus empirische Indizien dafür, ob man eine Variable im einen oder im anderen Sinn auffassen soll. Wir wollen abschließend solche Indizien aus unserer und anderen Untersuchungen diskutieren.

Wenn wir wirklich mit Schicht korrelierende psychische Fähigkeiten, die nicht zugleich auch Einflüsse schichtspezifischer Arten des Sprachgebrauchs abdecken, hätten kontrollieren wollen, so wäre es besser gewesen, die Intelligenz zu kontrollieren. Die drei Intelligenzgesamtmaße in der Gesamtstichprobe, aus der unsere Experimentalgruppe ausgewählt wurde, GESINT, VERBIT, NONVER korrelieren mit den vier Dimensionen der sozialen Schicht und dem Filterindex genau so hoch wie der WST (Tab. 22). Die Kontrolle sozialer Einflüsse überhaupt wäre also durch jedes der drei Intelligenzmaße genauso gesichert gewesen wie durch den WST. Was sich jedoch geändert hätte, wäre der Anteil des Einflusses schichtspezifischer Arten des Sprachgebrauchs am gesamten Einfluß überhaupt. Wir können sicher sein, daß er schon bei der verbalen Intelligenz VERBIT und der Gesamtintelligenz GESINT, die mit dem WST .60 bzw. .54 korrelieren, geringer gewesen wäre als beim Wortschatztest. Wir können ferner aus theoretischen Gründen sicher sein, daß der soziale Einfluß auf die nicht-verbale Intelligenz NONVER, die mit dem WST .38 korreliert, nur die Korrelation zwischen sozialer Schicht und psychischen Fertigkeiten enthält, die sich aus allgemeinen Unterschieden des soziokulturellen Anregungspotentials in verschiedenen Schichten ergeben und nur langfristig und vermittelt mit schichtspezifischen Arten des Sprachgebrauchs zusammenhängen; schließlich verlangen die nicht-verbale Subtests des PSB allesamt sprachfreie Leistungen des logischen Schließens (TEST 3 und 4), der räumlichen Vorstellung (TEST 7 und 8), des Rechnens (TEST 9) und der Wahrnehmungskonzentration (TEST 10). Wie wir an den Korrelationen gesehen haben, hängen in der Gesamtstichprobe alle Intelligenzge-

samtleistungen, auch VERBIT, nicht so hoch mit dem WST zusammen, wie man es nach dem gängigen Vorurteil des verbalen und daher mittelschichtorientierten, "bias" in Intelligenztests erwarten würde. Kurz: Von der Situation in der Gesamtstichprobe her wären GESINT und VERBIT für die Auswahl unserer Experimentalgruppe wahrscheinlich bessere, und NONVER wahrscheinlich die beste Kontrollvariable gewesen.

Hinzu kommt eine Besonderheit unserer Experimentalstichprobe der zweiten Messung. Wir haben dargelegt, daß wir ursprünglich die nicht-verbale Intelligenz zusätzlich zum Wortschatz kontrolliert hatten, daß jedoch durch einen Irrtum beim Testen das ursprüngliche Design geändert und die zusätzliche Kontrolle aufgegeben werden mußte. Wir haben uns damals - fast ohne Zögern - für die Aufgabe der Kontrolle von NONVER entschieden. Wir sehen jetzt, daß es ebenso diskutabel gewesen wäre, die Kontrolle von WST aufzugeben. - Wir haben in der Beschreibung der endgültig ausgewählten Untersuchungsgruppe der zweiten Messung dargelegt (siehe III.2.1.), daß der Zusammenhang zwischen VERBIT und WST in der Experimentalgruppe künstlich gewahrt, zwischen WST bzw. VERBIT und NONVER jedoch künstlich gelockert wurde; dies ist durch die Parallelisierung allein nicht notwendig, wie ein Vergleich mit der Experimentalgruppe der dritten Messung zeigt (siehe IV.2.) und zweifellos ein Ergebnis der Neugruppierung der Untersuchungsgruppe. Wir sehen aber jetzt, daß durch diese Tatsache wahrscheinlich der Anteil des Einflusses schichtspezifischer Arten des Sprachgebrauchs in der Kontrolle von WST zuungunsten des Anteils mit Schicht korrelierender psychischer Fähigkeiten über das Maß hinaus vergrößert wurde, das sich allein aus der Wahl der weniger günstigen Kontrollvariablen ergab. Die Korrelationen des WST betragen in der Stichprobe der zweiten Messung mit VERBIT .59, mit NONVER .04 (Tab. 22), in der Stichprobe der dritten Messung mit VERBIT .36, mit NONVER .24 (Tab. 167).

Ein Gegeneinwand gegen die Favorisierung von NONVER als alternativer Kontrollvariabler läßt sich jedoch aus unseren Daten entnehmen. Wir haben die Schichtvergleiche der zweiten Messung auch mit NONVER

und nicht-verbale Subtests des PSB als Kontrollvariablen durchgeführt und ebensowenig Schichtunterschiede feststellen können wie bei Kontrolle von WST (siehe III.5.2.2., Tab. 50). Die Schichtvergleiche mit alternativen Kontrollvariablen wurden jedoch mit einer Untersuchungsgruppe durchgeführt, die für die Kontrolle von WST ausgewählt worden war. Der Wechsel der Kontrollvariablen konnte sich nur noch in der Zusammenstellung anderer Paare innerhalb der gleichen Untersuchungsgruppe auswirken. Die entscheidenden Wirkungen der hier vorgeschlagenen alternativen Kontrollen bezieht sich jedoch auf die Auswahl der Vpn. aus der Gesamtstichprobe. Wir wollen daher zum Abschluß dieses Abschnitts die Kontrollvariablen untersuchen, die in anderen Untersuchungen verwandt wurden, die, anders als unsere Untersuchung, zu positiven Ergebnissen für die Theorie der linguistischen Codes geführt haben.

In Schlees (1973) Untersuchung über das passive Sprachverständnis bei Unter- und Mittelschichtkindern im Vorschulalter und im 4. Schuljahr wurden die Progressiven Matrizen von Raven als Kontrollvariable verwandt, ein Test also, der als der Idealfall eines "kulturfreien" nicht-verbale Tests angesehen wird. Hier bestätigt sich also unsere Vermutung, daß für den Test schichtspezifischer Strategien des Sprachgebrauchs am besten nicht-verbale Intelligenzleistungen kontrolliert werden. Auf der anderen Seite jedoch konnten die meisten Untersuchungen der Bernstein-Gruppe durch Strategien des Sprachgebrauchs bedingte Schichtunterschiede feststellen, obwohl sie verbale Leistungen kontrolliert hatten. Henderson (1970:21) fand, daß die Unterschicht eher die verb-bezogenen Formklassen der Sprache, die Mittelschicht eher die nominalen Formklassen der Sprache zur Beschreibung von Bildern benutzt; beide Schichtgruppen waren auf dem aktiven (Crichton Vocabulary Scale) und passiven (English Picture Peabody Vocabulary Test) Wortschatz parallelisiert. Robinson/Rackstraw (1972:76) zeigten, daß die Mittelschichtkinder Fragen häufiger beantworten und die Regeln eines Spiels besser erklären als Unterschichtkinder; beide Gruppen waren auf dem EPPVT parallelisiert. Wir verzichten darauf, die weiteren Untersuchungen der Bernstein-Gruppe, die in Bernstein et al. (1973)

zusammengestellt sind, aufzuzählen; alle - bis auf Hawkins (Kap. 4) Untersuchung, die offensichtlich keinerlei Kontrollen aufweist - kontrollieren eine der beiden verbalen Intelligenztests. Allerdings bestehen die Kontrollen in allen Bernsteinschen Untersuchungen nur aus annähernder Gleichheit der Gruppenmittelwerte und nicht - wie in unserer, Schlees und Sommers Untersuchung - in der Parallelisierung von Vpn.-Paaren auf der Kontrollvariablen. Schließlich ergeben sich in Sommers Untersuchung (1972), in der die verbale Intelligenz des HAWIK, ebenso wie die gesamte und die nicht-verbale Intelligenz des HAWIK kontrolliert werden, in keinem der drei Fälle für eine Mehrzahl von Variablen signifikante Schichtunterschiede. - Der Vergleich mit anderen Untersuchungen führt also zu keinem eindeutigen Schluß über die Angemessenheit der Kontrollvariablen. Wir können nur darauf hinweisen, daß die Alternative verbaler und nicht-verbaler Kontrollvariablen besonders in den Untersuchungen der Bernstein-Gruppe zu wenig beachtet worden ist.

1.3. Mängel der Operationalisierung und Schwierigkeiten der Diskriminierung zwischen aktivem und passivem Wortschatz.

In diesem Abschnitt behandeln wir in gewissem Sinne Konsequenzen der im vorangegangenen Abschnitt diskutierten Mängel. Wenn durch die Kontrolle des passiven Wortschatzes durch Strategien des Sprachgebrauchs bewirkte Unterschiede zwischen den Schichten schon ausgeschaltet sind, so wird nicht nur verständlich, daß in der aktiven Worterklärung Schichtunterschiede des Sprachgebrauchs nicht mehr auftauchen, sondern auch, daß die beiden Arten des Sprachgebrauchs, aktiver und passiver Wortschatz, wie wir sie gemessen haben, hoch miteinander korrelieren. In der Untersuchungsgruppe der zweiten Messung korrelieren der WST mit den Maßen der Worterklärung zwischen .50 und .60, während die Intelligenzgesamtmaße, die wir als alternative Kontrollvariable vorgeschlagen haben, zwischen .20 und .30 mit den Maßen der Worterklärung korrelieren (Tab. 108). Natürlich war eine hohe Korrelation zwischen den WST und der Worterklärung zu erwarten; daß sie jedoch in diesem Ausmaß über den ent-

sprechenden Korrelationen der Intelligenzgesamtmaße liegt, wirft die Frage auf, ob wir die angemessenste Operationalisierung für den aktiven Wortschatz gefunden und beide Aspekte des Wortgebrauchs genügend scharf getrennt haben.

Wir haben schon bei der Darstellung des Experimentalaufbaus (siehe II.1.) verschiedene Möglichkeiten der Operationalisierung des passiven und aktiven Wortschatzes diskutiert. Den passiven Wortschatz haben wir mit dem Frankfurter WST gemessen, der die Wortkenntnis selber wieder sprachlich durch die Wahl zwischen Worten erfaßt. Da in den Vorgaben des WST gewisse rein verbale "Fallen", wie Wort- und Klangassoziationen, liegen, wäre es möglich, daß in ihm die Kinder der Unterschicht schlechter abschneiden, als in einem sprachfreien Test des passiven Wortverständnisses, der mit der Wahl zwischen Bildern arbeitet. Aufgrund ihrer schichtspezifischen Einstellung zur Sprache würden sie eher assoziativ mit dem Stimulus verbundene Vorgaben wählen, und müßten über einen tatsächlich besseren Wortschatz verfügen, um zu einem mit den Mittelschichtkindern vergleichbaren Testwert zu gelangen. Nachdem wir von den Ergebnissen der dritten Messung und Schlees (1973) veranlaßt wurden, auch das Sprachverständnis als Kode-abhängig zu betrachten, gewinnt diese Interpretation an Wahrscheinlichkeit. Durch die Wahl des WST statt eines sprachfreien Tests als Kontrolle des passiven Wortschatzes hätten wir tatsächlich im passiven Wortschatz bessere Kinder aus der Unterschicht mit Kindern der Mittelschicht parallelisiert, und damit zugleich die Testbedingungen für eine durch schichtspezifische Strategien des Sprachgebrauchs bedingte Unterlegenheit der Unterschichtkinder in der zweiten Messung verschärft. Allerdings scheint diese Möglichkeit de facto nicht wirksam gewesen zu sein. Wir konnten in der ersten Messung (Frankfurter WST für Gesamtstichprobe, N=357) keinerlei durch den Test bedingte Schichtdiskriminierung feststellen, wenn auch mit der etwas kruden Methode der Korrelation zwischen Gesamtschwierigkeit und Schichtdiskriminierung der einzelnen Worte (siehe III.5.6.3.). Zudem stand die Alternative eines Bild-Vokabel-Tests praktisch nicht offen.

Den aktiven Wortschatz haben wir durch die Erklärung von Stimulusworten des Frankfurter WST gemessen. Eine Alternative wären frei oder experimentell stimulierte Sprachproben gewesen, deren absolute Anzahl von Worten, Type-Token-Ratios etc. wir hätten auszählen können; teilweise haben wir eine solche Auswertung in der der "sprachlichen Mittel der Erklärungen" (siehe III.6.) nachgeholt. Wir wollen alternative Operationalisierung des passiven und aktiven Wortschatzes aus den Untersuchungen Schlees (1973) und Sommers (1972) im Zusammenhang mit unseren Ergebnissen diskutieren.

Schlee (1973) maß das Sprachverhältnis durch das im Handeln ausgedrückte Verständnis von Instruktionen (Ausführung von Zeichnungen, Wahl von zu einer Erzählung passenden Bildern). Die in den Instruktionen gestellten Aufgaben für die jeweilige Altersgruppe (Vorschule, 4. Schuljahr) wurden so einfach ausgewählt, daß die Verständnisleistung nicht mit der Intelligenz, sondern mit dem gemessenen passiven Wortschatz korrelierte (Schlee 1973:60). Auch wenn die Gefahr, daß mit dem Instruktionsverständnis eher kognitive als sprachliche Leistungen gemessen werden, auf diese Weise empirisch ausgeschlossen ist, bleibt die Auswertung des sprachlichen Verständnisses zu unspezifisch. Schlee gab für jede korrekte Befolgung einer Anweisung einen Punkt und zählte die korrekt befolgten Anweisungen als gesamte Verständnisleistung zusammen. An keiner Stelle wird das spezifisch sprachliche Verständnis einzelner Worte, Sätze oder Konstruktionen gemessen. Mit Ausnahme der Zuordnung von Erzählungen und Bilder wird nur die pragmatische Sprachform des Befehls überprüft, die zudem durch die Testsituation und die Testbögen stark vorstrukturiert ist. Vor allem gibt es innerhalb der Testsituation keine alternativen Lösungen, an denen sich Verständnis und Nichtverständnis eindeutig trennen ließen, wie etwa in den in der Psycholinguistik angewandten Tests (Brown/Berko 1960, Brown/Fraser/Bellugi 1963, C. Chomsky 1969). Insgesamt scheint das von Schlee entwickelte Instrument viel zu krude für die Prüfung des Verständnisses von spezifischen Formen der Sprache, um als alternative Kontrollvariable für den Frankfurter WST verwandt werden zu können.

Sommer (1972) erhob in ihrer Untersuchung u.a. auch den aktiven und passiven Wortschatz, mit von unserer Untersuchung abweichenden Methoden. Der passive Wortschatz wurde mit Hilfe des (soweit dem Bericht entnehmbar: nicht für deutsche Sprecher neu geeichten) PPVT gemessen, der aktive Wortschatz mit Hilfe von offenen Fragen zum Alltagsleben der Kinder (Anzahl der Worte und TTR), von Satzergänzungstests, die Antonyme verlangten, und von zwei Untertests des HAWIK (WT: aktive Erklärung von 40 Testworten, die auf einer 3-Punkte-Skala bewertet wurden; GF: Oberbegriffe für zwei Stimulusworte, die ebenfalls auf einer 3-Punkte-Skala bewertet wurden). Beim Vergleich ohne Kontrolle von Intelligenzmaßen waren die Schulanfänger aus der Mittelschicht auf allen aktiven Maßen außer der TTR und auf den passiven Maßen den Schulanfängern aus der Unterschicht überlegen. Bei den drei Vergleichen mit Kontrolle der gesamten, der verbalen oder der nicht-verbalen Intelligenz ergab sich ein differenziertes Bild (Sommer 1972:40ff.): Im aktiven Wortschatz waren die Mittelschichtkinder nur dann überlegen, wenn man die besten der drei möglichen Antwortarten des WT-HAWIK (Synonyma) verglich. Im passiven Wortschatz waren die Unterschichtkinder überlegen, wenn man Kinder beider Schichten, die keinen Kindergarten besucht hatten und auf der nicht-verbalen Intelligenz parallelisiert waren, verglich; wenn man jedoch Kinder beider Schichten verglich, die den Kindergarten besucht hatten, so waren die Mittelschichtkinder überlegen, gleich ob man die verbale oder die nicht-verbale Intelligenz kontrollierte.¹⁾ Der Globaleindruck dieses Musters von Ergebnissen - und auch das Resümee der Autorin - ist, daß weder im aktiven noch im passiven Wortschatz bei Kontrolle der Intelligenz Unterschiede zwischen den Schichtgruppen bestehen. Insofern stimmen Sommers und unsere Ergebnisse voll überein.

1)-----

Dieses Ergebnis ist - nebenbei - interessant, weil es zeigt, daß die Wirkung des (traditionellen) Kindergartens eher in einer Vergrößerung der Chancenungleichheit besteht.

Die Übereinstimmung ist wichtig wegen der Unterschiede der Erhebungsmethoden. Soweit Sommers Erhebungsmethoden den aktiven Wortschatz betreffen, sind die von ihr verwandten offenen Fragen freier und produktiver als die von uns geforderte Bedeutungserklärung von Stimulusworten; soweit Sommer ähnliche Methoden wie wir verwendet - in den Tests WT und GF des HAWIK und im Bild-Vokabel-Test PPVT -, differenzierte sie weit weniger als wir in der Bewertung. Dennoch ergab sich eine gleiche Struktur von Ergebnissen. Das Versagen unserer Hypothesen kann sich also weder aus einer Rigidität des Untersuchungsansatzes (aktive Erklärungen von Stichworten), noch aus einer Über-Differenzierung der Auswertung ergeben haben.

Nach diesen Ergebnissen anderer Untersuchungen zum aktiven und passiven Wortschatz scheint es uns allein die Kombination beider von uns verwandten Erhebungsverfahren zu sein, die auf dem Gebiet der Operationalisierung für das Versagen unserer Hypothesen verantwortlich sein kann. Wir haben - wie keine der referierten Untersuchungen - das zu kontrollierende passive Verständnis und die zu prüfende aktive Erklärung mit dem gleichen Test bzw. Parallelformen gemessen. Die beiden Parallelformen des Frankfurter Wortschatztests korrelieren .90 (Anger/Bergmann/Hylla 1954). Der einzige Unterschied zwischen kontrollierter "passiver" und zu prüfender "aktiver" Leistung bestand darin, daß die passende Erklärung zum Stimuluswort einmal aus einer Reihe von Vorgaben ausgewählt, das andere Mal selber gebildet werden mußte. Wir hatten schon bei der Analyse der Unterschiede der Testformen und Messungen (siehe III.5.3.4., Tab. 63) Hinweise auf das "aktive" Element des Suchens im "passiven" Test und das "passive" Element der grundsätzlichen Wortkenntnis im "aktiven" Test hingewiesen: Bei der aktiven Erklärung ergaben sich zwischen den Testformen keine Unterschiede in der Güte der Erklärungen (B- und C-Kategorisierung), aber in der Anzahl der Worte, auf die überhaupt Erklärungen angestrebt wurden, die also in irgendeiner Weise "verstanden" worden waren (A-Kategorisierung). Wir können diesen Ansatz jetzt fortführen mit der Vermutung, daß zwar nicht Mängel in

der Operationalisierung des aktiven oder passiven Wortschatzes für unsere negativen Ergebnisse verantwortlich sein können, wohl aber die zu geringe Differenz zwischen den beiden Verfahren der Operationalisierung.

Zusammenfassung: Wir haben in Abschnitt 1. Ursachen für das Versagen unserer Hypothese in der Anlage des Experiments geprüft. Es liegt in der Natur der Sache, daß diese Ursachen keine zureichende Erklärung für das Versagen theoretisch begründeter Hypothesen bieten können. Jedoch können sie den Stellenwert unserer im Vergleich zu anderen Ergebnissen klären. In unserer Untersuchung wurde die Hypothese durch schichtspezifische Arten des Sprachgebrauchs bedingter Unterschiede zwischen Kindern der Unter- und Mittelschicht in der aktiven Worterklärung in mindestens dreierlei Hinsicht schärfer geprüft als in vergleichbaren Untersuchungen: Erstens war die soziale Distanz zwischen den Schichtgruppen nicht maximiert und die Leistungsdistanz zwischen den am engsten verwandten Gruppen am geringsten. Zweitens scheint durch die Wahl der Kontrollvariablen nicht nur die schichtspezifische Verteilung von Fähigkeiten ausgeglichen, sondern ein Teil des Effekts schichtenspezifischer Arten des Sprachgebrauchs schon eliminiert worden zu sein. Drittens waren die kontrollierte und die geprüfte Leistung sehr eng miteinander verwandt.

Schärfere oder weniger scharfe Testbedingungen sind jedoch kein Wert an sich. Wir haben daher zugleich diskutiert, welche der beiden Richtungen unter theoretischen Gesichtspunkten eingeschlagen werden kann. Der Vergleich sozialer Gruppen mit größerer Distanz scheint, sobald die Theorie - wie in dieser Untersuchung - aus einem rein explorativen Stadium heraus ist, keineswegs mehr sinnvoll. Die Wahl anderer als mit der Experimentalleistung direkt theoretisch verknüpfter Kontrollvariablen scheint uns jedoch besonders dann sehr sinnvoll, wenn sie vor allem die ungleiche "natürliche" Verteilung von Fähigkeiten ausgleichen und nicht selber schon die Effekte schichtspezifischer Arten des Sprachgebrauchs verdecken; wir haben versucht darzustellen, daß diese zunächst rein

analytische Trennung an empirischen Daten faßbar sein kann. Die zu enge Verwandtschaft der aktiven und passiven Testleistung in unserer Untersuchung schließlich scheint uns theoretisch nicht gefordert gewesen zu sein. Zusammen mit der Wahl der Kontrollvariablen scheint sie - auf der zunächst noch beschränkten Ebene der Experimentalanlage - für das Scheitern der Hypothesen verantwortlich zu sein.

2. Kritische Diskussion des Evokationsmodells im Rahmen des Wortschatzexperiments.

Die Verbindung der beiden Haupthypothesen unseres Experiments läßt sich mit Hilfe des Evokationsmodells herstellen. Es enthält die Annahme, daß die objektiv-sprachliche Form von Äußerungen des Senders, unabhängig von dem, was er tatsächlich meint oder versteht, die kognitiven Kategorisierungen des Empfängers bestimmt. Es bietet die Möglichkeit, Sprache und Denken über den Gebrauch von Sprache in sozialer Interaktion zu vermitteln. Die beiden Experimentalmessungen beziehen sich auf Schritte im Evokationsprozeß: die zweite Messung auf seine aktive Einleitung durch das Kind, die dritte Messung auf die beim Kind selber durch bestimmte sprachliche Formen evozierten Kategorien. Die Theorie der linguistischen Kodes sagt nun in diesem Evokationsmodell voraus, daß die Mittelschichtkinder aufgrund ihrer besseren Verbalisierung beim Zuhörer für sie günstigere Evokationsprozesse einleiten, und daß zugleich die Kinder beider Schichten mit den gleichen sprachlichen Formen die gleichen kognitiven Kategorien verbinden. Auf Dauer können beide Prozesse zusammen eine relative Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung unabhängig von ihren tatsächlichen Fähigkeiten erklären. Die Tatsache der schichtspezifischen Verteilung gemessener Intelligenz (Overmann 1966, 1970, Bosse u.a. 1972) und sogar das zunehmende Auseinanderklaffen schichtspezifischer Intelligenzwerte (Whiteman/Brown/Deutsch 1967) könnte im Rahmen des Evokationsmodells durch die Wirkung schichtspezifischer Arten des Sprachgebrauchs erklärt werden.

Statt der vorausgesagten Überlegenheit der Mittelschicht in der aktiven Erklärung ergab sich jedoch eine Gleichheit beider Schichten, und statt der vorausgesagten Gleichheit beider Schichten im passiven Verständnis z.T. eine Überlegenheit der Mittelschichtkinder. Wir wollen die Konsequenz der Widerlegung dieser Hypothesen zunächst für das Evokationsmodell diskutieren: Um welche weiteren empirischen Annahmen und Variablen muß es erweitert werden, damit die - unabhängig von unseren Experimentalergebnissen belegte - schichtspezifische Verteilung gemessener Intelligenz soziologisch

verständlich werden kann? In welcher Weise muß der Prozeß sozialer Interaktion differenziert werden, in dem Arten des Sprachgebrauchs die kognitive Entwicklung beeinflussen können?¹⁾ Erst nach dieser Diskussion werden wir uns den Konsequenzen der Ergebnisse für die die Voraussagen eigentlich motivierende Theorie der linguistischen Kodes zuwenden.

Die Hypothese der zweiten Messung, daß die Mittelschichtkinder trotz gleicher Wortkenntnis bessere aktive Worterklärungen geben als die Kinder der Unterschicht, bedeutet innerhalb des Evokationsmodells, daß die Kinder der Mittelschicht schon bei der Einleitung des Evokationsprozesses unabhängig von ihren Kenntnissen gegenüber den Unterschichtkindern im Vorteil sind. Die Bestätigung dieser Hypothese ist für die Theorie der linguistischen Kodes wichtig, die die Benachteiligung der Kinder der Unterschicht in der kognitiven Entwicklung auf die von ihnen internalisierte Strategie des Sprachgebrauchs zurück führen will, nicht jedoch für das Evokationsmodell, in dessen Rahmen wir die Chancen der Evokation in Abhängigkeit von der Sprachumwelt betrachten können. Innerhalb des Evokationsmodells diskutiert werden müssen vielmehr die Reaktionen der Kinder auf die Sprache der Umwelt, die die kumulative Benachteiligung der Unterschichtkinder nicht, und die Reaktionen des Erwachsenen auf die Äußerungen des Kindes, die die kumulative Benachteiligung der Unterschicht tatsächlich vermitteln sollen.

Die Hypothese der dritten Messung, daß die Kinder beider Schichten aus den gleichen sprachlichen Formen die gleichen Kategorisierungen evozieren, betrifft das Evokationsmodell unmittelbar. Sie konnte nur

¹⁾-----
Wir beschränken die Diskussion möglicher Erklärungen von Schichtunterschieden in der gemessenen Intelligenz auf das Evokationsmodell. Alternative Erklärungen, wie eine evtl. Schichtdiskriminierung durch die Testanforderungen, oder der schichtspezifische Charakter der "herrschenden" Kultur, sollen damit nicht ausgeschlossen werden.

z.T. bestätigt werden. Wir haben den nur beschränkten Stellenwert dieser partiellen Durchbrechung der Ausgangshypothese ausführlich diskutiert (siehe IV.3.4.3.). Aber auch wenn nur bestimmte Aspekte des passiven Verständnisses von schichtspezifischen Arten des Sprachgebrauchs abhängen, ist die Annahme des Evokationsmodells durchbrochen, daß sich allein aus den aktiven Verbalisierungen, nicht dem passiven Verständnis der Unterschichtkinder eine Benachteiligung in der kognitiven Entwicklung ergibt. Wir wollen im folgenden die Variablen im Evokationsprozeß auf der Seite des Zuhörers diskutieren, die trotz der festgestellten Unterschiede zwischen den Kindern beider Schichten im Verständnis die - empirisch ja unabhängig von unseren Ergebnissen gesicherte - Benachteiligung der Unterschichtkinder im Laufe der kognitiven Entwicklung erklären können.

Zunächst scheint das Evokationsmodell eine Grundtatsache sprachlicher Kommunikation zu bezeichnen: Die sprachliche Form der Äußerung des Senders bestimmt die evozierte Kategorie des Senders. Aber schon die Qualifikation "unabhängig von der tatsächlichen Meinung oder dem tatsächlichen Verständnis des Sprechers" bezeichnet tatsächlich eine empirische Variable, die der Empfänger, je nach der Gunst der außersprachlichen Umstände, manipulieren kann (Allert/v.Grote/Lappe 1972). Der Empfänger kann aufgrund von Situationshinweisen oder durch selbständige Extrapolation herausfinden, was ein Kind tatsächlich gemeint, aber nicht ausgedrückt hat; auf die gleiche Weise kann er das verbal leerlaufende überschüssige "Verständnis" eines Kindes aufdecken und seine Antwort entsprechend einrichten. Allerdings ist damit schon eine pädagogische Einstellung bezeichnet, die in alltäglichen Kommunikationen wohl selten wirksam ist. Das empirische Gewicht dieser Variablen für die vorausgesagte Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung durch ihre Strategie der eingeschränkten Verbalisierung dürfte daher gering, ihre Bedeutung für pädagogische Maßnahmen aber um so größer sein.

Aber auch abgesehen von der Möglichkeit der Ausbeutung außersprachlicher Informationskanäle enthält die Evokationstheorie eine vereinfachende Annahme über die Strategie der sprachlichen Reaktion des Zuhörers. Wir haben angenommen, daß der Zuhörer seine Antworten an dem aus der objektiv-sprachlichen Form erschließbaren Verständnis des Kindes orientiert und auf diese Weise Kinder mit verschiedenen Arten des Sprachgebrauchs unabhängig von ihrem tatsächlichen Verständnis unterschiedliche Informationseingaben erhalten. Diese Annahme ist zu einfach, weil unterschiedliche Strategien der Reaktion des Zuhörers denkbar sind und weil die Wahl der Strategie von der Schichtzugehörigkeit und der pädagogischen Rolle des Zuhörers abhängen kann. Ebenso wie die differenzierende, am kindlichen Sprecher orientierte ist eine egalitäre, am erwachsenen Zuhörer orientierte Reaktion des Zuhörers möglich. Beide Reaktionen können habitualisiert oder Teil einer bewußten pädagogischen Theorie sein. Robinson/Rackstraw (1972:113) konnten zeigen, daß die Antworten von Mütter beider Schichten - allerdings auf hypothetische Fragen des Kindes, die vom Interviewer gestellt wurden - sich nicht nach der Intelligenz ihrer Kinder unterschieden; in einer weiteren Untersuchung (Robinson/1973:232) zeigte sich, daß die Reaktionen der Mütter der Unterschicht - wiederum auf hypothetische Fragen des Kindes, die vom Interviewer gestellt wurden - über den Zeitraum von zwei Jahren relativ stabil waren, während die Mütter der Mittelschicht sich eher an - mutmaßliche oder tatsächliche - Entwicklungsfortschritte ihrer Kinder adaptieren. Wir können die Vermutung hinzufügen, daß eine egalitäre Reaktion wohl eher von primären, eine differenzierende Reaktion wohl eher von professionalisierten Sozialisationsagenten zu erwarten ist. Die langfristige Wirkung der Benachteiligung der Unterschichtkinder könnte nach diesen Differenzierungen des Evokationsmodells durch eine weniger flexible Reaktionsstrategie der Mütter der Unterschicht und eine schlechtere "Passung" zwischen differenzierten Reaktionen des Lehrers und den Erwartungen inflexibler Reaktionen auf seiten des Unterschichtkindes sich ergeben haben.

Eine weitere Vereinfachung betrifft die Kriterien, die dem Evokationsprozeß des Zuhörers zugrundeliegen. Wir haben ursprünglich angenommen, daß der Zuhörer aus der geäußerten Sprachform die objektiv enthaltenen Kategorisierungen herausfiltert. Ebenso aber kann der Zuhörer seine Einschätzung des Sprecher-Verständnisses an offener zu Tage liegenden Kriterien der sprachlichen Konvention und der sprachlichen Korrektheit orientieren; die Dimensionen der sprachlichen Form können sozial bewertet sein aufgrund populärer, aber keineswegs gesicherter Theorien über den Zusammenhang zwischen Sprechen und Denken. Die Benachteiligung der Unterschichtkinder durch die kumulativen Wirkungen eines sozialen Kommunikationsprozesses könnte sich, statt - wie wir angenommen haben - aus der schichtspezifischen Verbalisierungsstrategie der Unterschichtkinder, ebensogut aus der Selbstverifikation einer vulgarisierten Whorf- oder Bernstein-Hypothese ergeben haben. Wir haben (in III.6.) versucht, die analytische Trennung dieser beiden Ebenen an der Auswertung der "Erklärungsleistung" und der "sprachlichen Mittel der Erklärung" empirisch greifbar zu machen. Das Ergebnis, daß auf der Ebene der Erklärungsleistung keine, auf der Ebene der sprachlichen Mittel der Erklärung jedoch recht bedeutende Unterschiede zwischen den Kindern beider Schichten bestehen, zwingt die Evokationsannahme zu einer Verlagerung der Kriterien: vom objektiv kategorialen Gehalt zur sprachlichen Form der Äußerung. Allerdings ergab sich aus dem Muster der von den Kindern der Unter- und Mittelschicht bevorzugten sprachlichen Formen kein Anhaltspunkt für die Wirkung einer vulgarisierten Whorf- oder Bernstein-Theorie, mit Ausnahme der größeren Fehleranzahl in der Unterschicht. Im Gegenteil: die Unterschichtkinder bevorzugten eine Reihe von sprachlichen Formen, die man unter diesem Gesichtspunkt von den Mittelschichtkindern erwartet hätte (Indefinitpronomina, anaphorischen Pronominalgebrauch, Subordinationen, nominal ansetzende Erklärungen).

Für die langfristige Wirkung der Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung ist die Evokationstheorie auch

bei einer Verlagerung der Kriterien wichtig. Wir wissen, daß soziale Bewertungen das pädagogische Handeln des Lehrers unabhängig von den tatsächlichen Fähigkeiten des Kindes beeinflussen und auf die kognitive Entwicklung des Kindes andauernde Wirkungen zeigen, die sich auch unabhängig von den ursprünglich die Entwicklung des Kindes beeinflussenden Bewertungen durchsetzen, sich gleichsam verselbständigen (Rosenthal/Jacobson 1971). Allerdings kann die langfristige Wirkung dieser Bewertungen - nach unseren Daten - nicht durch mit Hilfe der Theorie Bernsteins legitimierte Mittelschichtvorurteile ausgelöst worden sein.

Eine letzte Modifikation der Evokationsannahme gründet sich auf den besseren Erfolgen vergleichbarer Untersuchungen. Robinson/Rackstraw (Robinson/Rackstraw 1972, Robinson 1973) untersuchten die Antworten von Müttern 5- und 7jähriger Kinder auf hypothetische Fragen des Kindes und die Antworten der Kinder auf Fragen und konnten in der sachlichen Angemessenheit der Antworten von Kindern und dem pädagogischen Entgegenkommen in den Antworten der Mütter (Nichtausweichen, keine autoritative Wiederholung der Frage als Feststellung, Berufung auf Prinzipien statt auf universelle Regelmäßigkeiten etc.) bedeutende Unterschiede zwischen den beiden Schichtgruppen feststellen. Fragen und Antworten sind Schritte in einer sozialen Kommunikation, die von der Evokation des kategorialen Verständnisses des Sprechers beim Hörer begleitet sind und die unmittelbar empirisch erhoben werden können. Gegenüber unserem Experiment hätte eine Untersuchung von Fragen und Antworten durch die objektiv soziale Form einer solchen Interaktion den Vorteil, daß der Zuhörer viel stärker auf das tatsächliche Verständnis des Sprechers als auf die nur beschränkten sprachlichen Formen einer Frage sich konzentrieren muß und daß seine taktischen Reaktionen als Variable erfaßt werden können. Zugleich könne man die langfristigen Wirkungen schichtspezifischer Strategien der Antworten der Erwachsenen auf die kognitive Entwicklung der Kinder leicht extrapolieren. Allerdings können Fragen als Auslöser des Evokationsprozesses nicht so leicht wie Bedeutungserklärungen erhoben werden; für die Theorie der linguistischen Codes ist jedoch - wie wir ausgeführt haben - der Nachweis entscheidend, daß

schichtspezifischer Strategien des Sprachgebrauchs vom Kinde selber internalisiert sind und seine Äußerungen unabhängig von seinen Kenntnissen beeinflussen.

Zusammenfassung: Wir haben in Abschnitt 2. die Konsequenzen der Widerlegung der beiden Ausgangshypothesen für das Evokationsmodell untersucht. Die Voraussage besserer aktiver Erklärungen der Mittelschichtkinder ist für die Theorie der linguistischen Codes entscheidend, nicht jedoch für das Evokationsmodell. In ihm kann angenommen werden, daß unterschiedliche Kategorisierungsprozesse auch durch Unterschiede der schichtspezifischen sprachlichen Anregung initiiert werden. Die Widerlegung der Voraussage gleichen passiven Verständnisses betrifft das Evokationsmodell unmittelbar. Sie zwingt zu einer Differenzierung der Annahmen über die Reaktionen des Zuhörers. Wir haben dargestellt, daß der Zuhörer auch außer-verbale Hinweise für den Schluß auf das Verständnis des Sprechers hinzuziehen kann, daß er differenzierende und egalitäre Strategien bei der Entscheidung über seine Reaktion befolgen kann, und daß er die objektiv vorliegenden Äußerungen des Sprechers unter zwei Kriterien, dem kategorialen Gehalt und der sprachlichen Form, beurteilen kann. Alle drei genannten Dimensionen können, in unterschiedlichem Maße, zu einer Benachteiligung der Unterschichtkinder innerhalb des Evokationsprozesses führen, selbst dann, wenn die Unterschichtkinder andere Arten des Verständnisses zeigen als die Mittelschichtkinder. Alle drei Dimensionen müssen wir als Variable in einem sozialen Interaktionsprozeß ansehen, in dem Sprache und Denken durch den Sprachgebrauch vermittelt werden. In dieser Arbeit haben wir das alternative Kriterium der sprachlichen Form ausführlich untersucht. Die Ergebnisse zeigen, daß starke Unterschiede zwischen den Schichten bestehen, die jedoch - bis auf die Grammatik - und Rechtschreibfehler - nicht mit ideologischen Vorstellungen über die Sprache der Unterschichtkinder in Zusammenhang gebracht werden können. Schließlich haben wir die Untersuchung von Fragen und Antworten als alternativen Zugang zu Evokationsprozessen, neben der Beurteilung von Worterklärung und Wortverständnis, vorgeschlagen.

3. Konsequenzen der Ergebnisse des Wortschatzexperimentes für die Theorie der linguistischen Kodes.

Die Theorie der linguistischen Kodes ist eine Theorie über die sozialen Determinanten des Gebrauchs der Sprache. Wenn wir als eine soziale Determinante die Schichtzugehörigkeit nehmen, sagt sie Unterschiede zwischen den Schichten im aktiven Sprachgebrauch, nicht aber in der passiven Verfügung der Sprache voraus. Keine dieser beiden Voraussagen konnte am Beispiel des Wortschatzes bestätigt werden. Erst recht konnte die angestrebte, auf dem Zusammenwirken beider Voraussagen ruhende Erklärung der Benachteiligung der Unterschichtkinder in gemessenen Intelligenzleistungen durch die sozialen Folgen eines schichtspezifischen Sprachgebrauchs nicht erbracht werden. In dieser Situation wollen wir über einige theoretische Dimensionen und Probleme diskutieren, die das Versagen der Theorie in diesem Experiment erklären können und mit denen die Theorie anhand der Ergebnisse dieses Experiments weiterentwickelt werden kann. Mit den Ergebnissen eines einzelnen Experiments, in dem nur bestimmte Aspekte eines Theoriegebäudes untersucht wurden (siehe I.1.) kann über den Wert der ganzen Theorie noch nicht entschieden werden. Aber es kann angegeben werden, welche Ableitungen modifiziert und welche zusätzlichen Annahmen eingeführt werden müssen. Unsere Ergebnisse berühren zwei strategische Punkte der Theorie der linguistischen Kodes, die wir im folgenden behandeln werden: die Differenzierung zwischen verschiedenen Ebenen sozialer Determination des Sprachgebrauchs (3.1.) und die Beeinflußbarkeit alltäglich wirksamer Arten des Sprachgebrauchs durch intentionale Erziehungsprozesse (3.2.); diese Darstellung verbleibt noch im Rahmen der Theorie linguistischer Kodes. Abschließend wollen wir dann kurz skizzieren, welche Gesichtspunkte der Theorie der linguistischen Kodes beibehalten werden sollten, selbst wenn sie in ihrer jetzigen Form als widerlegt gelten muß (3.3.).

3.1. Die Abhängigkeit des Sprachgebrauchs von typischen Sozialbeziehungen und von typischen Situationen als Problem bei der Erhebung schichtspezifischer Sprachproben.

Wir können zwei Ebenen der sozialen Determination des Sprachgebrauchs in der Theorie der linguistischen Kodes unterscheiden. Auf der ersten Ebene bestimmen typische Arten von Sozialbeziehungen, d.h. die allgemeinen Dimensionen von sozialen Rollen, eine typische Art des Sprachgebrauchs. Im Lauf der Primärsozialisation werden die Arten des Sprachgebrauchs von den Heranwachsenden internalisiert, in Abhängigkeit von den in der Familie vorherrschenden Formen der Sozialbeziehungen und - teils direkt, teils vermittelt über die Familie - in Abhängigkeit von den in den sozialen Subkulturen vorherrschenden Sozialbeziehungen. Ähnlich also wie andere grundlegende Persönlichkeitsdimensionen, z.B. kognitive Stile, Leistungsmotivation, Problemlösestrategien, werden die Arten des Sprachgebrauchs durch die in der Primärsozialisation erfahrenen Sozialbeziehungen determiniert und bilden in der weiteren Lebensgeschichte situationsüberdauernde Einstellungen des Individuums; ebenso wie jene gehören sie zu den allgemeinen Handlungsstrategien eines Individuums. Auf der zweiten Ebene bestimmen objektive Qualitäten und soziale Definitionen von Interaktionssituationen, d.h. die allgemeinen Dimensionen von sozialen Situationen, den Sprachgebrauch. Diese beiden Ebenen der sozialen Determination des Sprachgebrauchs wurden von Oevermann (1970:188-200) als die "komplexe" und die "elementaristische" Version der Theorie der linguistischen Kodes bezeichnet.

In soziale Situationen treten einerseits - wie es die "komplexe" Version darstellt - Individuen ein, die eine bestimmte Art des Sprachgebrauchs, eine bestimmte Einstellung zu Wahrnehmung und Interpretation der Situation, schon internalisiert haben, so daß eine Variation des Verhaltens und Sprechers in Abhängigkeit von der Vielfalt internalisierter Strategien des Sprechens und Handelns

der Teilnehmer zu erwarten ist.¹⁾ Soziale Situationen üben andererseits - wie es die "elementaristische" Version darstellt - von ihren objektiven Qualitäten, dem "setting", her einen Zwang zur Uniformität des Verhaltens aus und sind in vielen Fällen von einem Konsens fast aller Mitglieder der Gesellschaft beherrscht, so daß eine Gleichheit des Verhaltens und Sprechens unabhängig von den durch die persönliche Sozialisationsgeschichte bedingten Einstellungen der Teilnehmer zu erwarten ist. Auf der einen Seite können also in einigen Situationen allgemeine Strategien des Sprachgebrauchs durchschlagen und durch ihre unterschiedliche Wirkung die Unterschiede in der determinierenden Sozialisationsgeschichte der Sprecher offenlegen. Auf der anderen Seite können einige Situationen durch die objektiven Anforderungen und die Breite des sozialen Konsens so hoch strukturiert sein, daß es für lebensgeschichtlich erworbene Arten des Sprachgebrauchs keinen Spielraum, keine Alternativen in den sprachlichen Formen und Leistungen mehr gibt, durch die sie sich ausdrücken könnten. Ganz allgemein kann man sagen, daß die Empfindlichkeit von Situationen für Strategien des Sprachgebrauchs überhaupt eine empirische Variable ist.

Die Trennung dieser beiden Ebenen der sozialen Determination des Sprachgebrauchs hat Konsequenzen für die Erhebung schichtspezifischer Sprachproben. In den meisten sozio-linguistischen Untersuchungen - auch dem hier dargestellten Experiment - werden im Rahmen der "komplexen" Version der Theorie bestimmte Merkmale der "Sozialschicht" - zumeist auch, wie hier, in Form der Dichotomie - als Indikator für die lebensgeschichtlich erfahrenen typischen Formen der Sozialbeziehungen, und bestimmte Formen manifester Sprachproben als Indikator für die schichtspezifischen Strategien des Sprach-

¹⁾-----
Eine wichtige Ausnahme bilden Situationen der Sozialisation und des Lernens, in denen die Ungleichheit der Positionen notwendige Vorbedingung und die Gleichheit der Teilnehmer langfristiges Ziel der Interaktion ist. Diese Ausnahme ist für die analytische Trennung der "komplexen" und der "elementaristischen" Version des Kodex nicht systematisch relevant.

gebrauchs genommen. In den meisten sozio-linguistischen Untersuchungen - auch dieser - werden Sprachproben von Personen verschiedener Schichten wiederum in Situationen erhoben, über deren prinzipielle Abhängigkeit oder Unabhängigkeit von schichtspezifischen Arten des Sprachgebrauchs und über deren unterschiedliche Bedeutung für verschiedene Strategien des Sprachgebrauchs nichts bekannt ist. Die Unempfindlichkeit der Test-Situation für Strategien des Sprachgebrauchs überhaupt verringert jedoch die Chance für Sprachunterschiede, die als Indiz für Arten des Sprachgebrauchs i.S. der "komplexen" Version angesehen werden können, und die Empfindlichkeit der Test-Situation nur für spezifische Strategien des Sprachgebrauchs kann umgekehrt das Auftreten von Unterschieden der Sprache erleichtern, die als Indikator für unterschiedliche Arten des Sprachgebrauchs i.S. der "komplexen" Version angesehen werden. Wir wollen im folgenden zwei Untersuchungen darstellen, die den Einfluß der Testsituation auf das Auftreten von Unterschieden des Sprachgebrauchs zwischen Vpn. zweier Schichtgruppen überprüft haben, um aus diesen Untersuchungen Merkmale zu gewinnen, die die Empfindlichkeit von Testsituationen für Einflüsse des Sprachgebrauchs erfassen können, und um diskutieren zu können, wieweit möglicherweise die Unempfindlichkeit unserer Testsituation für Einflüsse des Sprachgebrauchs zu den negativen Ergebnissen beigetragen haben kann.

Beide Untersuchungen, über die wir referieren wollen, arbeiten mit der sprachlichen Auswertung von Interview-Antworten, die nach der Schwierigkeit der Interview-Fragen in unterschiedliche "Test-Situationen" eingeteilt wurden. Beide Untersuchungen variieren also die Situation bei den gleichen Personen. Lawton (1970) stellte 12- und 15jährigen Schülern in Einzelinterviews zuerst "deskriptive" (Ort und Gebäude der Schule), dann "abstrakte" (Zweck der Erziehung, Sinn der Strafe für Verbrecher) Fragen. Er stellte fest, daß die Unterlegenheit der Unterschicht im Vergleich zur Mittelschicht auf verschiedenen Indizes des "elaborierten" Sprachgebrauchs bei den Sprachproben auf "abstrakte" Fragen viel geringer ist als bei den Sprachproben auf "deskriptive" Fragen und daß keine

Unterschiede zwischen beiden Schichten in bezug auf den Inhalt der Antworten bestehen. Er gibt zwei Erklärungen für diesen Sachverhalt (1970:190): Erstens können durch den Charakter der "abstrakten" Fragen die Antworten viel stärker vorstrukturiert worden sein als durch die "deskriptiven"; die Fragen nach dem Sinn der Erziehung zwingt generell zum Gebrauch von Kausal- und Finalsätzen oder zur Wahl abstrakter Nomina oder ungewöhnlicher Verben und Adverbien, so daß auf den entsprechenden Maßen geringere Schichtunterschiede als bei "deskriptiven" Fragen auftauchen. Zweitens könnte die "Herausforderung" durch einen entgegenkommenden und freundlichen Interviewer die Schüler der Unterschicht in der Interview-Situation, in der eine Antwort ja verlangt und unmittelbar kontrolliert wurde, zu einem "elaborierten" Sprachgebrauch veranlaßt haben. Beide Erklärungen verweisen auf mögliche "elementaristische" Einflüsse auf den Sprachgebrauch: den objektiven Anspruch der Aufgabe und das Anspruchsniveau des Interaktionspartners.

Williams/Naremore (1969) teilten die Interview-Fragen nach dem Antwort-Spielraum ("probe-constraint") in minimal mit Ja-Nein, einer Benennung oder einer längeren Ausführung beantwortbare Fragen ein; auch die Ja-Nein- und die Benennungsfragen konnten mit längeren Ausführungen beantwortet werden. Auf Ja-Nein-Fragen antworteten Kinder beider Schichten mit Ja-Nein oder Ausführungen, die Unterschichtkinder jedoch häufiger mit erstem, die Mittelschichtkinder mit letztem; auf Benennungsfragen antworteten die Kinder beider Schichten mit einfachen Benennungen, ausführlichen Benennungen oder Ausführungen, die Unterschichtkinder jedoch häufiger mit erstem, die Mittelschichtkinder häufiger mit letztem; auf Ausführungsfragen antwor-

ten Kinder beider Schichten fast ausschließlich mit Ausführungen.¹⁾ Williams/Naremore schließen aus diesen Ergebnissen, daß die Kinder der Unterschicht nur minimal antworten, wenn dies möglich ist, aber dem Zwang der Fragestellung folgen, wenn dies notwendig ist. Beide Untersuchungen zeigen, daß der objektive Anspruch der Aufgabe als eine Variable der Situation i.S. der "elementaristischen" Version der Theorie der linguistischen Kodes über die Durchschlagskraft lebensgeschichtlich sozialisierter Strategien des Sprachgebrauchs i.S. der "komplexen" Version der Theorie der linguistischen Kodes entscheiden kann. Wenn es die Aufgabe verlangt, bedienen sich auch Sprecher des "restringierten" Kodes der manifesten Sprachformen des "elaborierten" Kodes und bewältigen die Aufgabe ebenso wie Sprecher des "elaborierten" Kodes. Gemessen am Ausgangspunkt ihres schichtspezifischen Sprachgebrauchs werden sie zwar zu größeren Anpassungsleistungen als die Sprecher der "elaborierten" Kodes gezwungen, die sie aber - als Sprecher ihrer Muttersprache und handlungsfähige Individuen - prinzipiell erbringen können.

1) -----
Dies scheint jedoch ein "Dacheffekt" zu sein: Für den Beweis, daß Kinder beider Schichten unter entsprechenden Testbedingungen gleiche Reaktionen zeigen, ist eine Experimentalanlage dann nicht zweckmäßig, wenn schon von der Logik der Testbedingung nur eine überhaupt sinnvolle Reaktion denkbar ist. Methodisch exakt müßten die Abweichungen von Zufalls- oder Durchschnittswerten anderer Populationen für jede Antwortform unter der Bedingung jeder Frageform für jede Schichtgruppe angegeben werden. Prinzipiell besteht bei der Untersuchung Williams/Naremore die Gefahr, daß unter "Testbedingungen" und "Antworten" gleiche oder sich logisch unmittelbar bedingende Sachverhalte doppelt klassifiziert werden. Die Ergebnisse auf Ja-Nein und auf Benennungs-Fragen scheinen jedoch diese Befürchtung zu entkräften. Wenn wir uns auf diese beiden Testbedingungen beschränken, zeigt sich immer noch eine Zunahme der "elaborierten Antwortformen" mit der minimal anspruchsvolleren Frage-Form, so daß wir die Untersuchung für unseren Zweck zitieren können.

In unseren Experimentalmessungen - der "aktiven Worterklärung" und der "kategorialen Differenzierung" - besteht nun ebenfalls die Möglichkeit, daß die Eindeutigkeit der Testsituation keine Differenzierung zwischen "komplexen" Strategien des Sprachgebrauchs mehr zugelassen hat, daß m.a.W. die Qualität der Testaufgaben selber die Testbedingungen für Schichtunterschiede verschärft hat. Wir haben in unserem Experiment die Variation von Situationen selber nicht geprüft. Deshalb kann die genannte Erklärungsmöglichkeit der negativen Ergebnisse nicht systematisch nachgeprüft werden; jedoch können wir einige Belege aus der Experimentalanlage und den Ergebnissen zitieren, die die genannte Erklärungsmöglichkeit unterstützen. Gegenüber einer derartigen nachträglichen Rechtfertigung der negativen Ergebnisse innerhalb der Theorie der linguistischen Kodes muß die Widerlegung ihrer Ausgangshypothesen natürlich der wichtigere Sachverhalt bleiben.

Zunächst wollen wir an zwei Verschärfungen der Testbedingungen erinnern (siehe 1.2.), die auch in diesem Zusammenhang wichtig sind: die beiden Schichtgruppen waren auf einer Testleistung parallelisiert,¹⁾ und sie waren mit der jeweiligen Parallelförm des Frankfurter Wortschatztests auf einer Testleistung parallelisiert, die mit der aktiven Erklärung der Stimulusworte der jeweils anderen Testform sehr eng verwandt war. Durch Unterschiede subkultureller Anregung bedingte Unterschiede des Wortschatzes waren damit ausgeschlossen. Einflüsse schichtspezifischer Arten des Sprachgebrauchs hätten sich zwischen den beiden Schichtgruppen immer noch in Unterschieden des kategorialen Gehalts, nicht nur der sprachlichen Form der Erklärung auswirken müssen. Die geringe Variation der Testleistungen mit der Sozialschicht, aber ihre auffällige Variation mit der Wortgruppe (siehe alle Analysen in III.5.4.) kann auch als ein Indiz für den geringen Spielraum möglicher Einflüsse schichtspezifischer Arten des Sprachgebrauchs bei der Erklärung gelesen werden: bei einer a-Priori Klassifika-

1)-----
Hier muß erwähnt werden, daß die beiden Schichtgruppen in der Untersuchung von Lawton, nicht aber der von Williams/Naremore auf der gemessenen Intelligenz parallelisiert waren.

tion der Stichworte ergaben sich keine Schichtunterschiede, aber starke Unterschiede zwischen diesen Gruppen in beiden Schichten; die Differenzen zwischen Verständnis und Explikation, die Profile der Bewertungen und die Anzahl der falschen Erklärungen - alles recht direkte Indikatoren für Arten des Sprachgebrauchs - variierten in beiden Schichten mit den Wortgruppen. Alle diese Ergebnisse zeigen, daß die entscheidenden Einflüsse auf die Testleistung in der Variation der objektiven Anforderung der Aufgabe selber lagen.

Zweitens ergaben sich zwischen der mündlichen und schriftlichen Messung keine Unterschiede, und zwar in beiden Schichtgruppen (siehe III.5.3.4.). Die Kommunikationsform kann jedoch nahezu als eine Variable der Testsituation selber angesehen werden; zumindest stellt sie einen "elementaristischen" Filter für den Einfluß "komplexer" Strategien des Sprachgebrauchs auf die Testleistung dar. Daß auch die Kommunikationsform keinen Effekt auf die Erklärungsleistungen zeigte, ist ein für die "Unempfindlichkeit" der Experimentalleistung für Einflüsse außerhalb des objektiven Anspruchs der Aufgabe ein sehr starkes Indiz, unabhängig von den Ergebnissen zum Schichteinfluß. Hinzukommt, daß alle Variablen der Analysen aus III.5.4., die mit der sozialen Schicht nicht variierten, auch mit der Kommunikationsform nicht variierten. Alle diese Ergebnisse weisen darauf hin, daß die in der zweiten Messung gemessene Erklärungsleistung vom objektiven Anspruch der Aufgabe so weit bestimmt war, daß schichtspezifische Strategien des Sprachgebrauchs nicht mehr durchschlagen konnten.

Drittens kann diese Erklärung der Ergebnisse der Erklärungsleistung auch durch die Ergebnisse zu den sprachlichen Mitteln der Erklärung unterstützt werden (siehe III.6.). Die Kinder der Unterschicht benutzten häufiger eine Reihe von sprachlichen Mitteln, die in allen bisherigen Untersuchungen als Zeichen eines "elaborierten Kodes" angesehen und häufiger bei den Kindern der Mittelschicht gefunden wurden. Wenn nun die Testsituation die Unterschichtkinder zu überproportionalen Anpassungsleistungen zwingt, dann kann man an ihrem Wechsel von "restringierten" zu "elaborierten" Sprachformen die Anstrengung der Erklärungsleistung messen, die ihnen nicht von ihrer schichtspezifischen Strategie des Sprachgebrauchs, sondern den Bedingungen der Testsituation abverlangt wurde. Wenn andererseits den Mittel-

schichtkindern die Erklärungsleistung nicht mehr an Anpassung abfordert, als ihnen ihre schichtspezifische Strategie des Sprachgebrauchs in jedem Fall nahegelegt hätte, dann kann man verstehen, daß sie auch zur Wahl typischer "elaborierter" Sprachformen nicht in außergewöhnlichem Maße gezwungen waren. Wir glauben m.a.W., daß man aus der Umkehrung der Schichttypik bei den sprachlichen Mitteln der Erklärung auf den Widerstand schließen kann, der den Unterschichtkindern bei der Erklärungsleistung aufgrund ihrer schichtspezifischen Strategie des Sprachgebrauchs entgegengestanden hat und den sie unter dem objektiven Anspruch der Aufgabe überwinden mußten.

Als ein viertes Indiz für eine derartige Erklärung der Ergebnisse der zweiten Messung sei daran erinnert, daß auch das Muster von unterschiedlichen und gleichwertigen Leistungen in der dritten Messung eine ähnliche Deutung nahelegte (siehe IV.3.5.). Im "kategorialen Präzisionstest" ergaben sich keine, im "differenzierten Verständnistest" signifikante Schichtunterschiede, die sich in der qualitativen Analyse auf die unterschiedlichen sprachlichen Formen der Stimuli und der Vorgaben sowie die größere semantische Variationsbreite der Vorgaben zurückführen ließ; zugleich war der "kategoriale Präzisionstest" in beiden Schichtgruppen einfacher als der "differenzierte Verständnistest" und korrelierten die analogen Maße der beiden Untertests in der Mittelschicht hoch, in der Unterschicht überhaupt nicht. Wir hatten als eine Erklärung vorgeschlagen, daß der "differenzierte Verständnistest" den Unterschichtkindern Ablenkungen eröffnet, die den Mittelschichtkindern aufgrund ihres schichtspezifischen Sprachgebrauchs fernliegen und daß dieser Unterschied sich nivelliert, sobald von der Testaufgabe her kein "Angriffspunkt" für schichtspezifische Strategien des Sprachgebrauchs mehr besteht. -

Wir sind uns bewußt, daß unsere Argumentation - als Rechtfertigung der vorliegenden Ergebnisse - hart am Rande einer Immunisierungsstrategie steht. Für mögliche zukünftige Untersuchungen der Abhängigkeit des Symbol- und Sprachgebrauchs von frühzeitig sozialisierten Handlungsstrategien enthält sie jedoch die Warnung, daß bei der Operationalisierung "komplexer" Einflüsse "elementaristische" Einflüsse nicht umgangen werden können und daß das Kräfteverhältnis beider vorgängig kalkuliert werden muß.

3.2. Nivellierung schichtspezifischer Arten des Sprachgebrauchs durch intentionale Elemente der familialen und schulischen Sozialisation.

Wir haben im vorangegangenen Abschnitt dargelegt, daß die Testleistungen der Kinder durch den "objektiven Anspruch der Aufgabe" so stark vorstrukturiert waren, daß sich lebensgeschichtlich erworbene Arten des Sprachgebrauchs auf ihnen nicht mehr abbilden konnten. Wir wollen in diesem Abschnitt untersuchen, welche Faktoren für die nivellierende Wirkung der Testsituation verantwortlich gemacht werden können. Der Begriff "objektiver Anspruch der Aufgaben" umfaßt zwei analytisch trennbare Ebenen: die dinglich-objektive Umwelt oder das "setting" der Testsituation und die sozial-objektive Stereotypisierung situationsspezifischer Erwartungen. Die Umwelt der Testsituation war durch die eindeutige Form der Testaufgabe, in der mündlichen Messung auch noch durch den Einsatz der Gegensprechanlage, stark standardisiert. Stereotype Erwartungen waren zwar nicht durch in der Testsituation präsente Interaktionspartner der Vpn. kontrolliert, aber sehr wohl durch den symbolischen Gehalt der objektiven Umgebung "Schule" repräsentiert. Wir vermuten, daß vor allem die in der "Schule" verkörperten situationstypischen Erwartungen mögliche Einflüsse schichtspezifischer Strategien des Sprachgebrauchs nivelliert haben und wollen für diese Vermutung - teils spekulativ, teils in Anlehnung an unsere empirischen Ergebnisse - im folgenden einige Belege geben.

Zunächst können wir für die zweite Messung die Wirkung einer "Erhebung in der Schule" mit einer "Erhebung im Elternhaus" direkt vergleichen. Wie in III.2.2. dargelegt, waren wir durch einen Irrtum bei der Durchführung des Experiments gezwungen, einige Kinder im Elternhaus nach Beginn der Ferien zu testen.¹⁾ Die Leistungen

¹⁾-----
Von den 66 Vpn. der zweiten Messung wurden 51 in der Schule, 15 im Elternhaus getestet (siehe Tab. 32). Die Vpn. der dritten Messung wurden alle in der Schule getestet.

der Kinder unterschieden sich zwischen beiden Situationen nur in der schriftlichen, durch das Fehlen der Gegensprechanlage für äußere Einflüsse empfänglicheren Messung: Insgesamt wurden die sprachlichen Mittel durch die Testsituation stärker als die Erklärungsleistung beeinflusst. Der Erhebungszeitpunkt zeigte in beiden Schichtgruppen ähnliche Effekte; bei den Maßen der B- und C-Kategorisierung, die die Güte der Erklärungen unabhängig von der Häufigkeit der Erklärungen erfaßten, ergab sich jedoch folgendes (Tab. 33a): die Kinder der Unterschicht zeigten in der Schule, die Kinder der Mittelschicht im Elternhaus bessere Leistungen. Ein solches Ergebnis kann genau die zusätzliche Leistungsherausforderung charakterisieren, die die Schule für die Unterschichtkinder darstellt.

Wir müssen jedoch zwei Bedeutungsaspekte des Begriffs "Schule" berücksichtigen. Man kann unter "Schule" die soziale Institution verstehen, in der Lernprozesse organisiert werden: auf diese Bedeutung haben wir uns bis jetzt beschränkt. Man kann unter "Schule" aber auch alle die Lernprozesse verstehen, die von alltäglichen Lebensvollzügen gelöst, "abstrakt" sind und die durch besondere Instanzen geplant und kontrolliert werden; in dieser Bedeutung kann auch das Elternhaus eine Schule sein. Wir sind an zwei Stellen auf Interpretationen gestoßen, die die Ergebnisse auf intentionale Elemente der elterlichen Erziehung in beiden Schichtgruppen zurückführten. Wir haben - beim Vergleich der vier in unserem Experimental-Sample enthaltenen Schichtgruppen - die signifikant größeren Differenzen zwischen Verständnis und Explikation in der Oberen Unterschicht und der Mittleren Mittelschicht als das Ergebnis einer größeren Risikobereitschaft dieser Kinder beim Erklären gedeutet, die sich aus einem verständnisvoll entgegenkommenden, quasi-pädagogischem Verhalten der Eltern ergeben haben könnte (siehe III.5.6.1.). Wir haben ferner den Einfluß der Ausbildung der Mutter auf die Testleistung in beiden Schichten so interpretiert, daß die Mutter ihre durch Berufstätigkeit nicht absorbierte

Ausbildung zum Aufbau pädagogischer Autorität in der Familie nutze (siehe III.5.6.2.). Beide Ergebnisse deuten darauf hin, daß schon in den Familien intentionale Elemente der Erziehung, wenn auch mit größeren Kosten für das Unterschichtkind, Unterschiede zwischen Kindern beider Schichten nivellieren.

Wenn wir die Widerlegung der Voraussagen der Theorie der linguistischen Kodes dadurch erklären, daß durch intentionale Elemente schulischer und familialer Sozialisation mögliche Wirkungen schichtspezifischer Strategien des Sprachgebrauchs nivelliert worden sind, dann können wir auch ein durchgängiges Ergebnis unseres Experiments besser interpretieren: Wo immer Unterschiede zwischen beiden Schichten auftauchten, bezogen sie sich eher auf passive als auf aktive Formen des Sprachgebrauchs; das Wortverständnis, nicht aber die Worterklärung scheint schichtspezifischen Strategien des Sprachgebrauchs unterworfen zu sein. In der aktiven Worterklärung fanden wir überhaupt keine Schichtunterschiede auf den Explikationsmaßen, aber einen nicht-signifikanten Überlegenheit der Mittelschichtkinder auf den Verständnismaßen (siehe III.5.3.2.). Im differenzierten Verständnistest der dritten Messung fanden wir eine signifikante Überlegenheit der Mittelschichtkinder (siehe IV.3.2.). Wenn nun durch pädagogische Einwirkung schichtspezifischer Strategien des Sprachgebrauchs nivelliert werden, so sind zuerst die unmittelbar kontrollierbaren aktiven Formen, allenfalls mittelbar die nicht beobachtbaren passiven Formen betroffen. Eine solche Interpretation wird zusätzlich durch die Ergebnisse Schlees (1973) unterstützt, der starke Unterschiede im Instruktionsverständnis zwischen Kindern der Unter- und Mittelschicht fand.

Wir wollen abschließend einige Konsequenzen unserer Interpretation andeuten. Daß durch eine schulische Vermittlung der "herrschenden Kultur" Unterschiede zwischen Kindern beider Schichten nivelliert werden, schließt ein, daß diese Nivellierung für Kinder beider Schichten mit unterschiedlichen Kosten verbunden ist. Sie hat sich ja erst bei der Einführung experimenteller Kontrollen ergeben; nach

wie vor bestehen signifikante Unterschiede zwischen Kindern beider Schichten im Wortschatz, sobald keine experimentellen Kontrollen eingeführt sind; hier liegt die Bedeutung der Ergebnisse der ersten Messung. Über die Mittel, die trotz unterschiedlicher Kosten gleiche Ergebnisse ermöglichen, können wir nur spekulieren: vermutlich spielt eine starke Motivation der Kinder ebenso eine Rolle wie eine implizit kompensatorische Didaktik der Lehrer. Ebenso müssen wir offenlassen, wieweit die Nivellierung der Schichtunterschiede soziale Herrschaft, als Verweigerung der Legitimität für non-konforme Bedürfnisse und Deutungsmuster, verbirgt, und wieweit sie ein notwendiger Schritt ist für eine gleichwertige Teilnahme aller an den sozialen Gütern, die einer Gesellschaft auf einer bestimmten sozial-evolutionären Entwicklungsstufe zur Verfügung stehen. Die genannten weiterreichenden Fragen müssen in einer Theorie der sozialen Konstitution und Organisation von Bildungsinhalten geklärt werden.

3.3. Allgemeine Gesichtspunkte der Theorie der linguistischen Kodes.

Wir haben in den bisherigen Abschnitten des vierten Kapitels versucht, auf verschiedenen Ebenen das Scheitern unserer Hypothesen zu erklären. Auf der Ebene der Anlage des Experiments könnte höchstens die Wahl einer Kontrollvariablen, die mit der Experimentalleistung zu eng verwandt ist, fehlende Schichtunterschiede verursacht haben. Auf der Ebene der Theorie des Experiments haben wir eine Reihe von Erklärungen vorgeschlagen. Zuerst sind - wie wir bei der Diskussion des Evokationsmodells dargelegt haben - zu viele intervenierende Variablen zwischen dem Sprachgebrauch des Kindes und seinen Lernerfolgen übersprungen worden. Zweitens glauben wir den Einfluß "elementaristischer" zugunsten "komplexer" Faktoren des Sprachgebrauchs unterschätzt zu haben. Schließlich haben wir - als einzige positive Erklärung - intentionale Elemente der schuli-

schen und familialen Erziehung für die Nivellierung der Schichtunterschiede in der aktiven Worterklärung und den Fortbestand von Schichtunterschieden im Sprachverständnis verantwortlich gemacht. Alle genannten Erklärungen bewegen sich im Rahmen der Anlage und der Theorie unseres Experiments. Eine weiterreichende Konsequenz ist es jedoch, die Theorie der linguistischen Kodes - angesichts der den Hypothesen diametral entgegengesetzten Hauptergebnisse und angesichts der Vielzahl von negativen Einzelanalysen - aufzugeben und nach alternativen, aber gleichwertigen Theorien zu suchen. Wir wollen dazu an die "essentials" der Theorie der linguistischen Kodes erinnern, die u.E., auch wenn sie widerlegt ist, als Desiderate der Theoriebildung verbleiben, und sie in den allgemeineren Begriffen kennzeichnen, in denen auch mögliche Alternativen beschreibbar sind.

Die entscheidenden Merkmale der Theorie der linguistischen Kodes lassen sich auf zwei Ebenen, dem makrosozialen Kontext und dem sozialisationstheoretischen Bezugspunkt der Theorie, darstellen. Erstens ist die Theorie der linguistischen Kodes als eine Sozialisationstheorie zugleich eine Theorie über die gesamtgesellschaftliche Reproduktion. Sie versucht die Prozesse aufzuschlüsseln, in denen sich sozial objektive Lebensbedingungen in sozial stereotype Arten der Weltauffassung und Lebenschancen der Individuen umsetzen. Auf dieser Ebene muß die Theorie der linguistischen Kodes in den makrosozialen Bezugspunkten sozialer Integration und sozialen Wandels gesehen werden. Theoretisch ist diese Betrachtung des Sozialisationsprozesses von Durkheim begründet worden; empirisch ist sie durch eine Vielzahl von Untersuchungen zur schichtspezifischen Verteilung gemessener Intelligenz und des Schulerfolgs motiviert.

Zweitens versucht die Theorie der linguistischen Kodes die Verbindung zwischen sozial objektiven Lebensbedingungen und individuellen Lebenschancen auf einer Reihe analytischer Ebenen herzustellen, die durch das Konzept "Arten des Sprachgebrauchs" vermittelt werden. Die sozialen Ursachen des Sprachgebrauchs sind auf einer all-

gemeinen Matrix- wie etwa Parsons "pattern variables"-klassifizierte Formen von Sozialbeziehungen: sie können sich auf jeder institutionellen Ebene, der Arbeits- und Berufswelt, der sozialen Schicht und der sozialen Subkultur, der Nachbarschaft, peer group oder Familie konkretisieren und auf jeder institutionellen Ebene einer Reihe von Funktionen (die Bernstein (1971) auch "Kontexte" nennt) dienen, der Kontrolle von Konformität, dem zweckorientiertem Umgang mit der Umwelt oder der wechselseitigen Befriedigung von Bedürfnissen. Die Konsequenzen des Sprachgebrauchs können in den individuellen, aber sozial typischen Kategorien der Wahrnehmung und Interpretation der dinglichen und sozialen Umwelt, in Arten der Handlungsplanung und Problemlösung, in den Fähigkeiten sozial angemessenen, autonomen Handelns gesehen werden; auch dies sind, wie die Formen der Sozialbeziehungen, abstrakte Dimensionen, die kategorial ausgearbeitet und empirisch gefüllt werden müssen.

Das Konzept "Arten des Sprachgebrauchs" bezeichnet eine allgemeine Einstellung, die an der Folge früher sozialisierender Interaktionen abgelesen, internalisiert und zunehmend stabilisiert werden und die später in den verschiedensten Interaktionssituationen in der gleichen Art die Handlungschancen für das Individuum eröffnen. Je nach den Formen der sozialisierenden Interaktion, die natürlich ihrerseits von den typischen Handlungsformen in der umgebenden Subkultur und der Berufs- und Arbeitswelt abhängen, können für Individuen typische Handlungspotentiale angenommen und, bei gegebenen Situationen, für Individuen bestimmte Reaktionen vorausgesagt werden. Für den Sprecher des "elaborierten Kode" bietet sich prinzipiell die Chance, den latenten Gehalt von Situationen auszudrücken und seine individuellen Handlungsmotive zugleich mit der Anerkennung sozialer Normen herauszustellen. Der Sprecher des "restringierten Kode" dagegen ist an die standardisierte Definition der Situation gebunden und sucht Rückhalt an kollektiv geteilten und rituell bekräftigten Bedeutungen; er empfindet die Leistungen des "elaborierten Kode" eher als redundant. Unterschiede in diesen prinzipiell sprachfreien Einstellungen zur Sprache entscheiden in vielen Fällen über die Chancen des Individuums, neue und fremde Situationen

zu bewältigen und sich in Bereichen außerhalb des Alltags in Beruf und Familie z.B. in politischen und bürokratischen Auseinandersetzungen, im Umgang mit Gütern der allgemeinen Kultur, in Auslandsreisen etc. zu orientieren.

Die Wahl des Konzepts "Arten des Sprachgebrauchs" kann auf beiden für die Theorie wesentlichen Ebenen begründet werden. Auf der sozialisationstheoretischen Ebene können wir zwei konkurrierende Motive unterscheiden: die Bedeutung des Gebrauchs von Sprache als eines Mittels für verschiedene psychische Leistungen, wie der Affektkontrolle, der Unabhängigkeit von kurzfristig wechselnden Umwelteinflüssen, der Verarbeitung von Normübertretungen und der Problemlösung; und die Bedeutung des Gebrauchs von Sprache in sozialer Interaktion. Das erste Motiv wird gestützt durch eine Vielzahl vor allem lerntheoretischer Untersuchungen der Leistungen der Sprache bei der Organisation der Wahrnehmung und Erinnerung und der Kombination konkret nicht zu verbindender Erfahrungspartikel (siehe v. Dewitz/Flaake/Härtel/Schweisfurth 1972). Durch dieses Motiv erhält die Sprache gleichsam eine psychologische Legitimation, die über den von schichtspezifischen Strategien des Sprachgebrauchs abhängigen Einsatz von Sprache soziologisch bedeutsam würde. Soziologisch gesehen bleibt die Wahl jedoch zufällig; jede durch psychologische Ergebnisse bedeutsame Fähigkeit, die zugleich mit sozialen Merkmalen variiert, könnte auf diese Weise als sozialisationstheoretischer Vermittlungsmechanismus gewählt werden. Das zweite Motiv wird gestützt durch theoretisch-soziologische Überlegungen zur Struktur intentionalen Handelns und zum Aufbau von Ich-Identität, die zuerst von Mead ausgeführt wurden (Oevermann 1970, Bernstein 1971). Mead (1934) analysierte, schematisch zusammengefaßt, daß gemeinsame Bedeutungen beim Sprecher implizit dieselben Handlungsreaktionen hervorrufen, die sie beim Zuhörer explizit hervorrufen und daß dies nur durch den Mechanismus der "vokalen Geste" sichergestellt werde; Mead analysierte weiter, daß Ich-Identität durch die Übernahme der Rolle zunächst einzelner sozial bedeutsamer Anderer, dann des verallgemeinerten Anderen auf-

gebaut wird und daß der Prozeß der Rollenübernahme aufgrund gemeinsamer - sprachlich gesicherter - Bedeutungen möglich ist.

Für die Wahl der "Arten des Sprachgebrauchs" als vermittelndem Konzept ist nun das erste, psychologisch-empirische Motiv subsidiär, das zweite, soziologisch-theoretische Motiv jedoch entscheidend. Durch dieses zweite Motiv kann das Ziel von Sozialisationsprozessen theoretisch als das autonom handlungsfähige, selbst-identische Subjekt expliziert und zugleich empirisch an Indikatoren des Sprachgebrauchs zugänglich gemacht werden. Während in der psychologischen Konzeption Sprache als Instrument für bestimmte Fähigkeiten angesehen wird, wird sie in der soziologischen Konzeption als Medium sozialen Handelns analysiert. Darüber hinaus ist es möglich, daß die psychischen Leistungen der Affektkontrolle, der Umwelt-Unabhängigkeit und der Problemlösung z.T. selbst als Leistungen des sozialen Handelns analysiert werden können.

Die sozialisationstheoretischen Motive für die Theorie der linguistischen Kodes sind weithin bekannt. Mindestens ebenso wichtig, aber kaum beachtet, sind die makrosoziologischen Motive. Das Konzept "Arten des Sprachgebrauchs" ist auf der makrosozialen Ebene vor allem deswegen wichtig, weil es - auf einer höheren Allgemeinheitsebene als der Ebene konkreter Handlungserwartungen und Handlungen - einen Komplex von Einstellungen bezeichnet, der, betrachtet man den Prozeß der Produktion, in weitgehender Abhängigkeit von der Sozialstruktur, betrachtet man das Produkt, in relativer Unabhängigkeit von der Sozialstruktur gesehen werden kann. Erst wenn - das ist die implizite These der Theorie der linguistischen Kodes - die Sozialisationstheorie auf dieser Ebene Konzepte entwickelt, hat sie für die makrosozialen analytischen Bezugspunkte der Integration bzw. Kontrolle der Reproduktion bzw. des Wandels Gewicht.

Die Aufgabe der Sozialisation wäre für die Sozialisationsagenten nicht zu bewältigen, wenn sie jede einzelne Verhaltenserwartung übertragen müßten; die Integration einer Gesellschaft könnte nicht

gewährleistet werden, wenn in jeder Interaktion die "Komplementarität der Erwartungen" kontrolliert werden müßte. Für den makrosozialen Bezugspunkt der Integration bzw. Kontrolle können sozialisationstheoretische Mechanismen Ausschließlichkeit beanspruchen, wenn man auch die institutionalisierten sekundären Formen der Kontrolle unter sie subsumiert. - Wenn wir uns die Gesellschaft als eine Summe einzelner Verhaltenserwartungen vorstellen, so wäre das Problem von Reproduktion bzw. Wandel gleichsam in zu kleine Einheiten zerlegt, um greifbar zu werden. Durch die soziale Genese und die Stabilität der allgemeinen Einstellungskomplexe ist jedoch die soziale Reproduktion der Gesellschaft garantiert, und durch das Aufeinandertreffen stabilisierter Einstellungskomplexe mit anderen als den ursprünglich sozialisierenden, "fremden" Sozialstrukturen der soziale Wandel der Gesellschaft. Für den makrosozialen Bezugspunkt des sozialen Wandels bzw. der sozialen Reproduktion können sozialisationstheoretische Mechanismen allerdings keine Ausschließlichkeit beanspruchen. Ebenso wie es naiv wäre, die Ursachen sozialen Wandels allein in mit der Veränderung individueller Einstellungen verbundenen Prozessen zu sehen, wäre es aber dogmatisch, diese Ursache des Wandels gegenüber der Organisation von und den Konflikten zwischen sozialen Gruppen zu ignorieren.

Je differenzierter eine Gesellschaft ist, desto notwendiger wird aus makrosozialen Gründen die Sozialisierung nicht nur von Verhaltenserwartungen, die an Rollen festgemacht sind, sondern von allgemeinen verhaltensgenerierenden Einstellungen, die an Individuen festgemacht sind. Der Differenzierung von Gesellschaften entspricht also nicht nur die Differenzierung, sondern die Hierarchisierung von Sozialisationsprozessen. Durkheims Analyse, daß die soziale Arbeitsteilung und die Genese von Individualität komplementäre Prozesse darstellen, kann - auf der Seite der für den Einzelnen aus sozialer Differenzierungsprozessen erwachsenden Chancen - von der Vielfalt der Rollenteilnahme zur autonomen Strukturierung der Rollenteilnahme ausgedehnt werden. Mit einem allgemeinen Satz, der nach den üblichen Trennungen individueller

und sozialer Anteile der Person nach dem Modell des "homo sociologicus" nicht mehr ganz trivial ist, kann man sagen, daß die Gesellschaft nicht nur aus Rollen, sondern auch aus Individuen besteht und daß das Handeln eines Menschen nicht nur aufgrund seiner Rolle, sondern auch aufgrund seiner sozialtypischen, individuellen Einstellung voraussagbar ist.

Die Erklärung der Reproduktion sozialer Ungleichheit wird erleichtert, wenn man anstelle der Internalisierung von Verhalten und Verhaltenserwartungen die Internalisierung von allgemeinen Einstellungen, die Verhalten und Verhaltenserwartungen produzieren, einführt. Aus der Annahme, daß soziale Unterschiede in der frühen Sozialisationsgeschichte durch die Bildung allgemeiner Einstellungen des Individuums langfristige Konsequenzen haben, wird verständlich, daß die Reproduktion sozialer Ungleichheit von den Individuen - auch und besonders denen, die im status quo benachteiligt sind - mitgetragen und vollzogen wird, und daß die Gesellschaft von der dauernden Durchsetzung und Legitimation der Ungleichheit entlastet ist. In einem Konzept der Gesellschaft als der Summe von mit Rollen verbundenen Handlungserwartungen hingegen muß mit einer Vielzahl von Annahmen erklärt werden, warum bestimmte soziale Ungleichheiten nicht wahrgenommen oder anerkannt werden: mit der Informationsüberforderung der Individuen, mit der Trennung von institutionellen Bereichen oder mit institutionalisierten Kontrollen in jedem Einzelfall. Durch die Annahme der Internalisierung allgemeiner verhaltensproduzierender Einstellungen wird zudem auch die Chance der Kompensation ungleicher sozialer Determination verständlich; sie liegt in der bewußten Bereitstellung derjenigen sozialen Formen, der Interaktion und der Aufschlüsselung der Umwelt, die unter günstigen Umständen latent wirken. Wir wollen aber wiederholen, daß für den makrosozialen Bezugspunkt von Reproduktion bzw. Wandel die primäre Sozialisation nur ein Mechanismus u.a. darstellt und entscheidende andere Mechanismen hier in den Herrschaftsverhältnissen, der Struktur des Arbeits- und Berufslebens und - nicht zuletzt - in der sozial organisierten öffentlichen Bildung liegen.

Wir haben versucht, die zwei wesentlichen Elemente der Theorie der linguistischen Kodes herauszuarbeiten: Auf der Ebene des makro-sozialen Kontexts muß sie als eine Theorie über Mechanismen der sozialen Reproduktion, auf der Ebene des sozialisationstheoretischen Bezugspunkts als eine Theorie über die Genese langfristig wirksamer, handlungserzeugender Einstellungskomplexe angesehen werden.¹⁾ Die negativen Ergebnisse unseres Experiments ziehen nun eine spezifische Ausführung einer solchen Theorie in Zweifel: Die Annahme, daß die individuellen Einstellungskomplexe in Arten des Sprachgebrauchs zu sehen sind, die typischerweise mit den in den beiden sozial-statistischen Gruppierungen der Unter- und Mittelschicht vorherrschenden Formen von Sozialbeziehungen verbunden sind. Die negativen Ergebnisse unseres Experiments können u.E. nicht die theoretischen Desiderate entkräften, die die Theorie der linguistischen Kodes in einer spezifischen Weise zu erfüllen versucht. Auf der theoretischen Ebene können wir uns drei Reaktionen auf die negativen Ergebnisse unseres Experiments denken, die sich sukzessive stärker von der ursprünglichen Konzeption der Theorie der linguistischen Kodes entfernen.

Als erstes kann die Verbindung zwischen Arten des Sprachgebrauchs und Formen von Sozialbeziehungen (bzw. sie operational definierenden sozialen Gruppen) theoretisch weiter differenziert werden. Bernstein hat verschiedene Formen familialer Sozialbeziehungen und Kontrollweisen analysiert und empirisch untersucht (Bernstein 1971, Cook-Gumperz 1973). Außerdem ließen sich die Formen verschiedener sozialer Situationen in ihrem Einfluß auf Arten des Sprachgebrauchs untersuchen. Analog zu den determinierenden sozialen Bedingungen müßten natürlich die Arten des Sprachgebrauchs selber theoretisch weiter differenziert werden. Kurz: Diese Strategie bedeutet eine intensive

¹⁾-----
Bernstein selber charakterisiert, in einem späteren Aufsatz seine Theorie auf diesen beiden Ebenen (1971:172): Essentially and briefly I have used Durkheim and Marx at the macro-level and Mead at the micro-level.

Prüfung der "elementaristischen" Version der Theorie der linguistischen Kodes, bevor sie wieder in ihrer "komplexen" Version (Oevermann 1970) für die Untersuchung der Reproduktion sozialer Ungleichheit fruchtbar werden kann.

Als zweites kann man alternative Konzepte handlungserzeugender Einstellungskomplexe theoretisch entwickeln und empirisch greifbar machen. Ansätze bietet das Konzept des "Habitus" (Bourdieu 1970, Bourdieu/Passeron 1973). Der Habitus ist ein für bestimmte soziale Gruppen typisches allgemeines Einstellungsmuster: man kann vom Habitus einer Epoche ("der Renaissance-Mensch"), einer Regional- oder Nationalgesellschaft, einer sozialen Klasse ("der Angestellte") oder eines Standes ("der preußische Junker"), einer Profession ("der Lehrer") oder eines biographischen Stadiums ("der Pennäler") sprechen. Man kann weiterhin nicht nur rekonstruierte Einstellungsmuster, sondern auch normativ wirksame Zielvorstellungen als Habitus betrachten: "christian gentleman", "der in Einsamkeit und Freiheit arbeitende Gelehrte". Bourdieu/Passeron sehen im Habitus ein "generatives Prinzip"; über die Vorstellung hinaus, daß der Habitus eine Vielzahl von Verhaltensweisen und -erwartungen produzieren kann und das Verhalten eines Individuums in einer Vielfalt von Situationen voraussagbar macht, hat der Terminus "generativ" jedoch keinen spezifischen Gehalt.

Gegenüber dem "Kode" ist der "Habitus" nicht in bezug auf ein spezifisches Medium, die Sprache, definiert, auf das die in beiden Konzepten erfaßten allgemeinen Einstellungen wirken. Das hat komplementäre Vor- und Nachteile. Die Vorteile liegen darin, daß der Habitus nicht mit bestimmten Verhaltensformen verwechselt werden kann, wie die Kodes mit grammatischen Konstruktionen, und daß mit ihm eine analytisch nicht auf bestimmte Dimensionen der Sozialstruktur beschränkte Vielfalt von Einstellungskomplexen erfaßt werden und dem Alltagsverständnis nachkonstruiert werden können. Die Nachteile liegen darin, daß die empirische Indizierung erschwert ist und die Gefahr der Lösung des Begriffs von analytisch klar definierten Bedingungen der Sozialstruktur aufkommt. Der theoretische

Punkt, an dem die Beziehungen beider Konzepte sich klären ließen, wäre das Verhältnis von Strategien des Sprachgebrauchs zu allgemeinen Strategien des sozialen Handelns. Auf dieser letzten Ebene müßten Habitus theoretisch definiert werden, damit das Konzept nicht in der empirischen Vielfalt und der Doppelung von Alltagsbegriffen verschwimmt.

Als drittes kann eine Theorie der sozialen Konstitution und Organisation von Bildungsinhalten (Bourdieu/Passeron 1973, Bernstein 1971 c, Young 1971) entwickelt werden, die die genannten theoretischen Desiderate erfüllt, zugleich die Wirkungsweisen sozialer Reproduktion und die Genese handlungserzeugender Einstellungskomplexe zu erklären. Bildungsinhalt i.S. einer solchen Theorie ist nicht nur der Kanon von Disziplinen und Curricula, sondern auch der Komplex nicht-kodifizierter Werte, Normen und Deutungen, der in jeder sozial organisierten Form von Bildung vermittelt wird. Eine solche Theorie verbindet religions-, wissens- und herrschaftssoziologische Elemente auf der einen Seite mit sozialisationstheoretischen Elementen auf der anderen Seite, die Frage der "Legitimierung" mit der Frage des "Erlernens" von Bildungsinhalten. Eine solche Theorie muß Aufgaben auf dem Gebiete theoretischer Grundlagen lösen, die Konsequenzen haben auch für die Bewertung politischer Reformen. Sie müßte als erstes klären, wieweit Bildungsinhalte aus sozial-historischen Herrschaftsverhältnissen niederschlagen, wieweit sie ferner allgemeine Kulturgüter der sozial-evolutionären Entwicklungsstufe einer Gesellschaft sind, wieweit sie schließlich Explikationen universell-antropologischer Kompetenzen sind. Sie könnte dann zeigen, ob durch den bekannten Kanon oder die latenten Gehalte sozial-organisierter Bildung spezifische Interessen privilegiert und spezifische Bedürfnisse unterdrückt sind und durch die Planung und Verwaltung sozial organisierter Bildung die sozial-naturwüchsige Determination von Bildungsinhalten gesteuert werden kann.

4. Praktische Folgerungen aus den Ergebnissen auf sozialpolitischem und pädagogischem Gebiet.

Ergebnisse der Sozialisationsforschung, die im Rahmen der Theorie der linguistischen Kodes gewonnen wurden, können - auf praktisch-sozialpolitischer Ebene - eine Erklärung für die Nicht-Realisierung des Prinzips der Chancengleichheit und - auf praktisch pädagogischer Ebene - Empfehlungen für die Kompensation durch die primäre Sozialisation bedingter ungleicher schulischer Bildungs- und damit sozialer Lebenschancen geben. Die Ungleichheit der Bildungschancen, auf deren Erklärung wir zuerst eingehen wollen, ist ein komplexes Bedingungsgefüge, in dem - sehen wir von den sozial nicht manipulierbaren konstitutionellen Variablen ab - Variablen der Primärsozialisation, der schulischen Sozialisation und der Verwaltung der herrschenden Kultur zusammenspielen. Wir können drei Ebenen der Analyse der Ungleichheit und entsprechender Programme des Ausgleichs von Bildungschancen unterscheiden, die jeweils bestimmte Bereiche dieses Bedingungsgefüges akzentuieren und die, so kann man sagen, auch die Abfolge der Diskussion um Chancengleichheit in der BRD nachzeichnen.

Die erste Ebene der Analyse bezieht sich auf die Mechanismen der Ungleichheit innerhalb der Schule. Sie untersucht besonders das Lehrerverhalten und die Organisationsstruktur des Schulwesens. Es konnte nachgewiesen werden, daß die Lehrer mit ihrer Notengebung z.T. schichtspezifisch diskriminieren und sich auf, gemessen am Kriterium der Leistung, illegitime Kriterien der "Gepflegtheit", des "guten Schülers" etc. verlassen (Latscha 1966, Steinkamp 1967, Preuß 1972); daß ferner das dreigliedrige Schulwesen zwischen sozialen Schichten, Konfessionen und Geschlechtern ungleich selektiert, und zwar in Land und Stadt und in verschiedenen Regionen in unterschiedlichem Maße (Peisert 1967, Popitz 1965). Entsprechende praktische Folgerungen betreffen die Objektivierung der Leistungsmessung (Tests) und die Beseitigung nicht aufgrund von Leistung selektiv wirkender

Differenzierungen des Schulwesens. Beide Maßnahmen laufen auf eine Rationalisierung der Schule im Sinne der Leistungsgerechtigkeit hinaus. Für schichtdiskriminierendes Lehrerverhalten, unabhängig von der Leistung, gibt es heute, d.h. ca. 5 Jahre nach den entsprechenden Untersuchungen, nur noch geringe empirische Anhaltspunkte (Bosse u.a. 1972); die Reform des dreigliedrigen Schulwesens ist von sozialer Anerkennung so weit entfernt wie von politischer Durchsetzung.

Die zweite und dritte Ebene der Analyse bezieht sich auf die ungleichen Voraussetzungen, mit denen Kinder aus verschiedenen sozialen Gruppen in die Schule kommen, ihr "kulturelles Kapital" (Bourdieu/ Passeron 1973; siehe III.5.6.5.), das in der Schule selektiv wirkt, obwohl die Schule selber nicht selektiert. Diese ungleichen Voraussetzungen werden auf der zweiten Ebene als individuelle, auf der dritten Ebene als kollektive Merkmale verstanden. Die zweite Ebene bezieht sich auf sozial produzierte Ungleichheiten der Intelligenz, der Leistungsmotivation und der kognitiven Stile; sie behandelt Mechanismen der Chancenungleichheit, die selbst dann wirken, wenn alle in der inneren Organisation der Schule wirksamen Mechanismen der Ungleichheit beseitigt sind und volle Leistungsgerechtigkeit in der Schule hergestellt ist. Solche Mechanismen können nur mit komplizierten statistischen Verfahren aufgedeckt werden; so ergab sich z.B., daß die Benotung der Lehrer mit der Sozialschicht nur noch insignifikant positiv korrelierte, sobald aber in einer Kovarianzanalyse die Intelligenzunterschiede zwischen den Schichten kontrolliert waren, wiederum signifikant positiv zusammenhing (Bosse u.a. 1972:162-182). Praktische Folgerungen betreffen den Ausgleich des in der Frühsozialisation wirksamen Einflusses sozialer Schichtzugehörigkeit auf die kognitiven und motivationalen Voraussetzungen schulischen Lernens durch vorschulische Programme. Sie sind bis

heute nur modellhaft realisiert.¹⁾

Die dritte Ebene der Analyse bezieht sich auf die ungleichen Voraussetzungen, die die Kinder als Mitglieder verschiedener sozialer Kulturen für die Aneignung der herrschenden Kultur mitbringen. Diese Analyseebene bezieht sich auf sozial typische Interessen und Erfahrungen und auf sozial typische Werte, Normen und Deutungen. Sie behandelt Mechanismen, die selbst bei Herstellung der Leistungsgerechtigkeit und der Schaffung gleicher kognitiver und motivationaler Bedingungen sozialen Handelns und schulischen Lernens über die ganze Schullaufbahn wirksam bleiben. Solche Mechanismen müßten in einer Theorie der sozialen Konstitution und Organisation von Bildungsinhalten genauer beschrieben werden; sie sind empirisch-wissenschaftlich kaum untersucht worden. Praktische Folgerungen betreffen die "innere Schulreform", die keineswegs eine innere Angelegenheit der Schule ist; als Postulat für ein Schulwesen, das auf dieser Ebene nicht sozial diskriminiert, kann man angeben, daß die Inhalte schulischen Lernens nur auf solchen Ebenen differenziert werden dürfen, die weder manifest noch latent mit sozialem Prestige belastet

¹⁾-----
Eine obligatorische Einrichtung der Vorschule vor der Reform des dreigliedrigen Schulwesens hätte jedoch eine Verstärkung der sozialen Legitimation von Ungleichheit zur Folge: Gegen die nachgewiesenen auf außerschulischen Kriterien ruhende Selektion zur weiterführenden Schule kann nur noch schwer ein Einwand erhoben werden, wenn "schon in der Vorschule alle optimal gefördert werden", wie sich aufgrund der Ergebnisse zur frühkindlichen Determination der Intelligenz ja jederzeit mit wissenschaftlicher Legitimation behaupten läßt. In einer spezifischen Weise ergäbe sich genau jene Konsequenz, die von den Kritikern der Chancengleichheit generell zugeschrieben wird: Differenzen des Schulerfolgs werden, obwohl sozial produziert, Individuen attribuiert und damit sozial legitimiert (Bourdieu/Passeron 1971). Zweierlei läßt sich daraus ersehen: zum ersten, daß die Strategie von Bildungsreformen unter dem Postulat der Chancengleichheit einer Logik folgen muß, die nicht mit der zeitlichen Folge von Schulstufen identisch ist; und zweitens, daß verschiedene sich gleichermaßen auf "Chancengleichheit" berufende Strategien der Bildungsreform nach ihrem Beitrag zur Legitimation von Chancenungleichheit bewertet werden können.

sind, und daß über die sozial-natürwüchsig legitimierten Bildungsinhalte politisch-administrativ entschieden werden kann. Wieweit diese Postulate nur idealtypische Fiktionen und wieweit sie realisierbare Modelle bezeichnen, müßte ebenfalls in einer Theorie der sozialen Konstitution und Organisation von Bildungsinhalten geprüft werden.

Die Ergebnisse unserer Untersuchungen sind für die Erklärung der Zusammenhänge auf der zweiten Ebene bedeutsam. Diese Zusammenhänge selber sind- wie wir gezeigt haben (siehe I.2.2.) - durch andere Untersuchungen ausreichend abgesichert. Es geht also nicht um einen Beleg für die Tatsache der Chancenungleichheit, sondern um eine Erklärung ihrer Mechanismen. Unser Experiment ist für eine Erklärung der Benachteiligung der Unterschichtkinder in der kognitiven Entwicklung durch ihre schichttypische Strategie des Sprachgebrauchs und den Nachweis der in dem Prozeß der Evokation wirksamen Mechanismen auf der Seite der Eltern wie Lehrer wichtig; es ist für die Errichtung vorschulischer Programme insofern bedeutsam, als wir in unserer Gruppe 10jähriger Kinder ja die langfristige Folge schichtspezifischer Strategien des Sprachgebrauchs prüfen wollen. Daß die Hypothesen nicht bestätigt werden konnten, bedeutet das Scheitern einer Erklärung, nicht aber eine Widerlegung der Chancenungleichheit.

Als inhaltliche Begründung für das Scheitern unserer Hypothesen haben wir dargelegt, daß durch intentionale Elemente familialer und schulischer Erziehung mögliche Wirkungen schichtspezifischer Arten des Sprachgebrauchs nivelliert worden sind. Ohne experimentelle Kontrollen bestehen signifikante Unterschiede in der Wortkenntnis zwischen Kindern der Unter- und der Mittelschicht - wir verweisen noch einmal auf die Ergebnisse der ersten Messung; daß aber bei gleichen Ausgangskenntnissen Kinder beider Schichten gleich gute Worterklärungen liefern, scheint uns ein Ergebnis intentionaler Erziehung zu sein, das durchaus mit größeren Kosten für die Unterschichtkinder verbunden sein kann. Diese Begründung der negativen Ergebnisse streicht die Bedeutung heraus, die eine Theorie sozialer

Konstitution und Organisation von Bildungsinhalten als Alternative zur Theorie linguistischer Kodes bei der Erklärung sozialer Chancenungleichheit hat.

Wir haben zu Beginn dieses Berichts (siehe I.1.) gesagt, daß praktisch-pädagogische Konsequenzen sich nicht allein aus der Bestätigung von Einzelhypothesen, sondern auch aus dem Stellenwert an sich praktisch belangloser Hypothesen in einem breiteren theoretischen Rahmen ergeben können. Eine solche indirekt-systematisch abgeleitete Konsequenz für die pädagogische Praxis hätte sich aus der simultanen Bestätigung unserer Hypothesen ergeben, daß der aktive Wortgebrauch, nicht aber das passive Wortverständnis von schichtspezifischen Strategien des Sprachgebrauchs abhängt. Der aktive Wortgebrauch hätte dann als der Angriffspunkt, das passive Wortverständnis als der Fixpunkt kompensatorischer Maßnahmen angesehen werden können. Da keine unserer Hypothesen bestätigt wurden, können wir nur noch einige an Einzelbeobachtungen der Untersuchung der sprachlichen Mittel gewonnene praktische Empfehlungen nachtragen. Am wichtigsten ist der Hinweis, daß der Evokationsmechanismus auch über die Beurteilung konventionell festgelegter Formen laufen kann: Grammatikfehler, Schreibfehler im Schriftlichen und stockender Ausdruck im Mündlichen überwiegen bei den Unterschichtkindern. Fernerhin ist der Hinweis wichtig, daß, wenn Kriterien des grammatisch frei wählbaren Ausdrucks im Spiele sind, sie nicht nach Vorurteilen einer volgarisierten Bernstein- oder Whorf-Theorie organisiert sind: gerade die Unterschichtkinder gebrauchten häufiger die "elaborierten" sprachlichen Formen.

Verzeichnis der benutzten und zitierten Literatur

- Allert, T./v.Grote, C./Lappe, L. (1972) Soziologische Diplom-Arbeit. Frankfurt. J.-W.-Goethe-Universität. FB 3.
- Anger, H./Bergmann, R./Hylla, E. (1954) Frankfurter Wortschatztest. Hochschule für Internationale Pädagogische Forschung.
- Anger, H./Bergmann, R. (1954) Entwicklung, Analyse und Eichung des Frankfurter Wortschatztests. In: Z.Exp. Angew. Psych. 2, 51-90.
- Arnold, G./Bower, G.H./Bobrow, S.A. (1972) Mediated Semantic Compatibility in Semantic Learning. In: J. Verbal Learning Verbal Behavior 11, 239-243.
- Bandura, A./Walters, R.H. (1963) Social Learning and Personality Development. New York (Holt, Rinehart, Winston)
- Bernstein, B. (1958) Some Sociological Determinants of Perception. In: Bernstein (1971b)
- Bernstein, B. (1959) A Public Language: Some Sociological Implications of a Linguistic Form. In: Bernstein (1971b)
- Bernstein, B. (1960) Language and Social Class. In: Bernstein (1971b)
- Bernstein, B. (1962a) Linguistic Codes, Hesitation Phenomena and Intelligence. In: Bernstein (1971b).
- Bernstein, B. (1962b) Social Class, Linguistic Codes and Grammatical Elements. In: Bernstein (1971b).
- Bernstein, B. (1965) A Socio-Linguistic Approach to social Learning. In: Bernstein (1971b).
- Bernstein (1969) A Critique of the Concept of Compensatory Education. In: Bernstein (1971b)
- Bernstein, B. (1971a) Social Class, Language and Socialisation. In: Bernstein (1971b)
- Bernstein, B. (1971b) Class, Codes and Control. Volume I. Theoretical Studies towards a Sociology of Language. London (Routledge & Kegan Paul)
- Bernstein, B. (ed.) (1973) Class, Codes and Control. Volume II. Applied Studies towards a Sociology of Language. London (Routledge & Kegan Paul)
- Bernstein, B. (1970) History of the Research. In: Brandis/Henderson (1970)
- Bernstein, B./Henderson (1961) Social Class differences in the Relevance of Language to Socialisation. In: Bernstein, B. (1973)
- Fosse, S./Kieper, M./Schmitt, M./Wienskowsky, P. (1971) Sozio-ökonomische Klassenlage, schichtenspezifisches Erziehungsmilieu und schulische Faktoren in ihrem Einfluß auf Intelligenz und Schulerfolg. Soziologische Diplom-Arbeit. J.-W.-Goethe-Universität. FB 3.
- Bourdieu, P. (1970) Zur Soziologie der symbolischen Formen. Frankfurt (Suhrkamp)

- Bourdieu, P./Passeron, J. (1971) Die Illusion der Chancengleichheit. Stuttgart (Klett).
- Bourdieu, P./Passeron, J. (1973) Grundlagen einer Theorie der symbolische Gewalt. Frankfurt (Suhrkamp).
- Brandis, W./Henderson, D. (1970) Social Class, Language and Communication. London (Routledge & Kegan Paul).
- Brown, R. (1956) Language and Categories. Appendix zu: J.S. Bruner/J.J. Goodnow/G.A. Austin (1956) A Study of Thinking. New York.
- Brown, R. (1957) Linguistic Determinism and the Parts of Speech In: J. Abn. Soc. Psychol. 55, 1-5.
- Brown, R. (1958) How shall a Thing be Called? In: Psychol. Rev. 65, 14-21.
- Brown, R./Berko, J. (1960) Word Association and the Acquisition of Grammar. In: Child Development 31, 1-14.
- Bruner, J. et al. (1971) Studien zur kognitiven Entwicklung. Stuttgart (Klett). Amerikanisches Original 1966.
- Bühler, H. (1972) Sprachbarrieren und Schulanfang. Eine praxisinguistische Untersuchung des Sprechens von sechs- bis achtjährigen. Weinheim (Beltz).
- Carson, A.S./Rabin, A.I. (1960) Verbal Comprehension and Communication in Negro and White Children. In: J. Educat. Psychol. 51, 47-51.
- Cazden, C.B. (1966) Subcultural Differences in Child Language: An Interdisciplinary Review. In: Merrill-Palmer-Quarterly 12, 185-219.
- Chodorkoff, B./Mussen, P. (1952) Qualitative Aspects of the Vocabulary of Normals and Schizophrenics. In: J. Consult. Psychol. 16, 43-48.
- Chomsky, C. S. (1969) The Acquisition of Syntax in Children from 5 to 10. Cambridge, Mass. (M.I.T.).
- Chomsky, N.A. (1969) Aspekte der Syntax-Theorie. Frankfurt (Suhrkamp). Amerikanische Erstausgabe 1965.
- Cook-Gumperz, J. (1973) Social Control and Socialization. A Study of Class Differences in the Language of Maternal Control. London (Routledge & Kegan Paul).
- Dahrendorf, R. (1963) Industrie- und Betriebssoziologie. Stuttgart (Th. Gabler)
- Deutsch, M. (1965) The Role of Social Class in Language Development and Cognition. In: Deutsch, M. and ass. (1967)
- Deutsch, M. (1967) and Associates: The Disadvantaged Child. Selected Papers. New York.
- Deutsch, M./Levinson, M./Brown, B.R./Cherry Peisach, E. (1964) Communication of Information in the Elementary School Classroom. Institute for Developmental Studies, Department of Psychiatry, New York Medical College, New York. Cooperative Research Project nr. 908.
- v. Dewitz/Flaake, K./Hartel, U./Schweissfurth, H. (1972) Soziologische Diplom-Arbeit. Frankfurt. J.-W.-Goethe-Universität. FB 3.

- Dornseiff, F. (1959) Der deutsche Wortschatz nach Sachgruppen; 5. Auflage. Berlin (de Gruyter).
- Entwisle, D. (1966) Developmental Sociolinguistics: A comparative Study in four Subcultural Settings. In: Sociometry 67.
- Entwisle, D./Forsyth, D.F./Muus, R. (1964) The Syntagmatic-Paradigmatic Shift in Children's Associations. In: J. Verb. Learn. Verb. Behavior 3, 19-29.
- Feifel, H./Lorge, I. (1950) Qualitative Differences in the Vocabulary Responses of Children. In: J. Educat. Psychol. 41, 1-18.
- Findlay, D.C./McGuire, C. (1957) Social Status and Abstract Behavior. In: J. Abn. Soc. Psychol. 54, 135-137.
- Francescato, G. (1973) Sprachstruktur und Spracherwerb beim Kinde. Stuttgart (Klett).
- Gloy, K. (1973) Zwei neuere Arbeiten der empirischen Soziolinguistik - eine Methodenkritik. In: Neue Sammlung, 85-99.
- Goffmann, E. (1970) Wir alle spielen Theater (Übersetzung von: The Presentation of Self in Everyday Life) München (Piper).
- Grimm, S. (1966) Die Bildungsabstinenz der Arbeiter. München (Barth).
- Hays, W.S. (1963) Statistics for Psychologists. New York.
- Hawkins, P.R. (1973) Social Class, the Nominal Group and Reference In: Bernstein, B. (ed.) (1973)
- Hess, R./Shipman, V. (1965) Early Experience and the Socialization of Cognitive Modes in Children. In: Child Development 36, 869-886.
- Hess, R./Shipman, V. (1967) Cognitive Elements in Maternal Behavior. In: Hill, J.P. (ed.) (1967) Minnesota Symposium on Child Psychology, Vol. I. Minneapolis.
- Henderson, D. (1970a) Social Class Differences in Form-Class Usage among five-year-old Children. In: Brandis/Henderson (1970).
- Henderson, D. (1970b) Contextual Specificity, Discretion, and Cognitive Socialization: With Special Reference to Language. In: Bernstein, B. (ed.) (1973).
- Hoffmann, M.L. (1960) Power Assertion and its Impact on the Child. In: Child Development 31, 129-143.
- Horn, W. (1969) Prüfungssystem für Schul- und Bildungsberatung (P-3-B) Göttingen (Hogrefe).
- Hurlburt, D. (1954) The Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meanings - Nouns, Verbs, Adjectives. In: J. Educat. Res. 47, 561-576.
- Jaeger, S. u.a. (1973) Sprechen und soziale Schicht. Frankfurt (Athenäum).
- Jencks, Ch. u.a. (1973) Chancengleichheit. Hamburg (Rowohlt).
- John, V.P. (1963) The Intellectual Development of Plum Children: Some Preliminary Findings. In: Amer. J. Orthopsychiatry 33, 813-822.
- John, V.P./Goldstein, L.S. (1964) The Social Context of Language Acquisition. In: Merrill-Palmer-Quarterly 10, 265-275.

- Katz, J.J./Fodor, J.A. (1963) The Structure of a Semantic Theory. In: *Language* 39, 170-210.
- Kruglov, I.P. (1953) Qualitative Differences in the Vocabulary Choices of Children as Revealed in a Multiple Choice Test. In: *J. Educat. Psychol.* 44, 229-243.
- Labov, W. (1966) Hyper-Correction by the Lower-Middle Class as a Factor in Linguistic Change, in: Bright, W. (ed.) (1966) *Sociolinguistics*. The Hague.
- Labov, W. (1970) The Study of Language in its Social Context. In: *Studium Generale* 23, 30-87.
- Latscha, F. (1966) Der Einfluß des Primarlehrers. In: Hess, F./Latscha, F./Schneider, W. (1966) *Die Ungleichheit der Bildungschancen. Soziale Schranken im Zugang zur Höheren Schule*. Olten/Freiburg (Walter).
- Lawton, D. (1970) *Soziale Klasse, Sprache und Erziehung*. Düsseldorf (Schwann). Engl. Ausgabe: *Social Class, Language and Education*. London 1968.
- Lenneberg, E.H. (1967) *Biological Foundations of Language*. New York (Wiley).
- McCarthy, D. (1954) Language Development in Children. Kap. 9 in: Carmichael, L. (ed.) (1954) *Manual of Child Psychology*, 2. Auflage, New York.
- McNeill, D. (1966) Developmental Psycholinguistics. In: Smith, F./Miller, G.A. (eds.) (1966) *The Genesis of Language. A Psycho-Linguistic Approach*. Cambridge, Mass. (M.I.T.)
- Mead, G.H. (1934) *Mind, Self, and Society*. Chicago.
- Menyuk, P. (1963) Syntactic Structures in the Language of Children. In: *Child Development* 34, 407-422.
- Menyuk, P. (1971) *The Acquisition and Development of Language*. Englewood Cliffs, New Jersey.
- Miller, D.R./Swanson, G.E. (1960) *Inner Conflict and Defense*. New York (Holt).
- Oevermann, U. (1966) Soziale Schichtung und Begabung. In: *Z. Päd. Beiheft* 6.
- Oevermann, U. (1968) Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse. In: Roth, H. (Hg.) (1968) *Begabung und Lernen*, Bd. 4 der Gutachten und Studien der Bildungskommission. Stuttgart (Klett).
- Oevermann, U. (1970) *Sprache und Soziale Herkunft. Ein Beitrag zur Analyse schichtenspezifischer Sozialisationsprozesse und ihrer Bedeutung für den Schulerfolg*. Berlin 1970 (edition suhrkamp, Frankfurt 1973).
- Oevermann, U./Krappmann, L./Kreppner, K. (1968) *Elternhaus und Schule. Projektvorschlag: Institut für Bildungsforschung Berlin*. Manuskript.
- Oevermann, U./Krappmann, L./Kreppner, K. (1972) *Forschungsbericht Sozialdaten. Institut für Bildungsforschung Berlin*. Manuskript.

- Oevermann, U./Krappman, L./Kreppner, K. (1973) Bemerkungen zur Diskussion der sogenannten "Kode"-Theorie. In: Linguistische Berichte 23, 59-69.
- Nelson, K. (1973) Structure and Strategy in Learning to Talk. Monographs of the Society for Research in Child Development, Serial No, 149, Vol. 38, nos. 1-2.
- Niepold, W. (1970) Sprache und soziale Schicht. Berlin (Spiess).
- Parsons, T./Bales, R. (1953) Family, Socialisation, and Interaction Process. New York (Free Press).
- Piaget, J. (1969) Nachahmung, Spiel und Traum. Stuttgart (Klett). Frz. Erstausgabe: La Formation du Symbole chez l'enfant. Neuchatel 1959.
- Peisert, H. (1967) Soziale Lage und Bildungschancen in Deutschland. München (Piper).
- Popitz, H. (1965) Die Ungleichheit der Chancen im Zugang zur höheren Schulbildung. In: Friedeburg, L.v. (Hg.) (1965) Jugend in der modernen Gesellschaft. Köln (Kiepenheuer & Witsch).
- Preuß, O. (1972) Soziale Herkunft und die Ungleichheit der Bildungschancen. Eine Untersuchung über das Eignungs-urteil des Grundschullehrers. Weinheim (Beltz).
- Rabin, A.I./King, G.F./Ehrmann, J.C. (1955) Vocabulary Performance of Short-Term and Long-Term Schizophrenics. In: J. Abn. Soc. Psychol. 50, 255-258.
- Robinson, W.P. (1973) Where do Children's Answers come from? In: Bernstein, B. (1973).
- Robinson, W.P./Rackstraw, S. (1972) A Question of Answers. Vol. I. London (Routledge & Kegan Paul).
- Rolff, H.G. (1967) Sozialisation und Auslese durch die Schule. Heidelberg (Quelle & Meyer).
- Rosenthal, R./Jacobson, L. (1971) Pygmalion im Unterricht. Weinheim (Beltz). Amer. Erstausgabe 1968.
- Russel, D.H./Saadeh, I.O. (1962) Qualitative Levels in Children's' Vocabularies. In: J. Educat. Psychol. 53, 170-174.
- Siegel, S. (1956) Nonparametric Statistics for the Behavioral Sciences. New York (McGraw-Hill).
- Siegel, I.W./Anderson, S.M./Shapiro, H. (1966) Categorization Behavior of Lower and Middle Class Negro Pre-school Children: Differences in Dealing with Familiar Objects. In: J. Negro Educat. 6, 218-229.
- Siller, J. (1957) Socio-Economic Status and Conceptual Thinking. In: J. Abn. Soc. Psychol. 55, 365-371.
- Sommer, M. (1972) Sozio-ökonomische Faktoren in der Sprache sechsjähriger Schulanfänger. Ratingen (Henn).
- Steinkamp, G. (1957) Die Rolle des Volksschullehrers im schulischen Selektionsprozeß. In: Hamburger Jahrbuch für Wirtschafts- und Gesellschaftspolitik 12, 302-324.
- Strodtbeck, F.T. (1967) The hidden Curriculum of the Middle Class Home. In: Passow, A.H./Goldberg, M./Tannenbaum, A. (eds.) (1967) Education of the Disadvantaged. New York (Holt, Rinehart & Winston), 244-260.

- Templin, M.C. (1957) Certain Language Skills in Children. Minneapolis.
- Thomas, D.R. (1962) Oral Language Sentence Structure and Vocabulary of Kindergarten Children Living in Low Socio-Economic Urban Areas. Detroit, Michigan. Dissertation: Wayne State University.
- Weinreich, U. (1970) Erkundungen zur Theorie der Semantik. Tübingen. Amer. Erstausgaben: Explorations in Semantic Theory. In: Current Trends in Linguistics, Vol. III. Den Haag (Mouton) 1966, 395-477.
- Werner, H./Kaplan, B. (1963) Symbol Formation. London.
- Wiederhold, K.A. (1971) Kindersprache und Sozialstatus. Eine empirisch-pädagogische Untersuchung zum Sprachstand und Sprachwandel bei Kindern des ersten und zweiten Schuljahres. Ratingen (Henn).
- Williams, F./Naremore, R. (1969) On the Functional Analysis of Social Class Differences in Modes of Speech. In: Speech Monographs XXXVI, 2, 76-102.
- Winer, B.J. (1962) Statistical Principles in Experimental Design. New York (McGraw-Hill).
- Whiteman, M./Brown, B./Deutsch, M. (1967) Some Effects of Social Class and Race on Children's Language and Intellectual Abilities. In: Deutsch, M. and Associates (1967)
- Whiteman, M./Deutsch, M. (1967) Social Disadvantage as Related to Intellectual and Language Development. In: Deutsch, M. and Associates (1967)
- Wygotsky, L.S. (1969) Denken und Sprechen. Frankfurt (Fischer). Russische Erstausgabe 1934.

Nachtrag

- Brown, R./Bellugi, U. (1964) Three Processes in the Child's Acquisition of Grammar, in: R. Brown (1970)
- Brown, R./Cazden, C./ Bellugi, U. (1969) The Child's Grammar from I to III, in: R. Brown (1970)
- Brown, R./Fraser, C./Bellugi, U. (1963) Control of grammar in Imitation, Comprehension and Production, in: R. Brown (1970)
- Brown, R. (1970) Psycholinguistics. New York (Free Press)
- Furth, H.G. (1966) Thinking without Language. New York (Free Press)
- Kohn, M. (1969) Class and Conformity. Homewood, Ill. (Dorsey)
- Lyons, J. (1968) Introduction to Theoretical Linguistics. Cambridge.
- Shultz, Th./Pilon, R. (1973) Development of the Ability to Detect Ambiguity, in: Child Development 44, 728-734.