

Eberhard Karls Universität Tübingen

Sonderforschungsbereich 441: Linguistische Datenstrukturen

Teilprojekt A2: Linguistische Theorien als Datentypen

Ergebnisbericht für die erste Phase 1999 – 2001

Prof. Dr. Uwe Mönnich

Universität Tübingen
Seminar für Sprachwissenschaft
Wilhelmstr. 113
72074 Tübingen
Tel: 07071/2974035
Email: um@sfs.uni-tuebingen.de

Dr. Stephan Kepser

Universität Tübingen
SFB 441
Nauklerstr. 35
72074 Tübingen
Tel: 07071/2972732
Email: kepser@sfs.uni-tuebingen.de

Stefanie Herrmann

Universität Tübingen
SFB 441
Nauklerstr. 35
72074 Tübingen
Tel: 07071/2977162
Email: herrmann@sfs.uni-tuebingen.de

Ergebnisbericht

Teilprojekt: A2

Thema: Linguistische Theorien als Datentypen

Leiter: Prof. Dr. Uwe Mönnich

Mitarbeiter: Dr. Stephan Kepser (wiss. Mitarbeiter)
Stefanie Herrmann (wiss. Mitarbeiterin)

Inhalt

1. Kenntnisstand bei der Antragstellung und Ausgangsfragestellung
2. Ergebnisse und ihre Bedeutung
3. Vergleiche mit Arbeiten außerhalb des Sonderforschungsbereichs und Reaktionen der wissenschaftlichen Öffentlichkeit
4. Offene Fragen
5. Veröffentlichungen und Manuskripte
6. Aktivitäten: Vorträge, Gastwissenschaftler
7. Verzeichnis der im Ergebnisbericht zitierten Literatur

1 Kenntnisstand bei der Antragstellung und Ausgangsfragestellung

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

Das vorliegende 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, werden in diesem Projekt behandelt.

Die zentrale Arbeitshypothese, die die Aktivitäten des Projekts in der ersten Phase des Sonderforschungsbereichs 441 geleitet hat, war inspiriert von dem systematischen Zusammenhang, der zwischen dem Paradigmenwechsel linguistischer Theorien und der Entwicklung der intendierten Datenstrukturen zu beobachten ist. Insbesondere sollte das Ziel verfolgt werden, das Postulat eines unter Komplexitätstheoretischen Gesichtspunkten praktikablen Grammatikalitätsbegriffs unter der Perspektive der relevanten Datentypen näher zu untersuchen.

Der oben genannte Paradigmenwechsel bezeichnet den Übergang von sogenannten *generierenden* Theorien zu *lizensierenden* Theorien in der Sprachwissenschaft. Eine generierende Theorie zeichnet sich, wie der Name schon sagt, dadurch aus, daß eine sprachliche Äußerung und ihre Analysen von der Theorie auch tatsächlich generiert werden. Eine solche Theorie wird typischerweise als Regelsystem gegeben, das aus einem Nukleus heraus durch eine endliche Anwendung der Regeln schließlich eine vollständige Analyse liefert. Der Prototyp solch einer Theorie ist eine kontextfreie Grammatik. In der Linguistik sind vor allem die Transformationsgrammatik (Chomsky 1965) und die Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum und Sag 1985) generierende Theorien. Die diesen Theorien zugrunde liegende Datenstruktur sind Bäume.

Interessanterweise fand sowohl in der chomskyschen Tradition als auch in der sich an GPSG orientierenden Linguistik im Laufe der 1980er Jahre ein Paradigmenwechsel weg von den generierenden Theorien und hin zu den lizensierenden Theorien statt. Dieser folgte dem Bedürfnis, die hinter den Regeln entdeckten allgemeinen Prinzipien abstrakt und allgemein ausdrücken zu können und nicht in jede einzelne Regel erneut kodieren zu müssen. Dadurch entstand in der chomskyschen Tradition die Prinzipien- und Parametertheorie (GB, (Chomsky 1981)) und auf der anderen Seite die Head-Driven Phrase Structure Grammar (HPSG, (Pollard und Sag 1987, Pollard und Sag 1994)). Als lizensierende Theorien zeichnen sich beide dadurch aus, daß die

Grammatik nur noch Wohlgeformtheitsbedingungen angibt, die als abstrakte Prinzipien formuliert sind. Eine grammatische Analyse gilt in diesem Paradigma als korrekt, wenn sie allen durch die Prinzipien gestellten Anforderungen genügt. Die Theorien selbst geben keinen Hinweis darauf, wie man an eine Analyse gelangt. Anders als bei den generierenden Theorien wird dies auch als unwichtig angesehen; bedeutend ist lediglich, daß die Wohlgeformtheitsbedingungen nicht verletzt werden. Entsprechend bilden Graphen und nicht mehr Bäume die zugrunde liegende Datenstruktur. Obwohl in der GB-Tradition der durch die freie Indizierung vollzogene Übergang von Bäumen zu Graphen nicht bewußt intendiert war, ist er doch ein Stück weit zwingend, wenn man das X-Bar-Schema als Prinzip und nicht als Regel auffaßt. Als Wohlgeformtheitsbedingung gelesen kann es nämlich gar nicht mehr zwischen echten Bäumen und nur baumähnlichen Graphen unterscheiden. In der HPSG wird der Übergang hin zu Graphen als Datenstrukturen bewußt vollzogen.

Eines der Ziele des Projekts ist es, diesen Paradigmenwechsel angemessen modelltheoretisch zu repräsentieren. Während für generierende Theorien mit freien oder initialen Algebren schon lange eine angemessene Modellierung zur Verfügung steht, ist mit der Theorie der Coalgebren (Aczel 1997, Jacobs und Rutten 1997, Rutten 2000) erst in den letzten Jahren ein geeigneter Kandidat für die Modellierung lizensierender Theorien sichtbar geworden. Coalgebren wurden zur Modellierung des sogenannten beobachtbaren Verhaltens versteckter Zustandsräume entwickelt und konnten somit bei der Modellierung objekt-orientierter Systeme erfolgreich eingesetzt werden (Reichel 1995, Abadi und Cardelli 1996). Das Projekt hatte sich daher zum Ziel gesetzt, eine coalgebraische Modellierung lizensierender Theorien zu leisten, um damit den Paradigmenwechsel auch in der Modelltheorie abbilden zu können.

Die Aufgabe einer „datengetriebenen“ Analyse des Grammatikalitätsbegriffs hatte in jüngster Zeit einen hohen Grad an Dringlichkeit gewonnen, da es sich gezeigt hatte, daß der Übergang von induktiv beschreibbaren Strukturen des generierenden Paradigmas zu coinduktiv beschreibbaren Strukturen des lizensierenden Paradigmas in vielen Fällen mit dem Verlust der Entscheidbarkeit des Grammatikalitätsbegriffs verbunden war. Es sei daran erinnert, daß nach den Ergebnissen von Rogers (1998) die Theorie der Prinzipien und Parameter ohne Indizierung zwar vollständig mit den Mitteln der monadischen Logik zweiter Ordnung über endlichen Bäumen formalisiert zu werden vermag und damit auch entscheidbar ist, die Berücksichtigung der Indizierung jedoch die Baumstruktur zerstört und mit den Mitteln der erweiterten logischen Sprache, die es gestattet, die für die Indizierung kennzeichnenden Relationen zwischen Baumknoten auszudrücken, eine Reduktion auf das Kachelungsproblem möglich macht. Die Reduktion auf dieses unentscheidbare Problem läuft somit auf den Nachweis hinaus, daß im Falle der (freien) Indizierung die Voraussetzungen erfüllt sind, sich von der Baumstruktur maximal zu entfernen, indem die für die Kachelung relevanten Nachbarschafts- und Randeigenschaften eine Baumzerlegung mit begrenzt vielen Komponenten nicht mehr gestattet.

Auf ein analoges Problem für die lizensierende „Schwester“ der auf Prinzipien und Pa-

rametern fußenden Theorie, die Head-Driven Phrase Structure Grammar war ebenfalls in der Begründung für das gegenwärtige Projekt hingewiesen worden. Auch für dieses linguistische Modell gelingt für die Modellierungseigenschaft eine Reduktion auf das Kachelungsproblem und damit der Nachweis ihrer Unentscheidbarkeit (King, Simov und Aldag 1999), während andererseits die schwächere Anforderung der Erfüllbarkeit und damit Konsistenz entschieden werden kann (Kepser 1994).

2 Ergebnisse und ihre Bedeutung

2.1 Reguläre Beschreibung von HPSG

Selbstverständlich lassen sich die oben aufgeführten Unentscheidbarkeitsresultate nicht durch den Übergang in eine andere Datenstruktur entkräften. Die Theorie einer Familie von Graphen, die das Modell einer formalisierten HPSG oder einer GB Grammatik bilden, wird nicht dadurch entscheidbar, daß man zunächst die Graphen in algebraische Strukturen baumartigen Charakters entfaltet und in einem zweiten Schritt die Erfüllbarkeit bzw. Modellierbarkeit von Aussagen auf den durch den Entfaltungsprozess gewonnenen Strukturen überprüft. Das Konzept der Baumähnlichkeit bietet sich jedoch als mächtiges Instrument an, um klarere Vorstellungen darüber zu gewinnen, welches Ausdrucksmittel in einer formalisierten linguistischen Theorie dafür verantwortlich ist, daß die aus berechnungstheoretischen und kognitiven Gründen wünschenswerte Eigenschaft der Entscheidbarkeit verloren geht.

Die fundamentale Idee hinter diesem Instrument ist die enge Beziehung, die zwischen baumähnlichen Strukturen und „wirklichen“ Bäumen besteht. Diese enge Beziehung sollte die Basis dafür liefern, daß viele Ergebnisse, die aus der formalen Sprachtheorie für den speziellen Fall der Bäume bekannt sind, sich auf den Fall von baumähnlichen Graphen übertragen lassen. In dieser Weise scheint das Konzept der Baumähnlichkeit eine systematische Verbindung mit Berechenbarkeitsfragen einzugehen.

Im Vorgriff auf das Arbeitsprogramm der kommenden Förderphase des SFB 441, das in dem Fortsetzungsantrag für das gegenwärtige Projekt näher erläutert wird, sei an dieser Stelle auf eine verwandte Problemstellung im Bereich der sogenannten semistrukturierten Daten hingewiesen. Das Programm, das sich unter dem Titel *Linguistische Theorien als Datenstrukturen* ausdrückt, war von Beginn an so konzipiert worden, daß neben der Parallelität zwischen linguistischen Paradigmenwechsel und der Entwicklung von abstrakten Datentypen zu objektorientierten Techniken das Postulat einer integrierten multimodalen und multidimensionalen Datenmodellierung sprachlicher Einheiten unter theoretischer und praktischer Perspektive umgesetzt werden sollte. Durch die unvorhersehbare Verzögerung bei der Implementierung von Dokumentationssystemen, deren kurzfristige Verfügbarkeit zur Zeit der Antragstellung in Aussicht gestellt war (LinguaLinks, EUDICO, Atlas), haben sich für die Umsetzung des praktischen Teils des Arbeitsprogramms unüberwindliche Schwierigkeiten

ergeben. Wir mußten uns daher bei der Beschreibung der ausgewählten grammatischen Phänomene des Warao, die zur Illustration unseres Konzepts einer unverkürzten Datenrepräsentation dienen sollten, auf ein Vorgehen im Sinne traditioneller ethnolinguistischer Dokumentationen beschränken. Die erzielten Resultate bilden den Gegenstand des Arbeitsberichts von Stefanie Herrmann (2001). Bei genauerer Betrachtung der Struktur der universellen Auszeichnungssprache XML haben sich allerdings Fragen ergeben, die es gestatten, das langfristige Forschungsprogramm, das in dem ursprünglichen Projektantrag skizziert wurde, auch unter dem Aspekt der theoretischen Grundlagen einer angemessenen Datenrepräsentation systematisch aufzugreifen und fortzuführen. Der Zusammenhang mit dem Konzept der Baumähnlichkeit ist dadurch gegeben, daß in Erweiterung des Kerns von XML Möglichkeiten eingebaut sind, allgemeine Graphenstrukturen zu repräsentieren (vgl. Abiteboul, Buneman und Suciu (2000)). Es stellt sich daher auch für die Struktur des WWW wieder das Problem, ob die Erfordernisse einer adäquaten Repräsentation so geartet sind, daß die durch die Basissysteme von XML spezifizierten Bäume in dem Maße „zerstört“ werden müssen, daß sie im technischen Sinne jede Baumähnlichkeit verlieren. Auf die sich im Rahmen dieser Zielsetzung ergebenden Kooperationsaspekte mit den aktuellen Aufgabenstellungen anderer Projekte wird ebenfalls im Fortsetzungsantrag zurückzukommen sein.

Der skizzierten Beziehung zwischen den von linguistischen Theorien intendierten Datenstrukturen und deren berechnungstheoretischen Eigenschaften, die sich aus der Baumähnlichkeit der intendierten Strukturen ergeben, wird mit Bezug auf die HPSG in dem Arbeitsbericht von Uwe Mönnich (2001) näher eingegangen. Sein auf den ersten Blick überraschendes Ergebnis lautet, daß die durch eine (endliche) Grammatik vom Typ der HPSG beschriebenen Graphen regulär sind. Genauer erläutert besagt dieses Ergebnis, daß sich die HPSG-Graphen mithilfe einer endlichen Anzahl von Operatoren zusammensetzen lassen, die ihrerseits die Knotenbeschriftungen einer regulären Baumfamilie bilden. Es ist bekannt, daß Graphen von der Art, wie sie im Rahmen der HPSG spezifiziert werden, sich durch Bäume repräsentieren lassen, deren Knoten nur Etikettierungen aus einer ausgewählten Familie von einfachen Graphenoperationen tragen. Es handelt sich bei diesen Operationen um Graphen mit einem einzigen Knoten (nullstellige Operationen), um die disjunkte Vereinigung, die einstellige Operation, die es erlaubt, neue Kanten zwischen Knoten zu kreieren, und schließlich um die einstellige Operation, die eine Veränderung von Knotenetikettierungen bewirkt. Aus der Tatsache, daß diese einfachen Operationen ausreichen, um für alle HPSG-Graphen einen Termausdruck erzeugen zu können, folgt noch nicht allein, daß die Menge der durch eine Grammatik dieses Typs beschriebenen Graphen durch eine reguläre Baumfamilie repräsentierbar ist. Um dieses Resultat zu erzielen, muß gemäß der Voraussetzungen des Myhill-Nerode Theorems gezeigt werden, daß die durch eine Grammatik vom Typ der HPSG beschriebene Graphenfamilie als Vereinigung von Kongruenzklassen einer Kongruenzrelation von endlichem Index darstellen läßt. Unter der Voraussetzung, daß mit den Mitteln der Kingschen Speciate Re-entrant Logic (SRL, (King 1989, King 1999)) eine adäquate Formulierung der konstitutiven Annah-

men des HPSG-Modells gelungen ist, läßt sich der erforderliche Nachweis für das Vorliegen einer Kongruenzrelation mit endlichem Index leicht führen. Da die Ausdrücke dieser Logik, die eine bestimmte Quantorentiefe nicht überschreiten, bis auf logische Äquivalenz nur eine endliche Familie bilden, muß nur noch gezeigt werden, daß die oben angegebenen einfachen Operationen sich invariant verhalten gegenüber diesen Äquivalenzklassen, die durch die SRL-Ausdrücke induziert sind. Eine Anwendung des klassischen modelltheoretische Resultats von Feferman und Vaught auf den Fall der HPSG-Graphen und ihrer logischen Beschreibungssprache schließt die letzte Lücke im Beweis.

Welche Konsequenzen folgen nun aus dieser Sachlage? Es wurde weiter oben schon angedeutet, daß die Motivation hinter dem Konzept der Baumähnlichkeit in der Übertragung berechnungstheoretischer Ergebnisse, die mit den Mitteln der formalen Sprachtheorie für den speziellen Fall algebraischer Baumstrukturen erzielt sind, auf die wesentlich komplexere Situation im Falle von Graphen besteht. Stellvertretend für eine Fülle weiterer algorithmischer Eigenschaften, die mit der regulären Beschreibung der HPSG-Strukturen verbunden sind, möge die Folgerung stehen, daß Formeln der monadischen Logik zweiter Ordnung, die über Graphen interpretiert werden, auf den Baumrepräsentationen, die aufgrund der angegebenen Operationen gebildet sind, in linearer Zeit verifizierbar sind. Weitere Ergebnisse sind in der Monographie von Downey und Fellows (1999) zusammengestellt, die während der abgelaufenen Förderphase erschienen ist.

2.2 Coalgebraische Modellierung von HPSG

Linguistische Theorien lassen sich in zwei verschiedene Arten einteilen. Auf der einen Seite stehen die sogenannten *generierenden* Theorien, in denen die Grammatik als ein System von Regeln eine Äußerung und deren Analyse tatsächlich generiert. Der Prototyp einer solchen Theorie ist eine kontextfreie Grammatik, Beispiele aus der Sprachwissenschaft sind etwa die frühen Arbeiten zur Transformationsgrammatik (Chomsky 1965) oder Generalized Phrase Structure Grammar (Gazdar et al. 1985). Auf der anderen Seite stehen die sogenannten *lizensierenden* Theorien. Theorien dieser Art generieren eine Analyse nicht. Statt dessen geben sie Wohlgeformtheitsbedingungen für Analysen an, die oftmals als abstrakte, generelle Prinzipien formuliert sind. Aus Sicht dieser Theorien ist es nicht wichtig, wie man zu der Analyse gelangt ist. Was zählt, ist, daß alle Wohlgeformtheitsbedingungen eingehalten sind.

Head-Driven Phrase Structure Grammar ist eine solche lizensierende Theorie. Nicht nur, daß die Grammatik aus einer Menge von universalen und sprachspezifischen Prinzipien besteht, die Autoren Pollard und Sag äußern auch explizit (auf S. 31), daß HPSG die Eigenschaften einer lizensierenden Theorie hat und haben soll. Wir halten es daher für entsprechend wichtig, daß die Art der Modellierung auch die grundsätzlichen Ideen einer lizensierenden Theorie in ihren Konzepten widerspiegelt. Insbesondere sollten

sich diese Konzepte auch von jenen unterscheiden, die bei der Formalisierung generierender Theorien Verwendung finden, um den Unterschied zwischen diesen Arten von linguistischen Theorien auch in der Art der Modellierung ausdrücken zu können. Dabei werden initiale Algebren schon längere Zeit als Standardmodelle für generierende Theorien verwendet, sodaß diese als Modelle für HPSG nicht so geeignet erscheinen. Statt dessen wird eine Art von Modellen benötigt, die zu der Eigenschaft lizensierender Theorien, über Wohlgeformtheitsbedingungen zu reden, paßt. Es geht also nicht um die exakte Art der internen Repräsentation einer Analyse, sondern um beobachtbare Merkmale der Repräsentation. Am Beispiel HPSG bedeutet dies, daß die exakte Form, wie Merkmalsstrukturen intern repräsentiert werden, relativ unwichtig ist. Da die Prinzipien von HPSG nur über beobachtbare Eigenschaften von Featurestrukturen wie Sorten und Merkmale und deren Verhältnisse reden, ist der passende Abstraktionsgrad in der dazu passenden Modelltheorie einer der ebenfalls nur über Eigenschaften spricht, und über die interne Repräsentation dieser Eigenschaften keine Annahmen macht.

Eine solche Art der Modellierung steht mit der Theorie der Coalgebren (Aczel 1997, Jacobs und Rutten 1997, Rutten 2000) zur Verfügung. Coalgebren stammen aus der theoretischen Informatik zur Beschreibung von Zustands-Übergangs-Systemen, bei denen man nur das äußere Verhalten der Systeme beobachten, nicht aber in die black-boxartigen Systeme hineinsehen kann. Da sie darüberhinaus mittlerweile häufig zur Modellierung von Graphen eingesetzt werden, bietet es sich an, wie im Antrag skizziert, zu versuchen, HPSG mithilfe von Coalgebren zu modellieren. Diese Modellierung ist in der Arbeit von Kepser (2000) gelungen. Dazu mußte eine Semantik von HPSG entwickelt werden, die der Modellierungsmethode durch Coalgebren entspricht. Insbesondere sogenannte Pfadgleichungen müssen in der richtigen Weise, nämlich nicht als Identitäten, sondern als beobachtbar gleiche Erscheinungsformen interpretiert werden. Damit kann man eine Klasse von HPSG-Coalgebren auszeichnen, die die wünschenswerte Eigenschaft besitzt, eine sogenannte finale Coalgebra zu beinhalten. Ein Kernpunkt ist der Beweis der Existenz einer finalen Coalgebra in der Klasse von HPSG-Coalgebren. Diese finale Coalgebra bietet sich aufgrund ihrer Eigenschaften als das Modell einer HPSG-Grammatik an. Sie sind konzeptionell adäquat, weil Coalgebren eine geeignete Form der Modellierung für lizensierende Theorien wie HPSG darstellen, und weil die Featurestrukturen von HPSG als spezielle Graphen durch Coalgebren besonders gut modelliert werden können. Finale Coalgebren sind ebenso im technischen Sinne adäquate Modelle, da sie als Modelle für Analysen korrekt, vollständig, kompakt und frei von virtuellen Ambiguitäten sind.

Damit ist im Rahmen der Zielsetzung des Projekts ein wesentlicher Beitrag geleistet, in dem zum ersten Mal gezeigt wurde, daß der Paradigmenwechsel von der generierenden Theorie GPSG hin zur lizensierenden Theorie HPSG auch modelltheoretisch adäquat erfaßt werden kann. Gleichzeitig wird damit auch eine wichtige Grundlage für die im Projekt A1 geleisteten Arbeiten zur HPSG (Richter, Sailer und Penn 1999, Richter 2000) geliefert.

2.3 Berechenbarkeit und Komplexität von RSRL

Stand im vorangehenden Abschnitt die modelltheoretische Seite von HPSG im Vordergrund, so ist es hier die logische Seite. Die RELATIONAL SPECIATE RE-ENTRANT LOGIC (im Nachfolgenden RSRL) ist eine von Richter, Sailer und Penn (Richter et al. 1999, Richter 2000) zur Formalisierung von HPSG entworfene Beschreibungslogik. Sie basiert auf der SPECIATE RE-ENTRANT LOGIC von King (1989, 1999). Beide sind als Featurelogiken Beschreibungslogiken, sprechen mithin also nicht über Wahrheit, sondern beschreiben linguistische Objekte. Daher ist die Denotation einer RSRL-Formel die Menge der Objekte aus dem Denotationsbereich, auf die die Formel zutrifft. Linguistische Objekte werden hauptsächlich durch Sorten und Merkmale beschrieben. Sorten beschreiben den Typ eines Objekts, z.B. *word*, *phrase* oder *subcat*. Merkmale beschreiben Teile von Objekten wie etwa die PHON (phonetisch/phonologischen) oder SYNSEM (syntaktischen und semantischen) Teile eines Objekts vom Typ *word*. Die beiden wesentlichen Erweiterungen von RSRL gegenüber SRL sind die Einführung von beliebigen Relationen und von Quantifikation. Beide sind für RSRL typisch und entsprechen in ihrer technischen Ausführung nicht ihren Pendanten in der klassischen Logik. Die gewählten Erweiterungen sind nach Ansicht von Richter für eine korrekte Formalisierung der Prinzipien von HPSG erforderlich, die Ausdruckstärke von SRL sei dafür zu schwach. Da die Erweiterungen recht spezifisch sind, stellen sich natürlicherweise die Fragen, wie RSRL sich zu klassischen Logiken verhält, und wo die Ausdruckstärke von RSRL einzuordnen ist. Die Antworten auf diese Fragen können helfen zu verstehen, ob für HPSG tatsächlich eine spezielle Logik erforderlich oder sinnvoll ist, oder ob man nicht auch mit einer klassischen Logik auskommt. Diese sind in der Regel besser bekannt und besser verstanden.

Ein erster Schritt zum Verständnis von RSRL, der in diesem Projekt geleistet wurde, ist die Betrachtung der Berechenbarkeit und Komplexität. Der Begriff der Komplexität einer Logik stammt aus der deskriptiven Komplexitätstheorie. Prinzipiell lassen sich Klassen von endlichen Strukturen auf zwei verschiedene Weisen definieren. Klassen von Strukturen können auf klassische Weise durch logische Axiomatisierungen definiert werden. Damit bestimmt die Ausdruckstärke der Logik, welche Klassen definierbar sind. Klassen endlicher Strukturen können aber auch durch Rechenmaschinen definiert werden, die berechnen können, ob eine als Kandidat vorgelegte Struktur zur Klasse gehört oder nicht. In diesem Fall ist es die Komplexität der Rechenmaschine, die bestimmt, welche Klassen von Strukturen definierbar sind. Die deskriptive Komplexitätstheorie bildet nun die Brücke zwischen diesen beiden Ansätzen, indem sie das Verhältnis von Ausdruckstärke einer Logik einerseits und Komplexität einer Rechenmaschine andererseits untersucht. Die Frage, die die deskriptive Komplexitätstheorie zu beantworten versucht, lautet also: Gegeben eine Klasse von Strukturen definiert in einer bestimmten Logik, was ist die minimale Komplexität einer Rechenmaschine, um diese Klasse zu entscheiden? Und gegeben eine Klasse von Strukturen, die von einer bestimmten Rechenmaschine entschieden wird, in welcher Logik kann diese Klasse

axiomatisiert werden.

Wir sind in unserer Arbeit (Kepser 2001) von durch RSRL axiomatisierbaren Strukturen ausgegangen und haben die zugehörige Komplexitätsklasse gesucht. Dabei haben wir zwei verschiedene Varianten von RSRL untersucht. In der originalen Definition von RSRL werden im Denotationsbereich sogenannte Ketten eingeführt. Eine Kette ist eine endliche Folge von Elementen aus dem Denotationsbereich. Eine Besonderheit von RSRL gegenüber klassischen Logiken besteht darin, daß die Denotation einer Variablen nicht nur ein Element im Denotationsbereich sein kann, sondern auch eine Kette, also ein Folge von Elementen. Auch Relationen können Relationen über Ketten, und nicht etwa nur über Elementen, sein. Daneben gibt es die kettenlose Variante von RSRL, in der ganz wie in klassischen Logiken die Denotation einer Variablen nur ein Element des Trägers ist, und Relationssymbole nur als Relationen über Elementen interpretiert werden. Die Unterscheidung zwischen der Variante mit Ketten und der ohne ist für die Komplexität wichtig.

Für RSRL ergibt sich das bemerkenswerte Resultat, daß die Gültigkeit einer Formel in einer endlichen RSRL-Struktur unentscheidbar ist. Kepser (1994) hat für SRL gezeigt, daß Konsistenz einer SRL-Formel entscheidbar ist, während King et al. (1999) bewiesen haben, daß der stärkere Begriff der Grammatikalität unentscheidbar ist. Erfüllbarkeit einer Formel der klassischen Prädikatenlogik erster Stufe ist bekanntlich unentscheidbar (Church 1936). Andererseits liegt nach einem Resultat von Immerman (1987) Prädikatenlogik mit transitiven Abschlüssen in LOGSPACE. D.h., gegeben eine Formel der Prädikatenlogik mit transitiven Abschlüssen und eine endliche Struktur, ist die Komplexität zur Berechnung, ob die Formel in der Struktur gilt oder nicht, LOGSPACE-beschränkt in der Größe der Struktur. Interessanterweise ist die entsprechende Frage für RSRL, also gegeben eine RSRL-Formel und eine endliche RSRL-Struktur, ist die Formel in der Struktur gültig, nicht einmal entscheidbar. Der Unentscheidbarkeitsbeweis beruht darauf, daß man in RSRL Post-Korrespondenzprobleme kodieren kann. Dabei spielen die Ketten eine entscheidende Rolle. Dies zeigt, wie folgenreich die Verwendung von Ketten ist. Schließlich gehört die Anforderung, daß die Gültigkeit einer Formel in einer *endlichen* Struktur entschieden werden könne, zu den grundlegendsten, die man machen kann.

Danach stellt sich die Frage, wie die Situation aussieht, wenn man in RSRL auf Ketten verzichtet. Das Ergebnis, das man für RSRL ohne Ketten erhält, liegt sehr viel näher an klassischen Resultaten: RSRL ohne Ketten liegt in der Komplexitätsklasse PTIME. Mit anderen Worten, jede in RSRL ohne Ketten axiomatisierbare Klasse von endlichen Strukturen ist durch eine PTIME-beschränkte Turingmaschine entscheidbar. Daß damit die Komplexität höher ist als für klassische Prädikatenlogik hängt mit der besonderen Art der Quantifikation in RSRL zusammen. Dieses Ergebnis wird erreicht, indem man RSRL-Strukturen als Eingaben für Turingmaschinen kodiert und dann ein Verfahren angibt, wie man für jede beliebige RSRL-Formel eine PTIME-beschränkte Turingmaschine erhält, die für eine Struktur als Eingabe die Denotation der Formel in der Struktur berechnet.

Insgesamt legen diese Ergebnisse wie auch schon das zur regulären Beschreibung von HPSG nahe, daß der im Antrag formulierte Zusammenhang zwischen dem Übergang von generierenden zur lizensierenden Theorien einerseits und dem Verlust der Berechenbarkeit wie behauptet auch existiert. Das Berechenbarkeitsergebnis zu RSRL (mit Ketten) spricht in diesem Zusammenhang eine deutliche Sprache; und man kann sich nicht einfach auf die Betrachtung von RSRL ohne Ketten zurückziehen, da damit ein der Hauptansprüche, nämlich HPSG angemessen beschreiben zu können, nicht mehr erfüllbar wäre. Die Bedeutung dieser Resultate für das Teilprojekt A1 muß in Anbetracht der Tatsache, daß RSRL in A1 entworfen wurde, nicht mehr betont werden.

2.4 Ethno-linguistische Daten des Warao

Neben der theoretischen Bedeutung sollte im Projekt auch die deskriptive Relevanz von Datenstrukturen untersucht werden. Dazu sollten ausgewählte Phänomene der amerindischen Sprache Warao mithilfe eines geeigneten objektorientierten Dokumentationssystems erfaßt und ethno-linguistisch aufbereitet werden. Leider mußten wir jedoch zu Beginn des Projekts feststellen, daß weder das im Antrag in Aussicht genommene Dokumentationssystem CELLAR noch ein Ersatz in geeigneter Weise zur Verfügung stehen. CELLAR vom Summer Institute of Linguistics ist in der erhältlichen Form zwar durch die Arbeit von Simons (1998) konzeptionell weit entwickelt, und die vorliegenden Konzepte entsprechen als objektorientierte auch unseren Vorstellungen. Leider sind diese aber nicht wirklich implementiert. Einige der Ideen sind in das Programm LinguaLinks vom SIL eingeflossen, das CELLAR implementieren soll, aber LinguaLinks hat zwei gravierende Nachteile: Erstens ist LinguaLinks kein Dokumentationssystem, es unterstützt zwar den Aufbau eines traditionellen, keineswegs multimodalen, Lexikons und die morphologische Analyse von Beispielsätzen, kann aber weder mit multimedialen Daten, etwa Bildern, Tondokumenten oder auch nur Corpora, umgehen, noch wird Multidimensionalität unterstützt. Zweitens wurde unserer Kenntnis nach die Entwicklung von LinguaLinks eingestellt. Damit kommt es als Dokumentationssystem nicht in Frage. Bei unserer Suche nach einem Ersatz sind wir zwar über die Annotationsseite vom Linguistic Data Consortium (LDC) (<http://morph ldc.upenn.edu/annotation>) auf einen ganzen Zoo von Annotationsstandards und -werkzeuge gestoßen. Die meisten sind jedoch für die Annotation jeweils ganz bestimmter Corpusdaten vorgesehen, und nicht für eine breite Dokumentation ethnolinguistischer Daten. Das System, das unseren Vorstellungen am nächsten kommt, ist das am Max-Planck-Institut für Psycholinguistik in Nijmegen in der Entwicklung befindliche EUDICO (<http://www.mpi.nl/world/tg/lapp/eudico/eudico.html>). Das Entwicklungsziel von EUDICO ist tatsächlich ein multimodales und multidimensionales Dokumentationssystem für (ethno-)linguistische Daten, sodaß dies weitestgehend unseren Wünschen entspricht. Die Entwicklung der einzelnen Komponenten ist jedoch enorm zeitaufwendig, sodaß gegenwärtig nur etwas mehr als ein Demonstrator von einigen Komponenten bereitgestellt wird. Wichtige Komponenten wie etwa

ein Lexikon sind noch gar nicht realisiert. Nicht zuletzt die im persönlichen Gespräch geäußerte Skepsis eines Gutachters, ob EUDICO in absehbarer Zeit einsetzbar sei, hat uns erkennen lassen, daß EUDICO gegenwärtig ein guter Kandidat für ein Dokumentationssystem ist, aber leider kein einsatzfähiges System.

Somit steht unserer Kenntnis nach kein Dokumentationssystem, das mit der Vielfalt der anfallenden Datentypen, mit der Tatsache, daß diese multimedial und multidimensional sind, angemessen umgehen könnte, zur Verfügung. Wir waren daher gezwungen, eigenständig eine sehr viel kleinere Lösung zu entwickeln. Angesichts der Tatsache, daß die Softwareentwickler am SIL die Entwicklung von LinguaLinks aufgegeben haben, und die Entwickler von EUDICO am MPI mit großem Team und einiger Vorbereitung immer noch kaum mehr als einen Demonstrator bereitstellen können, versteht es sich von allein, daß unser Projekt nicht imstande ist, ein solches Dokumentationssystem selbst zu implementieren. Da EUDICO andererseits als Basis des Dokumentationssystems des von der Volkswagenstiftung finanzierten Großprojekts zur Dokumentation bedrohter Sprachen (DOBES, <http://www.mpi.nl/world/DOBES>) verwendet werden wird, gehen wir davon aus, daß EUDICO schließlich zur Verfügung stehen wird. Wir haben uns daher entschieden, unsere Daten so zu speichern und aufzuarbeiten, daß sie später in EUDICO importiert werden können.

2.4.1 Linguistische Daten und Kontext

Wir gehen in diesem Projekt von einer breiteren Definition *linguistischer Daten* aus, die den direkten Kontext einer sprachlichen Äußerung ebenso berücksichtigt wie kulturelle und soziologische Verhältnisse und umweltbedingte Einflüsse. Unter Kontext wird hier nicht soziale, kulturelle, politische etc. „Zusatzinformation“ verstanden, die sich wie eine äußere Hülle und von ihm abtrennbar um das Kernphänomen (in unserem Falle abstrakt „die Sprache“ und konkret „das Warao“) legt; sondern es wird davon ausgegangen, daß dieser „Kontext“ untrennbar mit dem Kernphänomen selbst verbunden ist und daher für ein Verständnis desselben unentbehrlich bleibt (Latour 1999). Die Notwendigkeit eines solchen Vorgehens wird spätestens im Falle der Beforschung „exotischer“ Sprachen deutlich: Werden hier linguistische Beispiele „nackt“ präsentiert, leidet die Nachvollziehbarkeit in den meisten Fällen erheblich. Innerhalb des Projektes stehen solche Daten verschiedener Klassen und unterschiedlicher Typen aus Feldforschungsaufenthalten bei den Warao im westlichen Orinoko-Delta Venezuelas zur Verfügung. Sie dienen als Grundlage der Exemplifizierung formaler Inhalte des Projektes, wobei eine erschöpfende linguistische Beschreibung des Warao in diesem Rahmen nicht angestrebt ist.

Die Praxis der ethnolinguistischen Feldforschung unterscheidet sich von der linguistischen Labor-Situation dahingehend, daß die Möglichkeit, Variablen zu kontrollieren weitaus begrenzter ist. Die Zusammenarbeit mit lokalen Informanten gibt ihr zudem einen stark dialogischen Charakter (Dammann 1991). Wenig standardisierte Methoden wie die „teilnehmende Beobachtung“ und das offene Interview kommen bevorzugt zur

Anwendung. Der größte Teil der im Projekt verwendeten Daten wurde unter diesen Umständen im Zuge eines einjährigen ständigen Feldforschungsaufenthaltes 1998–1999 in einem großen Waraodorf (circa 400 Einwohner) im westlichen Orinoko-Delta in Venezuela erhoben.

2.4.2 Ergebnisse zu den Demonstrativpronomina

Als Beispiel einer linguistischen Analyse wurde eine erste Untersuchung des Systems der Demonstrativpronomina im Warao¹ durchgeführt. Die Daten zu den Demonstrativpronomina wurden während eines sechswöchigen zweiten Aufenthaltes im Sommer 2000 auf der Grundlage eines psycholinguistischen Fragebogens des Max-Planck-Institutes in Nijmegen (Wilkins 1999), ergänzt durch teilnehmende Beobachtung, erhoben. Der für sprachvergleichende Zwecke entwickelte Fragebogen testet Variablen, welche die Wahl von Demonstrativpronomina beeinflussen. Dabei wird versucht, sich auf die örtliche, nichtkontrastive Deixis zu beschränken. Etwa 5 Informanten wurden hierzu anhand von 25 nachzuspielenden Situationen befragt. Einerseits können in keiner Sprache alle im Fragebogen aufgeführten Variablen relevant sein, andererseits können einzelsprachlich neue Parameter hinzukommen. Es ist daher notwendig, die standardisierte Form der Befragung durch Beobachtungen und Gespräche zu ergänzen (vergl. Wilkins (1999)).

Für das Warao erlaubten gerade diese ergänzenden Methoden die Beobachtung idiomatischer Formen von Deixisdialogen. In Dialogen der Form A: „Tamaja?“ B: „Tai!“ (A: „Dieses (Ding)?“ B: „Ja, dieses (ist es, was ich meine)!“) kann die Wahl von *tai*, einem Pronomen, als diskursgesteuert gelten. Im Übrigen ist in dem dreigliedrigen System (*tamaja*, *amaja*, *otamaja*) die Variable *Abstand zum Sprecher* zentral. Vor diesem Hintergrund scheint eine saubere konzeptuelle Trennung örtlicher Deixis von diskursiver problematisch.

Weitere Kritik am Fragebogen des MPI hat im Rahmen einer Deixiskooperation im SFB² Frau Jungbluth geübt, die eine der grundlegenden Situationen örtlicher Deixis nur am Rande berücksichtigt fand: den Fall der Nicht-Übereinstimmung zwischen der Perspektive des Sprechers und der des Hörers (vergl. Ergebnisbericht B9). Die Variable *gemeinsame Sichtbarkeit des Objektes für Sprecher und Hörer*, die im Warao die Wahl zwischen Demonstrativpronomen und umschreibenden Formen steuert, kann als hierzu parallel angesehen werden. Eine ausführliche Darstellung der Ergebnisse ist dem Arbeitsbericht von Herrmann (2001) zu entnehmen.

¹Das vorliegende Projekt ist kein Deixisprojekt. Dennoch fand natürlich eine enge Kooperation mit den Deixisprojekten B8 und B9 statt.

²Neben den Projekten B8 und B9 des SFB 441 ist hieran auch eine Arbeitsgruppe aus der Japanologie beteiligt.

2.4.3 Archivierung und Annotation der verschiedene Datentypen

Im Zuge der Feldforschungsaufenthalte wurden zahlreiche Schrift, Ton- und Bilddaten erhoben. Insgesamt 1665 Lichtbilder, 19x72 Minuten Tonaufnahmen auf Minidisk, 8 Bände Tagebücher mit zahlreichen Zeichnungen und rund 2000 Karteikarten stehen zur Verfügung. Die elektronischen Erfassung dieser Vielfalt wirft neben technischen Fragen konzeptuelle auf: Zur inhaltlichen Vernetzung wurden 312 Schlagwörter zu Kultur und Sprache in einer *Autoritätsliste* systematisiert und auf dieser Grundlage zunächst Tagebucheinträge und Karteikarten annotiert. Als Ausgangspunkt diente der OCM-Code³. Dieses kulturvergleichend angelegt Instrument mußte auf die Warao-kultur und die spezifische Feldforschung hin abgeändert werden.⁴ Zusätzlich existiert eine Sprecherkartei mit 8 Einträgen zu den Informanten. Es stand in dieser Phase der Entwurf einer konzeptuellen Vernetzung der Daten im Vordergrund, daher wurde vor allem auf eine repräsentative Erfassung aller Datenklassen geachtet. Manche Verfahren sind in der Aufbereitung sehr zeitaufwendig: Die Transkription von Tonaufnahmen aus dem Feld beispielsweise kann bis zu einer Stunde pro aufgenommener Minute in Anspruch nehmen. Es schien daher sinnvoll, fremde publizierte Warao-Texte (Heinen, Wilbert und Rivero 1998, Heinen 1999, Lavadero 1991, Lavadero 1992, Lavadero 1994, Wilbert 1964, Wilbert 1969) einzubeziehen. In einigen Fällen wurden verschiedene Verfahren erprobt. So wurden beispielsweise 98 Photos auf einer Kodak-Photo CD archiviert und 31 von Hand eingescannt. Beide Vorgänge erwiesen sich für den Endzugang als gleich zeitaufwendig. Neben der manuellen Annotation nach thematischen Gruppen versuchten wir eine kontextbasierte Indizierung der Bilder auf der Grundlage von Tagebucheinträgen (Herrmann und Krüger 2000).

Auf der technischen Seite konnten wir aus den eingangs erläuterten Gründen nicht auf ein bestehendes Dokumentationssystem zurückgreifen. Um eine einfache Eingabe und Verarbeitung der Daten zu ermöglichen sowie den späteren Import in EUDICO vorzubereiten, haben wir uns entschieden, für Karteikarten und Bilder auf relationale Datenbanken, für Tonbandaufnahmen, Tagebücher und fremde Waraotexte auf XML-annotierte Corpora zurückzugreifen. Die einzelnen Datentypen sind damit zwar konzeptionell, nicht aber technisch miteinander verbunden. Die vorhandenen Benutzerschnittstellen sind, weil sie von uns selbst entwickelt werden mußten, dafür aber keine Ressourcen zur Verfügung stehen, entsprechend einfach. Insgesamt ist aber durch die Orientierung an etablierten konzeptionellen und technischen Standards und den von EUDICO versprochenen Importmöglichkeiten die zukünftige Verwendbarkeit und

³OCM (Outline of Cultural Materials) ist ein numerologisches System zur Kategorisierung kultureller Daten. Er wurde zuerst in den 1930er Jahren von dem Ethnologen George Murdock und seinen Kollegen entwickelt (Murdock, Ford, Hudson, Kennedy, Simmons und Whiting 1987). Unter anderem kommt er in den „Human Relations Area Files“ (ein Archiv kultureller Daten zu Vergleichszwecken) zur Anwendung und wird im Programm *LinguaLinks* verwendet.

⁴Theoretische Überlegungen führten zur Zusammenarbeit auch außerhalb des SFB 441. Zu nennen sind: Marburger Völkerkunde: Mark Münzel; Tübinger Völkerkunde: Volker Harms, Thomas Hauschild; IVIC, Caracas: Dieter Heinen. *VirtuGrade*: Karin Krüger.

Bereitstellbarkeit der Daten gewährleistet. Wir haben also keinen „Datenfriedhof“ geschaffen, sondern die Integration in große internationale Verbundprojekte vorbereitet.

3 Vergleiche mit Arbeiten außerhalb des SFB und Reaktionen der wissenschaftlichen Öffentlichkeit

Die Ergebnisse des Projekts konnten auf einer Reihe von Tagungen und Workshops vorgestellt werden und wurden dort sehr interessiert aufgenommen. Eine vollständige Liste aller Vorträge findet sich in Abschnitt 6.

Darüberhinaus führte die Arbeit zu wichtigen Kontakten mit internationalen Fachkollegen. Unsere intensive Auseinandersetzung mit ethno-linguistischen Dokumentationssystemen etwa führte dazu, daß Uwe Mönlich zum technischen Berater eines großen niederländischen typologischen Datenbankverbundprojektes berufen wurde. In diesem Rahmen haben wir auch Kontakte zu Prof. Liberman, dem Leiter des Linguistic Data Consortiums und des Institute for Research in Cognitive Science in Philadelphia geknüpft. Es wird eine gemeinsame Dagstuhl-Tagung im September zum Thema *semistrukturierte Daten* stattfinden, auf der sowohl Liberman als auch Mönlich vortragen werden. Mit unserem Vorschlag, coalgebraische Modelle auch im Bereich der Texttechnologie zu verwenden und damit strukturierte Dokumente zu modellieren, beteiligen wir uns an einer noch von der DFG zu genehmigenden distribuierten Forschergruppe *Texttechnologie* unter Leitung von Prof. Metzger in Bielefeld. Unsere Auseinandersetzung mit algebraischen Modellierungswerkzeugen dokumentiert sich in einem eingeladenen Panelvortrag von Uwe Mönlich bei der diesjährigen Konferenz *Formal Grammar – Mathematics of Language* in Helsinki in Zusammenhang mit der europäischen Sommerschule *Logic, Language, and Information*. Unter dem Titel *Algebraic Linguistics* haben wir in diesem Forschungsbereich auch eine Kooperation mit Prof. Keenan und Prof. Stabler des Linguistics Department der University of California at Los Angeles initiiert und erfolgreich Reisegelder der Humboldt-Stiftung eingeworben.

Die im Projekt geleistete coalgebraische Modellierung von HPSG steht im Gegensatz zu klassisch algebraischen Modellen von King (1999) und Pollard (1999). In Abschnitt 2.2 wird ausführlich dargelegt, warum eine coalgebraische Modellierung konzeptionell den Ideen der HPSG als einer lizensierenden Theorie weitaus näher kommt als algebraische Modelle. Interessanterweise besitzt die finale Coalgebra als Modell darüberhinaus technische Eigenschaften, die es sinnvoll erscheinen lassen, sie den Modellen von King oder Pollard vorzuziehen. Die Modelle von King oder Pollard sind im Gegensatz zur finalen Coalgebra nicht kompakt, sie besitzen Redundanzen. Und es können virtuelle Ambiguitäten auftreten, also Fälle, in denen es mehrere Analysen einer sprachlichen Äußerung gibt, ohne das diese verschiedenen Analysen unterschiedlichen Interpretationen entsprechen. Solche virtuellen Ambiguitäten gibt es in

der finalen Coalgebra nicht.

Die Arbeit über die Berechenbarkeit und Komplexität von RSRL entstand in Zusammenarbeit mit Frank Richter aus dem Projekt A1 und mit Prof. Gerald Penn von der University of Toronto.

Die ethno-linguistischen Arbeiten von Stefanie Herrmann führten zu einer Kooperation mit Prof. Gunter Senft vom Max-Planck-Institut in Nijmegen, wo sie im Oktober einen Vortrag halten wird. In der Ethnologie selbst kam es neben der Zusammenarbeit mit der hiesigen Völkerkunde (Prof. Hauschild) auch zu einer Kooperation mit Prof. Münzel von der Marburger Völkerkunde. Schriftlichen Daten werden auch in der Ethnologie nichtschriftliche Medien gegenüber bevorzugt behandelt. Bilder hingegen erfüllen meist rein illustrative Zwecke. Bildliche und tonliche Datentypen spielen jedoch während des Feldforschungsprozesses selbst eine äußerst wichtige Rolle. Multimediale Ansätze werden daher als wichtiges Thema der Zukunft begriffen.

Da die venezolanische Wissenschaft über verschwindend wenig Fachkräfte in den Bereichen Ethnologie und Linguistik verfügt, wird Arbeiten, die sich mit Sprache und Kultur der dortigen Ureinwohner beschäftigen allgemein sehr großes Interesse entgegengebracht. Im Feldforschungsgebiet selbst ist ein relativ großer Personenkreis (Angehörige der dortigen Kapuzinermission in Tucupita, Studierende der örtlichen technologischen Fachhochschule, Dorflehrer und Sozialarbeiter) interessiert an Material zur Waraosprache.

Im Rahmen der Kooperation mit anderen Instituten der Universität Tübingen wurden in der Vergleichenden Sprachwissenschaft zusammen mit Dr. Emmerich Weißhar mehrere Seminare über das Warao und ethno-linguistische Dokumentationssysteme veranstaltet.

Stephan Kepsers intensiver Einsatz für die Netzwerk- und Rechnersicherheit sowie Systemadministration für den gesamten SFB sind im Bericht des Projekts C1 dokumentiert.

4 Offene Fragen

Nachdem es gelungen ist, HPSG erfolgreich mit coalgebraischen Methoden zu modellieren, bietet es sich an, linguistische Theorien in der chomskyschen Tradition genauer zu untersuchen. An erster Stelle steht dabei die Frage, ob auch die Prinzipien- und Parametertheorie (GB) als lizensierende Theorie mithilfe von Coalgebren angemessen modellieren läßt, um damit auch den Paradigmenwechsel in der chomskyschen Tradition modelltheoretisch zu beleuchten. Es stellt sich aber auch die Frage, wie die weiteren Entwicklungen Minimalismus und Optimalitätstheorie zu betrachten sind. Dabei zeigen die Arbeiten von Mönnich (Michaelis, Mönnich und Morawietz 2000) zum Minimalismus, daß bei diesem wiederum ein Paradigmenwechsel, nun zurück zum generierenden Schema, stattgefunden hat und dieser durch algebraische Methoden model-

liert werden kann. Eine Einordnung und Formalisierung der Optimalitätstheorie steht in diesem Rahmen noch aus.

Wie schon in Abschnitt 2.1 erwähnt, stellt sich auch für Auszeichnungssprachen wie XML die Frage, auf welche Weise die durch Annotationen mit solch einer Sprache semi-strukturierten Dokumente korrekt modelliert werden können. Zwar sind die grundlegenden Strukturen Bäume, die Verzeigerungsmöglichkeit von XML erlauben aber, deutlich über Baumstrukturen hinauszugehen. Damit gilt es zu ergründen, ob die so strukturierten Dokumente baumähnlich genug sind, um algebraische Modellierungen zu verwenden, oder ob man es mit allgemeinen Graphen zu tun hat, die besser coalgebraisch modelliert werden können. Näheres hierzu wird im Fortsetzungsantrag beschrieben.

Speziell technisch stellt sich darüberhinaus die Frage, wie RSRL im Vergleich zu klassischen Logiken einzuordnen ist. Dazu könnte man in Komplementierung der in Abschnitt 2.3 behandelten Frage untersuchen, welche Berechenbarkeitsklassen durch RSRL axiomatisierbar sind. Daneben stellt sich die Frage, ob es eine klassische Logik gibt, in der RSRL kodiert werden kann, oder eine Logik, die sich in RSRL ausdrücken läßt. Letztlich geht es dabei um die Ausdruckstärke der Logik RSRL im klassischen Sinne.

5 Veröffentlichungen und Manuskripte

- Herrmann, Stefanie (2002): „Warao“. Erscheint in *Encyclopedia of Linguistics*, Chicago: Fitzroy Dearborn Publishers.
- Kepser, Stephan (2000): „A Coalgebraic Modelling of Head-Driven Phrase Structure Grammar“. In *Proceedings AMiLP2000*, Dirk Heylen, Anton Nijholt, G. Scollo (Hrsg.), S. 81–95.
- Kepser, Stephan (2001): „On the Complexity of RSRL“. In *Proceedings of FG-MOL2001*, Geert-Jan Kruijff, Larry Moss, Richard Oehrle (Hrsg), ENTCS 53, Kluwer.
- Michaelis, Jens, Uwe Mönlich und Frank Morawietz (2000): „Algebraic Description of Derivational Minimalism“. In *Proceedings AMiLP2000*, Dirk Heylen, Anton Nijholt, G. Scollo (Hrsg.), S. 125–141. In Verbindung mit SFB 340 A8.
- Mönlich, Uwe (2001): „Model-Theoretic Description of TAGs“. In *Proceedings of FG-MOL2001*, Geert-Jan Kruijff, Larry Moss, Richard Oehrle (Hrsg), ENTCS 53, Kluwer.

Manuskripte

- Herrmann, Stefanie und Karin Krüger (2000): *Context-Based Image Indexing*, Technischer Bericht, Universität Tübingen.

URL: <http://tcl.sfs.nphil.uni-tuebingen.de/herrmann/Doku/praes1.html>.

Herrmann, Stefanie (2001): *Demonstrativa im Warao: Bericht einer Feldstudie aus dem westlichen Orinoko-Delta*.

Kepser, Stephan (2000): *A Coalgebraic Modelling of Head-Driven Phrase Structure Grammar*.

Kepser, Stephan (2001): *Computability and Complexity of RSRL*.

Mönnich, Uwe (2001): *Regular Description of HPSG*.

6 Aktivitäten: Vorträge, Gastwissenschaftler

Vorträge

- 15.09.1998: Stefanie Herrmann, „First Impressions from Fieldwork-Research among the Warao in the Delta Centro Area“, Caracas, Venezuela. In Vorbereitung auf das Projekt.
- 24.8.1998: Stefanie Herrmann, „‘Ethnografía de la comunicación’ del idioma Warao en la parte occidental del Delta del Orinoco“, auf dem „Congreso Venezolano de Antropología y Sociología“, Mérida, Venezuela. In Vorbereitung auf das Projekt.
- 17.03.1999: Stefanie Herrmann, „Etnolingüística, Estructuras de Poder y Organización Social entre los Warao del Delta Occidental“, IVIC, Caracas, Venezuela.
- 18.5.1999: Stefanie Herrmann, „Cultura y Diferencias Lingüísticas del Warao, Area Bloque Delta Centro“, Caracas, Venezuela.
- 21.5.1999: Uwe Mönnich, „Linguistische Theorien als Datentypen“, Universität Bielefeld.
- 18.6.1999: Uwe Mönnich, „Descriptions of Cross Serial Dependencies“, Potsdam-Tübingen-Workshop über „Komplexitätstheoretische Eigenschaften von Grammatikformalismen“, Tübingen.
- 18.6.1999: Stephan Kepser, „Linguistic Theories as Data Types“, Potsdam-Tübingen-Workshop über „Komplexitätstheoretische Eigenschaften von Grammatikformalismen“, Tübingen.
- 25.6.1999: Stefanie Herrmann, „Datenlage für Warao“, SFB-Tag „Datenproblematik im Spannungsfeld zwischen Empirie und Theorie“, Tübingen.
- 1.7.1999: Frank Morawietz, „Logical and Automata-Theoretic Approaches to Context-Free Tree Sets“, FLoC '99 Workshop on Finite Model Theory and its Applications, Trento, Italien.
- 26.3.2000: Stephan Kepser, „Coalgebraic Modelling of Head Driven Phrase Structure Grammar“, Workshop Coalgebraic Methods in Computer Science, Berlin.
- 22.5.2000: Uwe Mönnich, „Algebraic Description of Derivational Minimalism“,

- AMiLP-Workshop, Iowa City, USA.
- 22.5.2000: Stephan Kepser, „Coalgebraic Modelling of Head Driven Phrase Structure Grammar“, AMiLP-Workshop, Iowa City, USA.
- 25.5.2000: Frank Morawietz, „Derivational Minimalism in Two Regular and Logical Steps“, Tag+5 Workshop, Paris, Frankreich.
- 15.7.2000: Stefanie Herrmann und Karin Krüger-Tielmann: „Context-Based Index Imaging“, VirtuGrade-Seminar, Tübingen.
- 3.10.2000: Frank Morawietz, „A Formal Approach to Non-Context-Free Phenomena“, Universität Passau.
- 13.10.2000: Stephan Kepser, „Coalgebraic Modelling of Head Driven Phrase Structure Grammar“, Procope-Workshop, Aix-en-Provence, Frankreich.
- 28.10.2000: Uwe Mönnich und Frank Morawietz, „Attribuierter Minimalismus“, Tübingen-Potsdam-Workshop, Tübingen.
- 3.11.2000: Stephan Kepser, „Coalgebraic Modelling of Head Driven Phrase Structure Grammar“, International HPSG-Workshop, Tübingen.
- 19.12.2000: Uwe Mönnich und Frank Morawietz, „A Two-Step Operational Description of TAGs“, Joint UPennsylvania (IRCS) – Uni Tübingen (SFB441) Workshop „Linguistic Data Structures“, Tübingen.
- 15.2.2001: Stefanie Herrmann, „Bilder und Erfahrungen einer Feldforschung bei den Warao im westlichen Orinokodelta, Venezuela“, Institut für Ethnologie, Universität Tübingen.
- 24.2.2001: Stefanie Herrmann und Stephan Kepser: „A Broader View on the Notion of a Linguistic Datum“, SFB 441-Tagung 2001, Tübingen.
- 6.5.2001: Uwe Mönnich, „Algebraic Description of TAGs“, Department of Linguistics, Universität Los Angeles, USA.
- 11.8.2001: Stephan Kepser, „On the Complexity of RSRL“, FG-MOL Konferenz, Helsinki, Finnland.
- 12.8.2001: Uwe Mönnich, „Model-Theoretic Description of TAGs“, FG-MOL Konferenz, Helsinki, Finnland.

Gastwissenschaftler

- 2.11.1999: Prof. Dr. Klaus Schulz, CIS, Universität München, „Abfrage von Baumdatenbanken und vollständige Antwortaggregate“ und „Ein System zur automatischen Analyse von Korrespondenz“.
- 6.12.1999: Dr. Dik Bakker, Universiteit Amsterdam, „A Language Typology Resource Centre: Linguistics Online“.
- 7.12.1999: Dr. Mily Crevels, Universiteit Amsterdam, „An Overview over Native Lan-

guages in Rondonia“.

27.3.2000: Peter Wittenburg, MPI für Psycholinguistik, Nijmegen, NL, „Tools for Ethno-Linguistic Data Processing“.

10.7.2000: Dr. Stefan Riezler, IMS, Universität Stuttgart, „Estimating Stochastic Constraint-Based Grammars from Unannotated Data for Statistical Disambiguation“ und Workshop „ECM Algorithmen“.

28.10.2000: Dr. Tom Cornell, Cymphonie Inc., USA, „Parsing and Grammar Engineering with Tree Automata“, Tübingen-Potsdam-Workshop, Tübingen.

7 Zitierte Literatur

Abadi, Martín und Luca Cardelli (1996): *A Theory of Objects*, Springer-Verlag.

Abiteboul, Serge, Peter Buneman und Dan Suciu (2000): *Data on the Web*, Morgan Kaufmann.

Aczel, Peter (1997): „The Initial Algebra and the Final Coalgebra Perspectives“, in Helmut Schwichtenberg, (Hrsg.), *Logic of Computation*, Springer-Verlag, S. 1–33.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*, MIT Press.

Chomsky, Noam (1981): *Lectures on Government and Binding*, Foris Publications, Dordrecht, Holland.

Church, Alonso (1936): „A Note on the Entscheidungsproblem“, *Journal of Symbolic Logic* **1**(40–41). Correction *ibid.* P. 101–102.

Dammann, Rüdiger (1991): *Die dialogische Praxis der Feldforschung : Der ethnographische Blick als Paradigma der Erkenntnisgewinnung*, Campus.

Downey, Rodney Graham und Michael R. Fellows (1999): *Parameterized Complexity*, Springer-Verlag.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum und Ivan Sag (1985): *Generalized Phrase Structure Grammar*, Basil Blackwell.

Heinen, Dieter (1999): *Vocabulario Warao Referente al Petróleo / Warao Petroleum Vocabulary*. Versión Trilingüe. (unveröffentlichtes Manuskript).

Heinen, Dieter, Werner Wilbert und Tirson Rivero (1998): „Idamo Kabuka. El “Viejo Corto”“, in *Antropológica*, Fundación La Salle. Suplemento No 6.

Immerman, Neil (1987): „Expressibility as a Complexity Measure: Results and Directions“, in *Second Structure in Complexity Theory Conference*, Computer Soc. of the IEEE, S. 194–202.

Jacobs, Bart und Jan Rutten (1997): „A Tutorial on (Co)Algebras and (Co)Induction“, *Bulletin of EATCS* **62**, 222–259.

- Kepser, Stephan (1994): *A Satisfiability Algorithm for a Typed Feature Logic*, Magisterarbeit, Seminar für Sprachwissenschaft, Universität Tübingen, Arbeitspapiere des SFB 340, Bericht Nr. 60.
- King, Paul John (1989): *A Logical Formalism for Head-Driven Phrase Structure Grammar*, Dissertation, University of Manchester.
- King, Paul John (1999): „Towards Truth in HPSG“, in Valia Kordoni, (Hrsg.), *Tübingen Studies in Head-Driven Phrase Structure Grammar, Vol 2*, Arbeitspapiere des SFB 340, Bericht Nr. 132, S. 301–352.
- King, Paul John, Kiril Ivanov Simov und Bjørn Aldag (1999): „The Complexity of Modellability in Finite and Computable Signatures of a Constraint Logic for Head-Driven Phrase Structure Grammar“, *Journal of Logic, Language and Information* **8**(1), 83–110.
- Latour, Bruno (1999): *Pandora's Hope : Essays on the Reality of Science Studies*, Harvard University Press.
- Lavandero, Julio Perez (1991): *Ajotejana, Mitos*, Hermanos Capuchinos.
- Lavandero, Julio Perez (1992): *Ajotejana, Relatos*, Hermanos Capuchinos.
- Lavandero, Julio Perez (1994): *Uaharaho. Ethos Narrativo*, Hermanos Capuchinos.
- Murdock, George P., Clellan S. Ford, Alfred E. Hudson, Raymond Kennedy, Leo W. Simmons und John W. M. Whiting (1987): *Outline of Cultural Materials*, 5. Auflage, Human Relations Area Files.
- Pollard, Carl (1999): „Strong Generative Capacity in HPSG“, in Gert Webelhuth, Jean-Pierre Koenig und Andreas Kathol, (Hrsg.), *Lexical and Constructional Aspects of Linguistic Explanation*, CSLI, S. 281–297.
- Pollard, Carl und Ivan A. Sag (1987): *Information Based Syntax and Semantics, Vol. 1: Fundamentals*, Lecture Notes Nr. 13, CSLI.
- Pollard, Carl und Ivan A. Sag (1994): *Head-Driven Phrase Structure Grammar*, University of Chicago Press.
- Reichel, Horst (1995): „An Approach to Object Semantics Based on Terminal Coalgebras“, *Mathematical Structures in Computer Science* **5**, 129–152.
- Richter, Frank (2000): *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*, Dissertation, SfS, Universität Tübingen.
- Richter, Frank, Manfred Sailer und Gerald Penn (1999): „A Formal Interpretation of Relations and Quantification in HPSG“, in Gosse Bouma, Erhard Hinrichs, Geert-Jan M. Kruijff und Richard T. Oehrle, (Hrsg.), *Constraints and Resources in Natural Language Syntax and Semantics*, CSLI Publications, S. 281–298.
- Rogers, James (1998): *A Descriptive Approach to Language-Theoretic Complexity*, CSLI Publications.

- Rutten, Jan (2000): „Universal Coalgebra: A Theory of Systems“, *Theoretical Computer Science* **260**(1–2), 3–80.
- Simons, Gary F. (1998): „The Nature of Linguistic Data and the Requirements of a Computing Environment for Linguistic Research“, in John M. Lawler und Helen Aristar Dry, (Hrsg.), *Using Computers in Linguistics: A Practical Guide*, Routledge, S. 10–25.
- Wilbert, Johannes (1964): *Warao Oral Literature*, Monograph Nr. 9, Instituto Caribe de Antropología y Sociología. Fundación La Salle de Ciencias Naturales.
- Wilbert, Johannes (1969): *Textos Folklóricos de los Indios Warao*, Latin American Studies Nr. 12, Latin American Center. University of California.
- Wilkins, David (1999): *Manual for the 1999 Field Season*, Language and Cognition Group of the Max Planck Institute for Psycholinguistics.