

**Ein experimenteller Vergleich von
Verfahren zur Bestimmung
individueller Begriffsstrukturen**

Harald Binder

**Diplomarbeit als Teil der
Diplomprüfung für Psychologen**

Universität Regensburg

21. Juli 2000

<i>INHALTSVERZEICHNIS</i>	2
---------------------------	---

Inhaltsverzeichnis

1 Einleitung	4
2 Der Begriff in der psychologischen Forschung	8
2.1 Definitionen von „Begriff“	8
2.2 Ansätze zur Struktur von Begriffen	9
2.2.1 Der klassische Ansatz	9
2.2.2 Die Prototypentheorie	12
2.2.3 Die formale Begriffsanalyse	14
2.3 Probleme der bisherigen psychologischen Ansätze	20
2.4 Verfahren zur Begriffsstruktur-Bestimmung	21
3 Ein meßtheoretischer Ansatz	23
3.1 Theoretische Vorüberlegungen	23
3.2 Der Ansatz von Heller	25
3.2.1 Formaler Rahmen	25
3.2.2 Anwendung in der Begriffsforschung	29
4 Befragungsverfahren	36
4.1 Komplettbefragung	37
4.2 QUERY	41
4.3 PSQUERY	44

<i>INHALTSVERZEICHNIS</i>	3
5 Experimenteller Vergleich	47
5.1 Fragestellung	47
5.2 Methode	49
5.2.1 Befragungssoftware	49
5.2.2 Ablauf der Befragung	53
5.2.3 Stichprobe	54
5.2.4 Unterschiedsmaß	54
5.2.5 Inferenzstatistik	57
5.3 Ergebnisse	59
5.3.1 Deskriptive Statistiken	59
5.3.2 Überblick über die Begriffsstrukturen	61
5.3.3 Darstellung ausgewählter Begriffsstrukturen	74
5.3.4 Hypothesentest	81
5.4 Diskussion	82
5.4.1 Kennwerte des PSQUERY-Verfahrens	82
5.4.2 Folgerungen	85
6 Zusammenfassung	91
Literaturverzeichnis	92
A Instruktion	96
B Deskriptive Statistiken aller Versuchspersonen	99

1 Einleitung

Die psychologische Forschung beschäftigt sich schon lange mit individuellen Begriffsstrukturen. So erforschte z.B. schon Hull (1920) die Bildung von Begriffen. Um die praktische Notwendigkeit zu illustrieren, sich mit der Individuumsspezifität von Begriffen und Verfahren zur Bestimmung dieser individuellen Strukturen zu beschäftigen, soll hier ein Anwendungsbeispiel vorgestellt werden. Ein eine Situation, in der man sich für individuelle Begriffsstrukturen interessieren könnte, ist z.B. folgende:

Ein Student im Grundstudium Psychologie hat eine Statistikvorlesung mit einer zugehörigen Übungsveranstaltung absolviert. Wie gut er die unterrichtete Materie beherrscht wird typischerweise mit Hilfe von schriftlichen Prüfungen getestet. Bei diesen sind verschiedene Aufgaben zu lösen, die in ihrem Spektrum möglichst die Lehrinhalte der Veranstaltung abdecken sollten. Das erfolgreiche Verständnis des Materials wird durch die Anzahl der gelösten Aufgaben operationalisiert. Dieser Punktwert eines Studenten wird meist nur für eine „bestanden oder nicht bestanden“-Entscheidung verwendet, die meist an einer minimal zu erreichenden Punktzahl festgemacht wird.

Manchmal ist es allerdings notwendig, genauere Informationen über das Wissen bzw. das Verständnis des Studenten zu bekommen. Besteht er z.B. eine Prüfung nicht und sucht einen Nachhilfelehrer auf, so muß dieser sich erst genau über den Wissens- bzw. Verständniszustand des Studenten klar werden, um anschließende Lerneinheiten planen zu können. Eine ähnliche Situation ergibt sich auch bei tutoriellen computergestützten Lernsystemen, die nach der Präsentation einer Lehreinheit immer den Wissens- bzw. Verständniszuwachs ermitteln müssen, um anschließend adaptiv weiteres Material präsentieren zu können.

Im Rahmen einer solchen genaueren Diagnose wäre es z.B. wünschenswert, zu wissen, ob der Student wichtige Begriffe aus dem Gegenstandsbereich richtig einordnen und richtig zueinander in Beziehung setzen kann. Der Beitrag der psychologischen Begriffsforschung könnte hier darin bestehen, eine geeignete Theorie über die Struktur individueller Begriffe zur Verfügung zu stellen und darauf aufbauend Mittel zur Bestimmung dieser individuellen Begriffsstrukturen zu bieten.

Die im Rahmen dieser Arbeit vorgestellten Konzepte und Theorien sollen also vorrangig unter dieser diagnostischen Perspektive daraufhin betrachtet werden, inwieweit sie derartigen Ansprüchen gerecht werden.

Doch auch schon bevor sich die psychologische Forschung mit individuellen Begriffsstrukturen beschäftigt hat, wurden in literarischer Form die Individuumsspezifität von Begriffen und die sich daraus möglicherweise ergebenden Probleme thematisiert.

Das folgende Textstück stammt aus dem (Kinder-)Buch „Through the Looking-Glass“ von dem Mathematiker und Logiker Lewis Carrol und soll zeigen, daß in der Belletristik das Thema der Individuumsspezifität von Begriffen manchem Autor schon für Kinder als relevant und aufbereitenenswert erschien. Das Textstück gibt einen Dialog zwischen den Charakteren Alice und Humpty Dumpty wieder.

“There’s glory for you!”

‘I don’t know what you mean by ”glory”,’ Alice said.

Humpty Dumpty smiled contemptuously. ‘Of course you don’t – till I tell you. I meant “there’s a nice knock-down argument for you!”’

‘But “glory” doesn’t mean “a nice knock-down argument”,’ Alice objected.

‘When I use a word,’ Humpty Dumpty said, in rather a scornful tone,
‘it means just what I choose it to mean – neither more nor less.’

‘The question is,’ said Alice, ‘whether you can make words mean so
many different things.’

Die Hauptfigur „Alice“ des Buchs reist durch eine bizarre Welt, in der Gesetzmäßigkeiten auf den Kopf gestellt sind und viele übliche Konventionen nicht mehr gelten. So auch in der hier zitierten Episode, in der die Figur Humpty Dumpty sich beim Gebrauch von Begriffen nicht durch die üblichen Definitionen dieser festlegen lassen will.

Lewis Carrol weist hier den (kindlichen) Leser auf ein Problem hin, das oft übersehen wird: Auch ganz alltägliche Begriffe werden nicht immer in exakter, allgemeingültiger Weise verwendet. Individuen richten sich, wenn sie Begriffe benutzen, meist eben nicht nach Lexika sondern verwenden ihre eigenen „Definitionen“, wenn auch meist nicht so willkürlich wie die Figur Humpty Dumpty in der zitierten Passage.

Doch wie sehen solche individuellen „Definitionen“ aus? Wie ist das „subjektive Lexikon“ eines Menschen, wie es Fillenbaum und Rapoport (1971) nennen, organisiert? Die Antwort auf derartige Fragen soll hier nicht in der Belletristik gesucht werden. Eine systematische Untersuchung individueller Begriffsstrukturen gibt es, wie schon angedeutet, in der psychologischen Forschung. Hier gibt es eine lange Tradition, die sich mit dieser Thematik beschäftigt hat.

Es gibt viele Forschungsrichtungen innerhalb der psychologischen Begriffsforschung und selbst wenn man für eine Überblicksdarstellung eine Auswahl der vorzustellenden Ansätze getroffen hat, so lassen sich diese unter verschiedenen Gesichtspunkten betrachten. Im Rahmen dieser Arbeit gilt das Interesse,

wie mit dem einleitenden Beispiel schon angedeutet, vor allem dem diagnostischen Einsatz der Ergebnisse der Begriffsforschung.

Zunächst aber ein kurzer Überblick über den Aufbau dieser Arbeit:

In Abschnitt 2 dieser Arbeit wird der „klassische Ansatz“, die „Prototypentheorie“ nach Rosch und die Kritik an diesen dargestellt. Mit dem Ansatz von Ganter und Wille werden Möglichkeiten zur mathematischen Formalisierung von Begriffsstrukturen eingeführt. Auf die angeführten psychologischen Ansätze aufbauend werden schließlich Verfahren zur Bestimmung von individuellen Begriffsstrukturen besprochen. Als Alternative zu den genannten Herangehensweisen wird in Abschnitt 3 der Ansatz von Heller vorgestellt. Im Rahmen dieses Ansatzes werden schließlich Befragungsverfahren zur Bestimmung individueller Begriffsstrukturen erläutert (Abschnitt 4) und zwei von diesen auf experimenteller Basis verglichen (Abschnitt 5).

2 Der Begriff in der psychologischen Forschung

2.1 Definitionen von „Begriff“

Die Frage, was man unter einem Begriff zu verstehen hat, ist das Erste, was im Rahmen der Erforschung individueller Begriffsstrukturen zu klären ist.

Eine Definition von „Begriff“, die sehr wenige Vorannahmen erfordert, verwendet z.B. Hull (1920). Er bezeichnet als Begriff eine durch Reizgeneralisierung gelernte gemeinsame Reaktion (response) auf verschiedene Reize. Diese Herangehensweise wird auch oft als „behavioristisch“ bezeichnet, da sie von einem Stimulus-Response-Paradigma ausgeht und keine weiteren Annahmen über kognitive Gegebenheiten trifft.

Laurenz und Margolis (1999), bezeichnen als „Begriff“ „subpropositionale mentale Repräsentationen“. Bei dieser Definition handelt es sich um eine negative in dem Sinne, daß nichts direkt über den „Begriff“ ausgesagt wird. Es wird nur festgelegt, daß alles, was (noch) nicht als „Proposition“ zu bezeichnen ist, „Begriff“ heißen soll. Die Autoren begründen eine derartige Definition damit, daß diese nicht nur den Ansprüchen für die Beschäftigung mit einfachen Begriffen wie z.B. „Brot“ genügen würde. Auch komplexe mentale Konstrukte wie z.B. der Begriff „Freiheit“ würden damit erfaßt. Problematisch bei dieser Definition ist allerdings, daß Begriffe wie „Proposition“ nicht weiter definiert werden. Die Definition stützt sich also auf möglicherweise umfangreiche, nicht definierte Konstrukte. Aus diesem Grund wird später in dieser Arbeit ein Ansatz gewählt, der von derartigen Definitionen und Annahmen nicht abhängt.

Weiterhin werden in dieser Arbeit die Bezeichnungen „Klasse“ und „Kategorie“ synonym mit „Begriff“ verwendet, vor allem wenn Probleme der Klassifizierung, d.h. die Frage ob ein Objekt unter einen Begriff fällt, betrachtet werden.

Wenn von verschiedenen Definitionen von „Begriff“ gesprochen wird, so ist es wichtig, zu bemerken, daß diese unterschiedlich umfangreiche bzw. starke Annahmen in Bezug auf die mentale Repräsentation von Begriffen machen. Für den Ansatz von Heller (1991, 2000), auf den sich diese Arbeit später beziehen wird, genügt die oben angeführte „behavioristische“ Definition als Grundlage. Weitergehende kognitive Konstrukte müssen bei dieser Herangehensweise nicht zwingend als existent angenommen, sie werden aber auch nicht ausgeschlossen.

Zunächst sollen aber Ansätze aus der kognitiven Psychologie dargestellt werden, die sich auf umfangreichere konzeptionelle Annahmen über Begriffe, d.h. auch auf umfangreichere Definitionen, stützen. Die Basis bildet in diesem Bereich der sogenannte „klassische Ansatz“. Beispielhaft für die Erweiterungen dieses Ansatzes und die neuere Forschung wird die „Prototypentheorie“ nach Rosch erläutert. Korrespondierend zum „klassischen Ansatz“ wird noch die „Formale Begriffsanalyse“ nach Ganter und Wille vorgestellt, die eine mathematische Formalisierung bietet, auf die z.B. der später darzustellenden Ansatz von Heller zurückgreift.

2.2 Ansätze zur Struktur von Begriffen

2.2.1 Der klassische Ansatz

Der sogenannte klassische Ansatz hat in der in der Geschichte der Philosophie, der Linguistik und der Psychologie eine lange Tradition. So beschäftigte sich schon Aristoteles mit der Bildung von Kategorien und Definitionen von Begriffen (siehe z.B. Röd, 1994).

Allerdings gibt es nicht *den* „klassischen Ansatz“. Dieser ist eher als Kernidee einer ganzen Reihe von Theorien und Schulen zu sehen. Die Anhänger dieser

gehen davon aus, daß Begriffe in Form von notwendigen und hinreichenden Bedingungen gespeichert sind. Diese Vorstellung läßt sich wohl noch am besten mit dem Begriff „subjektives Lexikon“ in Einklang bringen, da die postulierte Struktur einer Definition einem Lexikon nicht unähnlich ist. Der Erwerb von Begriffen besteht demnach aus einem Erlernen der Definitionskriterien die für eine alltagstaugliche Repräsentation eines Begriffs notwendig sind.

Der Vorteil solche Repräsentation ist es, daß sie es erlaubt, in jeder Situation bzw. für jedes Objekt zu überprüfen, ob ein bestimmter Begriff angewendet werden kann bzw. ob eine Objekt in eine bestimmte Klasse gehört. Da derartige Klassifikationsleistungen im menschlichen Alltag häufig auftreten, ist es wichtig, daß eine psychologische Theorie diese erklären kann.

Als Beispiel für die Definition eines Begriffes beschreibt der Linguist J. Katz (1972, S. 40) die Struktur des Begriffs „chair“:

„The English noun „chair“ can be decomposed into a set of concepts which might be represented by semantic markers in (4.10):

(4.10) object, physical, non-living, artifact, furniture, portable, something with legs, something with a back, something with a seat, seat for one.“

Die zur Definition verwendeten Begriffe lassen sich wiederum durch (elementarere) Begriffe definieren. Diese Zerlegung geht so weit, bis man auf der Ebene elementarer Sinneseindrücke angekommen ist. Dieses Postulat der Reduzierbarkeit auf Sinneseindrücke findet sich aber nicht in allen Varianten des klassischen Ansatzes. Es wird aber grundsätzlich von der Existenz von nicht mehr weiter reduzierbaren Grundbegriffen ausgegangen, die aber nicht unbedingt mit Sinneseindrücken gleichzusetzen sind.

Für das eben dargestellt Beispiel und viele weitere ist der klassische Ansatz sehr einleuchtend. Es gibt allerdings auch viele Ergebnisse der Begriffsforschung, die mit ihm nicht oder nur schwer erklärt werden können. Beispielsweise können Individuen, wenn sie nach Definitionen für Begriffe gefragt werden, für manche keine (expliziten) Definitionen angeben, die eine klare Abgrenzung ermöglichen würden. Ein Beispiel hierfür ist der Begriff „Liebe“, für den es sicherlich schwer ist, eine treffende Definition zu finden, die alle Erscheinungen umfaßt, die mit „Liebe“ bezeichnet werden, aber alle ausschließt, die man nicht mit „Liebe“ benennen würde. Es scheint also für einige Begriffe keine explizite Definition zu geben. Versucht man für solche kritischen Begriffe trotzdem eine Liste von Kriterien anzugeben, so lassen sich meist Gegenbeispiele finden, die nach einer verwendeten Kriterienliste anders klassifiziert würden, als eine Versuchsperson es angibt. Beim Fehlen von expliziten gültigen Definitionen stellt sich die Frage, ob es im Rahmen der internen Begriffsrepräsentation eines Individuums trotzdem eine Definition geben kann. Diese Probleme legen es nahe, nach einem Konzept für die Repräsentation von Begriffen zu suchen, das nicht die Existenz von Definitionen mit notwendigen und hinreichenden Kriterien annimmt.

Weiterhin zeigen Versuchspersonen nicht immer ein deterministisches Antwortverhalten, wenn es darum geht, Objekte Begriffen zuzuordnen. D.h. die Zuordnung variiert von Befragung zu Befragung. Solche Ergebnisse würden es nahelegen, daß individuelle Begriffe unscharf repräsentiert sind, so daß es einen Spielraum in ihrer Anwendung gibt. Derartige Verhältnisse ließen sich z.B. mit der „fuzzy logic“ beschreiben. Hier werden Bedingungen nicht in einer „Alles oder nichts“-Form verwendet. Müssen z.B. verschiedene Voraussetzungen erfüllt sein, damit ein Ereignis eintritt, so ist meist keine von diesen notwendig oder hinreichend. Vielmehr trägt jede einen Teil bei und erst

das Zusammentreffen von mehreren Bedingungen in ausreichend großer Ausprägung führt zum Auslösen des Ereignisses. Einen Überblick zur Konzeption der „fuzzy logic“ und ihren Einsatzmöglichkeiten gibt z.B. Kosko (1993).

Ein weiteres gut belegtes Ergebnis aus experimentellen Untersuchungen, das im Widerspruch zum klassischen Ansatz steht, sind die sogenannten „Typikalitätseffekte“:

Werden Versuchspersonen gebeten, für vorgegebene Objekte zu beurteilen, ob sie unter Begriffe fallen, ergeben sich für typische Vertreter einer Kategorie kürzere Antwortzeiten als für untypische (Rosch, 1973). Beispielsweise ist für den Begriff „Vogel“ das Rotkehlchen ein typischer Vertreter und der Geier ein untypischer. Bei Versuchen ergeben sich dementsprechend für die Einordnung von „Rotkehlchen“ kürzere Antwortzeiten als für „Geier“. Der klassische Ansatz kann auch für diesen Effekt keine Erklärung geben.

2.2.2 Die Prototypentheorie

Rosch, die die Typikalitätseffekte als Argument gegen den klassischen Ansatz vorbrachte entwickelte als Alternative die Prototypentheorie (Rosch, 1978). Allerdings gibt es auch hier verschiedene Weiterentwicklungen, so daß man eigentlich nicht von „der Prototypentheorie“ sprechen kann. Wie beim „klassischen Ansatz“ handelt es sich um eine ganze Gruppe von Theorien, deren Kerngedanken hier vorgestellt werden sollen.

„Die Prototypentheorie“ lehnt die Vorstellung des klassischen Ansatzes ab, daß Begriffe in Form von notwendigen und hinreichenden Bedingungen repräsentiert sind. Stattdessen wird angenommen, daß es für die Begriffe einen Mechanismus gibt, der analysiert, welche Eigenschaften Mitglieder einer Klasse *typischerweise* haben. Die zu einem Begriff gehörenden Eigenschaften werden

allerdings nicht als notwendig für alle Objekte, die unter den Begriff fallen, angesehen.

Das Erlernen von Begriffen besteht danach in der Herausbildung eines probabilistischen Beurteilungssystems, das man z.B. als Sammlung von Sätzen wie „Die meisten Vögel haben Federn“ oder „Fast alle Vögel können fliegen“ vorstellen. Ein Begriff beinhaltet also eine Verteilung von statistisch prominenten Merkmalen einer Kategorie, d.h. eine Sammlung von Merkmalen, für die man weiß, wie wahrscheinlich es ist, daß Vertreter einer Kategorie sie besitzen.

Soll also beurteilt werden, ob ein Objekt unter einen Begriff fällt, wird untersucht, welche den Begriff bestimmenden Merkmale das Objekt besitzt, und wie charakteristisch diese Merkmale für den Begriff sind. Für die Kategoriebildung und -verwendung wird also eine statistische Komponente angenommen. So könnte im oben genannten Beispiel des Begriffs „Vogel“ die Beurteilung für „Rotkehlchen“ folgendermaßen aussehen: „Vögel können fliegen, ein Rotkehlchen auch. Zudem hat ein Rotkehlchen wie die meisten Vögel Federn. Ein Rotkehlchen ist also ein Vogel“. Für einen Pinguin gestaltet sich diese fiktive Prozedur schwieriger: „Die meisten Vögel können fliegen, ein Pinguin nicht. Die meisten Vögel haben Federn, wie auch der Pinguin. Vögel legen Eier. Der Pinguin tut das auch, ist also ein Vogel“. Die statistische Bewertungsprozedur kommt also bei der Kategoriebeurteilung für das Rotkehlchen schneller zu einem Ergebnis als für den Pinguin. Dies entspricht dem experimentellen Ergebnis von kürzeren Kategoriebeurteilungszeiten für typischere Vertreter einer Kategorie (Rosch, 1973).

Da die Beurteilung von Kategoriezugehörigkeiten zudem probabilistisch abläuft, kann es die Prototypentheorie auch erklären, wenn Individuen Objekte bei mehrfacher Befragung nicht immer der gleichen Kategorie zuordnen.

Doch auch dieser Ansatz hat in manchen Bereichen Erklärungsprobleme: So müßten Begriffe, ähnlich wie im klassischen Ansatz, sehr weit ausgedehnt werden, um sehr untypische Vertreter noch zu erfassen. Eine solche Ausdehnung führt aber dann dazu, daß eigentliche Nicht-Mitglieder einer Kategorie als typischere Vertreter angesehen werden müßten als untypische Mitglieder. Laurence und Margolis (1999, S. 34) erläutern dies am fiktiven Beispiel eines „dreibeinigen, zahnlosen Albino-Tigers“, der trotz seiner Verunstaltungen immer noch unter den Begriff „Tiger“ fällt. Damit dies aber nach der Prototypentheorie möglich ist, müßten die Merkmale und ihre Wahrscheinlichkeitsverteilungen für den Begriff „Tiger“ so ausgedehnt werden, daß auch nicht-vierbeinige, zahnlose, streifenlose Lebewesen noch erfaßt werden. Dies würde dazu führen daß plötzlich auch Nicht-Mitglieder unter den Begriff „Tiger“ fallen.

Eine ausführlichere Darstellung der Prototypentheorie und der Kritik an dieser findet sich z.B. bei Laurence und Margolis (1999).

2.2.3 Die formale Begriffsanalyse

Bei der „Formalen Begriffsanalyse“ von Ganter und Wille (siehe z.B. Wille, 1987) handelt es sich um einen Ansatz aus der Mathematik, der versucht Formalismen für den Umgang mit Begriffen bereitzustellen. Es handelt sich also in erster Linie nicht um eine psychologische Theorie individueller Begriffsstrukturen. Durch eine mathematische Formalisierung soll allgemein ein Mittel zur Verfügung gestellt werden, um für Umgang mit Begriffen auf exakte Definitionen und genau spezifizierte Beziehungen zurückgreifen zu können. So stellt z.B. Wille (1987) verschiedene Anwendungsmöglichkeiten vor. Unter anderem sind Anwendungen zur systematischen Kategorisierung von Objekten

oder zur individuenübergreifenden Bereitstellung von Begriffsbeschreibungen denkbar.

Der Ansatz wird hier vorgestellt, da die später vorzustellende Alternative zu den bisher vorgestellten psychologischen Theorien ebenfalls auf eine mathematische Formalisierung zurückgreift und die Formale Begriffsanalyse wertvolle Hinweise gibt, wie geeignete mathematische Strukturen für den Umgang mit Begriffen aussehen könnten.

In Anlehnung an den oben vorgestellten „klassischen Ansatz“ wird eine Menge von Gegenständen G und eine Menge von für diese Gegenstände möglichen Merkmalen M angenommen. Wille (1987) verwendet in einem Beispiel als Gegenständsmenge die deutschen Bundespräsidenten von 1949 bis 1984. Also

$$G = \{\text{„Heuss“}, \text{„Lübke“}, \text{„Heinemann“}, \text{„Scheel“}, \text{„Carstens“}\}.$$

Die betrachtete Merkmalsmenge hängt von der gewünschten Anwendung ab. Im vorliegenden Beispiel wird

$$M = \{\text{„Antrittsalter} < 60\text{“}, \text{„Antrittsalter} \geq 60\text{“}, \\ \text{„eine Amtsperiode“}, \text{„zwei Amtsperioden“}, \\ \text{„Partei = CDU“}, \text{„Partei = SPD“}, \text{„Partei = FDP“}\}$$

verwendet.

Die Verbindung zwischen der Menge der Gegenstände G und der Menge der Merkmale M ist durch eine binäre Relation I zwischen G und M , d.h. $I \subseteq G \times M$, festgelegt.

I enthält also Kombinationen von je einem Gegenstand $g \in G$ und einem Merkmal $m \in M$. Wenn ein solches Tupel (g, m) in I enthalten ist, so wird das derart interpretiert, daß der Gegenstand g das Merkmal m besitzt. Ist ein Tupel (g, m)

mit $g \in G$ und $m \in M$ nicht in I enthalten so heißt dies, daß der Gegenstand g das Merkmal m nicht besitzt.

Für das obige Bundespräsidentenbeispiel würde die faktengetreue Relation I folgendermaßen aussehen:

$$\begin{aligned}
 I = & \{ („Heuss“, „Antrittsalter ≥ 60 “), \\
 & („Heuss“, „zwei Amtsperioden“), („Heuss“, „Partei = FDP“), \\
 & („Lübke“, „Antrittsalter ≥ 60 “), \\
 & („Lübke“, „zwei Amtsperioden“), („Lübke“, „Partei = CDU“), \\
 & („Heinemann“, „Antrittsalter ≥ 60 “), \\
 & („Heinemann“, „eine Amtsperioden“), („Heinemann“, „Partei = SPD“), \\
 & („Scheel“, „Antrittsalter < 60 “), \\
 & („Scheel“, „eine Amtsperiode“), („Scheel“, „Partei = FDP“), \\
 & („Carstens“, „Antrittsalter ≥ 60 “), \\
 & („Carstens“, „eine Amtsperiode“), („Carstens“, „Partei = CDU“) \}
 \end{aligned}$$

Die Relation I läßt sich auch als Kreuztabelle von Gegenständen und Merkmalen darstellen. Ist eine Gegenstand-Merkmal-Kombination in der Relation enthalten, so ist die zugehörige Zelle mit einem Kreuz gekennzeichnet. Tabelle 1 zeigt die Kreuztabelle für das Bundespräsidentenbeispiel.

Ein Tripel $\langle G, M, I \rangle$ mit einer Gegenstandsmenge G , einer Merkmalsmenge M und einer Relation I zwischen diesen wird *Kontext* genannt.

Eine Verbindung zwischen Mengen von Gegenständen und Mengen von Merkmalen innerhalb eines Kontextes $\langle G, M, I \rangle$ liefern die folgenden Ableitungsoperatoren:

$$X \mapsto X' := \{m \in M \mid gIm \ \forall g \in X\}$$

Tabelle 1: Kreuztabelle für Gegenstände (hier Bundespräsidenten) und Merkmale. Trägt ein Gegenstand ein Merkmal, ist die zugehörige Zelle durch ein \times gekennzeichnet.

Bundes- präsidenten	Antrittsalter		Amtsperioden		Partei		
	< 60	≥ 60	eine	zwei	CDU	SPD	FDP
Heuss		\times		\times			\times
Lübke		\times		\times	\times		
Heinemann		\times	\times			\times	
Scheel	\times		\times				\times
Carstens		\times	\times		\times		

$$Y \mapsto Y' := \{g \in G \mid gIm \ \forall m \in Y\}$$

wobei $X, Y' \subseteq G$ und $Y, X' \subseteq M$.

Diese Ableitungsoperatoren liefern also für eine Gegenstandsmenge X die Menge der Merkmale X' , die alle Gegenstände in X gemeinsam haben. Zum Beispiel für

$$X = \{\text{„Heuss“}, \text{„Lübke“}\}$$

ist

$$X' = \{\text{„Antrittsalter} > 60\text{“}, \text{„zwei Amtsperioden“}\}.$$

Umgekehrt beinhaltet für eine Merkmalsmenge Y die Gegenstandsmenge Y' alle Gegenstände, die die Merkmale in Y besitzen. Ist also z.B.

$$Y = \{\text{„eine Amtsperiode“}, \text{„Partei = FDP“}\},$$

so ist

$$Y' = \{\text{„Scheel“}\}.$$

Unter Verwendung der bisher eingeführten Formalismen kann nun definiert werden, was unter einem Begriff zu verstehen ist (Wille, 1987, S. 162):

Die mengensprachliche Definition eines Kontextes setzt einen Rahmen, in dem auch der Begriff, so wie ihn die traditionelle Begriffslehre versteht . . . , mengensprachlich festgelegt werden kann. Ein *Begriff* eines Kontextes $\langle G, M, I \rangle$ wird definiert als ein Paar (A, B) , für das $A \subseteq G$, $B \subseteq M$, $A' = B$ und $B' = A$ gilt. A heißt der *Umfang* und B der *Inhalt* des Begriffs (A, B) .

Im oben angegebenen Beispielkontext wären z.B. die Mengen {„Heinemann“, „Carstens“} und {„Antrittsalter ≥ 60 “, „eine Amtsperiode“} der Umfang und der Inhalt eines Begriffs, den man mit „Amtsantritt mit 60 oder älter bei einjähriger Amtsperiode“ umschreiben könnte.

Die Menge aller Begriffe wird mit $\mathcal{L}(G, M, I)$ bezeichnet.

Wille (1987) gibt noch weitere Eigenschaften von Kontexten und die Menge aller Begriffe an. Hier soll noch auf die Möglichkeit hingewiesen werden, $\mathcal{L}(G, M, I)$ graphisch darzustellen.

In Hasse-Diagrammen sind in einer netzwerkartigen Darstellung die Begriffe eines Kontextes als Knoten repräsentiert. Beinhaltet ein Begriff einen anderen, ohne daß es einen dazwischen stehenden Begriff gibt, werden die beiden zugehörigen Knoten mit einer gerichteten Kante verbunden. Der Begriff, der den anderen beinhaltet, wird weiter oben liegend gezeichnet. Das Merkmal, das einen zu einem Knoten gehörigen Begriff von dem nächsten über ihm liegenden Begriff/Knoten unterscheidet, wird neben den Knoten geschrieben. Beinhaltet der Umfang des Begriffs nur ein Objekt, so wird dieses ebenfalls beim zugehörigen Knoten notiert.

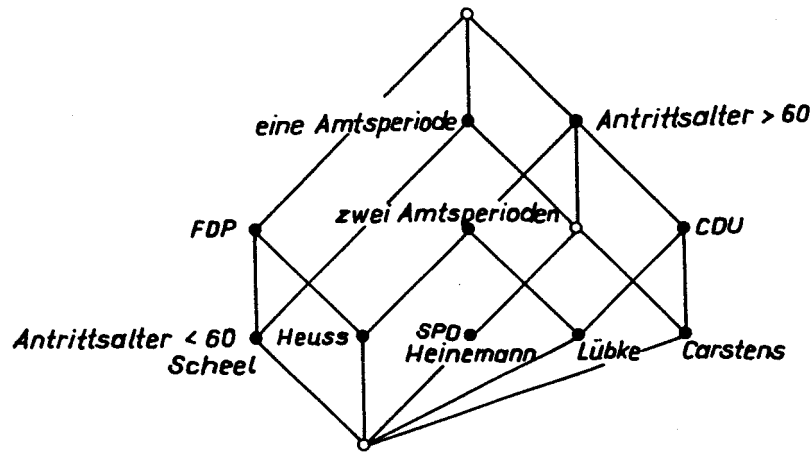


Abbildung 1: Graphische Darstellung der Menge aller Begriffe $\mathcal{L}(G, M, I)$ des Beispiel-Kontextes $\langle G, M, I \rangle$ (aus Wille, 1987)

Für das obige Beispiel ergibt sich die Darstellung in Abbildung 1.

Will man diese Formalismen auf individuelle Begriffsstrukturen übertragen, so stellt sich die Frage, ob individuelle Begriffe so strukturiert sind, wie es im Rahmen der Formalen Begriffsanalyse festgelegt wird. Kritisch ist hier, wie beim vorgestellten „klassischen Ansatz“, daß für Begriffe definierende Merkmale gefordert werden, die Begriffe scharf von anderen Begriffen abgrenzen. Die Problematik einer solchen Annahme und die empirische Evidenz gegen diese wurde oben schon angeführt. Es fehlt aber ganz allgemein auch für die Formale Begriffsanalyse eine Methode um prüfen zu können, ob die theoretischen Strukturen der psychologischen Realität entsprechen.

Allerdings ist, wie schon gesagt, die Formale Begriffsanalyse von Ganter und Wille in erster Linie nicht als Theorie der individuellen Begriffsstrukturen gedacht, sondern vielmehr als formaler Rahmen für den Umgang mit Begriffen.

2.3 Probleme der bisherigen psychologischen Ansätze

Anlässlich der oben dargestellten Probleme der psychologischen Begriffstheorien, wie z.B. der Prototypentheorie, wurden weitere Alternativ-Theorien entwickelt. Für eine Übersicht siehe z.B. Laurence und Margolis (1999).

Stellt man die Vielzahl von Ansätzen, die zur Repräsentation von Begriffen existieren zueinander in Beziehung, so zeigt sich, daß die Verfechter jeder Theorie experimentelle Ergebnisse anführen kann, die den jeweiligen eigenen Ansatz unterstützen und andere Ansätze möglichst widerlegen. Allerdings findet sich bisher kein Ansatz, zu dem es nicht auch experimentelle Ergebnisse gäbe, die ihm widersprechen. Laurence und Margolis (1999) stellen fest, daß es bisher noch keinen Ansatz gibt, der eine umfassende Erklärung für alle Phänomene, die im Umfeld der Repräsentation von Begriffen auftreten, liefern würde. Eine Entscheidung zwischen den konkurrierenden Ansätzen, die für eine diagnostische Anwendung nötig ist, fällt daher schwer.

Hilfreich könnte hier ein Rückgriff auf Ansätze der Wissenschaftstheorie sein, die Kriterien für die Bewertung von Theorien angeben. Beispielsweise unterscheidet der Ansatz des „Strukturalismus“ (Sneed, 1971, Stegmüller, 1973, 1974) bei Theorien zwischen Begriffen, die nur bestimmt werden können, wenn die jeweilige Theorie als gegeben vorausgesetzt wird, und Begriffen, die unabhängig von der Theorie bestimmt werden können. Würde man z.B. Parameter für die statistische Komponente der „Prototypentheorie“ angeben, so wäre ein Parameter, der die Wahrscheinlichkeit angibt, mit der eine Versuchsperson eine bestimmte Kategoriezuordnung vornimmt, außerhalb der Theorie bestimmbar, da das Zuordnungsverhalten und damit auch die relative Häufigkeit der Zuordnung zu einer Kategorie, auch ohne die Annahme der Theorie beobachtet werden kann. Ein Parameter, der beispielsweise die Gewichtung von

verschiedenen Kriterien beschreiben würde und sich aus den beobachteten Daten schätzen läßt, ist nur bestimmbar wenn die Theorie als gültig angenommen wird.

Um also zu entscheiden, ob eine Theorie für einen Gegenstandsbereich ein brauchbares Werkzeug ist, müssen die theoriespezifischen und die außertheoretischen Begriffe identifiziert werden. Auf Basis der außertheoretischen Begriffe kann dann überprüft werden, ob — in der Sprache des Strukturalismus — der Gegenstandsbereich ein Modell der Theorie ist, d.h. ob die Theorie ein passendes Werkzeug für den Gegenstandsbereich ist und für diesen verwendet werden kann.

Für die oben besprochenen Ansätze gibt es, orientiert man sich an der Übersicht von Westmeyer (1989), allerdings bisher noch keine solche strukturalistische Rekonstruktion. Deswegen wird später in dieser Arbeit auf einen meßtheoretischen Ansatz zurückgegriffen, der für die angedeuteten Überlegungen zugänglicher ist.

2.4 Verfahren zur Begriffsstruktur-Bestimmung

Im Rahmen der, mit dem Beispiel zum studentischen Statistikwissen angedeuteten, gewünschten diagnostischen Anwendung von Theorien über die Struktur von Begriffen, stellt sich die Frage, welche Implikationen diese bezüglich Verfahren zur Bestimmung individueller Begriffsstrukturen haben.

Folgt man dem klassischen Ansatz müßte jeder Begriff, den ein Individuum besitzt, durch eine Definition repräsentiert sein, die aus notwendigen und hinreichenden Kriterien besteht. Ein daraus abgeleitetes Verfahren zur Bestimmung von Begriffen würde das Abfragen der Definitionskriterien beinhalten. Das Problem hierbei ist allerdings, daß nicht zu kontrollieren ist, inwieweit Fakto-

ren wie z.B. sprachliche Ausdrucksfähigkeit hierbei einen Störfaktor darstellen. Es bleibt die Unklarheit, ob die Versuchsperson für einen Begriff wirklich eine Definition besitzt, wie es die Theorie behauptet, oder nur die Fähigkeit getestet wird, auf Nachfrage Definitionen zu generieren.

Ansätze wie die „Prototypentheorie“ motivieren Verfahren, die mit Ähnlichkeitsurteilen arbeiten. Gibt es nämlich tatsächlich einen statistischen Mechanismus für die Feststellung von Begriffseigenschaften und -zugehörigkeiten, so müßten Urteile über die Wahrscheinlichkeit der Zugehörigkeit eines Objektes zu einem Begriff und auch Ähnlichkeitsurteile möglich sein. Die direkte Erhebung von Ähnlichkeitsurteilen bringt allerdings Probleme mit sich. Es ist nämlich nicht klar, ob es zulässig ist aus diesen dann z.B. geometrische Räume abzuleiten, in denen die Begriffe liegen und in denen die Ähnlichkeit durch räumliche Nähe dargestellt wird. Eine Untersuchung von Tversky und Krantz (1969) hat beispielsweise gezeigt, daß die Anforderungen, die Ähnlichkeitsurteile für eine solche Anwendung erfüllen müßten teilweise systematisch verletzt sind. Eine detaillierte Beschreibung der meßtheoretischen Voraussetzungen, die für die Verwendung von Ähnlichkeitsmaßen erfüllt sein müssen, geben z.B. Krantz, Luce, Suppes und Tversky (1989).

Aufgrund dieser zahlreichen Probleme, die sich also nicht nur bei der Entscheidung zwischen den verschiedenen oben angedeuteten Ansätzen zur Begriffsrepräsentation sondern auch beim Einsatz von Verfahren zur Bestimmung von Begriffsstrukturen ergeben, stützt sich diese Arbeit auf eine andere, meßtheoretisch motivierte, Vorgehensweise.

3 Ein meßtheoretischer Ansatz

3.1 Theoretische Vorüberlegungen

Wie schon in der Einleitung angedeutet, wird in dieser Arbeit das Augenmerk auf diagnostische Aspekte von Begriffsstruktur-Theorien und Bestimmungsverfahren gelegt.

In der psychologischen Diagnostik wird meist versucht, bestimmte Eigenschaften eines Individuums zu quantifizieren, d.h. Zahlen zuzuweisen. Diesen grundlegenden Vorgang beschreiben Krantz, Luce, Suppes und Tversky (1971, S. 1) so:

„When measuring some attribute of a class of objects or events, we associate numbers (or familiar mathematical entities, such as vectors) with the objects in such a way that to properties of the attribute are faithfully represented as numerical properties.“

Es ist also nötig, die Eigenschaften der Attribute bzw. Objekte zu betrachten, die gemessen werden sollen. Im Falle der Messung durch Zuordnung von Zahlen werden von den zu messenden Attributen elementare Struktureigenschaften der Zahlen (z.B. Ordnung) erwartet. Dabei werden Forderungen (Axiome) angegeben, die erfüllt sein müssen, damit es eine Abbildung von der Menge der Objekte in die Menge der Zahlen gibt. Daß es eine solchen Abbildung gibt, wenn die Axiome erfüllt sind, besagt ein Repräsentationssatz. Da die Geltung der Axiome empirisch überprüfbar sein muß, sind diese qualitativ und auf Beobachtungen zurückführbar formuliert.

Überträgt man diese Situation auf den Forschungsbereich der Begriffsstruktur, so benötigt man passende „mathematische Objekte“, empirisch prüfbare

Bedingungen und einen Repräsentationssatz, der besagt, daß bei Gültigkeit der Bedingungen die interessierenden psychologischen Strukturen durch die mathematischen Objekte *repräsentiert* werden können. Die Suche nach passenden mathematischen Objekten gestaltet sich auf den ersten Blick schwierig. Zahlen und Vektoren erscheinen wenig geeignet für eine Repräsentation von Begriffen, v.a. da Begriffe nicht die nötigen Struktureigenschaften haben dürften. So läßt sich wohl z.B. keine Ordnungsrelation für sie finden.

Eine Anregung, wie eine passende Repräsentation aussehen könnte liefert die Wissensraumtheorie von Doignon und Falmagne (1999). Diese stellt ein Wissensgebiet durch eine Menge X von Aufgabentypen dar. Beispielsweise könnte das Wissensgebiet der Grundschulmathematik durch eine repräsentative Sammlung von Aufgabentypen beschrieben werden.

Der Wissenszustand eines Schülers wird durch eine Teilmenge von X dargestellt, eben die Menge von Aufgabentypen, die der Schüler lösen kann. Ob eine Schüler einen bestimmten Aufgabentypus lösen kann wird sicher teilweise davon abhängen, ob er auch andere Aufgaben beherrscht. Es existieren also gewisse Beziehungen zwischen den einzelnen Aufgabentypen aus X . Zum Beispiel wird ein Schüler, der nicht addieren kann, auch nicht subtrahieren können. Daraus folgt, daß es nur eine bestimmte Menge von Wissenszuständen gibt, da eben z.B. ein Schüler dessen Wissenszustand „Subtrahieren“ beinhaltet, auch „Addieren“ in seinem Wissenszustand haben wird. Die Menge der möglichen Wissenszustände, die jeweils Teilmengen von X sind, werden als Wissensstruktur bzw. unter gewissen Bedingungen als Wissensraum bezeichnet.

Dieser Ansatz ließe sich folgendermaßen auf individuelle Begriffsstrukturen übertragen:

Man betrachtet eine Menge X von Objekten, z.B. {„Jacke“, „Kleid“, „Rock“,

„Hose“}. Als Begriff faßt man dann Teilmengen von X auf. Zum Beispiel könnte im Rahmen des oben angegebenen X die Teilmenge {„Rock“, „Hose“} für den Begriff „Beinbekleidung“ stehen. Desweiteren benötigt man einen Repräsentationssatz, der empirische Bedingungen für das Antwortverhalten einer Versuchsperson angibt, bei deren Gültigkeit eine Repräsentation von Begriffen durch Teilmengen von X zulässig ist.

Genau den eben skizzierten Weg beschreitet der im Folgenden dargestellte Ansatz von Heller.

3.2 Der Ansatz von Heller

Der Ansatz von Heller (1991, 2000) basiert auf der oben angedeuteten, auf meßtheoretischen Überlegungen ausgehenden Vorgehensweise.

Von einem strukturalistischen Standpunkt ausgehend, wie er oben erläutert wurde, wird nun zuerst der formale Rahmen der Theorie vorgestellt. Anschließend wird gezeigt, wie sich dieser auf den Bereich der Repräsentation von Begriffen anwenden läßt.

3.2.1 Formaler Rahmen

Ein formaler Rahmen, der in Hinblick auf die Repräsentation von Begriffen entwickelt wird, muß zuerst definieren, was unter einem Begriff bzw. einer Begriffsstruktur zu verstehen ist. Heller (2000) führt folgende Definition ein:

Definition 1 *Eine Begriffsstruktur $\langle X, \sigma \rangle$ besteht aus einer nicht-leeren Menge X und einer Menge σ von Teilmengen von X , das heißt $\sigma \subset 2^X$, so daß $\emptyset, X \in \sigma$. Jede Begriffsstruktur auf X , die unter der Schnittmengenoperation abgeschlossen ist, heißt Begriffsraum auf X .*

Beispielsweise ist für $X = \{a, b, c, d\}$ $\langle X, \sigma \rangle$ mit

$$\sigma = \{\{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$$

ein Begriffsraum auf X .

Solche Begriffsräume lassen sich mit Hilfe von Hasse-Diagrammen graphisch darstellen:

In einer netzartigen Struktur wird jedes Element eines Begriffsraumes durch einen Knoten dargestellt. Knoten von Elementen mit weniger Objekten liegen weiter unten und umgekehrt. Umfaßt ein Element $A \in \sigma$ ein anderes Element B ($B \subset A$) und gibt es kein anderes Element aus σ , das B beinhaltet und Teilmenge von A ist, so werden die zu A und B gehörenden Knoten mit einer gerichteten Kante verbunden. D.h. A ist das kleinsten Element von σ , das B beinhaltet.

Abbildung 2 zeigt das Hasse-Diagramm für das oben angegebene Beispiel für einen Begriffsraum. In dieser Arbeit ist es, wie auch in meisten anderen Beiträgen zu dem hier vorgestellten Ansatz, Konvention, neben Knoten von Elementen von σ , die nur ein Objekt aus X beinhalten, den Namen dieses Objekts zu schreiben.

Wie sich später zeigen wird ist für die Anbindung an die experimentelle Beobachtung die Einführung der Entailment-Relation nützlich. Heller (2000) definiert diese folgendermaßen:

Definition 2 *Eine Relation Q von $2^X \setminus \{\emptyset\}$ nach X heißt Entailment(-Relation) auf X , wenn für alle $A, B \subset X$ und $c \in X$ folgende Bedingungen gelten:*

1. $a \in A$ impliziert AQa ,
2. AQb für alle $b \in B$ und BQc impliziert AQc (Transitivität)

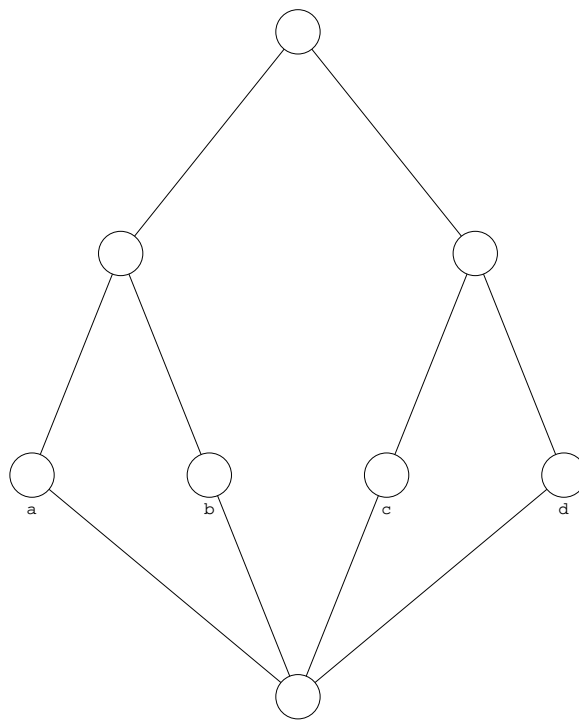


Abbildung 2: Hasse-Diagramm für den Beispiel-Begriffsraum $\langle X, \sigma \rangle$ (siehe Text)

Für $X = \{a, b, c, d\}$ wäre z.B.

$$\begin{aligned}
Q = & \{(\{a\}, a), (\{b\}, b), (\{c\}, c), (\{d\}, d), \\
& (\{a, b\}, a), (\{a, b\}, b), \\
& (\{a, c\}, a), (\{a, c\}, c), \\
& (\{a, d\}, a), (\{a, d\}, d), \\
& (\{b, c\}, b), (\{b, c\}, c), \\
& (\{b, d\}, b), (\{b, d\}, d), \\
& (\{c, d\}, c), (\{c, d\}, d), \\
& (\{a, b, c\}, a), (\{a, b, c\}, b), (\{a, b, c\}, c), \\
& (\{a, b, d\}, a), (\{a, b, d\}, b), (\{a, b, d\}, d), \\
& (\{a, c, d\}, a), (\{a, c, d\}, c), (\{a, c, d\}, d), \\
& (\{b, c, d\}, b), (\{b, c, d\}, c), (\{b, c, d\}, d), \\
& (\{a, b, c, d\}, a), (\{a, b, c, d\}, b), (\{a, b, c, d\}, c), (\{a, b, c, d\}, d), \\
& (\{a\}, c), (\{a\}, d), \\
& (\{c, d\}, b), \\
& (\{a\}, b), \\
& (\{a, b\}, c), (\{a, d\}, c), (\{a, b\}, d), (\{a, c\}, d), \\
& (\{a, c, d\}, b), \\
& (\{a, c\}, b), (\{a, d\}, b)\}
\end{aligned}$$

eine Entailment-Relation auf X .

Beachtenswert an diesem Beispiel ist, daß sich alle Elemente bis auf die in den letzten sechs Zeilen schon aus Bedingung 1 von Definition 2 ergeben. Das Element $\{\{a\}, b\}$ in der viertletzten Zeile ergibt sich nach Bedingung 2 von Definition 2, der Transitivität, daraus, daß Q die drei Elemente in den beiden

Zeilen davor enthält. Die Elemente in den nachfolgenden Zeilen ergeben sich aufgrund von Bedingung 2 aus den Elementen in den drei Zeilen davor und den Elementen in den darüberliegenden Zeilen, die aus Bedingung 1 hervorgegangen sind.

Die Verbindung für zwischen einem Begriffsraum und einer Entailment-Relation liefert folgendes Theorem von Heller (2000):

Theorem 1 Es gibt eine eindeutige Korrespondenz zwischen Entailments Q auf X und Begriffsräumen $\langle X, \sigma \rangle$ auf X , so daß für alle $A \subset X$ und $b \in X$ gilt:

$$AQb \text{ gdw } b \in \sigma(A),$$

wobei $\sigma(A)$ die kleinste Menge aus σ ist, die A enthält.

Der Beweis für dieses Theorem soll hier nicht angeführt werden. Er findet sich bei Heller (1991).

3.2.2 Anwendung in der Begriffsforschung

Um überprüfen zu können, ob der eben vorgestellte formale Rahmen eine brauchbare Theorie zur Bestimmung individueller Begriffsstrukturen darstellt, muß zuerst geklärt werden, was unter den jeweiligen theoretischen Begriffen in diesem konkreten Anwendungsfall zu verstehen ist. Die folgende Darstellung der empirischen Interpretation der theoretischen Begriffe orientiert sich an Heller (2000).

Die Menge X sei eine Menge von Objekten der Alltags- bzw. Begriffswelt. Beispielsweise könnte X eine Menge von Kleidungsgegenständen sein, mit $X = \{\text{„Bluse“}, \text{„Hose“}, \text{„Mantel“}, \text{„Weste“}\}$. Wichtig ist, daß X dann das Bezugssystem der weiteren Betrachtung individueller Begriffsstrukturen bildet.

Andere Objekte als die in X enthaltenen gehen nicht mehr in das Erhebungsverfahren ein. X stellt also eine Art Bezugsrahmen dar.

Ein Begriffsraum $\langle X, \sigma \rangle$ auf X wird als Raum der individuellen Begriffe eines Individuums interpretiert. σ beinhaltet dabei alle Begriffe, die den Bezugsrahmen X betreffen. Ein Element aus σ wird also als Begriff eines Individuums bezeichnet. Er wird durch die Menge der Objekte aus X , die er umfaßt, beschrieben, ist also auf X bezogen. Für den Bezugsrahmen X aus dem Beispiel wäre z.B. $\langle X, \sigma \rangle$ mit

$$\sigma = \{ \{ \text{„Bluse“} \}, \{ \text{„Hose“} \}, \{ \text{„Mantel“} \}, \{ \text{„Weste“} \}, \{ \text{„Bluse“}, \text{„Weste“} \}, \\ \{ \text{„Bluse“}, \text{„Hose“}, \text{„Weste“} \}, \{ \text{„Bluse“}, \text{„Hose“}, \text{„Mantel“}, \text{„Weste“} \} \}$$

ein Begriffsraum auf X .

In diesem fiktiven Begriffsraum könnte z.B. der Begriff $\{ \text{„Bluse“}, \text{„Weste“} \}$ für „Oberbekleidung“ stehen. Derartige „Sinn“-Interpretationen werden allerdings im weiteren Verlauf dieser Arbeit vermieden.

Das Hasse-Diagramm für diesen Begriffsraum zeigt Abbildung 3. An dieser Darstellungsart kann man auch eine weitere Interpretation des Begriffsraumes gut ablesen: Umfaßt ein Element A aus σ ein anderes Element B ($B \subset A$), so wird A als Oberbegriff von B bezeichnet.

Im Hasse-Diagramm erreicht man die Unterbegriffe eines Begriffs, indem man von dem zu ihm gehörenden Knoten alle Kantenzüge, die nach unten führen entlangfährt. Die Knoten auf diesen Pfaden sind die Unterbegriffe. Umgekehrt findet man die Oberbegriffe zu einem Begriff, indem man alle Knoten sucht, die auf Kantenzügen liegen, die von dem Knoten, der zu dem Begriff gehört, nach oben führen.

Bei dem Konstrukt des Begriffsraumes handelt es sich allerdings um etwas,

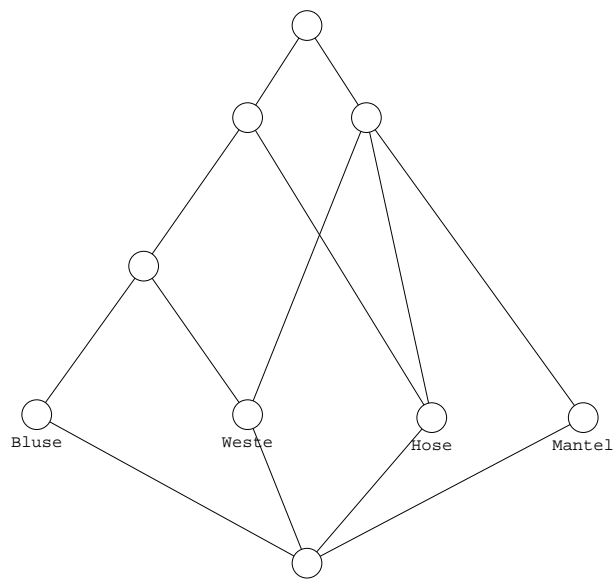


Abbildung 3: Hasse-Diagramm für den (fiktiven) Begriffsraum $\langle X, \sigma \rangle$ (siehe Text)

das, wenn überhaupt, nur beobachtet werden kann wenn die zugehörige Theorie vorausgesetzt wird. Für die Beurteilung, ob die vorgestellte Theorie auf den Problembereich der individuellen Begriffsstrukturen anwendbar ist, ist dieses Konstrukt deshalb untauglich. Was benötigt wird, ist ein Element der Theorie, für das sich überprüfen läßt, ob es ein passendes empirisches Äquivalent gibt, ohne die Theorie vorzusetzen. Diese Forderung erfüllt die oben eingeführte Entailment-Relation.

Doch wie läßt sich die Entailment-Relation zum Gebiet der individuellen Begriffsstrukturen so in Beziehung setzen, daß untersucht werden kann, ob sie tatsächlich vorliegt? Heller (2000) schlägt vor, dies dadurch zu tun, indem Individuen, deren Begriffsstrukturen untersucht werden sollen, Fragen vom folgenden Format vorgelegt werden:

„Haben die Begriffe in A etwas gemeinsam, was sie von dem Begriff b unterscheidet?“

wobei $A \subset X$ und $b \in X$.

Lautet die Antwort „Nein“, so wird definiert AQb . Ist die Antwort „Ja“, so wird $A\bar{Q}b$ gesetzt.

Für Menge A mit nur einem Element ist eine anderes Fragenformat nötig:

„Ist der Begriff A ein Oberbegriff von dem Begriff b ?“

Bei der Antwort „Nein“ wird $A\bar{Q}b$ definiert, bei „Ja“ AQb .

Dieser „Sonderfall“ wird im Folgenden nicht mehr gesondert aufgeführt.

Fragt man also beispielsweise

„Haben die Begriffe „Bluse“ und „Mantel“ etwas gemeinsam, was sie von dem Begriff „Hose“ unterscheidet?“

und erhält als Antwort „Ja“, so setzt man {„Bluse“, „Mantel“} \bar{Q} „Hose“.

Auf diese Weise erhält man eine empirische Relation, die daraufhin untersucht werden kann, ob es sich bei ihr um eine Entailment-Relation handelt, d.h. ob sie die in Definition 2 angegebenen Bedingungen erfüllt.

Die Struktur einer Entailment-Relation wurde schon im letzten Abschnitt anhand eines Beispiels erläutert. Werden nun für eine empirische Relation, die sich aus der Beantwortung von Frage im oben angegebenen Format ergibt, die Forderungen aus Definition 2 gestellt, so bedeutet dies für die beiden angegebenen Bedingungen Folgendes:

- Bedingung 1 besagt, daß eine Menge von Objekten, die ein bestimmtes Objekt beinhaltet, nichts gemeinsam haben kann, was sie von diesem bestimmten Objekt unterscheidet. Es darf also z.B. nicht gesagt werden: „Jacke und Hose haben etwas gemeinsam, was sie von Hose unterscheidet.“ Diese Forderung scheint trivial erfüllt zu sein. Eine Verletzung könnte beispielsweise dadurch auftreten, daß bei mehreren gemeinsam präsentiert Objekten Eigenschaften als gegeben gesehen werden, die bei der einzelnen Betrachtung von Objekten außer Acht bleiben. Würde eine Versuchsperson also im Stil von „Alle Objekte zusammengenommen haben die Eigenschaft xy, das Objekt einzeln hat sie nicht“, antworten, wäre die Forderung aus Bedingung 1 verletzt.
- Bedingung 2 besagt Folgendes: Wenn eine Menge von Objekten nichts gemeinsam hat, was sie von bestimmten dritten Objekten unterscheidet, und die Menge der dritten Objekte hat nichts gemeinsam, was sie von einem bestimmten Objekt unterscheidet, so haben auch die zuerst betrachteten Objekte nichts gemeinsam, was sie von diesem bestimmten Objekt unterscheidet.

Diese Transitivitätsbedingung wäre z.B. verletzt, wenn das befragte Individuum bei der Beantwortung der Fragen die Antwort-Kriterien wechselt. Geschieht dies, so ist nicht mehr sichergestellt, daß bei der Beurteilung des Verhältnisses von Objekten $A \subseteq X$ zu Objekten aus $B \subseteq X$ und von diesen zu einem Objekt $c \in X$ die gleichen Kriterien verwendet werden, wie bei der Beurteilung vom Verhältnis der Objekte A zu dem Objekt c . Wirkt sich so ein Kriterienwechsel auf das Antwortverhalten aus, ist die Forderung aus Bedingung 2 verletzt.

Handelt es sich bei der empirisch erhaltenen Relation um eine Entailment-Relation, so kann man nach Theorem 1 den zugehörigen Begriffsraum bestimmen.

Im Sinne der repräsentationalen Meßtheorie (Krantz, Luce, Suppes, Tversky, 1971) handelt es sich bei Theorem 1 also um einen Repräsentationssatz, da es besagt, daß unter gewissen (qualitativ formulierten) Bedingungen, die empirisch erfüllt sein müssen, eine Repräsentation von empirischen Gegebenheiten (in diesem Fall von Begriffen) durch mathematische Objekte (in diesem Fall Mengen) möglich ist.

Daß es sich bei den Elementen von σ eines Begriffsraumes dem Sinn nach wirklich um Begriffe handelt, wie sie zum Beispiel mit den Definitionen in Abschnitt 2.1 eingeführt wurden, soll noch illustriert werden.

Wenn ein Individuum einen Begriff $S \in \sigma$ besitzt, z.B. {„Bluse“, „Hose“, „Weste“} $\in \sigma$, so bedeutet das nach dem Ansatz von Heller Folgendes:

Für jedes Objekt $b \in S$, also z.B. „Weste“, wird der kleinste Begriff aus σ gesucht, der $S \setminus \{b\}$, hier also {„Bluse“, „Hose“}, umfaßt. Enthält dieser Begriff wiederum auch b , im Beispiel also „Weste“, so würde das Individuum sagen,

„Die Begriffe in $S \setminus \{b\}$ haben nichts gemeinsam, was sie von b un-

terscheidet.“

Der kleinste Begriff, der das $S \setminus \{b\}$ {„Bluse“, „Hose“} aus dem eben verwendeten Beispiel enthält, ist in unserem fiktiven Begriffsraum der Begriff {„Bluse“, „Hose“, „Weste“}. Dieser enthält aber wiederum „Weste“. Ein Individuum, das diesen Begriffsraum besitzt, sollte also sagen

„Bluse und Hose haben nichts gemeinsam, was sie von Weste unterscheidet unterscheidet.“,

Die Aussage

„Bluse und Weste haben nichts gemeinsam, was sie von Hose unterscheidet unterscheidet.“

kann allerdings *nicht* abgeleitet werden, da der kleinste Begriff, der {„Bluse“, „Hose“, „Weste“} \setminus {„Hose“}, also {„Bluse“, „Weste“}, umfaßt, {„Bluse“, „Weste“} ist. Dieser Begriff enthält aber „Hose“ nicht. Nach dem oben angegebenen „Sinn“ der Begriffe ist dies auch plausibel, da „Bluse“ und „Weste“ etwas gemeinsam haben, das sie von „Hose“ unterscheidet, nämlich, daß sie „Oberbekleidung“ sind.

Diese Erläuterung sollte zeigen, daß es sich bei den Elementen von σ , die aus einer möglichen Entailment-Relation abgeleitet werden, wirklich um das handelt, was man auch intuitiv unter „Begriffen“ versteht.

4 Befragungsverfahren

Will man nach dem Ansatz von Heller (1991, 2000) die Begriffsstruktur $\langle X, \sigma \rangle$ bezogen auf eine bestimmte Menge von Objekten X bestimmen, so bildet eine empirische Relation $Q \subseteq 2^X \times X$ den Ausgangspunkt. Diese erhält man, indem dem Individuum Fragen vom folgenden Format vorgelegt werden:

„Haben die Begriffe in A etwas gemeinsam, was sie von dem Begriff b unterscheidet?“

Lautet die Antwort „Nein“, so wird AQb gesetzt, bei „Ja“ $A\bar{Q}b$.

$A \subseteq X$ wird *Prämisse* und $b \in X$ *Konsequenz* genannt. Diese Bezeichnungen stammt aus dem Forschungsbereich der Wissensräume. Dort beantwortet ein Experte Fragen vom Typ „Wenn ein Schüler Aufgabentyp x nicht lösen kann, kann er dann auch Typ y nicht lösen?“. Aus Voraussetzungen (Prämisse) werden Folgerungen (Konsequenz) gezogen. Die Bezeichnungen sollen hier beibehalten werden, da die vorgestellten Befragungsverfahren meist im Bereich der Wissensräumforschung entwickelt wurden.

Für die vorgestellte, typischerweise mit einem Computer automatisierte Befragung gibt es verschiedene Herangehensweisen:

Die Erhebung der empirischen Relation Q erfordert eine Komplettbefragung, d.h. die Frage muß für alle nichtleeren $A \subseteq X$ und alle $b \in X$ gestellt werden. Aus dieser Vorgehensweise ergeben sich verschiedene Probleme, zu deren Lösung das Verfahren QUERY entwickelt wurde. Wiederum eine Erweiterung dieses Verfahrens ist das PSQUERY-Verfahren.

4.1 Komplettbefragung

Bei der Komplettbefragung wird dem Individuum, dessen Begriffsstruktur $\langle X, \sigma \rangle$ auf der Menge von Objekten X bestimmt werden soll, für jede Menge $A \subseteq X$ in Kombination mit jedem Objekt $b \in X$ die oben beschriebene Frage gestellt.

Die empirische Relation, die man daraus erhält, muß dahingehend überprüft werden, ob sie die Bedingungen für eine Entailment-Relation erfüllt (siehe Definition 2). Nur wenn dies der Fall ist, kann ein zu ihr korrespondierender Begriffsraum bestimmt werden.

Die Zahl der Fragen, die gestellt werden müssen, berechnet sich für die Zahl von Objekten n nach der Formel,

$$\sum_{k=1}^{n-1} \binom{n}{k} \cdot n.$$

In in der Praxis wird die Bedingung 1 aus Definition 2 als trivial erfüllt angesehen. Die Zahl der Fragen reduziert sich danach auf

$$\sum_{k=1}^{n-1} \binom{n}{k} \cdot (n - k),$$

da für jede Prämissengröße k (von 1 bis $n - 1$) jeweils $\binom{n}{k}$ verschiedene Prämissen möglich sind, für die es jeweils $(n - k)$ Konsequenzen gibt, die kein Objekt aus der Prämisse enthalten.

Tabelle 2 gibt einen Überblick über die Zahl der benötigten Fragen abhängig von der Zahl der Objekte im Bezugssystem. Zusätzlich ist die Zeit angegeben, die für diese Fragenzahl benötigt würde, wenn eine Antwortrate von vier Fragen pro Minute angenommen wird.

Für einen diagnostischen Einsatz, wie er in dem einleitenden Beispiel zum Statistikverständnis skizziert wurde, erscheint also höchstens eine Zahl von sechs

Tabelle 2: Zahl der nötigen Fragen bei einer Komplettbefragung für verschiedene Zahlen von Objekten in X und die dafür benötigte Zeit bei einer Antwortrate von vier Antworten/min

Objektzahl	6	8	10	12
Fragenanzahl	186	1016	5110	24564
Zeit (in Stunden)	.7	4.23	21.3	102.4

Objekten als noch akzeptabel, da für eine Befragungsdauer von mehr als einer Stunde mit starken Ermüdungserscheinungen zu rechnen ist. Will man das Befragungsverfahren zudem für diagnostische Zwecke, wie z.B. in dem einleitenden Statistik-Beispiel, verwenden, so sind mehr Objekte nötig. Zum Beispiel wird die Zahl der Objekte, die man benötigt um im Bereich der Statistik das Verständnis der relevanten Konzepte zu testen, weit höher als zehn liegen.

Weiterhin ist zu bedenken, daß Fragen mit sehr großer Prämissengröße (mehr als drei oder vier Objekte) möglicherweise nur noch schwer zu beantworten sind. Zudem könnten sich für diese Effekte verstärken, die zu Transitivitätsverletzungen führen können, wie z.B. der Kriterienwechsel.

Heller (2000) zeigt, daß sich auch zu einer empirischen Relation, die sich ergibt, wenn nur Fragen bis zu einer bestimmten Prämissengröße gestellt werden, unter bestimmten Bedingungen korrespondierende Begriffsräume bestimmen lassen. Er führt dazu eine modifizierte Form der Entailment-Relation ein, die nur noch Elemente bis zu einer bestimmten Prämissengröße beinhaltet. Die hieraus resultierenden Begriffsräume beinhalten allerdings möglicherweise mehr Begriffe als Räume, die sich aus einer Komplettbefragung ergeben hätten.

In der Anwendung (z.B. bei Ulbrich, 1997) wird typischerweise mit Prämissen

Tabelle 3: Zahl der Fragen für die Komplettbefragung bei verschiedenen Schranken für die Prämissengröße und verschiedenen Zahlen von Objekten

Prämissen- größe	Objektzahl			
	6	8	10	12
1	30	56	90	132
2	90	224	450	792
3	150	504	1290	2772
4	180	786	2550	6732
5	186	953	3810	12276

mit maximal zwei Objekten gearbeitet. Doch sind beliebige Beschränkungen denkbar. Einen Überblick über die Anzahl der der Fragen, die bei verschiedenen Grenzen für die Prämissengröße und verschiedenen Zahlen von Objekten benötigt werden, gibt Tabelle 3.

Wie zu sehen ist, ist selbst bei einer Beschränkung auf einen maximale Prämissengröße von zwei eine Befragung mit 12 Objekten zeitlich nur noch schwer realisierbar. Eine weitere Senkung der Fragenzahl ist nur noch dadurch möglich, daß Objekte gewählt werden, die zueinander in keiner „Oberbegriff-Unterbegriff“-Beziehung stehen. Dadurch kann man die Fragen mit einem Objekt in der Prämisse einsparen.

Wird schließlich auf die eine oder andere Weise eine empirische Relation Q erhoben, so ist zu prüfen, ob diese die Bedingungen einer Entailment-Relation, d.h. vor allem die Transitivität, erfüllt. Verletzen die Antworten der Versuchsperson z.B. die Transitivität, so läßt sich zu der empirischen Relation kein korrespondierender Begriffsraum bestimmen. Allerdings könnte es sein daß der Versuchsperson z.B. aufgrund von Ermüdung zufällige Antwortfehler unterlaufen. Möglicherweise würden die Antworten, die die Versuchsperson ohne

diese Fehler geben würde, die Transitivitätsforderung erfüllen.

Um Antwortfehler, die möglicherweise zu Transitivitätsverletzungen geführt haben, zu korrigieren, wurde z.B. bei Ulbrich (1997) die Befragung zweimal durchgeführt und die Fragen, die unterschiedlich beantwortet wurden, wurden ein drittes Mal zur Beantwortung vorgelegt. Die Ergebnisse aus den drei Befragungen werden zu einer einzigen empirischen Relation kombiniert, für die überprüft werden muß, ob sie die Bedingungen einer Entailment-Relation erfüllt. Ist dies der Fall, kann ein korrespondierender Begriffsraum bestimmt werden. Liegen nur wenige Verletzungen der Transitivität vor, die möglicherweise nur aus zufälligen Antwortfehlern resultieren, so kann versucht werden, die empirische Relation durch möglichst wenige Antwortumkehrungen in eine Entailment-Relation zu überführen. Korrespondierend zu dieser läßt sich dann ein Begriffsraum bestimmen.

Abgesehen davon, daß sich die insgesamt nötige Zahl an Fragen sich bei dieser Vorgehensweise mehr als verdoppelt, stellt sich die Frage, welcher Zahl an Transitivitätsverletzungen noch als durch zufällige Antwortfehler bedingt angesehen werden kann. Man sucht also nach einem Kriterium, das angibt, ob es noch zulässig ist, durch Antwortumkehrungen eine Entailment-Relation zu erzeugen, oder ob die Verletzungen der Transitivität als systematisch angesehen werden muß. Eine statistische Theorie, die eine solche Analyse von Verletzungen gestatten würde, gibt es bisher allerdings nicht.

Eine konzeptionell andere Herangehensweise zur weiteren Reduzierung der Fragenzahl und zum Umgang mit „zufälligen Antwortfehlern“ wird bei dem im Folgenden vorgestellten QUERY-Verfahren gewählt.

4.2 QUERY

Das Verfahren QUERY, das hier vorgestellt werden soll, wurde von Koppen und Doignon (1988; Koppen, 1993, 1994) vorgeschlagen. Parallel wurde das Verfahren von Dowling (1993a, 1993b; Müller, 1989) aufgegriffen und ausgearbeitet. Es stammt aus dem Gebiet der Wissensraum-Theorie und wurde zur Befragung von Experten zum Ermitteln von Wissensstrukturen entwickelt. Da aber zwischen Begriffs- und Wissensstrukturen eine eins-zu-eins-Korrespondenz besteht (Heller, 2000), kann es einfach auf das Gebiet der Begriffsbefragung übertragen werden.

Die Idee bei diesem Verfahren ist es, aus den Antworten der Versuchsperson Schlußfolgerungen zu ziehen. Das Ziehen von Schlußfolgerungen bei der Erhebung einer empirischen Relation Q wird dadurch möglich, daß *vorausgesetzt* wird, daß es sich um eine Entailment-Relation handelt. Wird diese sicher nicht triviale Annahme gemacht, so ergeben sich verschiedene Arten von Schlußfolgerungen (nach Kaluscha, 1994):

Die Fragen, für die die Antwort „Nein“ geschlußfolgert werden kann, fallen in zwei Gruppen:

1. Sei X eine Menge von Objekten so gehören der Menge der *Tautologien* alle Implikationen (A, b) an, für die gilt

$$A \subseteq X, A \neq \emptyset \quad \text{und} \quad b \in A$$

2. Sei Q die Relation, die die Fragen auf der Menge von Objekten X enthält, die mit „Nein“ beantwortet wurden, oder für die die Antwort „Nein“ geschlußfolgert wurde, dann ist jedes (A, b) mit $A \subseteq X$ und $b \in X$ eine Schlußfolgerungen nach dem *modus barbara* wenn gilt

$$(B, b) \in Q \quad \text{und} \quad \forall a \in B \quad (A, a) \in Q$$

Diese beiden Arten von Schlußfolgerungen greifen direkt auf die Bedingung 1 (für die Tautologien) und die Bedingung 2 (für den modus barbara) aus der Definition der Entailment-Relation (Definition 2) zurück.

Auch „Ja“-Antworten lassen sich aus den bisherigen tatsächlich abgegebenen und geschlußfolgerten Antworten eines Individuums ableiten. Derartige „Ja“-Schlußfolgerungen erhält man auf folgendem Weg: Für jede ungestellte Frage wird versuchsweise die Antwort „Nein“ angenommen. Aus dieser vorläufigen „Nein“-Antwort werden die möglichen „Nein“-Schlußfolgerungen gezogen. Wurde eine der Fragen, für die damit ein „Nein“ als Antwort geschlußfolgert wird, zuvor schon mit „Ja“ beantwortet (tatsächlich oder geschlußfolgert), so ergibt sich ein Widerspruch. Die einzige Möglichkeit um zu verhindern, daß durch eine spätere „Nein“-Antwort des Individuums auf die untersuchte Frage dieser Widerspruch tatsächlich entsteht, besteht darin, anzunehmen, daß sie mit „Ja“ beantwortet wird. Es wird also eine „Ja“-Schlußfolgerung gezogen.

Für diese Ermittlung der „Ja“-Schlußfolgerungen werden beide Bedingungen aus der Definition der Entailment-Relation für die empirische Relation als erfüllt angesehen.

Allgemein muß die Vorgehensweise, Bedingungen einer Entailment-Relation bei einer empirischen Relation als gegeben anzusehen, kritisch betrachtet werden. Auf die Abfrage der Tautologien wird schon bei den praktizierten Verfahren zur Komplettbefragung verzichtet. Die oben skizzierten zugrundeliegenden Annahmen können relativ unproblematisch gemacht werden. Die Schlußfolgerungen nach dem „modus barbara“, sind allerdings nicht als selbstverständlich zu akzeptieren. Sie setzen nämlich die Transitivität voraus, die eigentlich für eine empirische Relation gefordert wird, damit ein korrespondierender Begriffsraum bestimmt werden kann. Erfüllt also eine empirische Relation Q die Forderung der Transitivität nicht, so wird durch das QUERY-Verfahren trotz-

dem eine Relation geliefert, die die Bedingung erfüllt. Das liegt daran, daß das Verfahren so ausgelegt ist, daß Fragen, deren Beantwortung zu Intransitivitäten führen könnte, nicht gestellt werden und automatisch als Transitivitäts-konform beantwortet angenommen werden.

Vom Standpunkt der repräsentationalen Meßtheorie ist diese Vorgehensweise deshalb sehr fragwürdig, da die Bedingungen für die Zulässigkeit einer Repräsentation der Antworten eines Individuums durch mathematische Objekte nicht mehr überprüft werden können. So gesehen kann das QUERY-Verfahren also nur verwendet werden, wenn gesichert ist, daß die empirische Relation, die man eigentlich von einem Individuum erhalten würde, z.B. durch eine Komplettbefragung, die Bedingungen einer Entailment-Relation erfüllt.

Bei der Verwendung des QUERY-Verfahrens gibt es ein weiteres Problem:

Wie sich bei der Durchführung der Komplettbefragung (z.B. Ulbrich, 1997) gezeigt hat, unterlaufen Individuen bei der Befragung zufällige Antwortfehler. Diese können z.B. auf Ermüdung zurückzuführen sein. Bei der Komplettbefragung wird versucht, diese durch mehrfache Befragung zu isolieren. Beim QUERY-Verfahren werden dagegen auch aus fehlerhaften Antworten Schlußfolgerungen gezogen. Ein Korrektur ist nicht mehr möglich, da Fragen, die zu Widersprüchen führen könnten, nicht mehr gestellt werden. Gerade die Anfangsphase der Befragung, in der die Versuchsperson mit der Aufgabe sowieso noch nicht sehr vertraut ist, ist unter diesem Gesichtspunkt sehr problematisch, da in ihr meist besonders viele Schlußfolgerungen gezogen werden.

Aus zufälligen Antwortfehlern ergeben sich also mit dem QUERY-Verfahren bedingt durch die Schlußfolgerungen Begriffsräume, die sich vom eigentlichen Raum, wie man ihn z.B. durch eine Komplettbefragung hätte erhalten können, mehr oder weniger unterscheiden. Zum Beispiel vergleicht Ulbrich (1997) für jedes Individuum die Begriffsräume, die aus der Komplettbefragung und

dem QUERY-Verfahren resultieren. Es zeigt sich, daß die Räume aus beiden Verfahren nicht übereinstimmen. Auch Simulationsstudien zeigen, daß das QUERY-Verfahren anfällig gegenüber zufälligen Antwortfehlern ist (Schrepp und Held, 1995).

Ein Weg, um zu möglicherweise verlässlichen Ergebnissen zu kommen, besteht darin, zwei bis drei Befragungen durchführen, und die daraus resultierenden Ergebnisräume zu kombinieren. Abgesehen von der Frage, wie mehrere Ergebnisräume zu integrieren sind, ergibt sich daraus aber insgesamt eine starke Erhöhung der Zahl der nötigen Fragen, was die Vorteile des QUERY-Verfahrens gegenüber der Komplettbefragung wieder mindern würde.

Viele der Untersuchungen, die sich mit der Zuverlässigkeit des QUERY-Verfahrens beschäftigt haben (z.B. Kambouri, Koppen, Villano und Falmagne 1994; Schrepp und Held, 1995) kommen aufgrund dieser Probleme zu dem Schluß, daß das QUERY-Verfahren gegen zufällige Antwortfehler besser abgesichert werden müßte. Eine solche Erweiterung stellt das im Anschluß vorgestellte PSQUERY-Verfahren dar.

4.3 PSQUERY

Beim hier vorgestellten PSQUERY-Verfahren („PS“ für „pending status“) handelt es sich um eine Erweiterung des QUERY-Verfahrens. Die Idee, wie sie z.B. in Kambouri et al. (1994) dargestellt wird ist folgende:

Um das QUERY-Verfahren gegen zufällige Antwortfehler abzusichern, werden Antworten der Versuchsperson und die Schlußfolgerungen, die sich aus diesen nach den oben vorgestellten Regeln ergeben, nicht sofort als Elemente der empirischen Relation Q übernommen. Sie werden stattdessen in einen Zwischenspeicher geschrieben. Sie sind in einem „pending status“. Um aus

diesem Zwischenzustand in die Relation Q übernommen zu werden, müssen sie bestätigt werden. Wie diese Bestätigung auszusehen hat, ist prinzipiell in weiten Grenzen variierbar. Zum Beispiel kann man festlegen, daß eine (möglicherweise schlußgefolgerte) Antwort in die Relation Q übernommen wird, wenn sie ein zweites Mal in den Zwischenspeicher geschrieben werden würden. Diese Variante wird in den bisherigen Arbeiten zum PSQUERY-Verfahren vorgeschlagen und verwendet.

Eine genaue Ausarbeitung des PSQUERY-Algorithmus geben Cosyn und Thierry (in Vorbereitung) an:

Der Algorithmus verwaltet vier Speicher für Antworten: zwei für die als sicher angenommenen „Nein“- und „Ja“-Antworten (Q bzw. \bar{Q}) und zwei für „Nein“- und „Ja“-Antworten auf der „Warteliste“.

Eine zentrale Rolle spielt der *transitive Abschluß*. Diese Abschlußoperation erzeugt auf Basis einer Antwort-Relation eine Entailment-Relation, indem die dazu zusätzlich nötigen Antworten hinzugefügt werden.

Nach jeder „Nein“-Antwort des befragten Individuums wird diese vorläufig zu der Relation mit den „Nein“-Antworten hinzugefügt und der transitive Abschluß wird gebildet. Die soeben hinzugefügte Antwort und die aus dem Abschluß resultierenden (mit Schlußfolgerungen gleichzusetzenden) Antworten werden allerdings nicht in der Relation belassen. Sie werden in den Zwischenspeicher für „Nein“-Antworten geschrieben. Zuvor muß allerdings sichergestellt werden, daß keine der neuen Antworten sich im Zwischenspeicher für die „Ja“-Antworten befinden. Ist dies der Fall, werden die entsprechenden „Ja“-Antworten entfernt und die „Nein“-Antworten werden nicht in den Zwischenspeicher geschrieben. Wenn es keine derartigen Widersprüche gibt und „Nein“-Antworten, die in den zugehörigen Zwischenspeicher geschrieben wer-

den sollen, in diesem schon enthalten sind, werden diese in die endgültige Relation Q übernommen. Für diese muß anschließend ein transitiver Abschluß gebildet werden, damit sie auch weiterhin die Forderungen einer Entailment-Relation (siehe Definition 2) erfüllt.

Ähnlich wird mit „Ja“-Antworten und „Ja“-Antwort-Schlußfolgerungen verfahren. Allerdings werden aus „Ja“-Antworten direkt keine Schlußfolgerungen abgeleitet. Zu „Ja“-Schlußfolgerungen kommt es, wenn der Algorithmus feststellt, daß es für eine ungestellte Frage zu Widersprüchen kommen würde, wenn diese mit „Nein“ beantwortet werden würde. Die Frage wird dann als mit „Ja“ beantwortet gesetzt. Auf diesem Weg wird eine „Ja“-Schlußfolgerung gezogen.

Durch den transitiven Abschluß wird gewährleistet daß in jedem Schritt eine Entailment-Relation zur Verfügung steht und so z.B. die Befragung bei einer gewissen Prämissengröße abgebrochen werden und trotzdem einen Begriffsraum bestimmt werden kann.

Simulationsstudien im Bereich der Wissensraumforschung (Cosyn und Thiery, in Vorbereitung) haben gezeigt, daß dieses Verfahren, wie gewünscht, weniger anfällig für zufällige Antwortfehler ist.

5 Experimenteller Vergleich

5.1 Fragestellung

In den vorangegangenen Abschnitten wurden der Ansatz von Heller (1991, 2000) zu Bestimmung von individuellen Begriffsstrukturen und darauf aufbauend die Befragungsverfahren QUERY und PSQUERY vorgestellt. Das Verfahren QUERY wurde schon zu diesem Zweck eingesetzt und mit der Komplettbefragung verglichen (Ulbrich, 1997). Dabei zeigten sich die im letzten Abschnitt angedeuteten Probleme bei zufälligen Antwortfehlern, die dazu führen, daß bei der Verwendung des QUERY-Verfahrens möglicherweise mehrere Befragungen nötig sind.

Bei einem Einsatz als diagnostisches Instrument ist allerdings eine mehrfache Befragung unerwünscht. Es sollte eine erste Befragung genügen. Von dieser wird verlangt, daß ihre Ergebnisse hinreichend mit denen übereinstimmen, die eine zweite Befragung liefern würde. Das Verfahren PSQUERY könnte zu einer Lösung oder zumindest einer Verringerung der Probleme mit zufälligen Antwortfehlern beitragen. Es wurde bisher allerdings nur im Kontext der Wissensraumtheorie eingesetzt und zudem stammen die bisherigen Ergebnisse nur aus Simulationsstudien.

Eine empirische Untersuchung der Vorteile des PSQUERY-Verfahrens gegenüber dem QUERY-Verfahren im Bereich der Begriffsraum-Forschung steht also bisher noch aus. Diesen Beitrag will die vorliegende Arbeit liefern.

Für die experimentelle Untersuchung ergeben sich folgende Fragen bzw. Hypothesen

1. Sind die Räume, die mit dem PSQUERY-Verfahren aus einer Erst- und einer Zweitbefragung resultieren, ähnlicher als die, die man mit dem

QUERY-Verfahren erhält?

Es wird die Hypothese aufgestellt, daß dies der Fall ist.

2. Wie groß ist die Zahl der Fragen, die für das PSQUERY-Verfahren benötigt werden, im Vergleich zum QUERY-Verfahren?
3. Wie groß sind die Unterschiede, die sich zwischen zwei Befragungen für Begriffsräume aus dem PSQUERY-Verfahren ergeben?
4. Zeigen sich in den Daten Muster, die Rückschlüsse auf die Ursachen für die Unterschiedlichkeit der Begriffsräume aus zwei Befragungen geben können?

5.2 Methode

5.2.1 Befragungssoftware

Die bisherigen Untersuchungen zu Befragungsverfahren im Bereich der Begriffsstrukturen (z.B. Ulbrich, 1997) wurden mit einem auf dem PXL-System (Irtel, 1997) aufsetzenden und von Heller (2000) implementierten System, das die Komplettbefragung ermöglicht, und dem Programmpaket KQ (Kaluscha, 1994) aus dem Bereich der Wissensräume, das das QUERY-Verfahren zu Verfügung stellt, durchgeführt.

Für das PSQUERY-Verfahren stand bis zum heutigen Zeitpunkt keine öffentlich zugängliche Implementation zur Verfügung. Eine Möglichkeit wäre also gewesen, eines der bestehenden Systeme um das PSQUERY-Verfahren zu erweitern. Dies wurde verworfen, da die existenten Systeme weiterhin folgende Schwächen haben:

1. Sie sind auf MS-DOS bzw. auf bestimmte Windows-Varianten angewiesen. Eine Verwendung unter anderen Betriebssystemen oder auch nur neueren Windows-Varianten ist deswegen nur schwer oder nicht möglich.
2. Die Benutzerschnittstelle ist, an MS-DOS-Gegebenheiten ausgerichtet, textbasiert. Der Komfort einer grafischen Benutzeroberfläche fehlt. Dies könnte zu einer geringeren Benutzerakzeptanz führen.
3. Die Befragung selbst läuft teilweise sehr hastig ab, da die Versuchspersonen durch fehlende Rückfragen, Wartezeiten, etc. dazu neigen, die Fragen sehr schnell zu beantworten.

Aufgrund der angeführten Probleme entschied sich der Autor zu einer kompletten Neuimplementation. Diese zeichnet sich durch folgende Merkmale aus:

1. Die Benutzerschnittstelle entspricht den üblichen Standards und verwendet Fenster, Buttons, etc.
2. Nach jeder Antwort erfolgt eine Rückfrage und eine Pause, was zu einem weniger hastigen Antwortverhalten führen soll.
3. Der Benutzer hat die Möglichkeit, Fragen zu ignorieren und selbstbestimmt Pausen zu machen. Dies soll zu erhöhter Benutzerakzeptanz und Motivation führen.
4. Die verwendete Programmiersprache ist *Java*. Daraus resultiert eine Verwendbarkeit unter allen gängigen Betriebssystemen.
5. Es wurde Wert auf eine strenge Anwendung des objektorientierten Programmierparadigmas gelegt. So sind z.B. wichtige Komponenten des formalen Rahmens der Theorie, wie Mengen und ihre Elemente, Fragen und Antworten, Relationen (mit der Möglichkeit der Bildung eines transitiven Abschlusses) und Begriffs- und Wissensräume, durch eigenen Klassen repräsentiert.
6. Auch die Komponenten der Benutzerschnittstelle, der Protokollführung und allgemein der Dateiverwaltung wurden modular ausgelegt, so daß es leicht ist verschiedene Werkzeuge für den Umgang mit Begriffsdaten zu entwickeln. So wurden z.B. im Rahmen dieser Arbeit Programme zur Bestimmung von Begriffsräumen aus Entailment-Relationen und zur statistischen Auswertung der angefallenen Daten entwickelt.
7. Die vorhandenen Klassen für den Umgang mit Dateien wurden so ausgelegt, daß sie einen Datenaustausch mit anderen Softwarepaketen aus demselben Forschungsbereich ermöglichen.

Die Implementation des QUERY-Verfahrens wurde unter Rückgriff auf die Arbeiten von Dowling (1993a, 1993b) und Kaluscha (1994) realisiert. Der PSQUERY-Algorithmus wurde nach der Beschreibung von Cosyn und Thiery (in Vorbereitung) realisiert. Aufgrund von Rechenzeitproblemen mußte allerdings eine Implementierung des PSQUERY-Verfahrens gewählt werden, die mehr Fragen stellt als eigentlich nötig.

Eine Befragungssitzung läuft folgendermaßen ab:

Der Versuchsperson wird ein Begrüßungsbildschirm gezeigt, auf dem sie aufgefordert wird auf einen „Start“-Button zu klicken, um mit der Befragung zu beginnen. Die Fragen werden nun in dem Format präsentiert, wie es Abbildung 4 zeigt. Die Frage paßt sich dabei vom Text her der Anzahl der präsentierten Objekte an.

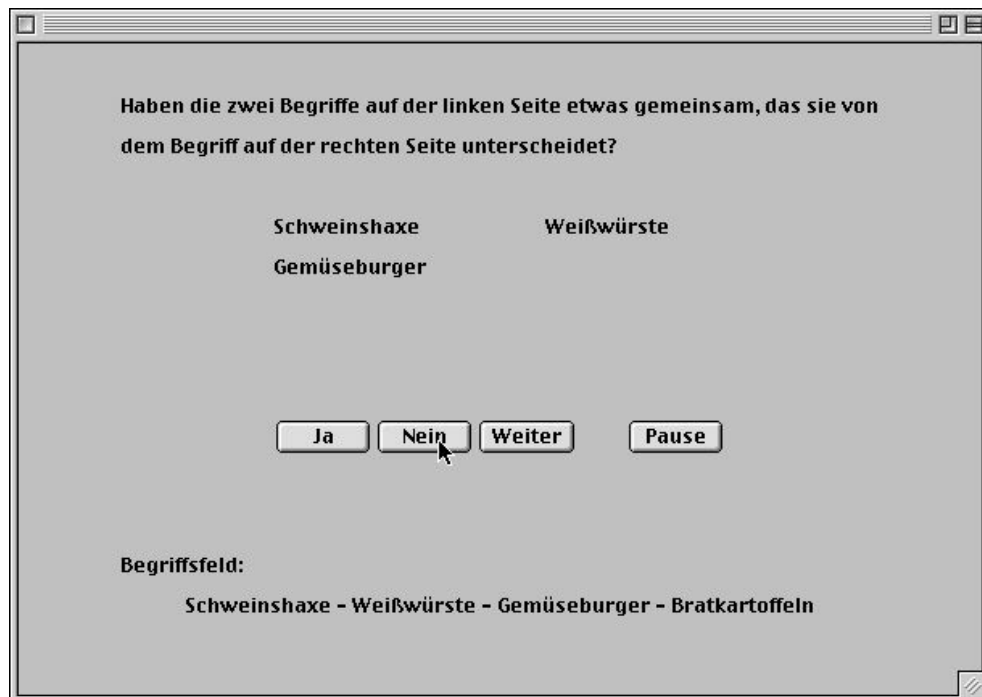


Abbildung 4: Anzeige einer Frage im Lauf der Vorbefragung

Die Versuchsperson, kann die Frage mit „Ja“ oder „Nein“ beantworten oder mit „Weiter“ zur nächsten Frage gehen, ohne die Frage zu beantworten. Es ist allerdings nicht sichergestellt, daß die nicht beantwortete Frage im Lauf der weiteren Befragung nicht wieder gestellt wird. Mit „Pause“ kann die Versuchsperson jederzeit selbstbestimmt auf eine Pausen-Anzeige umschalten, um nach einer beliebigen Zeit zur aktuellen Frage zurückzukehren.

Hat die Versuchsperson eine Frage mit „Ja“ oder „Nein“ beantwortet, erscheint eine Bestätigungs-Anzeige, wie sie exemplarisch in [Abbildung 5](#) dargestellt ist.

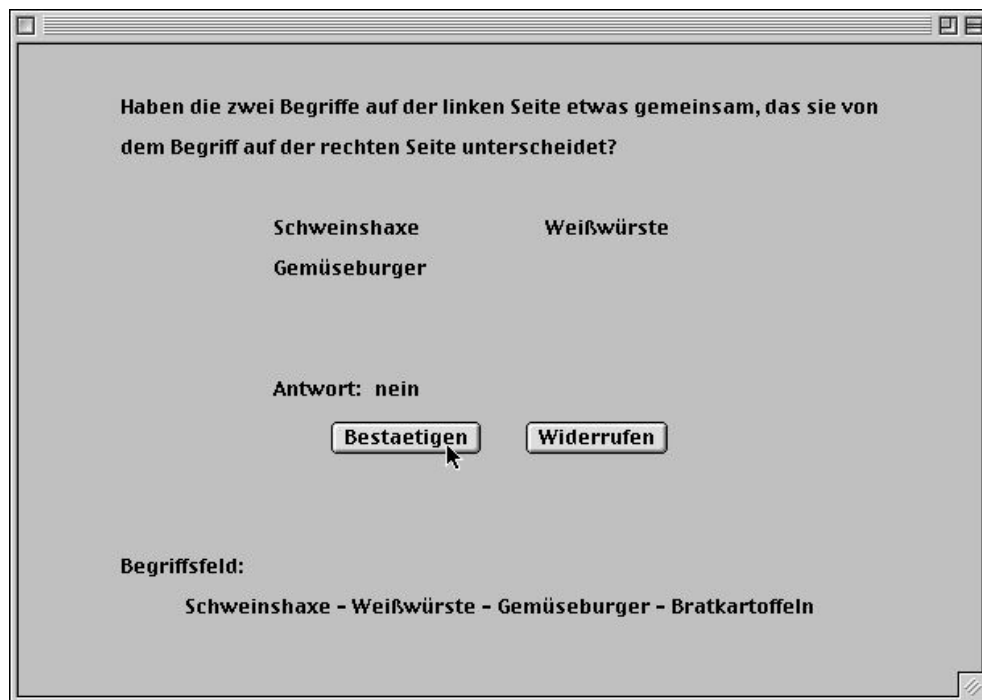


Abbildung 5: Bestätigungs-Anzeige, nachdem die angezeigte Frage mit „Nein“ beantwortet wurde

Die Versuchsperson kann mit „Bestätigen“ ihre Antwort bestätigen und zu nächsten Frage gelangen. Mit „Widerrufen“ kehrt sie zum Antwort-Dialog zurück und kann die Antwort nochmals beantworten. Nach der Beantwortung einer Frage erscheint für 1.5 Sekunden ein grauer Zwischenbildschirm

5.2.2 Ablauf der Befragung

Die Befragung erfolgte zu zwei Terminen. Zu jedem Termin wurden die Versuchspersonen mit beiden Befragungsverfahren (QUERY und PSQUERY) befragt. Da eine Eigenimplementation verwendet wurde, konnten beide Verfahren parallel zu Anwendung kommen, so daß die Versuchsperson nichts davon bemerkte.

Zu Beginn des ersten Termins erhielt sie eine schriftliche Instruktion (siehe Anhang A). In dieser wurde besonders darauf hingewiesen, das jeweilige Begriffsfeld vor Beginn der Befragung durchzugehen. An jedem Termin absolvierte die Versuchsperson zuerst eine Vorbefragung mit dem Begriffsfeld Nahrungsmittel mit vier Begriffen (siehe Tabelle 4), um mit der Aufgabe vertraut zu werden.

Tabelle 4: Begriffsfeld „Nahrungsmittel“ in der Vorbefragung

Schweinshaxe	Weißwürste
Gemüseburger	Bratkartoffeln

Anschließend fand jeweils die Hauptbefragung mit dem sechs Begriffe umfassenden Begriffsfeld Kleidung (siehe Tabelle 5) statt. Dieses wurde gewählt, da es sich schon in einer vorangegangenen Untersuchung (Ulbrich, 1997) als sehr arm an Konsistenzproblemen erwiesen hatte, und auch um die Vergleichbarkeit mit dieser Untersuchung zu gewährleisten. Allerdings wurden im Gegensatz zur Arbeit von Ulbrich (1997) nur sechs statt acht Begriffen verwendet, um die Gesamtbefragungszeit im Rahmen einer Stunde zu halten.

Im Anschluß an die zweite Befragung wurden die Versuchspersonen informell zu ihren Antwortkriterien befragt, um Anhaltspunkte für das Verständnis der Begriffsräume zu erhalten, auch wenn diese Arbeit soweit wie möglich auf der-

Tabelle 5: Begriffsfeld „Nahrungsmittel“ in der Hauptbefragung

Kleid	Hose
Jacke	Pullover
T-Shirt	Rock

artige Interpretationen verzichtet. Nach Ende der zweiten Sitzung wurden die Versuchspersonen weiterhin jeweils über Inhalt und Ziel der Untersuchung informiert. Wenn gewünscht, wurden ihnen auch ihre eigenen Begriffsräume gezeigt.

5.2.3 Stichprobe

Die Untersuchung wurde mit 34 Studenten der Universität Regensburg durchgeführt. Drei davon nahmen an Vorversuchen teil, 31 am Hauptversuch. Das Alter der 31 Versuchspersonen lag zwischen 19 und 43 Jahren (Median: 23), 24 sind weiblichen Geschlechts. 25 der 31 Teilnehmer waren Psychologiestudenten. Nur zwei gaben an, schon einmal an einer Untersuchung zum Thema „Begriffsstrukturen“ teilgenommen zu haben.

Die Befragungen fanden im Mai 2000 statt. Zwischen den beiden Sitzungen einer Versuchsperson lagen zwischen zwei und sieben Tage.

5.2.4 Unterschiedsmaß

Das Datenmaterial der oben beschriebenen Befragung besteht aus den jeweils zwei Begriffsräumen, die jedes der beiden Befragungsverfahren liefert. Benötigt wird also ein Maß, das den Unterschied zwischen den beiden Begriffsräumen aus je einem Verfahren quantifiziert.

In einer vorangegangenen Untersuchung (Ulbrich, 1997) wird das Differenzmaß $d(A, B)$ verwendet. Es berechnet sich für zwei Räume über deren Begriffsumfänge σ_1 und σ_2 aus dem Quotient aus der Mächtigkeit der symmetrischer Mengendifferenz (die Anzahl der Begriffe, die die beiden Räume jeweils nicht gemeinsam haben) und der Anzahl der Begriffe in beiden Räumen insgesamt:

$$d(\sigma_1, \sigma_2) = \frac{|\sigma_1 \Delta \sigma_2|}{|\sigma_1 \cup \sigma_2|} = \frac{|(\sigma_1 \setminus \sigma_2) \cup (\sigma_2 \setminus \sigma_1)|}{|\sigma_1 \cup \sigma_2|}$$

Der Wert dieses Quotienten liegt zwischen 0 und 1. Nach der Definition eines Begriffsraumes sind allerdings in jedem Raum die trivialen Begriffe (leerer Begriff und der Begriff, der alle Objekte aus X enthält) enthalten.

Da das Reizmaterial im vorliegenden Versuch so gewählt wurde, daß kein Objekt zu einem anderen in einer „Oberbegriff-Unterbegriff“-Beziehung steht, sind auch alle Begriffe, die jeweils nur ein Objekt enthalten in jedem Raum enthalten. D.h. jeder Raum umfaßt (bei n Objekten) mindestens $n + 2$ Begriffe. Für n Objekte (mit $n > 2$) ist also die maximale Differenz

$$d_{max}(n) = \frac{2^n - (n + 2)}{2^n}.$$

An dieser werden alle auftretenden Differenzen normiert, so daß sie theoretisch im ganzen Bereich zwischen 0 und 1 liegen könnten.

Um einen Eindruck zu vermitteln, in welchen Bereichen sich solche Differenzwerte bewegen, gibt Tabelle 6 einen Überblick. Gezeigt werden die Differenzwerte, die sich für den Vergleich von zwei Räumen über sechs Objekten ergeben, in Abhängigkeit von der Zahl der Begriffe in beiden Räume und der Zahl der Begriffe, in denen sie sich unterscheiden. Der Wertebereich der Differenzen in der Tabelle deutet schon darauf hin, daß wohl nicht mit Differenzen über .5 zu rechnen ist, da sich diese nur für Räume ergeben, die sich sehr stark unterscheiden. So sind zum Beispiel bei einer über zwei Befragungen

hinweg konstanten Zahl von 18 Begriffen in jedem Raum über 14 unterschiedliche Begriffe nötig, um solch einen Differenzwert zu erreichen. Ein derartiger Unterschied wäre aber inakzeptabel, d.h. die weitere Diskussion der Befragungsverfahren würde sich erübrigen. Tatsächlich zeigen sich z.B. bei Ulbrich (1997) mittlere Differenzwerte zwischen .23 und .29 über Befragungen hinweg.

Tabelle 6: Differenzwert für zwei Begriffsräumen über sechs Objekten in Abhängigkeit von der Zahl der Begriffe in beiden Räumen und der Zahl der verschiedenen Begriffe.

verschiedene Begriffe	Begriffe in Raum eins/zwei				
	12/12	14/12	16/12	18/18	22/18
1	.0476			.0317	
2	.0952	.0879		.0634	
3	.1429	.1319		.0952	
4	.1905	.1758	.1633	.1270	.1143
5	.2381	.2198	.2041	.1587	.1429
6	.2857	.2637	.2449	.1905	.1714
7	.3333	.3077	.2857	.2222	.2000
8	.3810	.3516	.3265	.2540	.2286
9		.3956	.3673	.2857	.2571
10		.4558	.4081	.3175	.2875
11			.4490	.3492	.3143
12			.4898	.3810	.3429
13				.4127	.3714
14				.4444	.4000

5.2.5 Inferenzstatistik

Für jede Versuchsperson i erhält man für jedes Befragungsverfahren eine normierte Differenz d_{QUERY_i} bzw. $d_{PSQUERY_i}$. Die Differenz dieser beiden Differenzen $d_{DIFF_i} = d_{QUERY_i} - d_{PSQUERY_i}$ bildet den Ausgangspunkt für die inferenzstatistische Betrachtung. Die korrespondierenden Zufallvariablen sollen im Folgenden mit δ_{QUERY_i} , $\delta_{PSQUERY_i}$ und δ_{DIFF_i} bezeichnet werden.

Die inhaltliche Nullhypothese, daß die beiden Verfahren QUERY bzw. PSQUERY sich nicht unterscheiden, führt zu der statistischen Nullhypothese, daß die Differenzwerte beider Verfahren δ_{QUERY_i} und $\delta_{PSQUERY_i}$ dieselbe Verteilung haben. Daraus läßt sich die testbare Nullhypothese ableiten, daß das Populationsmittel der Differenz zwischen diesen beiden Differenzen μ_{DIFF} gleich null ist. Da erwartet wird, daß das PSQUERY-Verfahren kleinere Differenzen liefert, wird eine gerichtete Null- und Alternativhypothese verwendet:

$$H_0 : \mu_{DIFF} \leq 0$$

$$H_1 : \mu_{DIFF} > 0$$

Um aufgrund eines Stichprobenmittels \bar{d}_{DIFF} entscheiden zu können, ob die Nullhypothese verworfen werden muß, müssen Verteilungsannahme über das Populationsmittel μ_{DIFF} gemacht werden.

Nach Rao (1973) ergibt sich nach dem „zentralen Grenzwertsatz“ für den Mittelwert von unabhängigen, gleich verteilten Zufallsvariablen für $n \rightarrow \infty$ die Normalverteilung. Es wird angenommen, daß es sich bei δ_{DIFF_i} um derartige Zufallsvariablen handelt. Für den Mittelwert μ_{DIFF} wird also für $n \rightarrow \infty$ eine Normalverteilungsannahme getroffen. Bei $n > 30$ kann diese Annahme schon als zulässig angesehen werden.

Da die Varianz $\sigma_{\bar{d}_{DIFF}}^2$ von \bar{d}_{DIFF} nicht bekannt ist und über

$$S_{\bar{d}_{DIFF}}^2 = \frac{1}{n-1} \sum_{i=1}^n (d_{DIFF_i} - \bar{d}_{DIFF})^2$$

geschätzt werden muß, ergibt sich für die statistische Prüfgröße eine t-Verteilung:

$$T = \frac{\bar{d}_{DIFF}}{S_{\bar{d}_{DIFF}}} \sqrt{n} \sim t(n-1)$$

Für 30 Freiheitsgraden (bei $n = 31$) ergibt sich für eine angestrebte Irrtumswahrscheinlichkeit $\alpha < .05$ ein kritischer Wert von 1.697 (siehe z.B. Rohatgi, 1976).

5.3 Ergebnisse

Ein Teil der Aufgabe dieser Arbeit besteht darin, Kennwerte für das PSQUERY-Verfahren zu ermitteln. In diesem Abschnitt werden deshalb zuerst Daten wie die durchschnittliche Zahl von Begriffen in den ermittelten Räumen, die mittlere Zahl der benötigten Fragen und die durchschnittlichen Differenzen zwischen den verschiedenen Begriffsräumen dargestellt. Die Ergebnisse für das QUERY-Verfahren, die parallel angegeben werden, bieten einen Referenzpunkt für die relative Beurteilung der Kennwerte des PSQUERY-Verfahrens.

Anschließend wird versucht, einen Überblick über die Vielfalt der individuellen Begriffsräume, die sich aus den Befragungen ergeben, zu vermitteln. Dazu werden exemplarisch Begriffsräume vorgestellt, die in den Räumen vieler anderer Versuchspersonen enthalten sind.

In der Fragestellung dieser Arbeit wird weiterhin Bezug genommen auf Muster in den Daten, die mögliche Schwächen und Verbesserungsmöglichkeiten des PSQUERY-Verfahrens aufzeigen. Deshalb wird im Anschluß ein „hypothesenkonformer“ Fall, d.h. das PSQUERY-Verfahren liefert hier die erwartete höhere Stabilität, und eine „nicht hypothesekonformer“ Fall dargestellt.

5.3.1 Deskriptive Statistiken

Bei der ersten Befragung benötigte das QUERY-Verfahren durchschnittlich 30.8 Fragen (SD: 5.96), das PSQUERY 85.1 Fragen (SD: 18.44), wobei aufgrund der oben angesprochenen Rechenzeitprobleme mehr Fragen gestellt werden mußten als nötig (Mittelwert: 111.6, SD: 25.80). Die vom QUERY-Verfahren erzeugten Räume enthalten durchschnittlich 17.7 Begriffe (SD: 4.55). PSQUERY liefert im Schnitt 16.2 Begriffe (SD: 5.06). Bei PSQUERY-Verfahren

mußten im Durchschnitt bei 7.96 Prozent (SD: 4.26) der reinen PSQUERY-Fragen aufgrund von Widersprüchlichkeiten Antworten aus dem Zwischenspeicher entfernt werden. Die Differenz zwischen den von QUERY und PSQUERY gelieferten Räumen beträgt im Mittel 0.179 (SD: 0.182).

Im Rahmen der zweiten Befragung waren für das QUERY-Verfahren im Mittel 30.6 Fragen (SD: 7.13) nötig, für das PSQUERY-Verfahren 84.5 Fragen (SD: 18.6). Insgesamt mußten im Durchschnitt 106.4 Fragen (SD: 20.4) gestellt werden. Die aus dem QUERY-Verfahren stammenden Räume der zweiten Befragung enthalten im Durchschnitt 17.5 Begriffe (SD: 4.23). Aus dem PSQUERY-Verfahren resultierten im Mittel 17.5 Begriffe (SD: 5.45). Aufgrund von Widersprüchlichkeiten mußten bei PSQUERY im Mittel bei 6.34 Prozent (SD: 5.14) der reinen PSQUERY-Fragen Antworten aus dem Zwischenspeicher entfernt werden. Die Differenz zwischen den vom QUERY-Verfahren und vom PSQUERY-Verfahren in der zweiten Befragung gelieferten Räumen beträgt im Mittel .190 (SD: .190).

Vergleicht man die aus jeweils einem Verfahren resultierenden Räume über beide Befragungen hinweg, so ergibt sich zwischen den QUERY-Räume eine durchschnittliche Differenz von .255 (SD: .103) und für die PSQUERY-Räume eine Differenz von 0.180 (SD: .100). Der mittlere Unterschied zwischen diesen beiden Differenzen beträgt .075 (SD: .125).

Abbildung 6 zeigt die mittleren Differenzen zwischen den vier Räumen, die aus der Befragung einer Person mit beiden Verfahren in zwei Sitzungen resultieren.

Die Daten der einzelnen Versuchspersonen, die in die Berechnung der hier angeführten Mittelwerte eingehen finden sich in Anhang B.

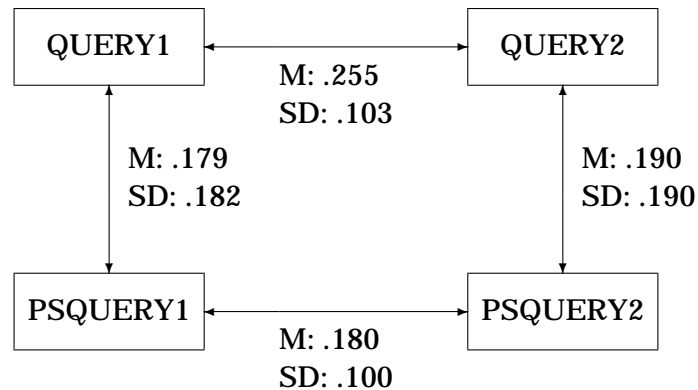


Abbildung 6: Differenzen zwischen den vier Begriffsräumen einer Person

5.3.2 Überblick über die Begriffsstrukturen

Im folgenden Abschnitt soll ein Überblick über die sich aus der Befragung ergebenden Begriffsräume gegeben werden.

Abbildung 7 zeigt eine Überblicksdarstellung der Begriffsräume aller Versuchspersonen, die sich bei der zweiten Befragung mit dem PSQUERY-Verfahren ergeben haben.

Geht man davon aus, daß die Versuchspersonen im Sinne einer latenten Konstrukts einen „wahren“ Begriffsräume besitzen, so dürften die Räume aus dieser Befragung diesen am wenigsten verfälscht wiedergeben, da einerseits das PSQUERY-Verfahren für die Korrektur von zufälligen Antwortfehlern ausgelegt ist und zu hoffen ist, daß die Individuen bei der zweiten Befragung schon geübter sind und deshalb weniger Fehler machen.

Die Überblicksdarstellung selbst ist folgendermaßen strukturiert: Jeder Raum ist durch einen Knoten dargestellt, der mit der Nummer der Versuchsperson beschriftet ist. Eine Ausnahme bilden identische Räume, die zusammen durch einen Knoten mit zwei oder mehr Nummern dargestellt sind. Beinhaltet ein Raum einen anderen, d.h. beinhaltet er alle Begriffe des anderen und noch

zusätzliche, so wird von ihm ein Pfeil zu dem Knoten des Raumes, den er beinhaltet, gezeichnet. Gibt es aber einen Raum, der zwischen diesen beiden liegt, d.h. der vom einen beinhaltet wird und gleichzeitig den anderen beinhaltet, so wird kein Pfeil gezeichnet.

Im Folgenden sollen einige prominente Räume aus der Überblicksdarstellung in Abbildung 7 gezeigt werden, um einen Überblick über die Bandbreite an empirisch vorkommenden Räumen zum Begriffsfeld „Kleidung“ zu geben. Um die Variabilität der Räume über die beiden Befragungen und Verfahren hinweg aufzeigen, wird zu jedem Raum noch ein zweiter Raum entweder aus der ersten PSQUERY-Befragung oder der ersten QUERY-Befragung angegeben.

Der Raum von Versuchsperson 29 (siehe Abbildung 8) ist in sehr vielen Räumen enthalten. Eine Besonderheit ist hier, daß sich in der ersten Befragung mit dem PSQUERY-Verfahren ein identischer Raum ergeben hat.

Der Raum, der aus dem QUERY-Verfahren in der ersten Befragung für diese Versuchsperson resultiert (siehe Abbildung 9) hat eine ähnliche Struktur wie die beiden PSQUERY-Räume. So beinhaltet er z.B. alle Begriffe, die sich auf „Rock“, „T-Shirt“ und „Kleid“ beziehen. Allerdings fehlen ihm die Begriffe {„Jacke“, „Hose“} und {„Pullover“, „Hose“}. Es ist z.B. vorstellbar, daß die Versuchsperson in einer frühen Phase der Befragung diese Begriffe nicht zur Anwendung gebracht hat, was dazu geführt hat, daß sie nur vom PSQUERY-Verfahren im späteren Verlauf der Befragung identifiziert werden konnten, nachdem das QUERY-Verfahren schon abgeschlossen war.

Ein weiterer PSQUERY-Raum aus der zweiten Befragung, der in vielen Räumen der anderen Versuchspersonen aus derselben Befragung enthalten ist, ist der von Versuchsperson 25 (siehe Abbildung 10).

Der PSQUERY-Raum den die erste Befragung für diese Versuchsperson liefert

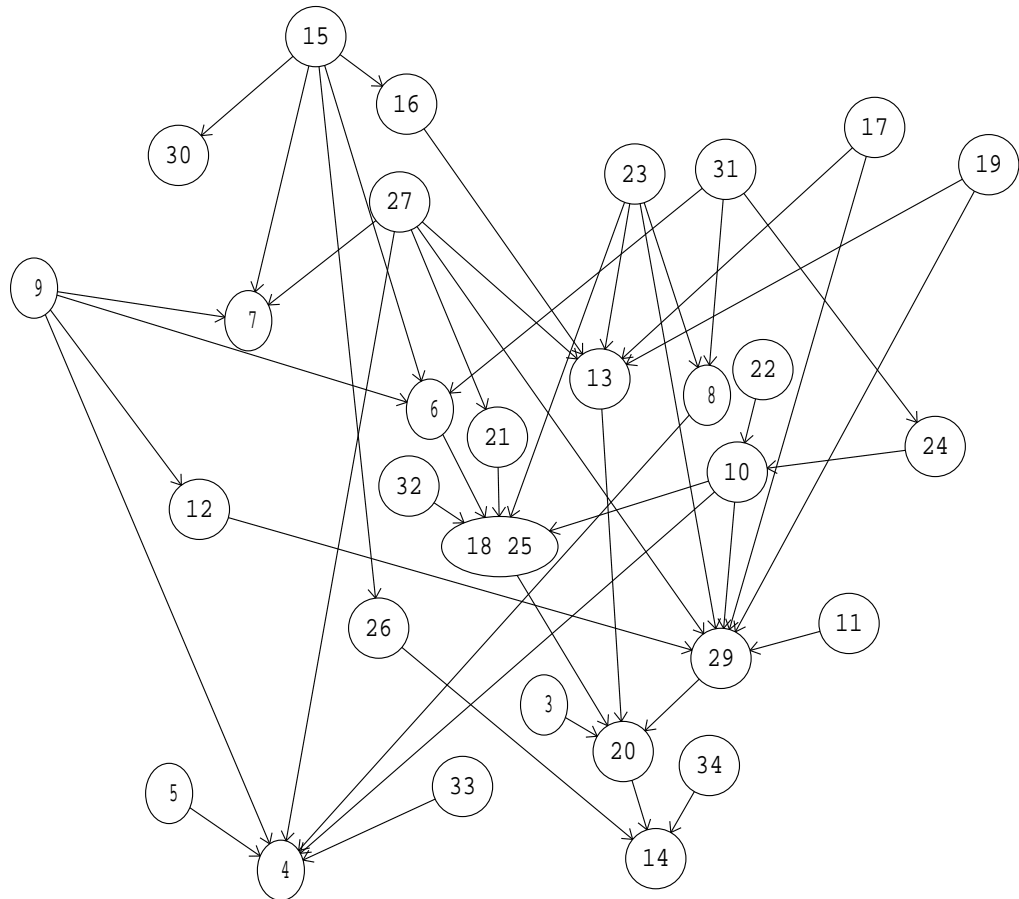


Abbildung 7: Überblicksdarstellung sämtlicher PSQUERY-Begriffsräume aus der zweiten Befragung

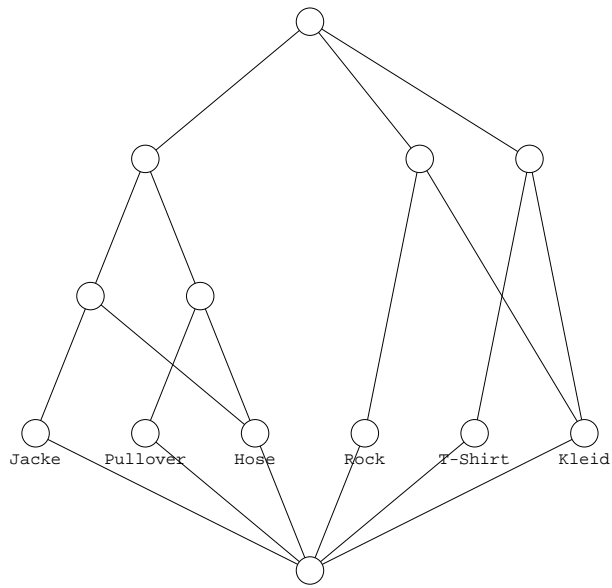


Abbildung 8: Begriffsraum der Versuchsperson 29 im zweiten Durchgang (PSQUERY-Verfahren)

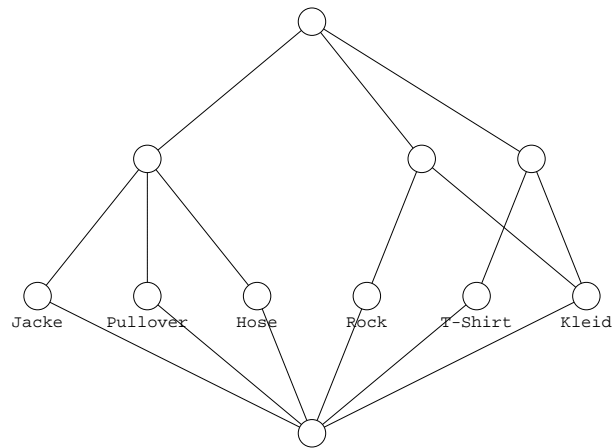


Abbildung 9: Begriffsraum der Versuchsperson 29 im ersten Durchgang (QUERY-Verfahren)

ist wiederum mit dem aus der zweiten Befragung identisch.

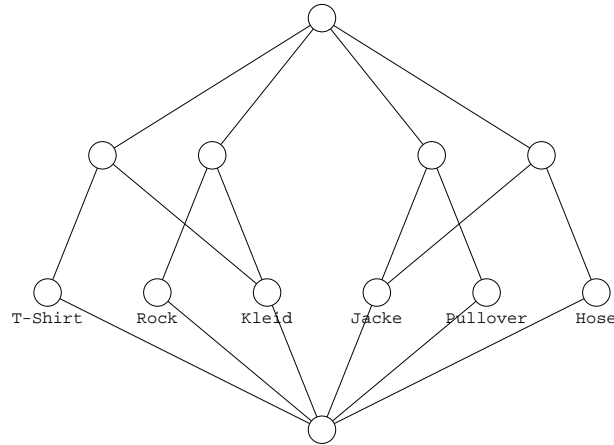


Abbildung 10: Begriffsraum der Versuchsperson 25 im zweiten Durchgang (PSQUERY-Verfahren)

Der Raum der aus der ersten Befragung dieser Versuchsperson mit dem QUERY-Verfahren resultiert (siehe Abbildung 11), ist in der Struktur den PSQUERY-Räumen sehr ähnlich. Allerdings beinhaltet er den zusätzlichen Begriff {„T-Shirt“, „Rock“, „Kleid“}. Statt dem Begriff {„Jacke“, „Hose“} enthält er {„Jacke“, „Pullover“, „Hose“}. Solche allgemeineren Begriffe treten z.B. auf, wenn eine Unterscheidung fälschlicherweise nicht getroffen wird, d.h. das Individuum irrtümlicherweise mit „Nein“ statt mit „Ja“ antwortet. Das PSQUERY-Verfahren ist auch hier auf den Ausgleich derartiger Fehler ausgelegt.

Ein dritter, mittelgroßer Raum, der Unterraum vieler anderer Räume ist, ist

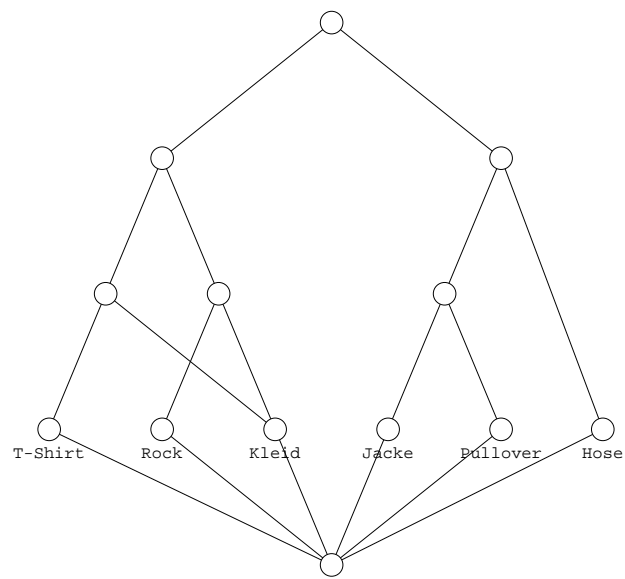


Abbildung 11: Begriffsraum der Versuchsperson 25 im ersten Durchgang (QUERY-Verfahren)

der von Versuchsperson 13 (zweite Befragung, PSQUERY-Verfahren), den Abbildung 12 zeigt.

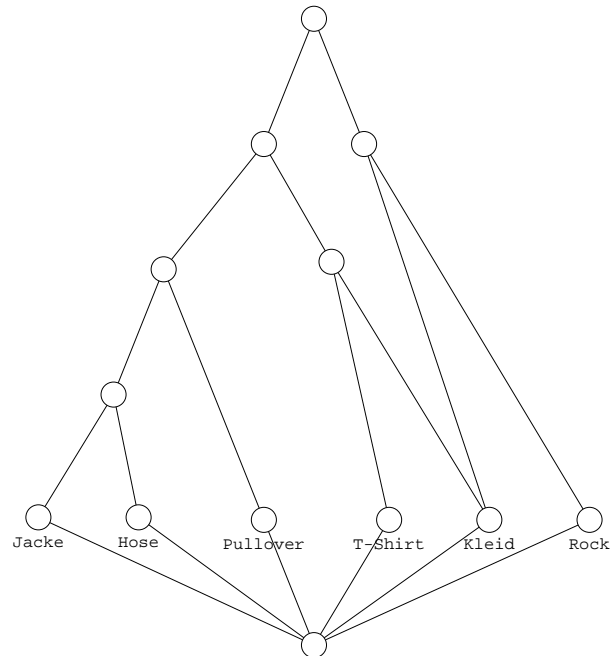


Abbildung 12: Begriffsraum der Versuchsperson 13 im zweiten Durchgang (PSQUERY-Verfahren)

Der PSQUERY-Raum aus der ersten Befragung (siehe Abbildung 13) enthält drei Begriffe, die dem Raum aus der zweiten Befragung fehlen. Diesem fehlt wiederum ein Begriff des Raumes aus der ersten Befragung, was insgesamt zu einem Differenzwert von 0.163 führt.

Für Versuchsperson 14 lieferte die zweite PSQUERY-Befragung einen sehr kleinen Raum (siehe Abbildung 14), der in den meisten anderen Räumen dieser Befragung enthalten ist. Der Raum enthält außer den trivialen Begriffen nur zwei weitere. Der Raum derselben Versuchsperson aus der ersten PSQUERY-

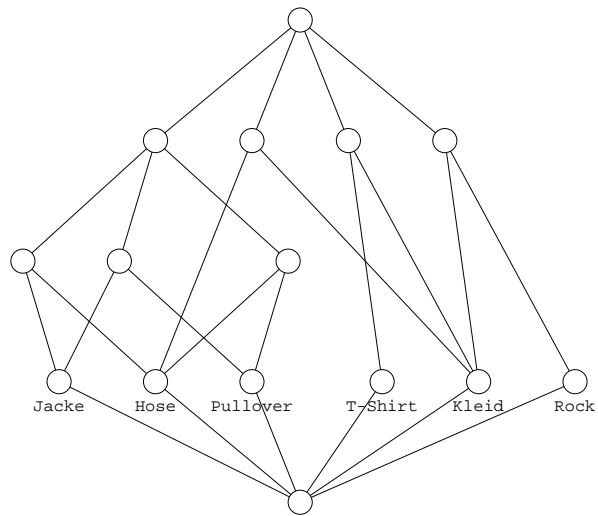


Abbildung 13: Begriffsraum der Versuchsperson 13 im ersten Durchgang (PSQUERY-Verfahren)

Befragung (siehe Abbildung 15) enthält dagegen zwei weitere Begriffe. Die Differenz zwischen den beiden Räumen beträgt danach also 0.276.

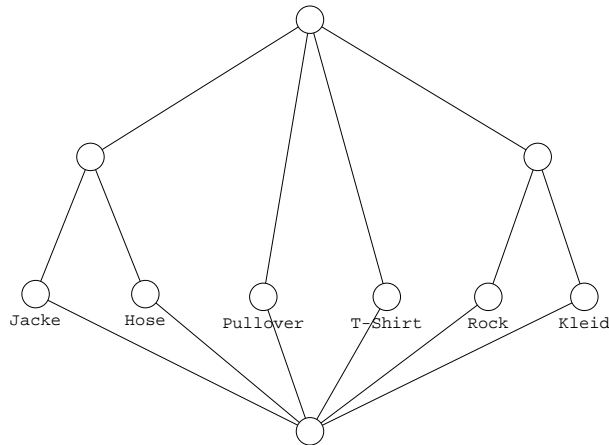


Abbildung 14: Begriffsraum der Versuchsperson 14 im zweiten Durchgang (PSQUERY-Verfahren)

Mit Ausnahme des Raums von Versuchsperson 30 enthalten alle Räume, die den Raum der Versuchsperson 14 nicht umfassen, den Raum von Versuchsperson 4. Diesen zweiten sehr kleine Raum zeigt Abbildung 16.

Der korrespondierende Raum aus der ersten PSQUERY-Befragung (siehe Abbildung 17) enthält den zusätzlichen Begriff {„Jacke“, „Pullover“, „Hose“}, was zu einem Differenzwert von 0.050 zum anderen PSQUERY-Raum führt.

Bei der zweiten Befragung mit dem PSQUERY-Verfahren enthalten also mit Ausnahme von Versuchsperson 30 alle Räume entweder den Raum der Ver-

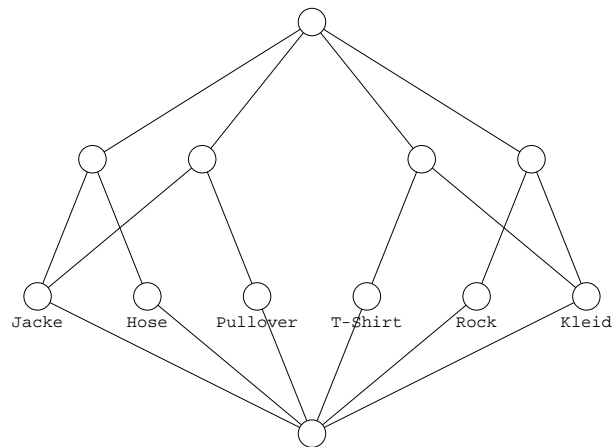


Abbildung 15: Begriffsraum der Versuchsperson 14 im ersten Durchgang (PSQUERY-Verfahren)

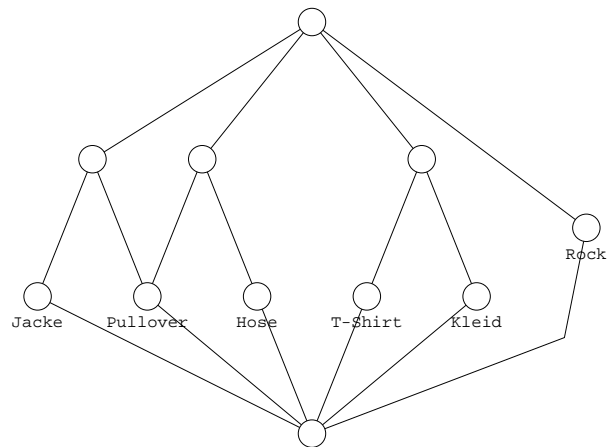


Abbildung 16: Begriffsraum der Versuchsperson 4 im zweiten Durchgang (PSQUERY-Verfahren)

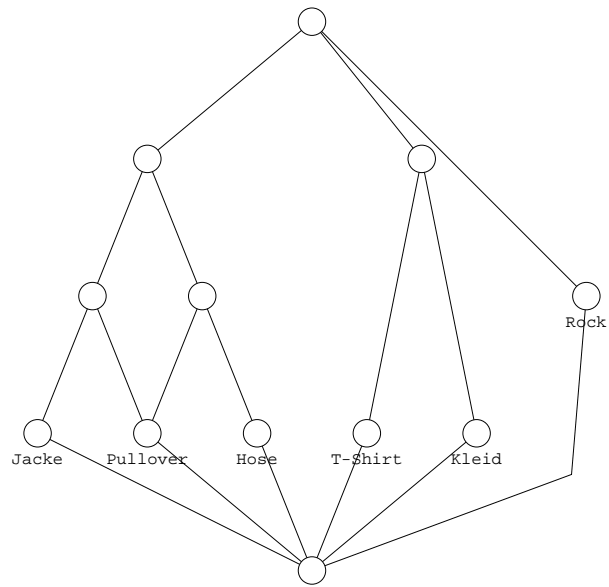


Abbildung 17: Begriffsraum der Versuchsperson 4 im ersten Durchgang (PSQUERY-Verfahren)

suchsperson 14 oder den der Versuchsperson 4. Das heißt, daß man die Versuchspersonen (bis auf Versuchsperson 30) in zwei Gruppen aufteilen kann:

Die Räume der einen Gruppe beinhalten mindestens die Begriffe {„Jacke“, „Hose“} und {„Rock“, „Kleid“}, die (nach informeller, nicht unbedingt repräsentativer Auskunft von Versuchspersonen) als „Außenbekleidung“ und „weibliche Bekleidung“ interpretiert werden könnten.

Die Räume der anderen Gruppe enthalten mindestens die Begriffe {„Jacke“, „Pullover“}, {„Pullover“, „Hose“} und {„T-Shirt“, „Kleid“}, die (nach Versuchspersonenauskunft) als „Kleidung für den Oberkörper“, „dicke Unterbekleidung“ und „leichte Überzieh-Bekleidung“ interpretiert werden könnten.

5.3.3 Darstellung ausgewählter Begriffsstrukturen

An dieser Stelle sollen exemplarisch die Daten von zwei Versuchspersonen ausführlicher dargestellt werden, auf die später im Lauf der Diskussion zurückgegriffen werden soll. Dazu wurden ein „hypothesenkonformer“ (Versuchsperson 25) und ein „nicht hypothesenkonformer“ Fall (Versuchsperson 9) ausgewählt.

Bei Versuchsperson 25 resultierten bei der ersten Befragung aus dem QUERY-Verfahren 13 Begriffe (bei 24 Fragen) und aus dem PSQUERY-Verfahren 12 Begriffe (bei 61 Fragen und 4 Widersprüchen). Die Differenz zwischen diesen beiden Räumen beträgt 0.137. Bei der zweiten Befragung benötigt das QUERY-Verfahren 21 Fragen und liefert 13 Begriffe. Das PSQUERY-Verfahren 52 Fragen bei 8 Widersprüchen und liefert 12 Begriffe. Die Differenz zwischen den beiden Räumen beträgt 0.0457. Die Differenz zwischen den beiden QUERY-Räumen beträgt 0.176. Die Differenz zwischen den beiden PSQUERY-Räumen ist 0. Sie sind identisch.

Die schon angeführte Abbildung 11 und Abbildung 18 zeigen die Hasse-Diagramme für die beiden QUERY-Räume der Versuchsperson 25. Die schon referierte Abbildung 10 zeigt den PSQUERY-Raum aus der ersten Befragung. Die Darstellung des Raumes aus der zweiten Befragung ist identisch.

Wie schon beschrieben beinhaltet der QUERY-Raum aus der ersten Befragung im Vergleich zu den PSQUERY-Räumen den zusätzlichen Begriff {„T-Shirt“, „Rock“, „Kleid“}. Statt dem Begriff {„Jacke“, „Hose“} enthält er {„Jacke“, „Pull-over“, „Hose“}, was auf eine irrtümlicherweise nicht getroffene Unterscheidung zurückzuführen sein könnte.

Der Begriffsraum, der aus der zweiten QUERY-Befragung resultiert, unterscheidet sich, wie es schon die niedrige Differenz von 0.0457 andeutet, nur geringfügig von den PSQUERY-Räumen. Er enthält nur den zusätzlichen Begriff {„T-Shirt“, „Rock“, „Jacke“, „Hose“}. Dieser könnte z.B. dadurch entstanden sein, daß den Objekten, die dieser Begriff enthält, vom QUERY-Verfahren aufgrund von zufälligen Antwortfehlern eine Gemeinsamkeit zugeschrieben wird.

Bei Versuchsperson 9 ergeben sich für die erste Befragung aus dem QUERY-Verfahren 25 Begriffe bei 43 Fragen. Das PSQUERY-Verfahren benötigt 110 Fragen bei 7 Widersprüchen und liefert 30 Begriffe. Die Differenz zwischen den beiden Räumen beträgt 0.229. Für die zweite Befragung ergeben sich beim QUERY-Verfahren 23 Begriffe bei 36 Fragen. Das PSQUERY-Verfahren benötigt bei 6 Widersprüchen 107 Fragen und liefert 21 Begriffe. Die Differenz zwischen den aus den beiden Verfahren stammenden Räumen beträgt 0.156. Über beide Befragungen hinweg betrachtet beträgt die Differenz zwischen den QUERY-Räumen 0.190 und die Differenz zwischen den PSQUERY-Räumen 0.336.

Die Begriffsräume der Versuchsperson 9 sind in Abbildung 19 bis 22 darge-

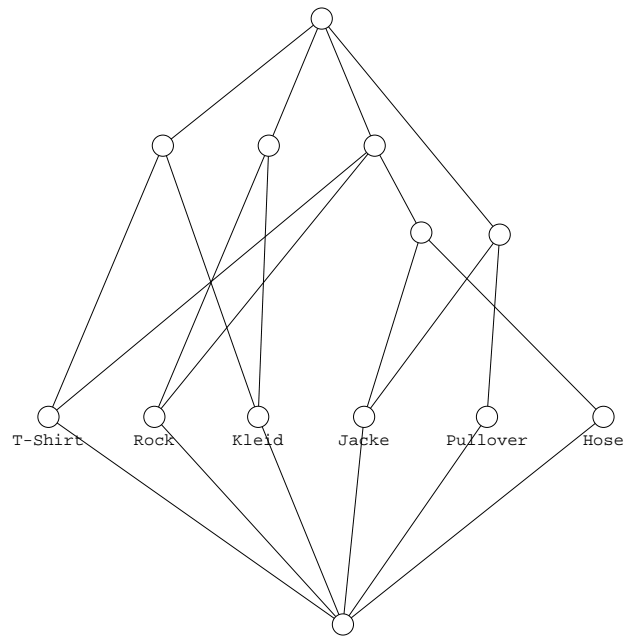


Abbildung 18: Begriffsraum der Versuchsperson 25 im zweiten Durchgang (QUERY-Verfahren)

stellt.

Die Vielzahl der Begriffe macht es schwer, visuell Ähnlichkeiten zwischen den Räumen zu erkennen. Wie der Differenzwert zwischen den beiden QUERY-Räumen von 0.190 andeutet, bestehen hier aber trotzdem große Ähnlichkeiten. Zum Beispiel liegen in beiden Räumen unter dem alle Objekte umfassenden Begriff jeweils drei Begriffe. Für den Begriff, der durch den linken der zugehörigen drei Knoten repräsentiert wird, werden die Objekte, die er umfaßt, in beiden Räumen durch zwei Begriffe unterschieden, die wiederum ähnlich strukturiert unterteilt werden.

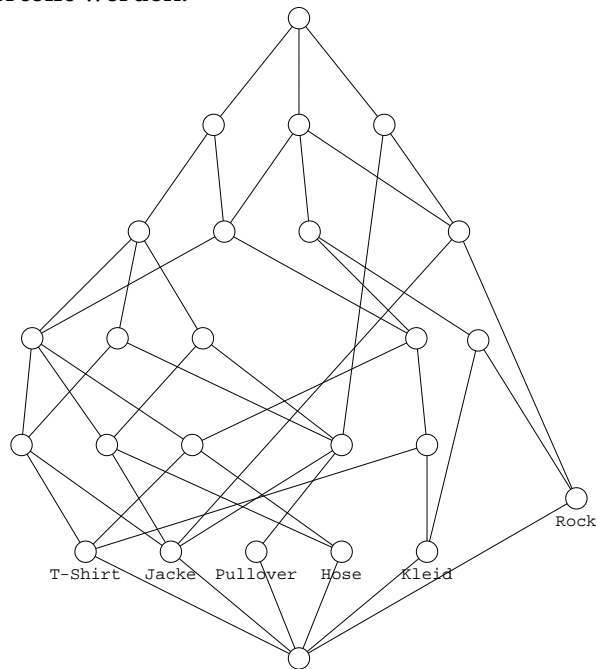


Abbildung 19: Begriffsraum der Versuchsperson 9 im ersten Durchgang (QUERY-Verfahren)

Allgemein ist zu den vier Räumen der Versuchsperson 9 anzumerken, daß

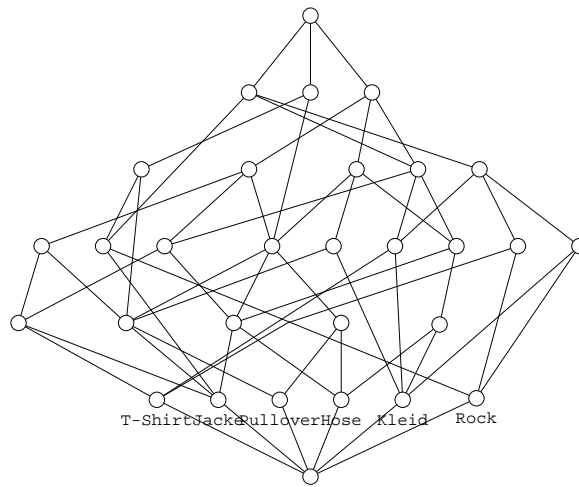


Abbildung 21: Begriffsraum der Versuchsperson 9 im ersten Durchgang (PSQUERY-Verfahren)

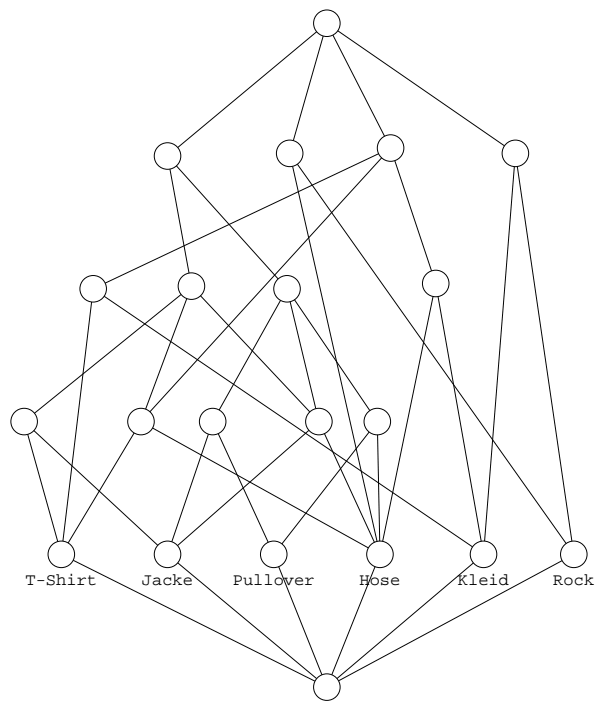


Abbildung 22: Begriffsraum der Versuchsperson 9 im zweiten Durchgang (PSQUERY-Verfahren)

es sich keineswegs um Produkte eines unsystematischen Antwortverhaltens handelt. Denn trotz großer visueller Unterschiede der Räume ergeben sich über die Befragungen hinweg vor allem für die QUERY-Räume gemessen am Stichprobenmittelwert relativ geringe Unterschiede. Bei der Versuchsperson scheint also tatsächlich eine gewisse Struktur vorzuliegen, die auch über die Befragungen hinweg erhalten bleibt.

5.3.4 Hypothesentest

Die statistische Nullhypothese ist, daß das Populationsmittel der Differenz zwischen den Raum-Differenzen beider Befragungsverfahren $\mu_{DIFF} \leq 0$ ist (siehe Abschnitt 5.2.5).

Die mittlere Differenz \bar{d}_{DIFF} beträgt für die vorliegende Stichprobe 0.075.

Die Varianz $\sigma_{\bar{d}_{DIFF}}^2$ ist unbekannt und muß über

$$S_{\bar{d}_{DIFF}}^2 = \frac{1}{n-1} \sum_{i=1}^n (d_{DIFF_i} - \bar{d}_{DIFF})^2 = 0.125$$

geschätzt werden muß. Für die Prüfgröße ergibt sich damit

$$T = \frac{\bar{d}_{DIFF}}{S_{\bar{d}_{DIFF}}} \sqrt{n} = 3.342 \sim t(30)$$

Bei 30 Freiheitsgraden (für $n = 31$) überschreitet der empirische Wert den kritischen Wert 1.697, der sich für eine Irrtumswahrscheinlichkeit $\alpha < .05$ ergibt.

Die Nullhypothese kann also auf dem 5 Prozent-Niveau verworfen werden.

5.4 Diskussion

5.4.1 Kennwerte des PSQUERY-Verfahrens

Das Anliegen dieser Arbeit ist es, die beiden Befragungsverfahren QUERY und PSQUERY in Bezug auf verschiedene Aspekte zu vergleichen und Kennwerte zur Verwendung des PSQUERY-Verfahrens für die Begriffsbefragung zu ermitteln.

Die erste, in der Fragestellung formulierte Annahme war, daß bei der zweimaligen Begriffsbefragung eines Individuums das PSQUERY-Verfahren ähnliche Räume liefert als das QUERY-Verfahren. Diese Hypothese erfährt aufgrund der vorliegenden Daten Unterstützung, da die gerichtete statistische Hypothese, daß kein Unterschied zwischen den Differenzen der Räume aus den beiden Verfahren besteht, verworfen werden kann ($\alpha = 0.05$). Besteht also z.B. im diagnostischen Einsatz der Wunsch, individuelle Begriffsstrukturen möglichst zuverlässig zu bestimmen, so ist nach den vorliegenden Ergebnissen dem PSQUERY-Verfahren der Vorzug vor dem QUERY-Verfahren zu geben. Mit der Implementation des PSQUERY-Verfahrens im Rahmen dieser Arbeit liegt nun auch ein Werkzeug hierfür vor. Beim Entwurf der Software wurde zudem auf Gesichtspunkte wie z.B. Motivationserhaltung bei der Befragung und Hinwirken auf längeres Überlegen pro Frage Wert gelegt.

Interessant ist auch ein absoluter Anhaltspunkt für die Zuverlässigkeit der individuellen Begriffsstrukturen, die aus dem PSQUERY-Verfahren resultieren.

Die mittlerer Differenz zwischen den PSQUERY-Räumen aus beiden Befragungen beträgt .180 (SD: .100). Tabelle 6 zeigt, daß bei derartigen Differenzen z.B. bei einer Zahl von 18 Begriffen bei der ersten und der zweiten Befragung mit fünf bis sechs unterschiedlichen Begriffen zu rechnen ist. Nimmt man z.B.

für jeden Raum drei eigene, im anderen nicht vorhandene Begriffe an, so sind das immerhin 30 Prozent der nicht-trivialen Begriffe. Ob derartige Abweichungen über mehrfache Befragungen hinweg tolerierbar sind, muß für den jeweiligen Anwendungsfall entschieden werden. Warum sich die Differenzen zwischen zwei mit dem PSQUERY-Verfahren bestimmten Begriffsstrukturen immer noch so groß sind, wird später noch zu diskutieren sein.

Ein weiterer wesentlicher, z.B. für die diagnostische Anwendung besonders bedeutsamer Kennwert ist die Zahl der Fragen, die für die Bestimmung einer individuellen Begriffsstruktur gestellt werden müssen. In der vorliegenden Untersuchung benötigte das PSQUERY-Verfahren deutlich mehr Fragen als das QUERY-Verfahren.

Unter der Annahme, daß das befragte Individuum vollkommen konsistent antwortet, wäre bei dem hier verwendeten Design für das PSQUERY-Verfahren höchstens die doppelte Anzahl der Fragen im Vergleich zum QUERY-Verfahren zu erwarten gewesen. Da nämlich beide Verfahren parallel laufen und zuerst das QUERY-Verfahren die zu stellenden Fragen bestimmt, werden beim PSQUERY-Verfahren in dieser Phase Antworten nur in den Zwischenspeicher geschrieben. Es gibt keine Bestätigungen, die Antworten aus dem Zwischenspeicher in die endgültige Relation überführen würden, da das QUERY-Verfahren keine derartigen „Überschneidungen“ zuläßt, da sie möglicherweise zu Widersprüchen führen könnten. Beim Ende des QUERY-Verfahrens ist also der Zwischenspeicher des PSQUERY-Verfahrens stark gefüllt. Würde man nun rein theoretisch alle QUERY-Fragen, die bisher gestellt wurden, nochmals beantworten lassen, so wäre, unter Annahme eines vollkommen konsistenten Antwortverhaltens, der Zwischenspeicher sicher geleert.

Die durchschnittliche PSQUERY-Fragenanzahl liegt allerdings über der doppelten QUERY-Fragenanzahl. Die möglichen Ursachen hierfür werden später dis-

kutiert. Zuerst ist aber die praktische Bedeutung dieses Ergebnisses von Interesse. Ein Ersetzen des QUERY-Verfahrens durch das PSQUERY-Verfahren zur Gewinnung zuverlässigerer Ergebnisse bedeutet danach also eine mehr als doppelt so lange Befragungszeit. Da diese in der praktischen Anwendung nicht beliebig ausgeweitet werden kann, wird man eine geringere Zahl von Objekten verwenden müssen. Wo mit dem QUERY-Verfahren noch Befragungen mit zehn Objekten realisierbar sind, wird beim PSQUERY-Verfahren die Obergrenze bei ca. sieben Objekten liegen. Bei dieser Zahl dürften sich noch akzeptable Befragungszeiten von unter einer Stunde ergeben, innerhalb derer nur mit moderaten Ermüdungseffekten etc. zu rechnen ist. Es muß also von Fall zu Fall abgewogen werden, ob mehr Wert auf eine kürzere Befragung bzw. eine Befragung mit mehr Objekten oder auf zuverlässigere Ergebnisse gelegt wird.

Eine mögliche Ursache für eine erhöhte Fragenzahl für das PSQUERY-Verfahren im Rahmen dieser Arbeit könnte natürlich darin liegen, daß das QUERY- und das PSQUERY-Verfahren parallel verwendet werden. Dadurch kann das PSQUERY-Verfahren solange das QUERY-Verfahren noch läuft keine eigenen Fragen stellen. Möglicherweise sind aber die Fragen, die das QUERY-Verfahren generiert, nicht optimal, um für das PSQUERY-Verfahren die Fragenzahl zu minimieren. Allerdings sind die Berechnungen zur Generierung von PSQUERY-Fragen sehr aufwendig, so daß in der momentanen Implementierung schon bei der ersten möglichen Frage, die gefunden wird, die Suche abgebrochen wird. Mit leistungsfähigeren Rechnern wäre allerdings eine eigenständige, optimierte PSQUERY-Fragengenerierung denkbar, die möglicherweise zu einer geringeren Fragenzahl führen könnte.

Wenn aber in der momentanen Implementation schon die doppelte QUERY-Fragenanzahl in Kauf genommen wird, so bestände rein theoretisch auch die

Möglichkeit, mit derselben Fragenzahl zweimal die QUERY-Befragung durchzuführen. Es wäre denkbar, daß sich die beiden daraus resultierenden Begriffsräume zu einer verlässlicheren individuellen Begriffstruktur integrieren ließen. Wenn dies gelänge, hätte das PSQUERY-Verfahren möglicherweise keine Bedeutung mehr. Das Problem, das die Umsetzung dieser Idee mit sichbringen würde, besteht allerdings darin, daß bei zweimaliger QUERY-Befragung dem Probanden wahrscheinlich gehäuft Fragen doppelt vorgelegt werden würden. Hier ist mit starken motivationalen Problemen zu rechnen.

5.4.2 Folgerungen

Beim Einsatz des PSQUERY-Verfahrens ergeben sich für ein Individuum über zwei Befragungen hinweg immer noch Differenzen zwischen den ermittelten Begriffsstrukturen. Dabei werden für das PSQUERY-Verfahren mehr als doppelt so viele Fragen benötigt wie für das QUERY-Verfahren. Eine mögliche Erklärung könnte die Zahl der Widersprüche liefern — also die Zahl der Fragen, bei denen das PSQUERY-Verfahren nach einer Antwort Einträge aus dem Zwischenspeicher entfernen muß, da sich Widersprüche ergeben.

Folgende Faktoren könnten für diese nicht zu vernachlässigende Größe (immerhin bei ca. 6 bis 7 Prozent der Antworten) verantwortlich sein:

1. Aufgrund von Ermüdung, Unkonzentriertheit, etc. kommt es zu (zufälligen) Antwortfehlern.
2. Die Versuchsperson beantwortet verschiedene Fragen (z.B. über die Zeit hinweg) nach verschiedenen Kriterien. Z.B. könnte es sein, daß manche Merkmale einmal berücksichtigt werden und einmal nicht.
3. Die empirische Relation, die die Versuchsperson eigentlich, d.h. bei einer

Befragung ohne Schlußfolgerungen, liefern würde, verletzt systematisch die Transitivitätsforderung.

Sollten die ersten beiden Faktoren ausschlaggebend sein, so müßte sich durch eine entsprechende Modifikation des Befragungsverfahrens die Zuverlässigkeit der Ergebnisse weiter erhöhen lassen. Durch eine bessere Gestaltung der Benutzerschnittstelle sollten sich z.B. Ermüdungseffekte verringern lassen. Zum Beispiel könnte den befragten Individuen die Möglichkeit geboten werden, die Fragen, die sie beantworten wollen/sollen, selbst zusammenzustellen. Das PSQUERY-Verfahren würde eine solche Modifikation zulassen. Zu denken wäre auch an die Möglichkeit, das Bestätigungskriterium, nach dem entschieden wird, ob eine Antwort aus dem Zwischenspeicher in die endgültige Relation übernommen wird, konservativer zu gestalten. Zum Beispiel könnte man eine zwei- oder mehrmalige Bestätigung verlangen. Dies würde aber dann auch die Fragenzahl erhöhen.

Allerdings wurden ähnliche Anstrengungen für die vorliegende Untersuchung schon unternommen: In der Instruktion wurde darum gebeten, sich vor der Untersuchung das Begriffsfeld zu betrachten. Es wurde deutlich gefordert, erst mit „Nein“ zu antworten, wenn wirklich *kein* passendes Unterscheidungskriterium gefunden werden kann. Der Zwang, jede Antwort explizit bestätigen zu müssen und eine darauf folgende Wartezeit sollte darüberhinaus hastiges bzw. unüberlegtes Antwortverhalten vermindern. Außerdem wurde die Benutzerschnittstelle komfortabler als bei vorangegangenen Untersuchungen gestaltet, z.B. mit der Möglichkeit, explizit Fragen abzulehnen und beliebig Pausen einschieben zu können. Gegenüber der vorangegangenen Untersuchung von Ulbrich (1997), bei der solche Maßnahmen nicht getroffen wurden, zeigt sich allerdings tendenziell keine Steigerung der Ähnlichkeit der Begriffsräume über zwei Befragungen hinweg durch diese Maßnahmen. Eine statistische

Prüfung ist auf Basis der vorliegenden Daten leider nicht möglich.

Angesichts dessen gewinnt das Argument an Bedeutung, daß die Relation, die eine Versuchsperson eigentlich bei einer Komplettbefragung liefern würde, systematisch die Transitivitätsforderung verletzt, d.h. es sich um keine Entailment-Relation handelt. Diese Vermutung läßt sich, da sowohl des QUERY-Verfahren als auch das PSQUERY-Verfahren die Transitivität als gegeben voraussetzen — PSQUERY soll eigentlich nur „zufällige“ Falschantworten abfangen — nicht überprüfen. Hier zeigt sich eindeutig der Nachteil von Verfahren, die immer ein Ergebnis liefern, gegenüber Vorgehensweisen, die z.B. eine Prüfung von Axiomen ermöglichen.

Auf rein deskriptiver Ebene lassen sich allerdings möglicherweise in den Daten Muster erkennen, die Hinweise zu Klärung der angesprochenen Fragen geben könnten.

Betrachtet man die Räume der beiden im Ergebnisteil vorgestellten Versuchspersonen 9 und 25, so läßt sich folgendes feststellen:

Bei der „hypothesenkonformen“ Versuchsperson 25 könnte es sich um einen Fall handeln, in dem wirklich eine Entailment-Relation vorhanden ist. Die QUERY-Räume unterscheiden sich untereinander und auch von den PSQUERY-Räumen. Sie könnten durch „zufällige“ falsche Antworten zustande gekommen sein. Abbildung 23 illustriert nochmals die Differenzen zwischen den vier Räumen, die für Versuchsperson 25 geliefert werden.

Diese Interpretation wird dadurch gestützt, daß die beiden PSQUERY-Räume identisch sind. Man könnte sagen, daß hier die Fehlerkorrektur des Verfahrens Wirkung gezeigt hat. Für das QUERY-Verfahren, das über keinen Mechanismus zur Fehlerkorrektur verfügt, ergeben sich, wie man es bei zufälligen Antwortfehlern erwarten würde, Differenzen über die beiden Befragungen

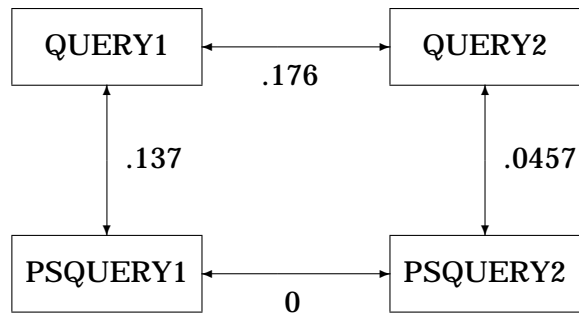


Abbildung 23: Differenzen zwischen den vier Begriffsräumen der Versuchsperson 25

hinweg.

Für die Versuchsperson 9 ergibt sich ein anderes Bild. Die Differenz zwischen den beiden QUERY-Räumen ist, wie die Differenz der jeweiligen PSQUERY-Räume zu den QUERY-Räumen aus der gleichen Befragung, relativ moderat. Die Differenz zwischen den beiden PSQUERY-Räumen ist allerdings sehr hoch. Abbildung 24 illustriert nochmals diese Differenzen zwischen den vier Räumen, die für Versuchsperson 25 geliefert werden.

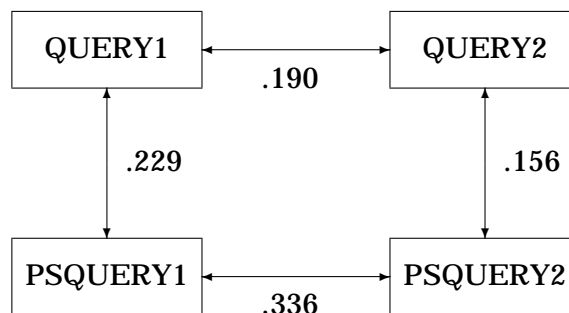


Abbildung 24: Differenzen zwischen den vier Begriffsräumen der Versuchsperson 9

Weiterhin fällt auf, daß alle Räume sehr viele Begriffe enthalten.

Ein mögliches Erklärungsmuster sieht folgendermaßen aus: Die Versuchsperson liefert stark inkonsistente/intransitive Antworten. Der QUERY-Algorithmus, der darauf ausgelegt ist, unter Annahme vollkommener Konsistenz die Fragenzahl zu minimieren, bringt die Befragung relativ schnell zum Abschluß. Da die Versuchsperson allerdings inkonsistent antwortet, d.h. Unterscheidungsmerkmale nicht immer gleichermaßen zur Anwendung bringt, ergibt sich eine größere Zahl von Begriffen. Zeichnet sich nämlich z.B. eine Gruppe von Objekten durch ein bestimmtes Merkmal aus und wird dieses nur bei einem Teil der Beurteilungen berücksichtigt, wird die Gruppe im Rahmen des QUERY-Verfahrens möglicherweise in Untergruppen mit Untergruppenmerkmalen aufgespaltet. Der PSQUERY-Algorithmus, der nun noch weitere Fragen stellt, wird „Opfer“ der Inkonsistenzen. Diese führen dazu, daß der resultierende Begriffsraum, den das Verfahren unter Annahme der Transitivität erzeugt, gegenüber dem QUERY-Räumen mit weiteren Begriffen angereichert wird.

Es scheint also zwischen den Versuchspersonen Unterschiede in Bezug auf die Transitivität zu geben, wie es die oben beschriebenen Datenmuster nahelegen. Die interessante Frage hierbei ist, ob es sich bei solchen Unterschieden, wenn man sie denn annimmt, um Personenfaktoren oder um Faktoren der Gestaltung der Befragung (möglicherweise in Interaktion mit Personenfaktoren) handelt.

Hier wäre eine Untersuchung denkbar, die durch systematische Variation von Befragungsbedingungen versucht, die oben beschriebenen Antwortmuster zu provozieren. Aus den Ergebnissen einer solchen Untersuchung wären dann Schlußfolgerungen für die Gestaltung der Begriffsbefragung ableitbar. So könnte es z.B. sein, daß Erkenntnisse gewonnen werden, auf welche Aspekte in der Instruktion besonders Wert gelegt werden muß. Auch Ergebnisse zur Gestaltung der Benutzerschnittstelle sind denkbar. Beispielsweise könnte es sein, daß der Begriffskontext, der bisher nur am unteren Rand des Fragedialogs

präsentiert wird, deutlicher ins Zentrum der Aufmerksamkeit gerückt werden muß.

Durch die Umsetzung solcher Empfehlungen könnte man schließlich zu einem Verfahren gelangen, das weniger mit dem Problem der Inkonsistenzen zu kämpfen hat und im diagnostischen Umfeld zuverlässigere Ergebnisse liefert. Die angedeuteten Untersuchungen gehen allerdings über den Rahmen dieser Arbeit hinaus, da z.B. die vorliegende Datenbasis zur Klärung der aufgeworfenen Fragen nicht ausreichen dürfte.

6 Zusammenfassung

In der psychologischen Forschung wurden viele konkurrierende Theorien zu individuellen Begriffsstrukturen und Verfahren zu deren Bestimmung vorgeschlagen. Diese Arbeit bezieht sich auf einen Ansatz, der durch eine meßtheoretische Herangehensweise eine Überprüfung seiner Eignung als Theorie individueller Begriffsstrukturen ermöglicht.

Die Datenbasis, die benötigt wird, um mit Hilfe dieses Ansatzes individuelle Begriffsstrukturen zu bestimmen, läßt sich mit verschiedenen Verfahren erheben. Zur Minimierung der Fragenzahl wurde das Verfahren QUERY entwickelt. Da sich gezeigt hat, daß dieses relativ anfällig gegenüber zufälligen Antwortfehlern ist, wurde es zum PSQUERY-Verfahren erweitert.

Ziel dieser Arbeit ist es, die beiden Verfahren dahingehend zu vergleichen, wie gut ihre Ergebnisse über zwei Befragungen hinweg übereinstimmen. Diese Zuverlässigkeit der Ergebnisse über die Zeit ist vor allem im diagnostischen Einsatz wichtig. Zudem sollen Kennwerte speziell für das neue PSQUERY-Verfahren ermittelt werden.

Den Ergebnissen einer zweimaligen Befragung von 31 Versuchspersonen mit beiden Verfahren folgend, kann die Hypothese, daß das PSQUERY-Verfahren keine zuverlässigeren Ergebnisse liefert, verworfen werden ($\alpha < 0.05$). Allerdings werden für das PSQUERY-Verfahren mehr als doppelt so viele Fragen benötigt. Bei der Wahl zwischen beiden Verfahren ist also ein Abwägen zwischen Zuverlässigkeit und Befragungsdauer nötig. Aus der Betrachtung der Antwortmuster ergeben sich Hinweise, daß für das PSQUERY-Verfahren und die Gestaltung der Befragung möglicherweise noch Spielraum für eine Verbesserung hinsichtlich der Zuverlässigkeit der Ergebnisse besteht.

Literatur

- [1] Cosyn, S. & Thiery, N. (in Vorbereitung). A practical procedure to build a knowledge structure. *Journal of Mathematical Psychology*.
- [2] Doignon, J.P., Falmagne, J.C. (1999). *Knowledge spaces*. Berlin: Springer.
- [3] Dowling, C.E. (1993a). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology*, 37, 21-48.
- [4] Dowling, C.E. (1993b). On the irredundant construction of knowledge spaces. *Journal of Mathematical Psychology*, 37, 49-62.
- [5] Fillenbaum, S. & Rapoport, A. (1971). *Structures in the subjective lexicon*. New York: Academic Press.
- [6] Heller, J. (1991). *Eine experimentelle und theoretische Untersuchung zur Begriffsbildung*. Dissertation, Universität Regensburg.
- [7] Heller, J. (2000). Representation and Assessment of Individual Semantic Knowledge. *Methods of Psychological Research – Online*, 5, 1-37.
- [8] Hull, C.L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, 28(1), 1-40.
- [9] Irtel, H. (1997). PXL: A library for psychological experiments on IBM PC type computers. *Spatial Vision*, 10, 467-469.
- [10] Kaluscha, R. (1994). *Ein effizienter Algorithmus zum Befragen von Experten*. Unveröff. Diplomarbeit, TU Braunschweig.

- [11] Kambouri, M., Koppen, M., Villano, M., Falmagne, J.C. (1994). Knowledge assessment: tapping human expertise by the QUERY routine. *International Journal of Human-Computer-Studies*, 40, 119-151.
- [12] Katz, J. (1972). *Semantic theory*. New York: Harper and Row.
- [13] Koppen, M. (1993). Extracting human expertise for constructing knowledge spaces: an algorithm. *Journal of Mathematical psychology*, 37, 1-20.
- [14] Koppen, M. (1994). The construction of knowledge spaces by querying experts. In G.H. Fischer & D. Laming (Hrsg.), *Contributions to Mathematical Psychology, Psychometrics and Methodology* (S. 137-147). New York: Springer.
- [15] Koppen, M. & Doignon, J.P. (1988). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, 34, 311-441.
- [16] Kosko, B. (1993), *Fuzzy-logisch*. Hamburg: Carlsen.
- [17] Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- [18] Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1989). *Foundations of measurement* (Vol. 2). New York: Academic Press.
- [19] Laurence, S. & Margolis, E. (1999). Concepts and Cognitive Science. In E. Margolis & S. Laurence (Hrsg.), *Concepts: core readings* (S. 3-81). Cambridge, Mass.: MIT Press.
- [20] Müller, C.E. (1989). A procedure for facilitating an expert's judgement on a set of rules. In E.E. Roskam (Hrsg.), *Mathematical Psychology in Progress*. (S. 157-170). Berlin: Springer.

- [21] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.
- [22] Rohatgi, V.K. (1976). *An introduction to probability theory and mathematical statistics*. New York: Wiley.
- [23] Röd, W. (1994). *Der Weg der Philosophie*. Erster Band. München: Beck.
- [24] Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. Moore (Hrsg.), *Cognitive Development and the Acquisition of Language* (S. 111-144). New York: Academic Press.
- [25] Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Hrsg.), *Cognition and Categorization* (S. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [26] Schrepp, M. & Held, T. (1995). A simulation study concerning the effect of errors on the establishment of knowledge spaces by querying experts. *Journal of Mathematical Psychology*, 39, 376-382.
- [27] Sneed, J.D. (1971). *The Logical Structure of Mathematical Physics*. Dordrecht: Reidel.
- [28] Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Theorie und Erfahrung*. 2. Halbband. Theoriestrukturen und Theoriendynamik. Berlin: Springer.
- [29] Stegmüller, W. (1974). Theoriendynamik und logisches Verständnis. In W. Diederich (hrsg.), *Theorien der Wissenschaftsgeschichte. Beiträge zur diachronen Wissenschaftstheorie* (S. 210-246). Frankfurt/Main: Suhrkamp.
- [30] Tversky, A. & Krantz, D.H. (1969). Similarity of schematic faces: a test of interdimensional additivity. *Perception and Psychophysics*, 5, 124-128.

- [31] Ulbrich, E. (1997). *Experimentelle Untersuchung struktureller Unterschiede bei der individuellen Begriffsbildung*. Unveröff. Diplomarbeit, Universität Regensburg.
- [32] Westmeyer, H. (Hrsg.). (1989). *Psychological Theories from a Structuralist Point of View*. Heidelberg: Springer.
- [33] Wille, R. (1987). Bedeutung von Begriffsverbänden. In B. Ganter, R. Wille, K.E. Wolff (Hrsg.), *Beiträge zur Begriffsanalyse* (S. 161-211). Mannheim: BI-Wissenschaftsverlag.

A Instruktion

Liebe Versuchsperson,

Zunächst einmal vielen Dank, daß Sie an folgendem Experiment teilnehmen wollen!

In dem Experiment werden Ihnen Fragen der Art

„Haben die x Begriffe auf der linken Seite etwas gemeinsam, das sie von dem Begriff auf der rechten Seite unterscheidet?“

gestellt.

Der Fragebildschirm sieht dabei in etwa so, wie auf dem folgenden Bild, aus:



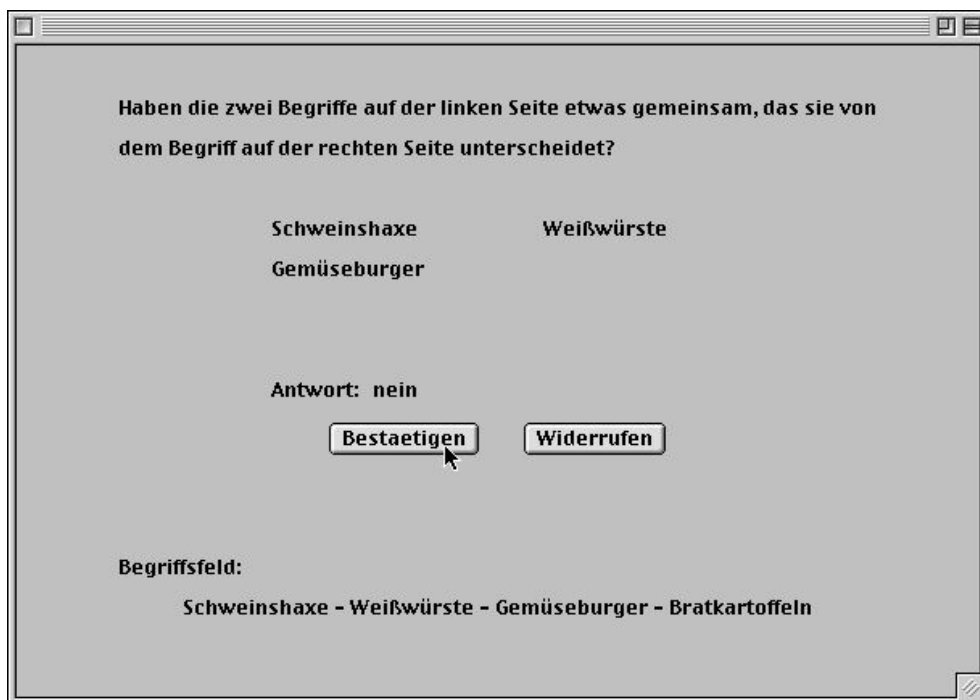
Dabei werden Ihnen auf der linken Seite bei jeder Frage jeweils mehrere Begriffe gezeigt, auf der rechten Seite einer.

Sie sollen sich überlegen, ob nach *ihrer Meinung* die Begriffe auf der linken Seite eine Gemeinsamkeit haben, die sie von dem Begriff auf der rechten Seite unterscheidet.

Wenn sie so eine Gemeinsamkeit gefunden haben, klicken Sie auf **Ja**.

Können Sie keine finden, klicken Sie auf **Nein**.

Ihre Antwort wird Ihnen dann noch einmal gezeigt. Sie können sie nun bestätigen oder widerrufen:



Wählen Sie **Bestätigen**, dann wird Ihnen nach einer kurzen Pause die nächste Frage gezeigt.

Wenn Sie **Widerrufen** wählen, können Sie die Frage noch einmal beantworten.

Wenn Sie eine Frage nicht beantworten wollen, wählen Sie **Weiter**.

Sie wird Ihnen aber möglicherweise später wieder präsentiert.

Mit **Pause** gelangen Sie zu einem **Pausebildschirm**, den Sie mit **Weiter** wieder verlassen können. Benutzen Sie diese Möglichkeit, um nach eigenem Empfinden Pausen zu machen!

Wenn Sie alle Fragen beantwortet haben erscheint ein **End-Bildschirm**. Bitte informieren Sie dann den **Versuchsleiter**.

Zunächst können Sie nun einen **Probeversuch** durchlaufen, um sich mit der Fragestellung vertraut zu machen. Dabei werden folgende vier Begriffe („Begriffsfeld“) verwendet:

Schweinshaxe, Weißwürste, Gemüseburger, Bratkartoffeln

Beim anschließenden **Hauptdurchgang** wird folgendes **Begriffsfeld** mit sechs Begriffen verwendet:

Kleid, Hose, Pullover, Jacke, T-Shirt, Rock

Bitte durchdenken Sie vor Beginn das jeweils zugehörige **Begriffsfeld**, um während des **Versuchs** zügiger antworten zu können.

Viel Vergnügen!

B Deskriptive Statistiken aller Versuchspersonen

Tabelle 7: Deskriptive Statistiken der Räume aller Versuchspersonen aus den beiden QUERY-Befragungen

Vp	Begriffe Raum 1	Begriffe Raum 2	Begriffe exklusiv Raum 1	Begriffe exklusiv Raum 2	Differenz
3	13	17	1	5	.2286
4	12	12	1	1	.0952
5	14	16	1	3	.1524
6	15	18	1	4	.1732
7	17	20	4	7	.3398
8	16	25	3	12	.4181
9	25	23	5	3	.1905
10	20	17	4	1	.1544
11	14	18	2	6	.2857
12	21	22	4	5	.2392
13	17	12	6	1	.2759
14	13	13	2	2	.1758
15	16	22	2	8	.3008
16	16	16	4	4	.2857
17	19	17	5	3	.2540

Tabelle 8: Deskriptive Statistiken der Räume aller Versuchspersonen aus den beiden QUERY-Befragungen (Fortsetzung)

Vp	Begriffe Raum 1	Begriffe Raum 2	Begriffe exklusiv Raum 1	Begriffe exklusiv Raum 2	Differenz
18	16	12	4	0	.1633
19	20	21	1	2	.0836
20	12	18	2	8	.3810
21	17	16	6	5	.3810
22	25	15	10	0	.2857
23	21	18	7	4	.3223
24	25	20	10	5	.3810
25	13	13	2	2	.1758
26	19	14	5	0	.1732
27	23	21	7	5	.3117
29	11	12	0	1	.0497
30	27	13	15	1	.4571
31	25	30	5	10	.3117
32	14	15	2	3	.1970
33	20	19	7	6	.3810
34	14	18	2	6	.2857

Tabelle 9: Deskriptive Statistiken der Räume aller Versuchspersonen aus den beiden PSQUERY-Befragungen

Vp	Begriffe Raum 1	Begriffe Raum 2	Begriffe exklusiv Raum 1	Begriffe exklusiv Raum 2	Differenz
3	13	19	1	7	.2857
4	12	11	1	0	.0497
5	14	17	0	3	.1106
6	16	14	3	1	.1524
7	14	13	3	2	.2116
8	12	14	0	2	.0879
9	30	21	12	3	.3361
10	15	16	2	3	.1843
11	15	16	3	4	.2581
12	17	16	3	2	.1732
13	15	13	3	1	.1633
14	12	10	2	0	.1039
15	15	28	2	15	.4518
16	15	15	2	2	.1524
17	14	25	0	11	.3223

Tabelle 10: Deskriptive Statistiken der Räume aller Versuchspersonen aus den beiden PSQUERY-Befragungen (Fortsetzung)

Vp	Begriffe Raum 1	Begriffe Raum 2	Begriffe exklusiv Raum 1	Begriffe exklusiv Raum 2	Differenz
18	15	12	3	0	.1270
19	21	23	2	4	.1558
20	11	11	0	0	0
21	11	14	0	3	.1371
22	21	17	5	1	.1805
23	16	24	1	9	.2857
24	18	17	3	2	.1633
25	12	12	0	0	0
26	14	16	2	4	.2286
27	25	26	4	5	.2017
29	13	13	0	0	0
30	18	14	5	1	.2143
31	23	26	3	6	.2099
32	15	19	1	5	.2017
33	15	20	1	6	.2286
34	13	14	2	3	.2116

Eidesstattliche Erklärung

Ich versichere hiermit, daß ich die vorliegende Arbeit mit dem Thema

Ein experimenteller Vergleich von
Verfahren zur Bestimmung
individueller Begriffsstrukturen

selbständig verfaßt und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall durch Angabe der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht.

Regensburg, den 21. Juli 2000

Harald Binder