

3.1 Allgemeine Angaben zum Teilprojekt A2

3.11 Thema

Linguistische Theorien als Datentypen

3.12 Fachgebiet und Arbeitsrichtung

Formale Linguistik, Vergleichende Sprachwissenschaft

3.13 Leiter

Prof. Dr. Uwe Mönnich

geb. 4. April 1939

Seminar für Sprachwissenschaft, Universität Tübingen Wilhelmstr. 113

D-72074 Tübingen

Tel.: 07071-2974035

Fax.: 07071-550520

Email: um@sfs.nphil.uni-tuebingen.de

Ist die Stelle des Leiters/der Leiterin des Projektes befristet?

(x) nein () ja, befristet bis zum _____

3.14 Aktenzeichen bei bisheriger Förderung in einem anderen Verfahren der DFG

Eine bisherige Förderung in einem anderen Verfahren der DFG liegt nicht vor.

3.15 Angaben zu Versuchen und Untersuchungen

In dem Teilprojekt sind vorgesehen

Untersuchungen am Menschen () ja (x) nein

klinische Studien im Bereich der

somatischen Zell- oder Gentherapie () ja (x) nein

Tierversuche () ja (x) nein

gentechnologische Untersuchungen () ja (x) nein

3.16 Bisherige und beantragte Förderung des Teilprojektes im Rahmen des Sonderforschungsbereichs

Haushaltsjahr	Personalkosten	Sächliche Verwaltungsausgaben	Investitionen	gesamt
1999	134.400,-	17.600,-	—	152.000,-
2000	134.400,-	8000,-	—	142.400,-
2001	147.600,-	8000,-	—	155.600,-

3.2 Zusammenfassung

Daten, insbesondere sprachliche Daten, sind nicht rein gegeben. Die jeweilige Struktur wird durch die theoretische Perspektive, unter der sie betrachtet werden, wesentlich bestimmt.

Das beantragte Projekt geht von der zentralen Hypothese aus, daß die aktuelle Diskussion über die Differenz leitender Paradigmen in der Sprachtheorie die Funktion der zugrundeliegenden Arten von Datentypen unterschlägt und durch die Beschränkung auf den Gegensatz zwischen charakteristischen formalen Verfahrensklassen (generative Erzeugungssysteme, Unifikation, modelltheoretische constraint satisfaction) den Aspekt der Datenstruktur aus dem Blickfeld verliert. Die mathematischen, wissenschaftssystematischen und praktischen Probleme, die sich aus einer methodisch reflektierten Behandlung des Datenbegriffes ergeben, sollen in dem Projekt in zeitlicher Reihenfolge behandelt werden.

In der ersten Phase steht vor dem Hintergrund einer formalen Explikation des Grammatikalitätsbegriffs der Gegensatz zwischen baumähnlichen Strukturen, die von Beginn an eine zentrale Rolle in der generativen Theorie gespielt haben, und graphenähnlichen Strukturen, die in der Form von attributwertigen Matrizen in der Head Driven Phrase Structure Grammar sich nahezu ausschließlicher Aufmerksamkeit erfreuen. Vor dem Hintergrund erster formaler Resultate soll die Vermutung erhärtet werden, daß das Postulat eines entscheidbaren Grammatikalitätsbegriffs "datengetrieben" ist und daß in einem Datentyp, der — wie es bei Graphen der Fall ist — nicht im Rahmen einer algebraischen Theorie abstrakter Datentypen zu erfassen ist, eine analoge Entscheidbarkeitsfrage für das Konzept der Grammatikalität sich nicht stellt. In einer zweiten Phase soll unter wissenschaftssystematischen Kriterien die Entwicklung alternativer Datenkonzepte in der generativen Theorie und konkurrierenden Modellen untersucht werden. Schließlich ist vorgesehen, die praktische Relevanz von linguistischen Beschreibungsverfahren, die sich an einem objektorientierten Datenbegriff ausrichten, durch die exemplarische Analyse eines amerindischen Korpus zu demonstrieren.

ren und deren Resultate mit Ergebnissen zu vergleichen, die sich linguistischen Methoden der klassischen Moderne verdanken.

3.3 Stand der Forschung

3.31 Datenstrukturen und Datenreflexion

Datenstrukturen, insbesondere in ihrer Ausprägung als Mengen, als Algebren und als Theorien, haben in der modernen Sprachwissenschaft in vielfacher Hinsicht eine wichtige Rolle gespielt. Die vertraute Unterscheidung von drei Adäquatheitsebenen aus der früheren Phase des generativen Paradigmas läßt sich unschwer mit den drei Typen von Datenstrukturen in eine methodologische Parallele setzen (Chomsky 1955). Eine ähnliche Aussage gilt für das Untersuchungsgebiet der formalen Sprachtheorie, in dem es sich als fruchtbar erwiesen hat, die schwache und starke generative Kapazität von Produktionssystemen von der Betrachtung der Komplexität logisch basierter Spezifikationssysteme zu trennen (Thomas 1997). Da in zeitgenössischen linguistischen Theorien Typendisziplinen eine wesentliche Funktion erfüllen, muß auch die Entwicklung der Typentheorie in der theoretischen Informatik in einem Forschungskontext berücksichtigt werden, der das Problem sprachlicher Datenstrukturen in seinen Mittelpunkt stellt. In der Theorie der Programmiersprachen sind verschiedene Begriffe von Typen herangezogen worden. Typen dienen dazu, Einheiten, die in den jeweiligen Programmierstilen vorkommen, zu klassifizieren. Im Laufe der Entwicklung hat auch hier ein Übergang von den Traditionen, in denen Typen als Mengen von Worten und als sog. freie Algebren aufgefaßt wurden, zu einem Konzept stattgefunden, nach dem Typen Modelle von Theorien klassifizieren (Goguen 1991).

Es ist aufschlußreich, auch den Paradigmenwechsel von generierenden zu lizenzierenden Theorien in der Linguistik unter dieser Perspektive der Entwicklung von Datenstrukturen nachzuzeichnen. Unter dem Einfluß der kognitivistischen "Revolution", nach der unsere geistigen Vermögen auf das syntaktische Manipulieren von Symbolen zu reduzieren und mit den Mitteln der Rekursionstheorie zu beschreiben sind, war zunächst die herrschende Auffassung, daß die Frage nach einer adäquaten Charakterisierung der Grammatikalität von Sätzen einer natürlichen Sprache die Frage nach einem Entscheidungsverfahren zwischen wohlgeformten und nicht-wohlgeformten Morphemketten ist (Putnam 1961). Nachdem es gelungen war, auch auf andere Datenstrukturen neben den natürlichen Zahlen und den Ketten über endlichen Alphabeten induktiv definierte Berechnungsverfahren anzugeben, stellte sich schnell ein innerer Zusammenhang heraus zwischen den scheinbar voneinander unabhängigen Begriffen der Berechenbarkeit, der Induktion und der charakteristischen Eigenschaft der Strukturen, auf die die beiden erstgenannten Begriffe sinnvoll angewandt werden können. Es handelt sich bei dieser Eigenschaft um die entsprechend der jeweiligen terminologischen Tradition als Anarchie, Freiheit oder Initialität bezeichnete Eigenschaft von Strukturen,

die insbesondere von der New Yorker Gruppe um Goguen, Thatcher, Wagner und Wright im Rahmen ihrer allgemeinen Theorie abstrakter Datenstrukturen in ihrer Wichtigkeit erkannt wurde (Meseguer und Goguen 1985).

Das Darstellungsmittel, das die weitestete Anwendung zur Repräsentation von linguistischen Strukturen, die im Umkreis von lizensierenden Theorien postuliert werden, gefunden hat, die Merkmalsstrukturen, lassen allerdings diese Eigenschaft der Anarchie vermissen (Aczel 1997). Welche Konsequenzen das für den Kognitivismus im oben angedeuteten Sinn hat, die Beantwortung dieser Frage wird Gegenstand des Teilprojektes sein. Unter dem Gesichtspunkt der Konvergenz der Entwicklungen, die sich für den Begriff der Datenstrukturen in formaler, wissenschaftssystematischer und deskriptiver Hinsicht abzeichnen, ist es jedoch von vorrangigem Interesse, daß die mathematische Charakterisierung von Merkmalsstrukturen ein Spezialfall des theoretischen Instruments darstellt, das sich ebenfalls zur Lösung des drängenden Problems heranziehen läßt, wie einer un verkürzten Konzeption sprachlicher Daten Rechnung zu tragen ist.

In der Präambel zu dem vorliegenden Antrag auf Einrichtung eines Sonderforschungsbereichs zum Thema "Linguistische Datenstrukturen" wird ausführlich begründet, welche Aufschlüsse von einer systematischen Berücksichtigung verschiedener linguistischer Datenklassen zu erwarten sind. Die Zielvorstellung des Sonderforschungsbereiches schließt in diesem Punkt an die Einsicht an, die besonders im Umfeld der Situationstheorie hervorgehoben worden ist (Barwise und Seligman 1997), daß neben dem strukturellen Aspekt sprachlicher Systeme ihr situierter oder sozialer Aspekt eine gleichrangige Rolle spielt. Es ist insbesondere dieser letzte Faktor, der in den Nachbarwissenschaften der Sprachwissenschaften einen Objektbegriff bestimmt hat, in dem nicht so sehr Wert auf dessen abstrakte Struktur, sondern auf seine indexikalischen, kontextabhängigen, d.h. im weitesten Sinne situationsbezogenen Eigenschaften gelegt wurde (B.C. Smith 1996).

Um den Linguisten bei der Feldarbeit ein Instrument an die Hand zu geben, das flexibel, und ausdrucksstark genug ist, um die Situationsabhängigkeit und den hohen Grad an wechselseitiger Abhängigkeit sprachlicher Daten angemessen zu repräsentieren, hat man am *Summer Institute for Linguistics* eine Systemumgebung entwickelt, die dem objektorientierten Ansatz verpflichtet ist (Rettig et al. 1993). Zwar erfreut sich das objektorientierte Paradigma einer großen Popularität und kann auf den Ebenen der Analyse von Problemen, des Programmentwurfs und der aktuellen Implementierung eine beachtliche Zahl von Erfolgen vorweisen, aber es handelte sich bei diesem Ansatz bislang eher um eine allgemeine Philosophie als um ein ausgearbeitetes mathematisches Modell. Insbesondere fehlte es an einer klaren Vorstellung eines semantischen Universums, das den mit dieser Philosophie verbundenen Berechnungsstil auf eine feste Grundlage stellt.

Diese Situation hat sich grundlegend geändert, seitdem mit den (finalen) Coalgebren bzw. den mit ihnen äquivalenten "versteckten" Algebren (*hidden algebras*) eine Kategorie gefunden wurde, die diese Funktion eines formalen Berechnungsmodells für die Objektorientierung erfüllt (Aczel 1997, Jacobs 1996, Reichel

1995). Bei ihnen handelt es sich um einen versteckten Zustandsraum, über den geeignete Operationen definiert sind. Wie oben angedeutet, lassen sich mithilfe der Coalgebren und ihrer Verbindung zu dem objektorientierten Paradigma nicht nur die systematischen Beziehungen zwischen verschiedenen Datentypen erfassen, sondern sie bilden außerdem den geeigneten formalen Hintergrund, um Merkmalsstrukturen universal zu charakterisieren (Rutten 1996). In ihnen sind gewissermaßen beide Entwicklungsstränge, die Differenzierung des Begriffs der Datenstruktur, ausgehend von einer einfachen Klassifikationsmenge in Richtung auf eine Klassifikation von Strukturmodellen durch Theorien, und die Ausbildung eines reflektierten Datenbegriffs als Grundlage der Grammatikforschung, an einem gemeinsamen Ziel zusammengelaufen. Aus diesem Zusammentreffen der skizzierten Entwicklungslinien des Datenbegriffs ergeben sich eine Reihe von Aufgabenstellungen und Hypothesen, unter denen sich einige durch ihre Nähe zur globalen Zielsetzung des Sonderforschungsbereichs und zu den Untersuchungsgebieten bestimmter Einzelprojekte auszeichnen und die Gegenstand des Projekts zum Thema "Linguistische Theorien als Datentypen" sein werden. Im einzelnen handelt es sich dabei um das formale Problem der Berechenbarkeit von Datenstrukturen, um das wissenschaftssystematische Problem des Paradigmenwechsels von generierenden zu lizensierenden linguistischen Theorien und des damit verknüpften Übergangs zu einem anderen Typ von Datenstrukturen und schließlich um das deskriptive Problem, das sich mit der Verwendung objektorientierter Programmierumgebungen für die linguistische Feldarbeit stellt.

3.32 Formale, theoretische und deskriptive Aspekte linguistischer Datenstrukturen

Angesichts des regen Forschungsinteresses, das der allgemeinen Problematik der Datenstrukturen entgegengebracht wird und die sich in einer für den Einzelnen nicht vollständig übersehbaren Fülle an Literatur niederschlägt, wird sich auch dieser Teil der Übersicht zum Stand der Forschung auf ausgewählte Fragestellungen konzentrieren, die mit den im Projekt verfolgten Zielen und den dabei angewandten Methoden in enger Verbindung stehen.

Berechenbarkeit von Datenstrukturen

Es war in der vorstehenden Übersicht darauf verwiesen worden, daß in der frühen Diskussion des Grammatikalitätsbegriffs die Entscheidbarkeit von Morphemketten eine wichtige Rolle spielte. Ein verwandter Ansatz beherrschte die Explikation des Konstrukts einer linguistischen Beschreibungsebene, die als Verkettungsalgebren im Sinne Rosenblooms (1950) und deren Beziehungen zueinander im Sinne von strukturerhaltenden Morphismen interpretiert wurden. In beiden Fällen macht sich der bestimmende Einfluß des linguistischen Positivismus und insbesondere Carnaps bemerkbar, nach dessen Auffassung auch auf die natürli-

che Sprache das Verfahren angewandt werden sollte, das im Falle der Logik auf der Basis von bestimmten undefinierten Grundbegriffen zur Ausbildung einer fruchtbaren deduktiven Wissenschaft geführt hatte. Sprache in dieser Auffassung wurde reduziert auf einen speziellen Typ von Algebra, in dessen Rahmen sich auf streng formale Weise das Churchsche Entscheidbarkeitstheorem herleiten ließ (Rosenbloom 1950).

Unter dem Aspekt der Berechenbarkeit von Datenstrukturen ist von besonderem Interesse dieser enge Zusammenhang von algebraischer Struktur und Entscheidbarkeit, der in der sprachwissenschaftlichen Methodologiedebatte nie explizit gemacht worden ist. Dieser Zusammenhang findet auf formaler Ebene seinen Ausdruck in dem Resultat, daß in kategorientheoretischer Sicht Berechenbarkeit und Initialität algebraischer Strukturen zwei Seiten derselben Münze sind (Meseguer und Goguen 1985, Barr und Wells 1995). Die algebraischen Voraussetzungen für dieses Resultat laufen in letzter Analyse darauf hinaus, daß sich die relevanten Strukturen in eindeutiger Weise aus Grundelementen und einer Menge von sogenannten Konstruktoren beschreiben lassen. Diese Voraussetzung ist nicht für die Ableitungsbäume erfüllt, die sich kontextsensitiven Produktionssystemen zuordnen lassen, und dieser Umstand liefert unter rekursionstheoretischer Perspektive eine nachträgliche Rechtfertigung für die Zurückweisung kontextsensitiver Grammatiken und die Entscheidung für die Anwendung von Transformationen, die Elemente von eindeutig lesbaren Strukturen in Elemente desselben Typs überführen (Postal 1964).

Angesichts der überwältigenden Fülle immer neuer Konstruktionsprinzipien, die in allen konkurrierenden Paradigmen linguistischer Modelle eingeführt wurden, hat sich ein grundlegender Wechsel von generierenden zu lizensierenden Theorien vollzogen. Auf der denotationellen Ebene der Grammatikformalismen lassen sich diese Prinzipien zwar mit dem algebraischen Strukturbegriff vereinbaren, aber das aus solchem Vorgehen resultierende Ergebnis ist von so hoher Komplexität, daß von einem erklärenden Anspruch nicht mehr die Rede sein kann. Die Konsequenz für die betrachteten Datenstrukturen, die aus dem Wechsel zu dem lizensierenden Konzept zu ziehen ist, besteht in der Aufgabe des bis dahin leitenden algebraischen Strukturbegriffs. Es kommt gemäß der neuen methodologischen Ausrichtung nicht mehr in erster Linie darauf an, einen schrittweisen Analyse- oder Generierungsprozess zur Erklärung des menschlichen Sprachverständnisses nachzuzeichnen, sondern durch präzise anzugebene allgemeine Prinzipien die Strukturen zu spezifizieren, die für natürliche Sprachen charakteristisch sind. Daß diese Strukturen dann noch die bis dahin im Zentrum des Interesses stehenden Entscheidbarkeitseigenschaften aufweisen, ist nicht zu erwarten, und Chomsky (1981) hat denn auch zunächst diesen Aspekt der Grammatiktheorie für irrelevant erklärt.

Unter kognitivistischen Annahmen ist diese Haltung unannehmbar. Das Problem, das sich damit stellt, kann in der Weise beschrieben werden, daß die scheinbar gegensätzlichen Forderungen einer lizensierenden Strukturspezifikation mit de-

nen einer generierenden Strukturproduktion miteinander zu verbinden sind. Der methodische Ausweg, der sich in dieser Situation anbietet, setzt an bei einer Komplexitätstheoretischen Untersuchung der Relation zwischen (der formalen Sprache der) prinziptheoretischen Spezifikationen und den durch sie lizenzierten Strukturen. Abgesehen von einigen Resultaten aus der endlichen Modelltheorie ist die Lage hier allerdings völlig unübersichtlich.

Die historisch erste Charakterisierung des Verhältnisses zwischen einer Klasse endlicher Strukturen und einer bestimmten Logik hatte die Beziehung zwischen der schwachen monadischen Logik zweiter Stufe und den regulären Baumfamilien zum Gegenstand (Doner 1970, Thatcher und Wright 1968). Vor dem Hintergrund der das generierende Paradigma bestimmenden Strukturvorstellungen, die oben skizziert wurden, und gestützt auf die Tatsache, daß die modelltheoretisch ausgezeichneten Baumfamilien nicht nur diesem Strukturkonzept entsprechen, sondern darüberhinaus unter der eingeschränkten Perspektive einer Verkettungsalgebra auch noch exakt die kontextfreien (Ketten-)Sprachen ergeben, lag es nahe, dieses methodische Instrumentarium auf die Prinzipien- und Parameter-Variante des lizenzierenden Paradigmas anzuwenden (Rogers 1994). Das Resultat dieses weit angelegten Versuchs ist kennzeichnend für die generelle Situation: Die überwiegende Zahl der postulierten Prinzipien läßt sich in diesem formalen Rahmen erfassen. Chomskys Verdacht gegenüber Komplexitätstheoretischen Betrachtungen, soweit sie nicht allein den Bereich der praktisch durchführbaren Berechenbarkeit betreffen, erweist sich damit als vorschnell. Es bleiben allerdings Spezifikationsmechanismen wie z.B. die freie Indexierung übrig, die zur Unentscheidbarkeit führen. Rogers erreicht dieses negative Resultat im Rahmen der starken monadischen Logik zweiter Stufe, indem er zeigt, wie sich die Indexierung auf das Kachelungsproblem reduzieren läßt (Börger et al. 1997). Die Frage, ob sich dieses Ergebnis auch schon im Rahmen der schwachen monadischen Logik zweiter Stufe erzielen läßt, scheint offen zu sein. Bei einer Beantwortung dieser Frage kommt es offenbar in entscheidender Weise darauf an, festzustellen, in welchem Ausmaß durch das Verfahren der Indexierung Baumstrukturen graphenähnlich werden (Seese 1991). Das Spektrum der strukturellen Möglichkeiten, innerhalb dessen das Problem angesiedelt ist, läßt sich durch zwei weitere Resultate angeben, die die Komplexität der Berechenbarkeit von bestimmten Problemen kennzeichnen. Wenn keine Strukturen ausgeschlossen werden, dann ist die Relation zwischen der Logik zweiter Ordnung mit einem Existenzquantor als Spezifikationsprache und ihren modellierten Strukturen entscheidbar, genauer: NP-vollständig (Fagin 1974). Bei einer Beschränkung auf graphenähnliche Strukturen ist dasselbe Verhältnis schon im Hinblick auf die monadische Logik zweiter Stufe unentscheidbar (Courcelle 1997). Courcelle erzielt sein Resultat wiederum durch eine Reduktion auf das Kachelungsproblem, ein Verfahren, das auch Blackburn und Spaan (1993) anwenden, um die Unentscheidbarkeit einer modallogischen Spezifikationsprache über graphenähnlichen Strukturen zu beweisen. Auf dieses Ergebnis stützen sich wiederum King et al. (1997), um die Unentscheidbarkeit einer bestimmten

Formalisierung der HPSG nachzuweisen.

Aus der geschilderten Forschungslage im Hinblick auf die formalen Eigenschaften von Datenstrukturen, die in der Grammatiktheorie im Vordergrund des Interesses gestanden haben, ergeben sich eine solche Fülle von Konjekturen und methodischen Lösungsstrategien zu ihrer Bestätigung, daß sie die Möglichkeiten eines Einzelprojekts sprengen. In dem Abschnitt über die verfolgten Ziele des Projekts wird daher nur der Ausschnitt aufgenommen, der auf der einen Seite eng mit dem übergreifenden Thema des Sonderforschungsbereichs verzahnt ist und auf der anderen sich an den besonderen Konsequenzen ausrichtet, die sich aus der Leitidee ergeben, linguistische Theorien als Datentypen zu betrachten.

Wissenschaftssystematische Entfaltung von Datenstrukturen

In der allgemeinen Begründung der Thematik des Sonderforschungsbereichs wird die Entwicklung skizziert, in deren Verlauf der idealisierte Objektbegriff der modernen Sprachwissenschaft unter erheblichen Druck geraten ist. Diese Entwicklung soll hier nicht noch einmal nachgezeichnet, sondern in einem kurzen kritischen Bericht soll die Forschungslage in einem Punkt ergänzt werden, der die Grundlage für die sich daran knüpfenden Fragestellung bildet. Aus der Perspektive einer allgemeinen Objekttheorie sprachlicher Phänomene, die ihr besonderes Augenmerk auf die Angebote (halb-)formaler Beschreibungskategorien sprachlicher Strukturen richtet, fällt auf, daß es sich in diesem Bereich selbst unter idealtypischen Prämissen keineswegs um eine Dichotomie handelt. Es scheint vielmehr so zu sein, daß unter wissenschaftssystematischen Gesichtspunkten den verschiedenen grammatiktheoretisch fundierten Beschreibungsstrategien mindestens vier Datentypen entsprechen und daß deren Verhältnis zueinander dringender Klärung bedarf. Neben den generierenden und lizensierenden Varianten und den ihnen korrespondierenden abstrakten Datenstrukturen und Modellklassen spielen in der gegenwärtigen Diskussion nicht nur die Datentypen eine prominente Rolle, die von den Verfechtern eines Erkenntnisgegenstands der Sprachwissenschaft, der nicht der geschichtlichen und eurozentrischen Erosion anheimgefallen ist, vehement eingeklagt werden (Jäger 1993), sondern man begegnet daneben noch einer weiteren Kategorie, die in erster Annäherung mit einer objektivierten Informationsstruktur verglichen werden kann.

Die Problematik einer solchen Informationsstruktur spiegelt eine Situation auf der Ebene der wissenschaftssystematischen Analyse, die weiter oben schon im Zusammenhang mit den formalen Eigenschaften von Datenstrukturen thematisiert wurde und die charakteristisch für das lizensierende Paradigma ist. Die Motivation für dieses Paradigma entsprang in erster Linie der Abkehr von einer inflationären regelorientierten Analyse einzelner Konstruktionen und dem Versuch, allgemeine deklarative Beschränkungen zu formulieren, die in ihrer Gesamtheit den Grammatikalitätsbegriff definieren. Diese Definition wird dabei durchaus im klassischen modell-theoretischen Sinn verstanden, d. h. grammatisch sind alle jene

Strukturen, die die Beschränkungen (constraints) semantisch erfüllen (Johnson und Lappin 1997). In Analogie zu dem Fall der endlichen Bäume und Graphen als ausgezeichnete Klassen endlicher Modelle drängt sich die Frage auf, ob es sich bei den alle spezifizierenden Beschränkungen erfüllenden Strukturen um (abstrakte) Repräsentanten der sprachlichen Objekte handelt oder ob diese Strukturen nur für (einen Teilausschnitt an) Informationen stehen, die in das Blickfeld der jeweiligen lizensierenden Theorie geraten (King und Pollard 1997). Unter der Voraussetzung, daß es sich bei den repräsentierenden Strukturen unabhängig von ihrer Deutung um Merkmalsstrukturen handelt, die nach gängigen Annahmen das intendierte Modell lizensierender Theorien bilden, stellt sich das Dilemma, wie derselbe Datentyp — eine bestimmte Klasse etikettierter Graphen — mit zwei grundlegend divergierenden Interpretationen zu vereinbaren ist. Folgt man dem Vorschlag von Pereira und Shieber (1984), nach dem den Merkmalsstrukturen als Informationsträgern eine bereichstheoretische Deutung zu unterlegen ist, lassen sich diese Strukturen in einem präzisen technischen Sinn sogar als beste Lösung für das Problem auszeichnen, wie die Übereinstimmung des Informationsgehalts von Substrukturen zu repräsentieren ist (Moshier 1995). Sieht man in diesen Strukturen jedoch Objektrepräsentanten, scheint allein eine Charakterisierung als terminale Coalgebra sich anzubieten, die ebenfalls in einem präzisen technischen Sinn einen optimalen Status beanspruchen kann.

Beide Auffassungen der Merkmalsstrukturen als des zentralen Repräsentationsmediums lizensierender Theorien gehen so weit auseinander, daß eine harmonisierende Interpretation kaum vorstellbar ist, Ebenfalls scheint der Versuch, die informations- bzw. objektbezogenen Lesarten verschiedenen semantischen Theorien zuzuordnen, von vornherein zum Scheitern verurteilt. Die Semantik von Grammatikformalismen ist nämlich durch den Gegensatz zwischen operationalen und denotationellen Ansätzen gekennzeichnet, und es finden sich dementsprechend in der Literatur Hinweise, die informationsorientierte Lesart der Merkmalsstrukturen dem operationalen Ansatz zuzuschlagen (King und Pollard 1997). Den Hintergrund dieses Vorgehens bildet die Tatsache, daß zu dem Informationsbegriff wesentlich die Eigenschaft einer partiellen Ordnung gehört, die es erlaubt, Informationsgehalte miteinander zu vergleichen. Die Verbindung zu der Kategorie der operationalen Semantik, die über einen metaphorischen Vergleich hinausginge, würde allerdings voraussetzen, daß Merkmalsstrukturen sich nicht nur im Hinblick auf ihren Informationsgehalt in eine Ordnung bringen, sondern sich durch einen schrittweise Prozess nach dem Modell der operationalen Grundidee induktiv erzeugen lassen.

Diese Möglichkeit ist jedoch dadurch ausgeschlossen, daß es sich bei den Merkmalsstrukturen nur um quasi-freie Strukturen handelt, die Induktion nicht zulassen, wie oben im Zusammenhang mit den Berechenbarkeitsproblemen schon hervorgehoben wurde. Wenn es daher richtig ist, daß beide Datentypen nur im Rahmen einer denotationellen Konzeption von Semantik zu verstehen sind. Es besteht dringender Bedarf an einer Hypothese, die diesen Konflikt auf der Ebe-

ne der Datenstrukturen in wissenschaftssystematischer Hinsicht zu entschärfen erlaubt.

Der hohe Grad an Formalisierung, der im Bereich der Modelle linguistischer Gegenstandskonstitution erzielt worden ist, die sprachliche Daten von ihren kommunikativen, historischen und soziokulturellen Bezügen isoliert betrachten, hat die Diskussion des Verhältnisses zwischen bestimmten Typen der Theoriebildung und abstrakten Datentypen überhaupt erst möglich gemacht. In dem Moment, wo die genannten Faktoren, die mit dem Sprachvermögen im engeren Sinn auf vielfältige Weise interagieren, zu konstituiven Teilen des linguistischen Erkenntnisobjekts werden, geschieht dies beim augenblicklichen Stand der Modellbildung unter Verzicht auf einen kohärenzstiftenden Datentyp. Da die Probleme einer expliziten Modellbildung sich am deutlichsten in den Teilgebieten der Sprachwissenschaft bemerkbar machen, in denen aufgrund der besonderen Datenlage eine Beschränkung auf intuitive Sprecherurteile z.B. ausgeschlossen und ohne eine Berücksichtigung soziokultureller Rahmenbedingungen, historischer Überlieferung und lokaler Konzeptualisierungsstrategien eine Interpretation sprachlicher Belege unmöglich ist, erfolgt die kritische Bestandsaufnahme zur Frage eines angemessenen Datentyps erst im nächsten Unterabschnitt. Auch dort wird sich das Ziel des Einzelprojekts, den Zusammenhang zwischen Paradigmen in der Grammatiktheorie und bestimmten abstrakten Repräsentationsformen des intendierten Gegenstandsbereichs aufzuhellen, eine Perspektive auf die derzeitige Forschungslage eröffnen, die eine überprüfbare Hypothese über einen Datentyp nahelegt, der keine Dimension der sprachlichen Wirklichkeit ausblendet.

Deskriptive Adäquatheit von Datenstrukturen

Der überwiegende Teil der (Teil-)Beschreibungen von Sprachen aus dem amerikanischen Gebiet ist der Tradition des amerikanischen Strukturalismus verpflichtet, wenn man von den Missionarsgrammatiken absieht, bei denen nicht die Struktur der jeweiligen Sprache, sondern pädagogische Ziele im Vordergrund stehen. Im Falle des Warao, einer Sprache, die hauptsächlich im venezolanischen Teil des Orinocodeltas gesprochen wird, befinden wir uns in der glücklichen Lage, über zwei Darstellungen zu verfügen, deren Autoren zwar beide Priester sind, die aber eindeutig den verschiedenen Beschreibungstraditionen der Missionarsgrammatiken bzw. der strukturalistischen Grammatiken zuzuordnen sind (de Olea 1928, Osborn 1966/67). Während de Olea seine grammatischen Informationen rein kontrastiv an ihrem Verhältnis zum Spanischen orientiert, bemüht sich Osborn, die Morphophonemik und die beiden Hauptwortarten mit ihren Flexionsformen klassifikatorisch zu erfassen. In der letztgenannten Arbeit lassen sich die Definitionen, die z.B. im dritten Kapitel für spezielle Morphemklassen gegeben werden, unschwer auf ihren methodologischen Hintergrund zurückzuführen. Die einzelnen Klassen beruhen alle auf operationalisierten Verfahren der kombinatorischen Klassifikation und gehören damit zu der ursprünglichen Konzeption, nach der es

sich bei Datentypen um unstrukturierte Mengen oder Zusammenfassungen handelt.

Die am Ende des letzten Abschnitts erwähnte Verbindung zu einem Modell linguistischer Gegenstandskonstitution, das in seinen Erkenntnishorizont bewußt Dimensionen aufnimmt, die sich den kommunikativen, historischen und soziokulturellen Faktoren der Sprache verdanken, wird daher erst in dem Moment etabliert, wo auf dem Fundament einer sorgfältigen konzeptuellen Analyse des Beschreibungsobjekts ein anderes Datenformat gewählt wird. Die Entscheidung im Akademischen Rechenzentrum des *Summer Institute of Linguistics* für eine objektorientierte Umgebung, die sowohl den Linguisten bei der Feldarbeit unterstützen als auch die Dokumentation der Resultate nach international vereinbarten Standards garantieren soll, stellt eine Entscheidung für eine Datenstruktur dar, in der linguistische Theorien den Typ bilden, nach dem ganze Modelle als Elemente des Objektbereichs zu Klassen zusammengefaßt werden. Die Begründung für diesen behaupteten Zusammenhang leitet sich her aus der Einsicht, daß die allgemeinen Grundsätze der objektorientierten Programmierphilosophie, wie sie vom SIL verfolgt werden, durch ihre Deutung im Rahmen der coalgebraischen Systemtheorie bzw. in der damit äquivalenten Theorie der Versteckten Algebren eine präzise semantische Basis erhalten.

Der objektorientierte Ansatz geht aus von einer intuitiven Korrespondenz zwischen der Simulation eines Modells der realen oder fiktiven Welt in einem Programm und dem Modell selbst. Das Modell besteht nach dieser Auffassung aus Objekten, die bestimmte Eigenschaften (Attribute) haben, auf die wiederum bestimmte Methoden angewandt werden können. Die Objekte lassen sich zu Klassen zusammenfassen, die selbst in verzweigten Ordnungsrelationen zueinander stehen. Von besonderem Interesse ist dabei die Vererbungsrelation, bei der Objekte "späterer Generationen" Eigenschaften und Methoden der Erblasser übernehmen. Dabei kann es durchaus zu Modifikationen kommen, die es erforderlich machen, Objektidentität und Methoden- und Eigenschaftsäquivalenz voneinander zu trennen. Eng verbunden mit diesem Erfordernis ist der Begriff der Dateneinkapselung, der der intuitiven Einsicht Rechnung trägt, daß Objekte eine Innen- und eine Außenansicht bieten, die für die Klassenbildungen relevant sind. Diese Angaben zur Objektorientierung, mit denen keineswegs die Absicht einer repräsentativen Charakterisierung verbunden ist, sollen nur dazu dienen, die vom SIL entwickelte Programmierumgebung in ihrem Bezug zu unserer Fragestellung einordnen zu können (Abadí und Cardelli 1996, von Gudenberg 1993).

Das in diesem Zusammenhang zu erwähnende System CELLAR baut auf einer konzeptuellen Analyse der Erfordernisse auf, die sich aus den grundlegenden Eigenschaften der Objekte, mit denen der Sprachwissenschaftler konfrontiert ist, ergeben. In dem System CELLAR und der darauf aufbauenden Software *Lingua-Links*, die den Sprachwissenschaftler bei der Feldarbeit unterstützen soll, werden alle Einheiten mit den Mitteln einer konzeptuellen Spezifikationsprache (*Conceptual Modeling Language*) definiert. Obwohl die Syntax der konzeptuellen Sprache

in der Form von Bedingungen angegeben ist, die den Aufbau graphenähnlicher Strukturen regeln, ist ihre strukturelle Verwandtheit mit der coalgebraischen Spezifikationsprache (*Coalgebraic Specification Language*), die in Amsterdam und Dresden entwickelt wird, unübersehbar (Hensel et al. 1997, Simons und Thomson 1995). Die Familienähnlichkeit ist die Basis für die Teilhypothese des Projekts, daß in der coalgebraischen Systemtheorie ein adäquater Datentyp für die konzeptuelle linguistische Modellierung vorliegt und daß die für die coalgebraische Spezifikationsprache etablierten multidimensionalen Repräsentationsformen auf das konzeptuelle Medium übertragbar sind. Neben der Objektorientierung ist beiden formalen Modellierungsvorschlägen gemein, daß sie den Vorteil unterstreichen, der in der Möglichkeit beschlossen ist, Klassen von Objekten unter verschiedenen Blickwinkeln und Perspektiven (*views*) zu betrachten und auf diesem Weg zu einer mehrdimensionalen Repräsentation des Objektbereichs zu gelangen (Simons 1994, Jacobs 1997).

Dieses Potential des Perspektivenwechsels bildet auf der inhaltlichen Seite die Grundlage, bei der linguistischen Gegenstandskonstitution nicht nur den strukturellen Aspekt zu berücksichtigen, sondern über das Mittel des Perspektivenwechsels auch den kommunikativen, historischen und soziokulturellen Attributkomplexen natürlicher Sprache eigenständige Darstellungsformen zu widmen. Die Entwicklung eines erweiterten Objektbegriffs im Verein mit dem Aufbau eines adäquaten Spezifikationsmediums in einem Bereich der Linguistik, der sich traditionell fast ausschließlich auf deskriptive Probleme konzentrierte, haben daher die systematische Grundlage geschaffen, um einen Datentyp zu etablieren, der in einem entscheidenden Punkt über die abstrakten Datenstrukturen aus der Universalen Algebra hinausgeht, insofern er die Mehrdimensionalität des linguistischen Erkenntnisobjekts von vorneherein integriert. Da die Perspektivenvielfalt auf der formalen Seite darüberhinaus die Abbildung der Analysen in verschiedene Formate international vereinbarter Notationskonventionen gestattet, bietet sich die beim SIL geschaffene Systemumgebung dazu an, die Beschreibung des Warao nicht nur in einen theoretischen Rahmen einzubetten, der die methodologischen Restriktionen des amerikanischen Strukturalismus überwindet, sondern durch die Präsentation der deskriptiven Resultate in einer Vielzahl von etablierten Notationsstandards deren Zugänglichkeit für die Fachwelt sicherzustellen.

Die kritische Bestandsaufnahme der Forschungslage, die mit dem Thema der Datenstrukturen in ihrer Funktion als Repräsentanten von Typen linguistischer Gegenstandskonstitution verbunden ist, sollte dazu dienen, den Aufschlußwert zu belegen, der sich aus dieser Sicht auf das übergreifende Forschungsziel des Sonderforschungsbereichs ergibt, und in Parallelführung mit diesem Nachweis den inneren Zusammenhang deutlich machen, der die formalen, theoretischen und deskriptiven Problemfelder datenstruktureller Provenienz miteinander verbindet. Es war daher nötig, über einen neutralen Bericht zum Stand der Forschung hinauszugehen und die in diesem Gebiet erzielten Erkenntnisse gemäß der Leitidee dieses Einzelprojekts zu bewerten. Der vorstehende Abschnitt stellt infolge die-

ser Ausrichtung in einigen Passagen einen Vorgriff auf die Aufgabenbeschreibung in dem übernächsten Abschnitt dar. Bei der dort durchgeführten erneuten Reise mit dem Queneauschen Bus wird allerdings die Route in ihrer Gesamtheit weniger prominent sein, und es wird stattdessen den einzelnen Stationen eine erhöhte Aufmerksamkeit gewidmet werden.

3.4 Eigene Vorarbeiten

Das beantragte Einzelprojekt befindet sich in der glücklichen Lage, daß aufgrund der Vorarbeiten der Projektmitarbeiter, der Berater und der Mitarbeiter an dem vom Antragsteller geleiteten Arbeitsbereich am Seminar für Sprachwissenschaft sichergestellt ist, daß das vorzustellende Arbeitsprogramm ohne Verzögerung umgesetzt werden kann. Der Antragsteller selbst hat sich seit Jahren sowohl mit dem Untersuchungsfeld (co-)algebraischer Semantiken für Grammatikformalismen als auch mit methodischen Fragen der Beschreibung nicht-indogermanischer Sprachen beschäftigt. In der Arbeit von 1996 wird der Versuch unternommen, der kombinatorischen Komplexität in algebraischen Systemen mit Operationssymbolen höherer Ordnung mit kategorientheoretischen Mitteln zu begegnen. Komplementär zu diesem Verfahren ist der Vorschlag zu sehen, der ein Jahr später vorgelegt wurde und der die Probleme, die sich jenseits des kontext-freien Bereichs ergeben, mit dem Instrument eindeutiger Homomorphismen aus freien Termalgebren zu lösen unternimmt. In der Publikation von 1998 kommen wiederum kategorientheoretische Prinzipien zur Anwendung, die eine alternative Konzeptualisierung der Baumadjunktionsgrammatiken auf letztlich algebraischer Grundlage gestatten. Als das Papier von 1994 konzipiert wurde, war dem Verfasser nicht klar, daß mit der dort angegebenen situationstheoretischen Deutung der Merkmalsstrukturen ein Schritt in Richtung einer coalgebraischen Systemtheorie getan wurde. Bei einer genaueren Analyse des von Aczel (1990) entwickelten Modells für die Situationstheorie wird nämlich deutlich, daß es sich bei diesem Modell um eine Coalgebra handelt, obwohl die Konstruktion nicht in dieser Terminologie vorgetragen wird. Die deskriptiven Interessen des Antragstellers sind durch die methodologische Abhandlung (Mönnich 1998b) dokumentiert. Stephan Kepser hat in seiner Magisterarbeit (1994) den ersten Nachweis geführt, daß die Erfüllbarkeit einer bestimmten Formalisierung der HPSG entscheidbar ist. Seine spätere Publikation, die er zusammen mit Klaus Schulz vorgelegt hat, gehört ebenfalls in das Untersuchungsgebiet des lizensierenden Theorieparadigmas. Seine Erfahrung in diesem Bereich wird ergänzt durch eine Reihe von Tom Cornells Arbeiten (1997a,b, im Erscheinen), die der Frage gewidmet sind, ob es sich bei dem minimalistischen Programm Chomskys um eine Rückkehr zum generierenden Paradigma handelt und in welchem Maß die Einordnung in ein bestimmtes theoretisches Modell unabhängig von dem Stil seiner Formalisierung ist. Stefanie Herrman wird sich zum Zeitpunkt des Beginns des geplanten Sonderforschungsbereichs ein Jahr lang mit der Erhebung von Daten des Warao befaßt haben. Sie wird dabei von

einem der besten Kenner dieser Sprache, Dieter Heinen, Direktor des Anthropologischen Seminars am Instituto Venezolano de Investigaciones Científicas, beraten und hat einen ersten kurzen Bericht über ihr Dissertationsvorhaben vorgetragen (Herrman 1998). Sie wird sich bei ihrer Projektarbeit auf die Hilfe von Emmerich Weißhar stützen, dem Ethnolinguisten am hiesigen Seminar für Vergleichende Sprachwissenschaft, der ein Übungsbuch zum Warao verfaßt hat (Weißhar 1986). Teil des Dokumentationssystems CELLAR, das für die Aufnahme der Warao-Daten zugrundegelegt werden soll, ist eine für linguistische Zwecke konzipierte Modellierungssprache. Diese Sprache scheint sich zu eignen für eine Übertragung in ein Deduktionssystem. Für diese Aufgabe ist Stephan Kepser aufgrund seiner Dissertation einschlägig ausgewiesen (Kepser 1998).

3.41 Veröffentlichungen

- CORNELL, T.L. (1997a): 'Derivational and representational views of minimalist syntactic calculi.' In: *Proceedings of LACL '97*, Nancy, France.
- CORNELL, T.L. (1997b): 'A type-logical perspective on minimalist derivations.' In: *Proceedings of the Formal Grammar Conference*, Aix-en-Provence, France.
- HERRMAN, S. (1998): 'El proyecto de una "etnografía de la comunicación del idioma Warao" en la parte occidental del delta del Orinoco.' erscheint in den Akten des *Congreso Nacional de Antropología: Hacia la Antropología de Siglo XXI*, Mérida, Venezuela.
- KEPSEK, S. (1994): 'A Satisfiability Algorithm for a Typed Feature Logic.' Tech. rep., Sonderforschungsbereich 340, Tübingen, Arbeitsbericht Nr. 60.
- KEPSEK, S. (1998): *Combination of Constraint Systems*. Ph.D. thesis, Universität München, Diss.
- KEPSEK, S. & K. SCHULZ (1996): 'Combination of Constraint Systems II: Rational Amalgamation.' In: E. Freuder (ed.), *Principles and Practice of Constraint Programming*, vol. CP '96 of *Lecture Notes in Computer Science 1118*, pp. 282–296, Springer.
- MAUNER, G., V.A. FROMKIN & T.L. CORNELL (1993): 'Comprehension and Acceptability Judgments in Agrammatism: Disruption in the Syntax of Referential Dependency.' *Brain and Language* 45: 340–370.
- MÖNNICH, U. (1994): 'Zur deklarativen Repräsentation grammatischer Datenstrukturen.' Ms., Universität Tübingen.
- MÖNNICH, U. (1996): 'Sprachliche Skizzen.' Ms., Universität Tübingen.
- MÖNNICH, U. (1997): 'On Cloning Context-Freeness.' Erscheint in: H.P. Kolb & U. Mönnich (eds.), *The Mathematics of Syntactic Structure*, Berlin: de Gruyter.
- MÖNNICH, U. (1998a): 'The Logic of Internal and External Perceptions.' erscheint in einem Sonderheft *Semantik der Linguistischen Berichte*.
- MÖNNICH, U. (1998b): 'TAGs M-Constructed.' In: *Proceedings of the TAG+ Workshop*, Philadelphia.

WEISSHAR, E. (1986): 'Warao: Lese- und Sprachprogramm.' *Arbeitspapiere der Tübinger Ethnolinguistik III*.

3.5 Ziele, Methoden, Arbeitsprogramm und Zeitplan

3.51 Ziele und Methoden

Den Mittelpunkt des Projekts bildet die Hypothese, daß aus einer systematischen Analyse der vorausgesetzten Datentypen Aufschlüsse über Paradigmenwechsel in der Sprachtheorie zu erzielen sind. Es werden drei Klassen von Datenstrukturen in der Absicht unterschieden, sie als charakteristisch für strukturalistische, generierende und lizensierende Ansätze linguistischer Theoriebildung nachzuweisen. Bei den drei Arten von Datentypen, deren Fruchtbarkeit für eine grundlegende Differenzierung linguistischer Erkenntnisobjekte postuliert wird, handelt es sich um Klassen, (freie) Algebren und (terminale) Coalgebren. Soweit die formalen Eigenschaften dieser Strukturen Gegenstand der Untersuchung sind, werden die Methoden aus den einschlägigen Gebieten der naiven Mengenlehre, der Komplexitätstheorie und der Algebra angewandt. Auf rein formalem Weg sollen Behauptungen überprüft werden, die auf der Basis kognitivistischer Annahmen über die Berechenbarkeit verschiedener Modelle linguistischer Beschreibungen aufgestellt worden sind, wobei im Sinne der Leitidee des Projekts eine solche Überprüfung ihren Ausgang von den modellspezifischen Datentypen zu nehmen hat. Ein Versuch, der sich die Bestätigung einer Annahme über die Affinität von Theorie- und (abstrakten) Datenklassen zum Ziel gesetzt hat, kann sich nicht auf das Arsenal formaler Instrumentarien beschränken, sondern wird durch eine wissenschaftssystematische Analyse der Konsequenzen, die sich aus den Axiomen der jeweiligen linguistischen Forschungsprogramme ergeben, zu ergänzen und zu relativieren sein. Eine Deckungsgleichheit der Identitätskriterien für Daten- und Theorieklassen ist nicht zu erwarten, und eine wichtige Aufgabe wird darin bestehen, die Sprengkraft entdeckter Divergenzen zwischen Formen des linguistischen Erkenntnisgegenstands und Formen des theoretischen Zugriffs zu bewerten. Darüberhinaus ist nicht zu erwarten, daß für alle Aspekte der Konzeptualisierung sprachlicher Gebilde geeignete Datenstrukturen entwickelt worden sind. Der Erfolg des Projekts wird daher wesentlich davon abhängen, ob seine Anschlußhypothese sich bestätigen läßt: daß nämlich in dem coalgebraischen Denotationsbereich der lizensierenden Theorievariante ein Datentyp zur Verfügung steht, der sowohl den deskriptiven Ansprüchen des Gebiets der Sprachwissenschaft, das sich mit nicht-indogermanischen Sprachfamilien beschäftigt, als auch den Traditionen in der Gegenstandsbestimmung der Sprachwissenschaft genügt, die in den kommunikativen, historischen und soziokulturellen Faktoren essentielle Merkmale des menschlichen Sprachvermögens sehen. Die Anschlußhypothese schließt an jüngste Resultate der algebraischen Systemtheorie an, die entscheidende Fortschritte

in der semantischen Fundierung objektorientierter Analysen erbracht haben. Das Forschungsziel in diesem Projektteil läßt sich in dem Sinn präzisieren, daß zwischen der coalgebraischen Spezifikationsprache und der für linguistische Zwecke entwickelten konzeptuellen Modellierungssprache eine Übersetzungsrelation erstellt und die Ausdruckskraft des linguistischen Beschreibungsinstrumentes an einer geeigneten amerindischen Sprache, dem Warao, evaluiert werden soll. Die für dieses Teilziel eingesetzten Methoden gehören einerseits zu den Techniken der (syntaktischen) Theorieinterpretation und andererseits zu dem klassischen Kanon der ethno-linguistischen Datenerhebung und Dateninterpretation.

3.52 Arbeitsprogramm

Zur Berechenbarkeit verschiedener Typen von Datenstrukturen

Um sich eine Vorstellung von der Nützlichkeit von Bäumen zu machen, sei kurz ein Beispiel betrachtet, das auf den ersten Blick keinen Bezug zu der Frage nach möglichen Typen von Datenstrukturen hat, die in der Sprachwissenschaft eine prominente Rolle spielen. Bei dem Beispiel handelt es sich um eine Sortierungs-routine, die sich in jedem Standardwerk über algorithmische Verfahren beschrieben findet und darin besteht, auf einem binären Baum eine “Abpflückordnung” zu definieren, die es erlaubt, eine Liste von Eingaben in geordneter Form wieder auszugeben. Bei dem *treесort* genannten Verfahren wird eine Liste von Zahlen z. B. in einen binären Suchbaum transformiert und jede der eingegebenen Zahlen wird wieder ausgegeben, sobald ein bestimmtes Durchlaufverfahren ein zweites Mal bei ihr ankommt. Die Zahlen auf der Ausgabeliste erscheinen dann in ansteigender Folge. Die Struktur der Suchroutine ist wesentlich von dem Format bestimmt, in dem die Eingabedaten dem Durchlaufverfahren präsentiert werden. Bei binären Bäumen handelt es sich um eine rekursive Struktur, die entweder leer ist oder sich aus einer Wurzel und zwei Unterbäumen aufbaut. Das ist der tiefere Grund dafür, warum sich bestimmte Durchlaufverfahren, die Strukturen in einem festgelegten Sinn abarbeiten, auf Bäumen relativ einfach spezifizieren lassen (Harel 1993).

Es ist keineswegs so, daß Suchroutinen, die auf Abpflückordnungen beruhen, sich nur für Datenstrukturen definieren lassen, die den Charakter von abstrakten Datenstrukturen in dem Verständnis haben, das sie mit freien Algebren gleichsetzt. Wenn dieser Zusammenhang so strikt wäre, bestünde keine Hoffnung, die bekannten Entscheidbarkeitsresultate, die sich auf Bäume und Graphen beziehen, in linguistisch relevanter Weise zu verfeinern. Wie in dem Bericht zum Stand der Forschung unterstrichen wurde, ist die Theorie der endlichen Bäume entscheidbar und die Theorie der endlichen Graphen unentscheidbar, wobei beide Theorien in derselben Sprache notiert sind. Es wurde außerdem darauf hingewiesen, daß sich die Rektions- und Bindungstheorie weitgehend im Rahmen der Theorie endlicher Bäume formalisieren läßt, die in der monadischen Logik zweiter Ordnung

spezifiziert ist.

Dieses Resultat steht in starkem Gegensatz zu dem Trend, der mit dem Wechsel von generierenden Produktionssystemen zu lizensierenden oder *constraint*-basierten Formalismen einsetzte. Mit der ursprünglichen Definition von Komplexitätsklassen auf der Grundlage der Chomsky-Hierarchie war im generierenden Theorieparadigma die Erwartung verbunden, daß mit diesen Klassen wesentliche Eigenschaften der Struktur natürlicher Sprachen erfaßt seien. Diese Erwartung war von der Idee inspiriert, daß die natürlichen Sprachen nur einen generativen Mechanismus begrenzter Ausdrucksstärke benötigten und sie auf diesem Weg von einer reichen Klasse nicht-natürlicher Sprachen in empirisch signifikanter Weise abgegrenzt werden könnten. Dieses Forschungsprogramm wurde in der Theoriebildung aufgegeben, die Sprachen über die Angaben von Beschränkungen definiert, die von Strukturanalysen der Ausdrücke dieser Sprachen zu erfüllen sind. Der Grund für die Aufgabe des ursprünglichen Forschungsprogramms ist auf der einen Seite in der Tatsache zu sehen, daß die zur Bestimmung der Komplexität lizensierter Modellklassen notwendigen Methoden teilweise erst entwickelt werden mußten und daß auf der anderen Seite mit dem Paradigmenwechsel in der Theoriebildung ein prononciertes Desinteresse an Fragen generativer Kapazität von Formalismen einherging, das von der Überzeugung gespeist war, mit der Reduktion des Spektrums natürlicher Sprachen auf endlich wenige Klassen deren rekursionstheoretische Eigenschaften vernachlässigen zu können (Chomsky 1981). Vor diesem forschungsstrategischen Hintergrund ist der Nachweis in Rogers (1994), daß die Rektions- und Bindungstheorie wahrscheinlich eine entscheidbare Formalisierung gestattet, von besonderem Interesse und war der Auslöser für eine Reihe forschungsintensiver Unternehmen mit dem Ziel, die von Rogers gelassenen Lücken zu schließen. Die von ihm als Grundlage gewählte Theorie endlicher Bäume entspricht mit den von ihr lizensierten Modellklassen genau den kontextfreien Sprachen, was zu unmittelbaren Problemen mit der Behandlung der empirisch gut abgesicherten Kreuzabhängigkeiten in Sprachen wie Bamba-ra, Holländisch oder Schweizerdeutsch führt. Die andere Lücke hängt mit der freien Indizierung zusammen, die, zumindest in der Darstellung von Rogers, die Entscheidbarkeit der von ihm zugrundegelegten formalen Theorie sprengt.

Ein Schwerpunkt in diesem Bereich des Arbeitsprogramms wird daher in dem Versuch bestehen, die von Rogers hinterlassenen Probleme, die mit der Komplexitätstheoretischen Einschätzung lizensierender Theorien verbunden sind, einer formal abgesicherten und linguistisch korrekten Lösung zuzuführen. Im Fall der Erweiterung des kontextfreien Bereichs sei ein möglicher Ausweg skizziert. Es ist seit Montagues Vorschlägen zu einer algebraisch formulierten Syntax bekannt, daß die generative Kapazität neben dem Regelformat in entscheidendem Maße von den syntaktischen Operationen abhängt, die von einer Grammatiktheorie zugelassen werden, eine Einsicht, die in virtuoser Weise im kategorialgrammatischen Umfeld umgesetzt wird. Formuliert man die Regeln in der Weise, daß diese Operationen zunächst nur benannt und ihre Ausführung einem weiteren Schritt

überlassen wird, entstehen als Ergebnis der ersten Phase dieses schrittweisen Verfahrens zunächst nur reguläre Bäume. Die Vermutung ist nun, daß die Ausführung jener Klasse von Operationen, die für die Darstellung der Kreuzabhängigkeiten ausreicht, wieder auf der Basis jener Theorie möglich ist, von der Rogers ausgegangen ist. Sollte sich diese Vermutung als richtig erweisen, würde die Form der deklarativen Theoriebildung nicht verlassen, sondern nur zu einem zweistufigen Ansatz verfeinert: In der ersten Stufe werden mit den Mitteln derselben Theorie endlicher Bäume die regulären Baumfamilien definiert, und auf der zweiten Stufe werden unter Ausnutzung desselben definitorischen Verfahrens diese Baumfamilien in eine Konfiguration überführt, die die Kreuzabhängigkeiten in deskriptiv angemessener Weise repräsentiert.

Im Fall der freien Indizierung entsteht für die Formalisierung der auf Prinzipien und Parameter fußenden Theorie ein Problem, das sie mit ihrer lizensierenden "Schwester", der Head Driven Phrase Structure Grammar, teilt. Auch für die HPSG gibt es einen bisher nicht veröffentlichten Beweis, daß sie zumindest in der von King vorgeschlagenen Formalisierung unentscheidbar ist (King et al. 1997). Angesichts der geometrischen Nähe zu dem zweidimensionalen Charakter der Zeit- und Platzeigenschaften von Turingmaschinenberechnungen ist es nicht verwunderlich, daß beide Beweise auf einer Reduktion auf das unentscheidbare Parkettierungs- oder Kachelungsproblem beruhen. Das Problem besteht darin, für einen bestimmten Ausschnitt der Ebene zu untersuchen, ob er mit einer endlichen Auswahl an Kacheltypen so belegt werden kann, daß bestimmte Rand- und Nachbareigenschaften erfüllt sind. Man kann sich diese Nachbarschafts- und Randbedingungen in der Weise vergegenwärtigen, daß die Kacheln an den Stellen, wo sie "vertikal" oder "horizontal" zusammenstoßen, dieselbe Farbe aufweisen müssen. Die Reduktion auf dieses unentscheidbare Problem läuft somit auf den Nachweis hinaus, daß im Fall der freien Indexierung die Möglichkeit bereitsteht, sich von der Baumstruktur maximal zu entfernen, indem die für die Kachelung relevanten Nachbarschafts- und Randeigenschaften eine Baumzerlegung mit begrenzt vielen Komponenten nicht mehr gestatten. In dem Moment, wo eine derartige Zerlegung noch erreichbar ist, bleibt die entsprechende Theorie entscheidbar, obwohl Kantenzusammenführungen nicht mehr verboten sind. Das Forschungsproblem, das sich angesichts der beschriebenen Situation stellt, besteht darin zu untersuchen, ob es spezifisch linguistische Gründe gibt, die gegen eine vollständige Aufgabe der Baumstruktur sprechen. Sollte sich eine positive Antwort mit Hilfe der grundlegenden Prinzipien sei es der Rektions- und Bindungstheorie, sei es der Head Driven Phrase Structure Grammar verteidigen lassen, wäre die Entscheidbarkeit der entsprechenden Theorie wieder gesichert, da man es dann nur noch mit Graphen zu tun hätte, die eine hierarchische, baumähnliche Struktur aufweisen und sich rekursiv aus gewissen primitiven Graphen durch Anwendung geeigneter Operationen aufbauen lassen. Auf Strukturen dieser Art ließen sich wieder Suchverfahren anwenden, von denen dieser Abschnitt seinen Ausgang nahm.

Zur theoretischen Relevanz von Datenstrukturen

Die zentrale Hypothese des Projekts zieht eine Parallele zwischen den verschiedenen Paradigmen der generierenden und lizensierenden Theoriebildung und den formalen Eigenschaften der sie kennzeichnenden Datentypen. Für das Paradigma der lizensierenden Theorievariante wird eine Form der Gegenstandskonstitution postuliert, die mit den Techniken der coalgebraischen Systemtheorie präzisiert zu werden vermag, und es wird dieses Postulat durch eine Zusatzhypothese ergänzt, nach der die Techniken der Systemtheorie die Formalisation einer Gegenstandskonzeption erlauben, die sprachliche Objekte nicht auf ihre strukturellen Faktoren reduziert, sondern in diesen Objekten das Resultat des Zusammenwirkens des Sprachvermögens mit anderen Vermögen des Geistes sieht, Vermögen, die den kommunikativen, historischen und soziokulturellen Komponenten menschlichen Verhaltens zugrundeliegen. Beide Hypothesen im Zusammenhang mit der der Systemtheorie zugewiesenen Rolle haben jene Teile des Arbeitsprogramms zur Folge, die den theoretischen Aspekt von (Typen von) Datenstrukturen berühren. Insbesondere handelt es sich dabei um den Nachweis, daß das intendierte Modell lizensierender Theorien, die Merkmalsstrukturen, mit den Mitteln der Systemtheorie beschreibbar sind, um die Diskussion der "richtigen" Auffassung von Merkmalsstrukturen und um die Integration eines mehrdimensionalen Objektbegriffs in den Rahmen der Systemtheorie.

Da die Begriffe, mit denen die Systemtheorie arbeitet, relativ neu sind, wird es auch eine Aufgabe des Projekts sein, eine an linguistischen Bedürfnissen orientierte Präsentation der Theorie zu erarbeiten. In der Hoffnung, daß die folgenden Bemerkungen dennoch eine Idee von der geplanten Durchführung des Arbeitsprogramms vermitteln, sei zumindest der Versuch gemacht, die Behandlung der Merkmalsstrukturen zu erläutern. Es wurde oben daran erinnert, das sich binäre, etikettierte Bäume durch die Angabe von zwei "Konstruktoren" beschreiben lassen: Ein solcher Baum ist entweder leer oder trägt ein Etikett an seiner Wurzel und verzweigt sich in zwei Unterbäume. Diese Form der algebraischen Spezifikation hat ein kanonisches Modell, das aus den geschlossenen Termausdrücken konstruiert werden kann, d.h. den Ausdrücken, die sich durch iterative Anwendung der Konstruktoroperationen rekursiv generieren lassen. Diese Menge der Termausdrücke ist isomorph zu den endlichen Bäumen. Dual zu diesem Vorgehen beruht die coalgebraische Spezifikation auf sogenannten Destruktoroperationen. Sie beinhalten, welche Operationen oder Beobachtungen auf dem Datentyp der Bäume vorliegen, nämlich die Entfernung eines Etiketts oder die Wahl einer rechten oder linken Verzweigung an einem Knoten. Diese Beobachtungen machen keine Aussagen über die interne Struktur der "Baumwelt" und erlauben zu ihr nur einen beschränkten Zugang über die mit den Destruktoren verbundenen Manifestationen. Der entscheidende Unterschied zwischen der algebraischen und der coalgebraischen Spezifikation liegt darin beschlossen, daß in dem einen Fall die Operationen in die Baumwelt hineingehen, während sie in dem anderen Fall

die umgekehrte Richtung haben. Diesem Umstand entspricht, daß das kanonische Modell der algebraischen Spezifikation eine initiale Algebra ist, während das kanonische Modell der coalgebraischen Spezifikation eine terminale Coalgebra bildet.

Der Schritt von dem Bild des “konstruktiven” Aufbaus zu der “destruktiven” Beobachtung schafft die entscheidende Voraussetzung, um dem Problem zu begegnen, daß das graphenähnliche Erscheinungsbild der Merkmalsstrukturen infolge der Identifizierungsmöglichkeit von Substrukturen keine induktive Definition zuläßt. Obwohl diese Möglichkeit ausgeschlossen ist, lassen sich auch an einer Merkmalsstruktur Beobachtungen vornehmen wie z.B., daß ein Knoten einen bestimmten Wert hat und daß aus einem Knoten eine bestimmte Anzahl an Attributen hinausführen, daß dieses in deterministischer Weise geschieht etc. Eine Kennzeichnung der Merkmalsstrukturen im Rahmen der Systemtheorie, die ihren Ausgang von Destruktoren nimmt, ist bisher in der formal-linguistischen Literatur nicht versucht worden. Angesichts dieser Forschungslage und wegen des zentralen Aufschlußwerts, die einer coalgebraischen Behandlung der Merkmalsstrukturen für die These zukommt, daß zwischen dem lizensierenden Theorieparadigma und dem systemtheoretischen Datentyp ein innerer Zusammenhang besteht, ist eine eigene Untersuchung der skizzierten semantischen Modellierungsstrategie unumgänglich.

In diesem Zusammenhang ist ebenfalls das Verhältnis zwischen einer — an dieser Stelle supponierten — coalgebraischen Charakterisierung der Merkmalsstrukturen und der damit konkurrierenden bereichstheoretischen Kennzeichnung durch Moshier (1995) zu untersuchen, die weiter oben bereits erwähnt wurde. Die Klärung des Verhältnisses ist deswegen von grundsätzlicher Bedeutung, weil sich an den Interpretationsvarianten der Merkmalsstrukturen eine intensive wissenschaftstheoretische Diskussion entzündet hat, die um die Frage kreist, ob es sich bei der aktuellen Version der Head Driven Phrase Structure Grammar (Pollard und Sag 1994) um eine Mischform zwischen einer lizensierenden und einer mit derivationellen “Schlacken” behafteten Theorie handelt. Für eine Beantwortung dieser Frage bietet offenbar die Kategorientheorie die angemessene Abstraktionsebene. Nach vorläufigen Analysen läßt sich Moshiers Kennzeichnung mit den Mitteln der coalgebraischen Systemtheorie wiederholen. Sollten sich diese Analysen bestätigen, besteht zwischen beiden Charakterisierungen der Merkmalsstrukturen zumindest konzeptuell kein Unterschied, und die gegenwärtige wissenschaftstheoretische Diskussion würde “nur” die ontologischen Eigenschaften der jeweiligen Kategorie betreffen, auf der die entsprechende Coalgebra definiert ist.

Grundlage für die Vermutung, daß mit dem coalgebraischen Typ von Datenstrukturen auch eine systematische Möglichkeit bereitsteht, mehrere Komponenten sprachlicher Objekte in ihren aufeinander bezogenen Manifestationen zu berücksichtigen, stützt sich auf die Einsicht, daß coalgebraische Spezifikationen eng verwandt sind mit vorliegenden objektorientierten Analysen (Jacobs 1997). Ein wichtiger methodischer Vorteil dieser Analysen ist der Umstand, daß sie

es gestatten, dasselbe Objekt unter verschiedenen Perspektiven mit den jeweils relevanten Attributen und Methoden darzustellen. Welcher Fortschritt mit diesem Verfahren verbunden ist, läßt sich unmittelbar an den visuellen Modellen ablesen, die das objektorientierte System CELLAR des *Summer Institute of Linguistics* für die Repräsentation sprachlicher Gebilde anbietet (Simons 1994). Die wissenschaftstheoretisch unbefriedigende Situation, die in dem Moment entsteht, wo verschiedene Richtungen in der Sprachwissenschaft den gemeinsamen empirischen Gegenstand unter einander komplementären Perspektiven konzeptualisieren und als Folge dieser Divergenz des Erkenntnisinteresses ganze Datenbereiche ausblenden, ist in der Präambel zu diesem Antrag auf Einrichtung eines neuen Sonderforschungsbereiches dargelegt. Die technische Möglichkeit als solche, der linguistischen Gegenstandskonstitution eine "liberale" formale Plattform anzubieten, wird den Streit darüber nicht schlichten, ob das menschliche Sprachvermögen am vorteilhaftesten in seinen isolierten Manifestationen oder mit mehr Gewinn im Verein mit anderen kognitiven Funktionen zu untersuchen ist. Die Tatsache eines theoretisch fundierten Ansatzes, der von vorneherein die Multidimensionalität seines Objektbereichs vorsieht, sollte jedoch die rationale Basis für eine Verständigung liefern. Es wird daher ein weiteres Anliegen des Einzelprojektes sein, dem Unterschied zwischen Objektidentität und Eigenschaftsäquivalenz besondere Aufmerksamkeit zu widmen.

Zur deskriptiven Relevanz von Datenstrukturen

In mittelalterlicher Sprechweise ist bisher nur in formalem Sinn von Datenstrukturtypen die Rede gewesen. Datenstrukturen haben daneben auch eine materiale Seite, die in der ihnen entsprechenden Spezifikationsprache zum Ausdruck kommt. Sollte die Unterhypothese des gegenwärtigen Einzelprojektes sich bestätigen, daß die Objekte der coalgebraischen Systemtheorie nicht nur den geeigneten Datentyp für die lizenzierende Theoriebildung abgeben, sondern daß mit ihnen auch eine integrative Auffassung linguistischer Gegenstandskonstitution semantische Gestalt gewinnt, dann bietet sich in syntaktischer Wende das Programm an, die Spezifikationsprache der Systemtheorie deskriptiv zu evaluieren. Die Spezifikationsprache erhält durch diese Wende den Status eines erklärungsadäquaten Notationssystems, das durch den Beleg sprachlicher Phänomene falsifiziert wäre, die sich mit seinen Ausdrucksmitteln nicht angemessen erfassen lassen.

Als Grundlage für dieses deskriptive Unternehmen ist im Hinblick auf das beschriebene Forschungsziel das System CELLAR mit seiner objektorientierten konzeptuellen Modellierungssprache gewählt worden. Da die oben erwähnte coalgebraische Spezifikationsprache eng mit objektorientierten Analysen verwandt ist und diese Verwandtschaft sich auch im Fall der konzeptuellen Modellierungssprache beobachten läßt, ist auf der "materialen" Ebene der Anschluß an das Arbeitsprogramm zum systemtheoretischen Datentyp gewährleistet. Die Orientierung an diesen Datentyp und das mit ihm verbundene Forschungsvorhaben schließt die

Benutzung anderer, in der Feldarbeit bewährter Programmwerkzeuge aus. So verbietet sich z.B. die Übernahme der Entscheidung, die im Max-Planck-Institut in Nijmegen für das System SHOEBBOX getroffen wurde, da dieses System nicht nur einem anderen Datenkonzept verpflichtet, sondern auch in seiner Anbindung an international vereinbarte Standardnotationen sehr eingeschränkt ist,—eine Einschränkung, die die Dokumentation und den Austausch von Beschreibungsergebnissen erheblich erschwert, wenn nicht unmöglich macht.

Zu dem Gegenstand der Beschreibung, dem Warao, wurden weiter oben die Gründe angeführt, die zu seiner Wahl geführt haben. Da es sich um eine nach bisherigen Erkenntnissen genealogisch isolierte Sprache handelt, deren vorliegende Analysen der strukturalistischen Methodologie folgen, bietet sie sich als idealer Prüfstein für das Beschreibungspotential eines objektorientierten Ansatzes an. Durch die Feldarbeit der vorgesehenen Mitarbeiterin und die guten Kontakte zu ausgewiesenen Kennern des Warao ist mit Verzögerungen in der Durchführung des Arbeitsprogramms nicht zu rechnen. Der Schwerpunkt der Untersuchung soll dabei auf einer systematischen Erfassung auch der Faktoren liegen, die zu den schon mehrfach erwähnten kommunikativen und soziokulturellen Dimensionen gehören. So lassen z.B. erste Feldergebnisse vermuten, daß Perzeptionsausdrücke mit Evidentialfunktion mit der üblichen Sinnesartenhierarchie nicht vereinbar sind, ein Faktum, das auf soziokulturelle Einflüsse zurückzuführen sein dürfte.

Das Ziel dieses Projektteils besteht somit in der Erstellung einer repräsentativen Datenbank für das Warao, die gleichzeitig der Evaluation des benutzten semantisch fundierten Notationsformats dient. Dabei sollen zumindest die Kriterien erfüllt sein, an denen Simons (1993) die Qualität von Dokumentationsumgebungen mißt. Nach ihm muß der multilinguale Charakter der zu bearbeitenden Daten berücksichtigt werden können, dann muß das System in der Lage sein, hierarchische Strukturen arbiträrer Tiefe aufzubauen. Drittens ist es notwendig, den multidimensionalen Daten die verschiedensten Arten von Analysen und Informationen zuordnen zu können. Ferner fordert die starke Integration des Datenmaterials das Speichern und Verfolgen assoziativer Verknüpfungen zwischen verwandten Daten, und obendrein sollen die Daten in einem konventionellen Format präsentiert werden können. Unter Beachtung dieser Kriterien ist beabsichtigt:

1. ein einfaches Wörterbuch oder ein komplexeres Lexikon entweder direkt oder aus interlinearisierten Texten heraus zu erstellen;
2. Texte zu interlinearisieren, so daß neue Worte automatisch ins Wörterbuch übernommen werden;
3. die Texte nach grammatikalischen Gesichtspunkten zu sortieren und dann zu analysieren;
4. kulturelle Notizen einzugeben und in einer Datei komplementär zum Text abzulegen.

Zur Datenverwaltung und Analyse trägt bei, daß das gewählte Programm

1. einen Texteditor besitzt, mit dem sich die Texte eingeben und editieren lassen;
2. die Möglichkeit bietet, sehr schnelle Suchvorgänge durchzuführen;
3. durch eine rigorose Selektionsoption dem Benutzer erlaubt, nur die Einträge zu sehen, die zu bestimmten Kriterien passen;
4. weiterhin eine 'Flash-Kartenfunktion hat, die dem Sprachlerner eine Hilfe ist; und
5. das Numerieren und Interlinearisieren von Text ermöglicht.

Damit sind keineswegs alle Möglichkeiten aufgelistet, die von dem System CELLAR unterstützt werden. In der Durchführung dieses Teils des Arbeitsprogramms wird kontinuierlich zu überprüfen sein, ob die deskriptiven Resultate dem praktischen Ziel einer repräsentativen Dokumentation und dem theoretischen Ziel einer Überprüfung der dem objektorientierten System hypothetisch unterstellten Ausdruckstärke gerecht werden.

3.53 Zeitplan

1. Jahr Arbeit an folgenden Schwerpunkten:

- Beginn der Untersuchungen zu Entscheidungsfragen baumähnlicher Strukturen,
- zur Formalisierung der konzeptuellen Modellierungssprache und
- Auswertung der ersten Felddaten des Warao.
- Installation und Einarbeitung in die Systeme CELLAR und Lingua-Links.

2. Jahr

- Fortsetzung der Untersuchungen zu Entscheidungsfragen mit besonderer Berücksichtigung von graphenähnlichen Merkmalstrukturen.
- Wissenschaftssystematische Analyse informationstheoretischer Modelle für lizensierende Theorien.
- Konstruktion einer formalen Semantik für die Spezifikationssprache von CELLAR.
- Beginn der Entwicklung einer strukturierten Dokumentation ausgewählter Phänomene des Warao.

3. Jahr

- Zusammenfassung der Ergebnisse zu formalen Eigenschaften von Datentypen.

- Ausarbeitung des Konzepts mehrdimensionaler Gegenstandskonstitution in der Sprachtheorie.
- Untersuchung der CELLAR-Sprache als Deduktionssystem.
- Vorläufiger Abschluß und Integration der deskriptiven Resultate zum Warao.

3.6 Stellung innerhalb des Programms des Sonderforschungsbereichs

Linguistische Datenstrukturen lassen sich unter verschiedenen Gesichtspunkten zu Einheiten zusammenfassen. Daten, die zu einem Phänomenbereich gehören, die einem Areal zuzuordnen sind, die einer historischen Phase entstammen oder den Sprachgebrauch einer sozialen Schicht dokumentieren, sind z.B. als empirische Basis für die Verifikation oder die Falsifikation einer bestimmten Hypothese geeignet. Das beantragte Einzelprojekt abstrahiert von diesen inhaltlichen Gesichtspunkten und betrachtet stattdessen die strukturellen Gemeinsamkeiten, die für die implizite Gegenstandskonstitution ganzer Theorieparadigmen kennzeichnend sind.

Innerhalb des geplanten Sonderforschungsbereichs besteht eine enge Verbindung zum Projekt A1, da die formalen Eigenschaften von Annotationssystemen durch die Einschränkung auf intendierte Modellklassen beeinflusst werden. Insbesondere vereint beide Projekte das Interesse an den Konsequenzen, die aus dem Grad der Strukturähnlichkeit der betrachteten Modelle zu endlichen Bäumen resultieren. Enge inhaltliche Berührungspunkte ergeben sich ebenfalls zum Projekt A3. In beiden Vorhaben spielt der Grammatikalitätsbegriff eine herausragende Rolle, wobei im Nachbarprojekt der Schwerpunkt auf den Status introspektiver Daten gelegt wird, während das gegenwärtige Projekt der Frage nachgeht, unter welchen Voraussetzungen überhaupt eine nachprüfbar Unterscheidung zwischen "grammatisch" und "ungrammatisch" garantiert ist. In dem Problembereich der Multidimensionalität sprachlicher Gegenstände gibt es Affinitäten zu fast allen Projekten des phänomenorientierten B-Bereichs. Parallelen bestehen insbesondere zum Projekt B3 auf dem Gebiet der evidentiellen Funktion von Perzeptionsverben. Es richtet sich das Interesse im Projekt Ehrich-Reis allerdings vornehmlich auf die semantisch-syntaktische Rolle dieses Phänomens, während bei der Beschreibung dieser Verwendungsweise im Warao ein kommunikativer Faktor im Vordergrund steht, für den es in der einschlägigen Literatur bisher nur Belege aus westaustralischen Sprachen gibt.

Außerhalb des geplanten Sonderforschungsbereichs sind Kooperationen mit folgenden Fachkollegen vereinbart:

- Prof. Dr. Peter Aczel, Manchester
- Prof. Dr. Dieter Heinen, Caracas

- Prof. Dr. Edward Keenan, Los Angeles
- Prof. Dr. Lawrence Moss, Bloomington
- Prof. Dr. Klaus Schulz, München
- Dr. Jean Véronis, Aix-en-Marseille

3.7 Ergänzungsausstattung für das Teilprojekt

es bedeuten:

PK: Personalbedarf und –kosten (Begründung vgl. 3.71)

SV: Sächliche Verwaltungsausgaben (Begründung vgl. 3.72)

I: Investitionen (Geräte über 20.000; Begründung vgl. 3.73)

PK	1999			2000			2001		
	Verg.-Gr.	Anzahl	Betrag in DM	Verg.-Gr.	Anzahl	Betrag in DM	Verg.-Gr.	Anzahl	Betrag in DM
	BAT IIa	1	100.800,-	BAT IIa	1	100.800,-	BAT IIa	1	100.800,-
	Wiss.HK	1	33.600,-	Wiss.HK	1	33.600,-	BAT IIa/2	1	46.800,-
	zusammen	2	134.400,-	zusammen	2	134.400,-	zusammen	1,5	147.600,-
SV	Kostenkategorie oder Kennziffer		Betrag in DM	Kostenkategorie oder Kennziffer		Betrag in DM	Kostenkategorie oder Kennziffer		Betrag in DM
	Kleingeräte		9000,-	Kleingeräte			Kleingeräte		
	Verbrauchsmittel		2600,-	Verbrauchsmittel		2000,-	Verbrauchsmittel		2000,-
	Kopien		1000,-	Kopien		1000,-	Kopien		1000,-
	Reisemittel		5000,-	Reisemittel		5000,-	Reisemittel		5000,-
	zusammen		17.600,-	zusammen		8000,-	zusammen		8000,-
I	Mittel für Invest. insgesamt:			Mittel für Invest. insgesamt:			Mittel für Invest. insgesamt:		

3.71 Begründung des Personalbedarfs

	Name, akad. Grad, Dienststellung	engeres Fach des Mitarbeiters	Institut der Hochschule oder der außeruniv. Einrichtung	Mitarbeit im Teil- projekt in Stunden pro Woche	auf d. Stelle im SFB tätig seit	derzeitg. Einstu- fung in BAT...
<i>Grund- ausstattung</i>						
3.71.1 wissenschaftl. Mitarbeiter (einschl. Hilfskräfte)	1. U. Mönnich, Prof. Dr.	Theor. Com- puterlinguistik, Vergleichende Sprachwiss.	SfS	8		
	2. T. Cornell, PhD, Wiss. Ang.	Formale Linguistik	SfS	8		
	3. E. Weißhar, Dr., Akad. Rat	Ethnoling.	Seminar für Vergl. Sprachwiss.	2		
3.71.2 nichtwissen- schaftliche Mitarbeiter	4. P. Rhein, Stud. HK	Angewandte Computerlin- guistik	SfS	10		
<i>Ergänzungs- ausstattung</i>						
3.71.3 wissenschaftl. Mitarbeiter (einschl. Hilfskräfte)	5. St. Kepser, Dr.	Theor. Com- puterlinguistik	CIS, Univ. München	38,5		Ila
	6. St. Herrman, MA	Ethnoling.	Forschungs- stipendium Venezuela	21		Wiss. HK
3.71.4 nichtwissen- schaftliche Mitarbeiter						

1. Der Antragsteller wird das Teilprojekt leiten und die damit verbundenen Forschungsarbeiten betreuen. Neben seiner Lehrtätigkeit und der Anleitung von Abschlusarbeiten wird er sich hauptsächlich auf den Teil des Arbeits-

programms konzentrieren, in dem die Berechenbarkeit von Datenstrukturen untersucht wird.

2. Der Schwerpunkt der Interessen von Dr. Thomas Cornell liegt seit seiner Promotion bei Ed Stabler in Los Angeles über Baumbeschreibungssprachen auf dem Gebiet der Formalisierung generierender linguistischer Theorien. Th. Cornell kann daher nicht nur als einschlägig ausgewiesen gelten für die Mitarbeit am geplanten Einzelprojekt, sondern sein besonderes Interessengebiet, das in mehreren Publikationen zu diesem Thema dokumentiert ist, stellt eine wünschenswerte Ergänzung der von St. Kepser eingebrachten Erfahrung dar, da in der Verteilung der Hauptarbeitsgebiete dieser vorgesehenen wissenschaftlichen Mitarbeiter die Entwicklung von dem Konzept abstrakter Datenstrukturen zu der Auffassung von Theorien als Datentypen in angemessener Weise repräsentiert ist.
3. Bei der Bewertung der Daten des Warao wird Frau Herrman von Dr. Emmerich Weißhar unterstützt werden, dem Ethnolinguisten am hiesigen Seminar für Vergleichende Sprachwissenschaft, der als Berater für das Projekt hat gewonnen werden können und der seit vielen Jahren über intensive eigene Erfahrungen im mittel- und südamerikanischen Sprachraum verfügt.
4. In dem Arbeitsbereich des Antragstellers am Seminar für Sprachwissenschaft ist Peter Rhein als Hilfskraft beschäftigt. Herr Rhein ist aufgrund seiner früheren Tätigkeit am Rechenzentrum der Universität Tübingen bestens vorbereitet, um Frau Herrman bei der Installierung und technischen Betreuung der in Aussicht genommenen Systeme LinguaLinks und CEL-LAR zu unterstützen.
5. Die beantragte BAT IIa-Stelle ist mit einem Linguisten / einer Linguistin mit Schwerpunkt in der Theoretischen Computerlinguistik zu besetzen. Erforderlich sind neben vertieften Kenntnissen in der Formalisierung zeitgenössischer linguistischer Theorien eine gründliche Ausbildung im Bereich der Kategorientheorie und der Universellen Algebra, um die Anwendung der Methoden der coalgebraischen Systemtheorie auf die mathematische Beschreibung umsetzen zu können. Mit Dr. Stephan Kepser ist für die beschriebene Aufgabe ein Mitarbeiter gewonnen, der die Anforderungen in idealer Weise erfüllt. Seit seiner Magisterarbeit zu Entscheidbarkeit und Erfüllbarkeit von einem von Paul King entwickelten Formalismus der HPSG hat er sich kontinuierlich mit formalen Eigenschaften von lizensierenden linguistischen Theorien beschäftigt und ist in diesem Sommer mit einer Dissertation über allgemeine Methoden für die Kombination von Spezialverfahren zur Deduktionsunterstützung in München promoviert worden.
6. Um die Leistungen der vom *Summer Institute of Linguistics* entwickelten Systemumgebung testen zu können und um in der Lage zu sein, die Not-

wendigkeit des Übergangs zu einer objektorientierten Umgebung empirisch zu überprüfen, ist die Stelle einer geprüften wissenschaftlichen Hilfskraft durch eine Linguistin / einen Linguisten zu besetzen, die über Erfahrungen in der deskriptiven Analyse von Sprachen verfügt, die von den vertrauten Mustern der indogermanischen Sprachfamilie genealogisch unabhängig sind. Mit Frau Stefanie Herrman ist für diese Aufgabe eine in Tübingen ausgebildete Ethnolinguistin gewonnen worden, die sich zur Zeit mit einem Stipendium im Orinocodelta in Venezuela aufhält. Sie beschäftigt sich dort mit dem Warao, das nach bisherigen Erkenntnissen als isolierte Sprache klassifiziert werden muß. Durch die Feldarbeit von St. Herrmann würde das Projekt über neues Material in einer Sprache verfügen, deren Erfassung im oben erwähnten CELLAR-System bisher nicht versucht worden ist und die daher in besonderem Maße geeignet erscheint, den im theoretischen Hauptteil des Teilprojekts thematisierten Begriff eines elaborierten Datentyps im Hinblick auf seine deskriptive Fruchtbarkeit zu evaluieren. Es wird beantragt, die Stelle von Frau Herrman im dritten Jahr auf BAT IIa/2 anzuheben.

3.72 Aufgliederung und Begründung der Sächlichen Verwaltungsausgaben (nach Haushaltsjahren)

	1999	2000	2001
Für Sächliche Verwaltungsausgaben – stehen als <u>Grundausstattung</u> voraussichtlich zur Verfügung	47.000,-	2000,-	2000,-
– werden als <u>Ergänzungsausstattung</u> beantragt (entspricht den Summen „Sächliche Verwaltungsausgaben“ in Übersicht 3.7)	17.600,-	8000,-	8000,-

Grundausstattung

Der Antragsteller wird jährlich 2000,- DM aus seinen Sachmitteln für das Teilprojekt verwenden und darüberhinaus einen PC mit zwei Festplatten für die Erstellung der amerindischen Sprachdokumentation durch die geprüfte wissenschaftliche Hilfskraft finanzieren (ca. DM 5000.-). Außerdem ist aus seinem Bücheretat für ca. DM 40000.- Literatur zur formalen Linguistik und Sprachtypologie erworben worden.

Ergänzungsausstattung

Kleingeräte (515) Vorstudien haben ergeben, daß sowohl die Schlußsysteme für endliche Bäume als auch das Werkzeug, das zur Zeit in Amsterdam und Dresden für die im Hauptteil des Antrags erwähnte objektorientierte Spezifikationsprache entwickelt wird, extrem rechenintensiv sind. Der zur Beschaffung vorgeschlagene LINUX PC stellt angesichts dieser Situation eine Kompromißlösung dar, die zur Voraussetzung hat, daß ein Teil der rechenintensiveren Aufgaben auf einem der Zentralrechner des Tübinger Rechenzentrums durchgeführt werden kann. Sollte ein im Moment der Antragstellung noch nicht entschiedener Pool aus Wissenschaftler-Arbeitsplatz-Rechnern eingerichtet werden — in dem für den Antragsteller ein PC im Wert von ca. 4.500,- DM vorgesehen ist —, würde sich der erbetene Beitrag der DFG zur Anschaffung von Kleingeräten von 9.000,- DM auf die Hälfte reduzieren.

Verbrauchsmittel (522) Dem Arbeitsbereich des Antragstellers steht ein leistungsfähiges Aufnahmegerät zur Verfügung. Der überwiegende Teil des erbetenen Betrags wird für die Anschaffung von Kassetten benötigt. Im ersten Jahr muß für die objektorientierten Systeme LinguaLinks und CELLAR vom Summer Institute of Linguistics eine Lizenz und eine Zweitlizenz erworben werden.

Reisemittel (527) Neben dem Besuch von Tagungen zum Themenschwerpunkt des Einzelprojekts (ESSLLI, EACL, ACL, MOL, LACL) soll aus den Reismitteln jährlich auch ein mehrwöchiger Aufenthalt in Venezuela finanziert werden, um die Materialien zum Warao zu überprüfen und zu ergänzen.

Kopien (531 b) Die in dieser Rubrik beantragten Mittel sind ausschließlich dazu bestimmt, Arbeitspapiere des Projekts drucken zu lassen. Die Kosten für Xerox-Kopien werden aus Mitteln der Grundausrüstung des Antragstellers bestritten.

3.73 Investitionen

(Geräte über DM 20.000 brutto und Fahrzeuge)
Es werden keine Investitionsmittel beantragt.

Zitierte Literatur

- ABADI, M. & L. CARDELLI (1996): *A Theory of Objects*. New York: Springer.
- ACZEL, P. (1988): ‘Non-Well-Founded Sets.’ *CSLI Lecture Notes* 14.
- ACZEL, P. (1990): ‘Replacement Systems and the Axiomatization of Situation Theory.’ In: R. Cooper & al. (eds.), *Situation Theory and Its Applications*, vol. I, pp. 3–31, Stanford: CSLI.
- ACZEL, P. (1997): ‘The Initial Algebra and the Final Coalgebra Perspectives.’ In: H. Schwichtenberg (ed.), *Logic of Computation*, pp. 1–33, Berlin: Springer.
- BARR, M. & C. WELLS (1995): *Category Theory for Computing Science*. London: Prentice Hall.
- BARWISE, J. & L. MOSS (1996): *Vicious Circles*. Stanford: CSLI.
- BARWISE, J. & J. SELIGMAN (1997): *Information Flow*. Cambridge: Cambridge University Press.
- BIRD, R. & O. DE MOOR (1997): *Algebra of Programming*. London: Prentice Hall.
- BLACKBURN, P. & E. SPAAN (1993): ‘A Modal Perspective on the Computational Complexity of Attribute Value Grammar.’ *Journal of Logic, Language and Information* 2: 129–169.
- BÖRGER, E., E. GRÄDEL & Y. GUREVICH (1997): *The Classical Decision Problem*. Berlin: Springer.
- CHOMSKY, N. (1955/1975): *The Logical Structure of Linguistic Theory*. New York: Plenum Press.
- CHOMSKY, N. (1981): *Lectures on Government and Binding*. Dordrecht: Foris.
- CHOMSKY, N. (1995): *The Minimalist Program*. Cambridge, MA: MIT Press.
- COHN, P. (1965): *Universal Algebra*. New York: Harper & Row.
- COURCELLE, B. (1997): ‘The Expression of Graph Properties and Graph Transformations in Monadic Second-Order Logic.’ In: G. Rozenberg (ed.), *Handbook of Graph Grammars and Computing by Graph Transformation*, pp. 313–400, Singapore: World Scientific.
- DEVLIN, K. (1997): *Goodbye, Descartes*. New York: John Wiley.
- DONER, J.C. (1970): ‘Tree Acceptors and Some of Their Applications.’ *J. Comp. Syst. Sci.* 4: 406–451.
- FAGIN, R. (1974): ‘Generalized First-Order Spectra and Polynomial-Time Recognizable Sets.’ In: R. Karp (ed.), *Complexity of Computation*, SIAM-AMS Proceedings 7, pp. 43–73, New York.
- FEFERMAN, S. (1992): ‘A New Approach to Abstract Data Types II. Computation on ADTs as Ordinary Computation.’ In: E. Börger & al. (eds.), *Computer Science Logic*, vol. 5 of *Lecture Notes in Computer Science* 626, pp. 79–95, Springer.

- GIBBONS, J. (1994): 'An Initial-Algebra Approach to Directed Acyclic Graphs.' In: B. Möller (ed.), *Mathematics of Program Construction*, Lecture Notes in Computer Science 947, pp. 282–303, Springer.
- GOGUEN, J. & G. MALCOM (1997): 'A Hidden Agenda.' Technical Report CS97-538, Oxford.
- GOGUEN, J.A. (1991): 'Types As Theories.' In: G.M. Reed & al. (eds.), *Topology and Category in Computer Science*, pp. 357–390, Oxford: Clarendon Press.
- GOLDFARB, C. (1990): *The SGML Handbook*. Oxford: Oxford University Press.
- VON GUDENBERG, J.W. (1993): *Objektorientiert Programmieren von Anfang an*. Mannheim: Bibliographisches Institut.
- HAREL, D. (1993): *Algorithmics*. Reading, MA: Addison-Wesley.
- HENSEL, U., M. HUISMAN, B. JACOBS & H. TEWS (1997): 'Reasoning about Classes in Object-Oriented Languages: Logical Models and Tools.' Tech. Rep. R9718, CSI, Nijmegen, Report.
- JACOBS, B. (1996): 'Objects and Classes, Co-Algebraically.' In: B. Freitag & al. (eds.), *Object-Oriented Programming with Parallelism and Persistence*, pp. 83–103, Dordrecht: Kluwer.
- JACOBS, B. (1997): 'Invariants, Bisimulations and the Correctness of Coalgebraic Refinements.' In: M. Johnson (ed.), *Algebraic Methodology and Software Technology*, Lecture Notes in Computer Science 1349, pp. 276–291, Springer.
- JACOBS, B. & J. RUTTEN (1997): 'A Tutorial on (Co-)Algebras and (Co-)Induction.' *EATCS Bulletin* **62**: 222–259.
- JOHNSON, D. & S. LAPPIN (1997): 'A Critique of the Minimalist Program.' *Linguistics and Philosophy* **20**: 273–333.
- KING, P. & C. POLLARD (1997): 'A Formalism for HPSG '94.' Tech. rep., Universität Tübingen.
- KING, P., K. SIMOV & B. ALDAG (1997): 'The Complexity of Modelability in Finite and Computable Signatures of a Constraint Logic for Head-Driven Phrase Structure Grammar.' Universität Tübingen, Seminar für Sprachwissenschaft.
- LANGENDOEN, D.T. & G. SIMONS (1995): 'A Rationale for the TEI Recommendations for Feature-Structure Markup.' *Computers and the Humanities* **29**: 191–209.
- LAUSEN, G. & G. VOSSEN (1997): *Models and Languages of Object-Oriented Databases*. Harlow: Addison-Wesley.
- MESEGUER, J. & J. GOGUEN (1985): 'Initiality, Induction, and Computability.' In: M. Nivat & J. Reynolds (eds.), *Algebraic Methods in Semantics*, pp. 459–541, Cambridge: Cambridge University Press.
- MOSHIER, A. (1995): 'A Rational Reconstruction of the Domain of Feature Structures.' *J. of Logic, Language and Information* **4**: 111–143.

- OSBORN, H.A. (1966a): 'Warao I: Phonology and Morphophonemics.' *Int. Journal of American Linguistics* **32**: 108–123.
- OSBORN, H.A. (1966b): 'Warao II: Nouns, Relationals, and Demonstratives.' *Int. Journal of American Linguistics* **32**: 253–261.
- OSBORN, H.A. (1967): 'Warao III: Verbs and Suffixes.' *Int. Journal of American Linguistics* **33**: 46–64.
- PEREIRA, F. & S. SHIEBER (1984): 'The Semantics of Grammar Formalisms Seen as Computer Languages.' In: *Proceedings COLING*, vol. 10, pp. 123–129.
- POLLARD, C. & I. SAG (1994): *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- POSTAL, P. (1964): *Constituent Structure*. Bloomington: Mouton.
- PUTNAM, H. (1961): 'Some Issues in the Theory of Grammar.' In: *Proc. of Symposia in Applied Mathematics*, vol. 12, pp. 25–42.
- REICHEL, H. (1995): 'An Approach to Object Semantics Based on Terminal Co-Algebras.' *Mathematical Structures in Computer Science* **5**: 129–152.
- RETTIG, M., G. SIMONS & J. THOMSON (1993): 'Extended Objects.' *Communications of the ACM* **36**: 19–24.
- ROGERS, J. (1994): *Studies in the Logic of Trees with Applications to Grammar Formalism*. Ph.D. thesis, University of Delaware, PhD Thesis.
- ROOSENBLUM, P. (1950): *The Elements of Mathematical Logic*. New York: Dover.
- RUTTEN, J. (1996): 'Universal Coalgebra: A Theory of Systems.' Report CS-R9652, CWI, Amsterdam, Report.
- SEESE, D. (1991): 'The Structure of the Models of Decidable Monadic Theories of Graphs.' *Annals of Pure and Applied Logic* **53**: 169–195.
- SEESE, D. (1992): 'Interpretability and Tree Automata: A Simple Way to Solve Algorithmic Problems on Graphs Closely Related to Trees.' In: M. Nivat & A. Podelski (eds.), *Tree Automata and Languages*, pp. 83–114, Amsterdam: North Holland.
- SGML Standard (1986): *Standard Generalized Markup Language (SGML)*. International Organization for Standardization (ISO), Genf.
- SIMONS, G. (1994): 'Conceptual Modeling vs. Visual Modeling: A Technological Key to Building Consensus.' Tech. rep., Summer Institute of Linguistics, Working Paper.
- SIMONS, G. (1996): 'Implementing the TEI's Feature-Structure Markup by Direct Mapping to the Objects and Attributes of an Object-Oriented Database System.' *Research in Humanities Computing* **5**: 220–242.
- SIMONS, G. (1998): 'The Nature of Linguistic Data and the Requirements of a Computing Environment for Linguistic Research.' In: J. Lawler & H. Dry (eds.), *Computers and the Ordinary Working Linguist*, Berlin: de Gruyter.
- SIMONS, G. & J. THOMSON (1995): 'Multilingual Data Processing in the CEL-

- LAR Environment.’ Tech. rep., Summer Institute of Linguistics, Working Paper.
- SMITH, B.C. (1996): *On the Origin of Objects*. Cambridge, MA: MIT Press.
- SPERBERG-MCQUEEN, C.M. & L. BURNARD (eds.) (1994): *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford: Text Encoding Initiative.
- THATCHER, J.B. & J. WRIGHT (1968): ‘Generalized Finite Automata with an Application to a Decision Problem of Second Order Logic.’ *Math. Systems Theory* **2**: 57–82.
- THOMAS, W. (1997): ‘Languages, Automata, and Logic.’ In: G. Rozenberg & A. Salomaa (eds.), *Handbook of Formal Languages*, vol. 3 : Beyond Words, pp. 389–455, Springer.
- ZOOK, K. (1996): *LinguaLinks and CELLAR Programming*. Dallas: Academic Computing Department at the Summer Institute of Linguistics (SIL).