Ergebnisbericht

Teilprojekt:	**A8**
Thema:	**Syntaktische Domänen für die Grammatiktheorie und die maschinelle Verarbeitung natürlicher Sprache**
Leiter:	**Prof. Dr. Uwe Mönnich**
Mitarbeiter:	**Hans-Peter Kolb** **Frank Morawietz** **Tom Cornell**

Inhaltsverzeichnis

1 Kenntnisstand bei der letzten Antragstellung und Ausgangsfragestellung

Ziel des Projekts war die Entwicklung eines formalen Rahmens, der es erlaubt, syntaktische Analysen im Rahmen des "Prinzipien- und Parameter"-basierten Ansatzes der generativen Grammatik (im folgenden: PP-Theorie) mit Hilfe von binären Relationen zwischen syntaktischen Objekten auf der Basis einer möglichst schwachen Logik auszudrücken.

Die PP-Theorie erschien dabei nicht nur wegen ihres breiten empirischen Bereichs, sondern auch hinsichtlich ihrer universalistischen und modularen Ausgangspunkte als linguistische Grundlage für die Verarbeitung natürlicher Sprache besonders geeignet. Gleichzeitig beschränkten sich alle Versuche, sie formal zu behandeln, methodisch auf zwei konkurrierende Ansätze, die beide mit großen Nachteilen behaftet sind: Entweder wurde eine *fixierte Version* der Theorie ohne Berücksichtigung von alternativen Vorschlägen oder gar möglichen künftigen Entwicklungen formalisiert; oder es wurde eine starke Metatheorie, wie etwa die Prädikatenlogik—eventuell selbst höherer Stufe—benützt, in deren Rahmen gerade wegen ihrer Allgemeinheit die speziellen Eigenschaften natürlich-sprachlicher Grammatiken nur in sehr unzureichendem Maße strukturell reflektiert, geschweige denn ausgenützt werden können, wodurch solche Formalisierungen nur begrenzten Wert für die Entwicklung der linguistischen Theorie haben. Zudem stehen allgemeine Entscheidungsverfahren für derartige starke Logiken aus prinzipiellen Gründen nicht zur Verfügung.

Im Gegensatz dazu sollte dieses Projekt eine Metatheorie für die Formalisierung (PP-basierter) syntaktischer Theorien entwickeln, die sowohl die spezifischen Eigenschaften syntaktischer Objekte als auch Entscheidbarkeits- und Effizienzerfordernisse inhärent reflektiert, m.a.W., eine Metatheorie, die auf der Basis einer möglichst schwachen Logik von den primitiven Begriffen der grammatischen Theorie direkt Gebrauch macht—und sie damit erst bedeutungsvoll zur Verfügung stellt.

Dazu mussten vor allem diejenigen formalen Eigenschaften untersucht werden, die die PP-Theorie von anderen modernen Grammatikmodellen unterscheidet, wobei die Frage der konfigurationellen Lizensierung der grammatischen Entitäten und ihrer Lokalitätsbedingungen eine zentrale Rolle spielen sollte, da das PP-Modell gerade an diesem Punkt durch den globalen Charakter seiner Wohlgeformtheitsbedingungen deutlich stärker ist als andere Modelle. Die Erwartung war, daß die daraus entstehenden Probleme insbesondere bezüglich der Übertragbarkeit syntaktischer Analysen durch die Installierung von syntaktischen Domänen im Sinne von Kracht [27] als Basis der Metatheorie des PP-Modells auf ein traktierbares Maß würden zurückgebracht werden können.

Da grammatische Formalismen gesehen werden können als Spezialfälle von Computersprachen, boten sich die Konzepte und Techniken der denotationellen Se-

mantik an, um eine genaue Spezifikation der Bedeutung dieser Formalismen auszuarbeiten. Sie sollten es erlauben, Fragen der Korrektheit derartig spezifizierter Grammatiken ebenso Beachtung zu schenken, wie dem Problem der Vergleichbarkeit und—eventuellen—Übersetzbarkeit zwischen verschiedenen Grammatikformalismen. Begreift man die PP-Theorie als ein System zum "Rechnen" mit partiell spezifizierten Elementen, so wird unmittelbar deutlich, dass hier insbesondere die Bereichstheorie von Nutzen ist, deren Leitidee sich dahingehend zusammenfassen lässt, daß die Bedeutung von Konstrukten einer Programmiersprache durch partielle, stetige Funktionen zwischen wiederum partiell spezifizierten Daten oder Berechnungszuständen gegeben sind.
Unmittelbar zu Beginn des Förderungszeitraums wurden Rogers' [42] Forschungsergebnisse bekannt, dessen metatheoretische Behandlung von (insbesondere PP-basierten) Grammatiktheorien natürlicher Sprache auf der Basis der schwachen monadischen Prädikatenlogik zweiter Stufe mit multiplen Nachfolgerfunktionen (in folgenden kurz $S\omega S$ oder MSO genannt) von einer ähnlichen Analyse der Lage ausgeht und äusserst elegante Teillösungen für eine Reihe der oben genannten Desiderata bietet. Gleichzeitig lässt sie aber systematisch und an interessanten Stellen zu wünschen übrig. Es lag daher nahe, MSO und MSO-basierten Analysen einen prominenten Platz im Arbeitsprogramm einzuräumen, zumal der auf syntaktischen Domänen fussende Ansatz bei weniger Transparenz denselben Beschränkungen unterliegt (vergl. Kracht [28] und die kurze Diskussion in Abschnitt 2). Auch eignet sich der MSO-Ansatz vorzüglich für eine bereichstheoretische Aufarbeitung.
Hauptziel blieb dabei die Entwicklung eines linguistisch fundierten Systems zur Formulierung syntaktischer Theorien, das einerseits stark genug ist, um weitere theoretische Entwicklung (insbesondere im Rahmen der PP-Theorie) zu erlauben, andererseits aber schwach genug, um formalen Wildwuchs—in linguistisch interessanter Weise—zu beschränken, und dessen formale Eigenschaften wohldefiniert und von einer Art sind, die effiziente Verarbeitung zumindest nicht von vorneherein ausschliesst.

2 Angewandte Methoden

Projekt A8 steht in der Tradition der modelltheoretischen Behandlung grammatischer Formalismen. Bei der Annäherung an seinen Gegenstand bedient es sich einer in dieser Form ungewöhnlichen Kombination bereichstheoretischer, algebraischer und logischer Methoden.

[Bereichstheorie:] Auf der einen Seite steht der bereichstheoretische Ansatz, der von den für die linguistische Theorie relevanten Datenstrukturen und der sie bestimmenden Informationsordnung ausgeht, um den Begriff von Interpretationbereichen in konsistenter Weise zu formulieren.

Ein Bereich im mathematischen Sinn ist eine Struktur, die ein präzises Modell für den Informationszuwachs bei Berechnungsvorgängen liefert. Die Vorstellung, die das Modell bestimmt, richtet sich auf ein Verfahren, bei dem schrittweise eine bessere Annäherung an das Resultat der Berechnung erzielt wird. Da das Resultat entweder in endlich vielen Schritten erreicht wird oder sich als eine unendliche Kette von Annäherungen darstellt, sollte der Bereich so gestaltet sein, dass sich unendliche Folgen von Annäherungen durch ein Element des Bereichs wiedergeben lassen, d.h. der Bereich sollte über die notwendigen Vollständigkeitseigenschaften verfügen.
Das Verhältnis zwischen Bereichen und ihrer logischen Betrachtung kann in gegenläufiger Weise Gegenstand von Untersuchungen sein. Auf der einen Seite stellt sich die Aufgabe, für bestimmte Klassen von Bereichen eine formale Sprache und einen Beweiskalkül zu entwerfen. Auf der anderen Seite stellt sich die modelltheoretische Aufgabe, mathematische Strukturen vorgegebener logischer Systeme als Bereiche zu präsentieren. Es ist dieser zweite Ansatz, der für unser logisch orientiertes Ziel, eine deklarative Semantik für grammatische Systeme zu finden, während der laufenden Phase des Sonderforschungsbereiches im Vordergrund stand. Da für uns nicht die Perspektive bestimmend war, für die im Rahmen einer operationalen oder derivationellen Theorie der schrittweise Aufbau und, damit verbunden, der wachsende Informationsgehalt grammatischer Repräsentationen charakteristisch ist, sondern wir uns in erster Linie auf die Uniformität konzentriert haben, die Modelle mit der gleichen Struktureigenschaft verbindet, haben wir uns auf das Problem konzentriert, wie derartige Modellklassen—im konkreten Fall konfigurationell charakterisierte Baumklassen—als Elemente eines Bereichs aufzufassen sind. Dies ist im skizzierten methodischen Rahmen durch Anwendung klassischer modelltheoretischer Techniken in vielen Situationen möglich, ohne dabei das Bild von Zuständen mit grösserem oder schwächerem Informationsgehalt aus den Augen zu verlieren.

[Algebraische Sprachtheorie:] Die Betrachtung von abstrakten Automaten und von kontextfreien Sprachen bezieht sich üblicherweise auf Halbgruppen mit einem Einselement. Das Ziel der algebraischen Sprachtheorie besteht darin, eine Erweiterung des Untersuchungsbereichs auf andere algebraische Systeme vorzubereiten. Durch dieses methodische Vorgehen lassen sich die Verhältnisse auf der Ebene der Halbgruppen im nachhinein als Spezialfälle allgemeiner Strukturzusammenhänge begreifen.
Dieses Programm einer Erweiterung der ursprünglich (zeichen-)kettenbezogenen formalen Sprachtheorie erhält durch einen Vergleich zwischen Halbgruppen und monadischen Algebren eine besonders eindrückliche Motivation. Sei Σ ein endliches Vokabular. Eine kontextfreie Sprache lässt sich durch ein Regelsystem charakterisieren, in dem die Ausdrücke, die für Einheiten des Hilfsvokabulars nach Vorgabe des Systems substituierbar sind, Elemente der freien Halbgruppe $H(\Sigma, N)$ bilden, die durch das (End-)Vokabular Σ und das Hilfsvokabular N

erzeugt ist. Das Regelsystem bestimmt einen generativen Prozess, in dem für jede Einheit des Hilfsvokabulars eine kontextfreie Sprache simultan erzeugt wird. In vergleichbarer Weise sind die regulären Sprachen durch ein Regelsystem definierbar, in dem die ersetzenden Ausdrücke Elemente der (absolut) freien Algebra $M(\Sigma, N)$ bilden, in der die Endsymbole aus Σ die Rolle von monadischen Operatoren spielen und deren erzeugende Menge sich aus den Hilfssymbolen N und der leeren "Kette" λ zusammensetzt. Unter dieser allgemeinen algebraischen Perspektive lassen sich bekannte Resultate aus dem Gebiet der regulären Sprachen übertragen, sobald die erforderliche Angleichung an die jeweils relevante Signatur der gemeinten Algebra explizit gemacht ist. So haben z.B. die endlichen (Ketten-)Automaten ihre Entsprechung in den endlichen Baumautomaten, eine (endliche) Äquivalenzrelation, die gegenüber monadischen Operatoren links- oder rechtsinvariant ist, entspricht einer Äquivalenzrelation, die gegenüber der binären Verkettung invariant ist, und kontextfreie Sprachen erscheinen in dieser Betrachtungsweise als Vereinigungen endlicher Kongruenzklassen, deren Invarianz unter Berücksichtigung der Halbgruppensignatur spezifiziert ist.
In der algebraischen Sprachtheorie wird die Absicht verfolgt, den grössten methodischen Nutzen aus dieser Generalisierung auf eine arbiträre Signatur zu ziehen. Ersetzungsregeln stellen in diesem theoretischen Rahmen Substitutionsanweisungen dar, die es erlauben, Ausdrücke der jeweils vorgegebenen Signatur schrittweise zu erweitern. In ihrer bildlichen Darstellung erscheinen diese Ausdrücke als Bäume, deren Knoten durch Symbole des Operatorbereichs etikettiert werden und die in dieser Repräsentationsform die Voraussetzung bieten, unmittelbar an die Begrifflichkeit der starken generativen Kapazität grammatischer Formalismen anzuknüpfen. Da sich die Ausdrücke bzw. ihre isomorphen Vertreter in Baumform als Elemente einer freien Algebra auffassen lassen, können sie in beliebigen semantischen Bereichen derselben Signatur interpretiert werden. Diese strukturverträgliche Beziehung bildet die Voraussetzung dafür, semantische Probleme auf die Ebene der Bäume anzuheben und sie dort mit Methoden zu lösen, die sich an die monadische Theorie der regulären Sprachen anlehnen.

[Schwache Logiken:] Am konkretesten ist ein methodischer Ansatz, der von einer schwachen Logik mit bekannten Eigenschaften ausgeht und versucht, sie—unter Erhalt der relevanten Eigenschaften—zu einem adäquaten formalen System für linguistische Theoriebildung auszubauen.
Eine solche Logik ist die schwache monadische Prädikatenlogik zweiter Stufe mit mulitplen Nachfolgerfunktionen (MSO). Rogers [42] verdanken wir die Einsicht, daß sie sich in vielerlei Hinsicht vorzüglich als realistische formale Grundlage für die linguistische Theoriebildung eignet, obwohl ihre angenehmen formalen Eigenschaften, nämlich ihre Entscheidbarkeit und die Möglichkeit zur Definition von Finitheit von Mengen, durch eine gewisse Ausdrucksschwäche erkauft ist: Seit den späten 60er Jahren (vergl. Doner [14], Thatcher&Wright [50], Rabin [41]) ist bekannt, dass MSO mit der Klasse der regulären Baumautomaten bzw.

-grammatiken und damit der Klasse der kontextfreien Sprachen äquivalent ist.
Ein ähnliches Resultat besteht auch für die Modallogik-basierten Systeme, wie sie etwa von M. Kracht oder von P. Blackburn u.a. entwickelt worden sind: Krachts [28] *Coding Theorem* besagt, dass jede in seiner modalen, bzw. dynamischen "Orientierungssprache" *Olt* definierbare Verteilung von Merkmalen über endlich verzweigende Bäume in einer kontextfreien Grammatik kodiert werden kann. Da einerseits mittels *Olt* beliebige syntaktische Domänen mit explizit definierbarem Generator dargestellt werden können, andererseits auf finit verzweigenden Bäumen jeder definierbare Generator explizit definierbar ist, ist auch eine domänenbasierte Metatheorie mit Bäumen als intendierten Modellen kontextfrei.
Beide Ansätze zeigen, dass kontextfreie Metatheorien es durchaus erlauben, viele der interessanteren Generalisierungen des PP-Ansatzes auszudrücken. Insbesondere im MSO-Fall sind dabei durch die Strukturorientierung des Formalismus Formalisierungen von ungekannter Transparenz möglich, und einige der Nicht-Definierbarkeitsresultate von Rogers wie die zur *freien Indizierung* oder zur *Kopiertheorie der Bewegung* scheinen selbst lange und erbitterte Diskussionen in der PP-Literatur aus rein formalen Gründen zu entscheiden. Andererseits ist offensichtlich, dass die Formalismen in ihrer reinen Form zu schwach sind, um einige der exotischeren Sprachen—wie etwa das Deutsche oder das Niederländische—angemessen zu erfassen.
In diesem methodischen Umfeld geht es nun um die Suche nach der "richtigen" schwachen Logik (also der Logik mit genau der für linguistische Anwendungen nötigen Ausdruckskraft) durch kontrollierte Erweiterungen der bestehenden schwachen Formalismen. A8 hat sich dabei in dieser Projektphase auf MSO konzentriert.

[Descriptive Complexity Theory] Die traditionelle Brücke zwischen algebraischer und logischer Betrachtungsweise bildet die *deskriptive Komplexitätstheorie*, die das Verhältnis zwischen einem (logischen) Beschreibungssystem und dem (algebraischen) Bereich des damit Beschreibbaren untersucht. Allerdings hält dieses Forschungsgebiet seit seinen Anfängen (Büchi [5]; Elgot [15]) die beiden Ansätze streng geschieden und benützt algebraische Methoden "nur" zur Erforschung der Eigenschaften des jeweiligen logischen Systems.
Projekt A8 geht hier insofern einen Schritt weiter, als es Ergebnisse aus der algebraischen Sprachtheorie etwa auch benützt, um die Ausdruckskraft des logischen Beschreibungssystems zu erweitern, ohne die eigentliche Logik zu verändern.

[Automaten, Entscheidbarkeit, Verarbeitung:] Durch die Reformulierung des Entscheidbarkeitsbeweises von Rabin/Doner/Thatcher&Wright für MSO auf finiten Mengen direkt mit einer Induktion auf Baumautomaten und ihren Konstruktionen zeigt sich sein konstruktiver Charakter für beliebig MSO–definierbare Signaturen. MSO Formeln sind erfüllbar genau dann, wenn ihre erfüllenden Variablenzuweisungen von einem Baumautomaten erkannt werden. Der Beweis nutzt dies aus, indem jede MSO Formel induktiv in Baumautomaten übersetzt wird.

Wenn die vom Automaten erkannte Sprache nicht leer ist, hat man gleichzeitig gezeigt, daß der existentielle Abschluß der Formel ein Theorem darstellt. Im Detail müssen also als Induktionsverankerung die Relationen der Signatur als Baumautomaten spezifiziert werden, die dann gemäß dem Formelaufbau zu einem neuen Automaten kombiniert werden.
Die Effektivität des Beweises kann zur Implementation eines Entscheidungsverfahrens für die Erfüllbarkeit von MSO Formeln ausgenützt werden. Da linguistische Verarbeitung für Constraint-basierte Grammatikformalismen als Theorembeweisen betrachtet werden kann, können damit etwaige Spezifikationen in MSO effektiv getestet und visualisiert werden.

3 Ergebnisse und ihre Bedeutung

Die Theorie der Semantik von Grammatikformalismen ist durch eine Dichothomie gekennzeichnet, die ihren Ausdruck u.a. in folgenden Begriffspaaren findet:

denotational	/	operational
deklarativ	/	prozedural
logisch	/	computationell

Die operationale Semantik ist durch einen stufenweisen effektiven Prozeß beschrieben, der, ausgehend von einem syntaktischen Objekt, seine Bedeutung generiert. Die deklarative Semantik in einer ihrer Ausprägungen ist definiert durch die Interpretation des syntaktischen Objekts als eines Systems von Gleichungen, dessen Lösung in einem geeigneten Bedeutungsraum gesucht wird. Im idealen Fall sind beide Arten von Semantik äquivalent, d.h. die Resultate des stufenweisen Prozesses sind identisch mit der Lösungsmenge. Das klassische Beispiel für diese Äquivalenz ist die duale Charakterisierung der kontextfreien Sprachen einmal als Herleitungsverfahren auf der Basis der Produktionsregeln und einmal als Fixpunkt eines Systems von Gleichungen in einer Halbgruppe. Ideale Verhältnisse, die den Beweis solcher Zusammenhänge gestatten, sind jedoch eher die Ausnahme, und die Aufgabe der deklarativen Semantik besteht darin, eine stabile Grundlage für die Deutung von Formalismen zu liefern, mit denen man bis dahin geleitet von einem intuitiven Verständnis eher sorglos umgegangen ist. Als Teilschritte auf diesem Weg hat Projekt A8 Ergebnisse auf den folgenden Gebieten vorzuweisen:[1]

1. Bereichstheorie und Linguistik: Mönnich [31, 32], Morawietz [34]
2. Formale Eigenschaften linguistischer Theorien: Kolb [24, 25], Grefe & Kracht [17], Kim et.al. [22], Cornell [12, 13]
3. Schwache Logiken und algebraische Sprachtheorie

[1]Im Interesse einer einheitlichen Zitierweise haben wir die besprochenen Schriften des Projektes A8—neben der Auflistung in Abschnitt 6—auch in die allgemeine Literaturliste aufgenommen.

– Linguistische Aspekte: Mönnich [33], Kolb [26], Rogers [43]

– Verarbeitungsaspekte: Morawietz [35], Morawietz & Cornell [36, 37], Basin & Podelski

Linguistik und Bereichstheorie

In einer grundlegenden Arbeit haben Pereira und Shieber für grammatische Systeme, in denen Merkmalsstrukturen eine zentrale Rolle spielen, eine denotationelle Semantik angegeben, die auf den Techniken der von Scott und Strachey entwickelten Bereichstheorie aufbaut. Mit einer rigorosen denotationellen Bedeutungsexplikation sind eine Reihe von Vorteilen verbunden, die geeignet sind, Einsichten in die generative Kapazität linguistischer Formalismen fruchtbar zu ergänzen. So erleichtert dieser Ansatz unter anderem die Aufgabe, die durch einen Formalismus spezifizierten Datenstrukturen zu erhellen und die Korrektheit von Grammatiken zu beweisen. Darüber hinaus lassen sich durch dieses methodische Vorgehen Techniken aus anderen Bereichen, wie z.B. der Programmiersprachensemantik, für die Analyse von Grammatikformalismen nutzen und eine Generalisierung verschiedener Resultate erreichen, die sich auf die Entscheidbarkeit, Normalformen und Abschlußeigenschaften von grammatischen Systemen beziehen. Schließlich erlauben die strukturverträglichen Beziehungen zwischen verschiedenen denotationellen Bereichen den Vergleich zwischen verschiedenen Typen von Semantiken und in besonders gelagerten Fällen den Nachweis von universellen Interpretationen.

In Mönnich [31] wird nun gezeigt, daß die deklarative Interpretation der Merkmalsstrukturen auch in einem situationstheoretischen Rahmen präsentiert werden kann. Er baut auf der bekannten informationstheoretischen Deutung der Merkmalsstrukturen auf und zeigt, daß sie als die syntaktische Seite eines situationstheoretisch motivierten Klassifikationsbereichs angesehen werden kann. Vor dem Hintergrund dieser Interpretation ergibt sich ein überraschend einfacher Vollständigkeitsbeweis für die Beschreibung von Merkmalsstrukturen durch Pfadgleichungen und eine Reduktion einer spezifischen Form von Informationsübertragung im Sinne von Dretske auf den Begriff einer Scott-stetigen Funktion. Dieser situationstheoretische Ansatz zur deklarativen Deutung illustriert exemplarisch das Potential, das in einer Kombination semantischer und konstruktiver, i.e. informationstheoretischer, Methoden für die präzise Spezifikation grammatischer Datenstrukturen verborgen ist.

Trotz der unabweisbaren Vorteile eines denotationellen Ansatzes scheint es bisher nur für merkmalsbezogene Grammatiksysteme den Versuch zu geben, die Möglichkeiten der Bereichstheorie nutzbar zu machen. Ein Grund für diese Zurückhaltung dürfte darin liegen, daß im allgemeinen Fall es unmöglich ist, für linguistisch relevante Bereiche die notwendigen Vollständigkeitseigenschaften nachzuweisen. Einen Ausweg verspricht hier die logisch basierte Theorie der

Baumsprachen, die den Ausgangspunkt von Mönnich [32] bildet. Dort wird nachgewiesen, dass durch eine Anwendung klassischer modelltheoretischer Methoden sich eine bereichstheoretische Interpretation für die schwache monadische Logik zweiter Stufe mit mehreren Nachfolgerfunktionen gewinnen lässt. Wie gezeigt wird, legt die Äquivalenz zwischen kontextfreien Sprachen und Formeln der schwachen monadischen Logik zweiter Ordnung es nahe, eine Familie kontextfreier Sprachen im modelltheoretischen Sinn als einen Typ der genannten logischen Sprache zu betrachten. Die Lösung besteht darin, einen Bereich in der Form zu konstruieren, dass in ihm die Erfüllbarkeitsmengen von Bäumen, die eine Formel dieser schwachen Spezifikationssprache zweiter Ordnung wahr machen, die Rolle von kompakten Elementen übernehmen. Im Gegensatz zu dem inkrementellen Informationszuwachs, der für den derivationellen Ansatz kennzeichnend ist, legt sich für das modelltheoretische Vorgehen eine Halbordnung nahe, bei der eine Erfüllbarkeitsmenge L eine bessere Approximation bildet als eine Erfüllbarkeitsmenge L', wenn L in L' enthalten ist und damit weniger Alternativen offenlässt. Die erforderliche Vollständigkeitseigenschaft, die es erlaubt, unendlichen Sequenzen von Approximationen einen eindeutigen Wert zuzuweisen, wird durch eine geeignete Filterkonstruktion gesichert. Die Baummengen, deren Blätter die kontextfreien Sprachen bilden, spielen hierbei die Rolle von kompakten Elementen. Die Filterkonstruktion hat die angenehme Eigenschaft, dass für nahezu alle linguistisch interessanten Definitionsformen sich die Erfüllbarkeit der entsprechend konstruierten Formeln der grammatischen Spezifikationssprache von dem ursprünglichen auf das Filtermodell überträgt. Somit erhalten zumindest jene Begriffe aus der PP-Theorie eine Deutung informationstheoretischen Gehalts, die sich mit den Mitteln dieser Logik explizieren lassen. Mit dieser Deutung liegt freilich noch keine verbindliche deklarative Interpretation vor, die dann auch für diese Strukturen Korrektheit, Universalität der Interpretation und Effizienz der Implementation in dem Maße vorbereiten würde, wie dies schon für andere, berechnungstheoretisch besser untersuchte Ansätze geschehen ist.

Eine weitere, klassischere Anwendung der Bereichstheorie ist Morawietz [34], der sie als Basis für die (Re-)Definition seiner relationalen Extension der monadischen Prädikatenlogik 2.Stufe $\mathcal{R}(\mathcal{MSO})$, auf die unten näher eingegangen wird, benützt, was eine klarere Sicht auf infinite Berechnungen erlaubt, die einen Problemfall für die auf Höhfeld und Smolka beruhende ursprüngliche Definition darstellen. Gleichzeitig vereinfacht diese Definition auch die Anwendung Constraint–logischer Optimierungs- und Erweiterungstechniken.

Formale Eigenschaften linguistischer Theorien

Die Basis jeder formalen Behandlung linguistischer Theorien ist die Klärung der von ihnen postulierten Strukturen. Für PP-orientierten Theorien scheint dieses Problem schnell gelöst: Repräsentationell beschränken sich die Analysen scheinbar auf Bäume, derivationell auf Folgen von Bäumen. Bezieht man allerdings Konstukte wie Segmente und Kategorien, Indizes oder Ketten mit in die Betrachtung ein, so wird das Bild schnell komplizierter.

Insbesondere eine kohärente Explikation der "Segment/Kategorie"-Dichotomie, die seit ihrer Lancierung in Chomsky [6] für nicht-enden-wollende Verwirrung in der linguistischen Literatur sorgt, ist ein echtes Desiderat. Bis zu jenem Zeitpunkt waren die "Kategorien", also die linguistisch relevanten Komponenten eines Strukturbaumes, die mit Namen versehenen Knoten der Struktur; die "linguistischen" Relationen waren einfach Spezialfälle der "geometrischen". Der *"Barriers"*-Ansatz jedoch entkoppelt die linguistischen Relationen auf Kategorien von den geometrischen Relationen auf Knoten in einer Weise, die sich auf einem Standardbaum schwer darstellen lässt.
Kolb [25] zeigt nun, dass "Dominanz" zwischen Knoten nicht in der gebräuchlichen Art zu einer Relation auf Kategorien generalisiert werden kann. Vielmehr hat die geometrische Relation zwei Entsprechungen in der Welt der Kategorien, nämlich diejenige, die *alle* Segmente, d.h. Knoten, einer Kategorie in ihrem Bereich fordert ('I-Dominanz'), und die, für die jeweils ein Segment genügt ('C-Dominanz'). Beide Relationen beziehen auch nicht Knoten aufeinander, sondern sind strikt auf Kategorien beschränkt. Als Entsprechungen der geometrischen "Dominanz"-Relation und als Träger wichtiger abgeleiteter Relationen wie etwa C-Kommando, die ursprünglich über die geometrische Relation definiert sind, ist zu fordern, dass beide die definierende Eigenschaft dieser Relation erben: Beide sollten Baumordnungen sein. Das ist aber auf ein und demselben Baum unmöglich. Es ist allerdings nicht schwer, vom geometrischen "Dominanz"-Baum auf eindeutige Weise einen "I-Dominanz"- und einen "C-Dominanz"-Baum abzuleiten, die die jeweiligen Dominanzverhältnisse zwischen Kategorien—nun durchaus wieder im Sinne der geometrischen Relation—direkt repräsentieren. Dies geschieht durch Einschränkung der "Dominanz"-Baumordnung auf je einen Repräsentanten jeder Kategorie: das geometrisch tiefste Segment im Fall der "I"-Dominanz, das geometrisch höchste im "C"-Fall.
Der umgekehrte Weg, vom $\langle I, C\rangle$-Baumpaar zum geometrischen Baum mit Kategorie/Segment-Unterscheidung ist auf interessante Weise unterdeterminiert: Die relative hierarchische Ordnung multipler Adjunkte an ein und derselben Kategorie ist in $\langle I, C\rangle$ nicht repräsentiert. Für jedes $\langle I, C\rangle$-Paar erfüllt dann auch eine unendliche Äquivalenzklasse [T] von Bäumen die relevanten Beschränkungen. Alle Elemente von [T] sind allerdings isomorph bezüglich Relationen zwischen Kategorien—und da nach Chomsky [6] nur diese *linguistische* Entitäten sind, kann gefahrlos (und effektiv) das kleinste Element als Repräsentant von [T] gewählt werden.
Damit ist allerdings auch gezeigt, dass selbst die Rolle als Hilfsbegriff, die die geometrische "Dominanz"-Relation seit der Einführung der Kategorie/Segment-Unterscheidung in der linguistischen Argumentation spielt, unnötig—und unter

"minimalistischen" Vorzeichen damit unerwünscht ist: Der geometrische Baum enthält "zu viel Information" über abstrakte Knoten, die keine linguistische Rolle haben. Kolb [25] schlägt denn auch vor, den kombinierten Segment/Kategorie-Baum ganz zu eliminieren, und stattdessen das $\langle I, C\rangle$-Baumpaar als Träger der strukturellen Eigenschaften syntaktischer Repräsentationen zu installieren.

Zu einer ähnlichen Konklusion kommen unabhängig auch Grefe&Kracht [17]. Allerdings identifizieren sie "C-Dominanz" und die "geometrische" Dominanzrelation, wodurch sie zwar einerseits den ursprünglichen Segment-Baum zurückerhalten, andererseits aber die linguistisch erwünschte Beschränkung auf Kategorien nicht leisten können. Sie zeigen allerdings überzeugend, dass auch auf derartigen hybriden Bäumen syntaktische Domänen mit den gewünschten Eigenschaften definiert werden können.

Chomskys [9] "Minimalisisches Programm" scheint auf den ersten Blick viele Axiome der generativen Linguistik der letzten zwanzig Jahre in Frage zu stellen. Grund genug, um es in Zusammenarbeit mit den anderen syntaktisch orientierten Projekten A3 und A7 einer kritische Untersuchung zu unterziehen. In Kim, Kolb, Müller&Sternefeld [22] werden die Grundlagen aufgearbeitet und diverse Schwachpunkte aufgezeigt, die dann auch zu einer substantiellen Revision des Programms in Chomsky [10] geführt haben. Diese Zusammenarbeit wurde im Folgenden in Form einer gemeinsamen Arbeitsgruppe der drei Projekte bis heute fortgesetzt.

Cornell [12] beschäftigt sich mit der generativen Komplexität des minimalistischen Programms. Ein derivationelles System wird präsentiert, das sich weitgehend an Stablers Formalisierung [46, 47] (einiger Ideen) des minimalistischen Programms orientiert. Zweck der Übung ist zu zeigen, dass selbst eine extrem eingeschränkte Version dieses Systems Grammatiken erlaubt, die ausserhalb der kontextfreien Sprachklasse fallende Sprachen generieren—in diesem Fall die Sprache $\mathcal{L}_{\mathrm{ww}} = \{\mathrm{ww} | \mathrm{w} \in \{0,1\}^*\}$.
Wiewohl eine Kunstsprache, ist sie doch die Basis der Beweise in Huybregts [20] und Shieber [45] für die Nicht-Kontext-Freiheit natürlicher Sprache. Auch kann für sie eine vollständige Grammatik angegeben werden, die sie beweisbar generiert. Obwohl der Formalismus also beweisbar stark genug ist, um Grammatiken für (zumindest einige) mild kontext-sensitive Sprachen zu erlauben, bleibt er doch unter Verarbeitungsaspekten attraktiv. Insbesondere sein im Prinzip linearer Verbrauch lexikalischer Resourcen macht ein Entscheidbarkeitsresultat wahrscheinlich. Dies wiederum könnte zu einem Verfahren führen, mit den entscheidbare Logiken wie etwa die von Rogers [42] oder Kracht [28] über ihre heutige strikt kontextfreie Ausdruckskraft hinaus erweitert werden können, ohne die Entscheidbarkeitseigenschaft zu verlieren.
Uns geht es aber zunächst um die Bestimmung des minimalen technischen Apparats, der für die Ableitung von Sätzen der Sprache $\mathcal{L}_{\mathrm{ww}}$ in einem MERGE & MOVE

Kalkül à la Chomsky benötigt wird. Es zeigt sich, dass tatsächlich nur sehr wenig nötig ist. Insbesondere können wir auf sukzessiv zyklische Bewegung verzichten—Kopfbewegung reicht aus. Das wiederum bedeutet, dass wir im "linearen" Fragment des minimalistischen Systems arbeiten können: Jede Bewegungsoperation konsumiert Resourcen und keine Operation fügt Resourcen hinzu. Auch ist nur ein Merkmalstyp nötig, die kategorialen Merkmale—doch sicherlich die minimalen minimalistischen Merkmale—und ihr Widerpart auf der *checking*-Seite, die Selektionsmerkmale. Wir arbeiten also im Prinzip mit einer kategorialen Grammatik mit Bewegung.

Aus einer an der Programmiersprachensemantik orientierten Perspektive erscheint der seit Jahrzehnten in der linguistischen Literatur tobende Streit zwischen "Repräsentationalisten" und "Derivationalisten" leicht—manchmal etwas zu leicht, wie ein Blick in Abschnitt 5 und den Fortsetzungsantrag des Projekts zeigt—als ein Streit um des Kaisers Bart. Derivationelle und repräsentationelle Präsentationen einer syntaktischen Theorie sind schliesslich im Prinzip äquivalent. Dass hier dennoch Raum für Argumentation bleibt, zeigen Kolb [24] und Cornell [13].

Kolb [24] diskutiert die Identifikation des Gegenstands der generativen Linguistik, *I-Sprache*, mit einer generativen Prozedur, wie sie Chomsky in einer Reihe von Beiträgen (etwa [8, 11, 9]) vorgeschlagen hat. Gezeigt wird, dass dieser Ansatz in Verbindung mit der weitverbreiteten Ansicht, dass die *Funktion* I-Sprache nicht notwendigerweise rekursiv ist, unhaltbar ist.
Wenn nämlich UG die Spezifikation jeder rekursiv aufzählbaren Menge von Strukturbeschreibungen erlaubt, erlaubt es auch die Spezifikation von—im mathematischen Sinn—unentscheidbaren Sprachen. Da UG selbst eine I-Sprache ist (nämlich S_0) incorporiert sie unter diesem Ansatz eine generative Prozedur. Daraus folgt unmittelbar, dass jede ausgereifte (I-)Sprache entweder ein computationelles System von Turing-Kraft erbt, also selbst unentscheidbar wird, oder von einem solchen abgeleitet ist. Chomsky verweist in diesem Zusammenhang auf *"computational tricks"*, die dafür sorgen, dass *"unusable parts of the language are simply not used"* [7, p. 49]. Aber in einem algorithmisch spezifizierten, prozeduralen System gibt es im Prinzip nur zwei Wege mit "computationellen Tricks" zu arbeiten: direkter Einbau in den Algorithmus oder Trennung von Spezifikation und Verarbeitung.
Im ersten Fall wäre es die Aufgabe des Sprachlernsystems, die generative Prozedur im Übergang von S_0 in den stabilen Zustand so zu verändern, dass genau die Restriktionen eingeführt werden, die die Komplexität der Prozedur hinreichend reduzieren, aber gleichzeitig sicherstellen, dass die korrekte Sprache generiert wird. Das ist aber nur unter zwei Voraussetzungen möglich: Entweder das Sprachlernsystem hat vom jeweiligen I-Sprachzustand S_n unabhängige Information über die Zielsprache zur Verfügung—dann ist I-Sprache als theoretische

Entität unabhängig von Sprachlern- und -verabeitungssystemen redundant und überflüssig; oder die komplexitätsreduzierenden Faktoren sind universell—dann induzieren sie eine obere Schranke für die Komplexität *jeder* I-Sprache und sollten demnach schon in S_O enthalten sein.
Aber auch die Trennung der "reinen" generativen Prozedur I-Sprache vom trickreichen unabhängigen Verarbeitungssystem bringt keine Lösung: Um nämlich die *computational tricks* überhaupt irgendeine Wirkung zeitigen zu lassen, müsste das Verarbeitungssystem Strukturbeschreibungen zumindest partiell selbst aufbauen. Die I-Sprache wäre reduziert zu einer "Subroutine" des Verarbeitungssystems. Das Entscheidbarkeitsproblem impliziert dann, dass die (partiellen) Strukturbeschreibungen, auf denen I-Sprache operiert, strikt wohlgeformt sein müssten. Wiederum muss das Verarbeitungssystem entweder die I-Sprache selbst oder eine unabhängige Spezifikation der verarbeiteten Sprache inkorporieren, die konzeptuelle Unterscheidung zwischen Kompetenz und Performanz verschwindet, das Forschungsprogramm der generativen Grammatik wird ernsthaft unterminiert.
Akzeptiert man also das—selbst alles andere als unumstrittene—Postulat, dass I-Sprache rekursiv aufzählbar ist, dann ist das Studium von I-Sprache unabhängig von E-Sprache und Verarbeitungsmechanismen nur sinnvoll, wenn sie nicht als generative Prozedur, sondern als abstrakte Spezifikation verstanden wird. Nur diese Sicht der Dinge lässt die Möglichkeit offen, eine wohldefinierte Beziehung herzustellen zwischen der vollständigen, aber möglicherweise nicht entscheidbaren Axiomatisierung der I-Sprache und dem computationell traktablen, aber unvollständigen Verarbeitungssystem.

Cornell [13] untersucht detailliert die Strukturen, die nötig sind, um das Lokalitätsproblem für X_0-Ketten in einem repräsentationellen Rahmen zu lösen. Dies geschieht in einer Weise, die die starken Verbindungen dieses Ansatzes mit der derivationellen Charakterisierung von $\mathcal{L}_{\text{ww}}$ in [12] herausarbeitet. Diese Parallelen deuten darauf hin, dass ein Äquivalenzbeweis für einen (entsprechend eingeschränkten) minimalistischen Kalkül durchaus im Bereich des Möglichen liegt. Der Schlüssel zu diesem Problem ist natürlich die Relation zwischen der Operation MOVE und der Definition wohlgeformter syntaktischer Ketten. Wenn wir zeigen können, dass $\mathcal{L}_{\text{ww}}$ durch ein System von Wohlgeformtheitsbedingungen charakterisiert werden kann, wobei insbesondere die Bedingungen bezüglich X_0-Ketten von Interesse sind, so sind wir in einer guten Position, um einen deduktiven Kalkül zu entwerfen, für den der "derivationelle Minimalismus" die Beweistheorie und der "repräsentationelle Minimalismus" die Modelltheorie liefert.

Schwache Logiken und algebraische Sprachtheorie

Linguistische Aspekte. Die monadische Logik zweiter Ordnung erfreut sich in jüngster Zeit einer verstärkten Aufmerksamkeit in der theoretischen Linguistik. Die beschreibungstechnischen Mittel dieser Sprache erlauben es, zentrale Begriffe der generativen Theorie Chomskys zu erfassen.

Aber auch grammatische Theorien wie etwa GPSG, die gemeinhin als formal besser erforscht gelten, können von einer MSO-basierten Formalisierung profitieren. So zeigt Rogers [43], dass zwei notorisch schwierig zu fassende GPSG-Konstrukte, nämlich *Feature Specification Defaults (FSD)* und die *Exhaustive Constant Partial Ordering (ECPO)*-Eigenschaft, in MSO eine natürliche Behandlung erfahren, die gerade von der Zweitstufigkeit der Logik abhängt. Die Möglichkeit, über Mengen von Knoten zu quantifizieren, wird ebenso ausgenützt, um die FSDs als statische Eigenschaft von syntaktischen Bäumen zu erhalten, wie sie die Formalisierung von ECPO als universales Prinzip natürlicher Sprache—also als Eigenschaft der Klasse von Bäumen, die natürliche Sprache analysieren—zulässt.

Angesichts des bekannten Resultats, daß genau die kontextfreien Sprachen durch die monadische Logik zweiter Ordnung mit mehreren Nachfolgern charakterisiert sind, stellt sich allerdings das Problem, wie durch eine Erweiterung des logischen Apparats nicht-kontextfreie Phänomene wie Kreuzabhängigkeiten, die in natürlichen Sprachen gut belegt sind, zu beschreiben sind. Ein erster Schritt zur Lösung dieses Problems wird in Moennich [33] im Rahmen freier algebraischer Theorien gemacht.
In diesen algebraischen Theorien werden Operationen höherer Ordnung eingeführt, die die Operationen der ursprünglichen Algebra als Argumente nehmen. Auf diese Weise werden auch mehrstellige Funktionenvariablen, die in einem grammatischen Produktionssystem oder in einem äquivalenten Gleichungssystem schrittweise präzisiert werden, zu Argumenten der neuen Operationen höherer Ordnung. Anschaulich gesprochen bedeutet dies, dass im Fall von baumartigen Datenstrukturen bei dem schrittweisen Präzisierungsprozess nicht mehr ganze Teilbäume umstrukturiert werden, sondern dass nur noch einzelne variable Blätter durch Teilbäume grösseren Informationsgehalts ersetzt werden. Wir haben es hier daher mit einem Verfahren zu tun, das bei monadischen Bäumen zu den vertrauten rechts- oder linkslinearen Grammatiken und bei den Verkettungsstrukturen zu den kontextfreien Grammatiken in Chomsky-Normalform führt. Die oben erwähnte Äquivalenz zwischen der schwachen monadischen Logik zweiter Ordnung mit mehreren Nachfolgern und den kontextfreien Sprachen bezieht sich nun genau auf die Blattmengen von Baumfamilien, die durch dieses eingeschränkte Verfahren der Substitution von variablen Etikettierungen an Endknoten entstanden sind. Dies bedeutet m.a.W., dass auch die schrittweisen Auswertungen von mehrstelligen Funktionen, die sich der Umstrukturierung ganzer Teilbäume verdanken, nach einer Übersetzung in eine algebraische Theorie mit höherstufigen Operationen plötzlich als Werte von Argumentvariablen erscheinen und damit wieder in den Definitionsbereich der speziellen Logik zweiter Stufe geraten, mit der erfolgreich zentrale Begriffe der Prinzipien- und Parametertheorie haben expliziert werden können. Die Strukturzuweisungen zu den einzelnen Ketten einer

solchen Sprache sehen nach ihrer Übersetzung in das höherstufige Vokabular der algebraischen Theorie wesentlich anders aus, als dies in ihrer nicht-kontextfreien Inkarnation der Fall ist. Dabei werden jedoch die Beziehungen, die zur Nicht-Kontextfreiheit führen, nicht etwa als solche beseitigt, sie erscheinen nur in einem anderen Gewand. Technisch ausgedrückt, bedeutet dies im einzelnen, dass zwischen den "höheren" und den intendierten Strukturen ein eindeutig bestimmter Homomorphismus besteht und dass in der anderen Richtung konfigurationelle Verhältnisse auf den intendierten Strukturen—zumindest soweit sie sich durch endliche Baumautomaten überprüfen lassen—in einem effektiv angebbaren Verfahren auch auf den höheren Strukturen wiederfindbar sind. Der Homomorphismus, dessen eindeutige Existenz gewährleistet ist, ist allerdings an eine Berechnungsprozedur gebunden, für die sich in der theoretischen Informatik nach einem Vorschlag von Plotkin die Bezeichnung *call by value* eingebürgert hat. Dies ist der Grund dafür, dass im Ausgang von der Signatur von Halbgruppen wir als homomorphe Bilder der höheren Strukturen nicht die indizierten, sondern Fischers [16] IO-Macrosprachen erhalten. Für linguistische Belange ist dieses Resultat jedoch mit keinen Einschränkungen verbunden, da alle bekannten nicht-kontextfreien Phänomene natürlicher Sprachen im Spektrum der schwachen und starken Kapazität von Fischers Sprachfamilie liegen.[2]

Das erstaunliche Resultat aus Mönnich [33] bildet den Ausgangspunkt von Kolb [26], wo das dort beschriebene Verfahren an einem protoypischen nicht-kontextfreien Phänomen natürlicher Sprache, nämlich dem schweizerdeutschen Verbkomplex (vergl. Huybregts [20]) erprobt wird. Diese Konstruktion ist insbesondere im Vergleich mit ihren niederländischen und deutschen Entsprechungen von besonderem Interesse. Obwohl letztere nämlich als Kettensprachen kontextfrei (nämlich von der Form $a^n b^n$ bzw. ww^{-1}) sind, deutet alles darauf hin, dass ein systematischer Zusammenhang zwischen den syntaktischen Strukturen der drei Konstruktionen besteht und dass diese in allen drei Fällen keine *erkennbare* Menge von Bäumen bilden. Diesen Zusammenhang transparent machen und die jeweiligen Unterschiede mit einfachen Mitteln ableiten zu können, ist eines der überzeugenderen Argumente für eine nicht-konstruktionsgebundene generative (Transformations-)Grammatik.
Kolb [26] zeigt nun anhand einer (vereinfachten) minimalistischen Analyse dieser Konstruktionen, dass die 'natürlichen' Strukturbeschreibungen dieses Phänomens ebenso wie ihre systematische Variation durch eine Macro-Baumgrammatik erfasst werden können. Die durch die mittels regulärer Kodierung dieser Macro-Grammatik abgeleitete Grammatik generierten Bäume bilden definitionsgemäss eine *erkennbare* Menge und sind somit durch eine Menge von MSO-Formeln—und damit konstruktionsunabhängig—zu beschreiben. Allerdings enthalten diese

[2] ...mit der einen möglichen, aber nicht wahrscheinlichen Ausnahme eines obskuren skandinavischen Extraktionsphänomens, das laut Miller [30] selbst ausserhalb der indizierten Sprachen liegt.

Formeln drei Sorten von Konstanten, die nicht standardmässig zum Inventar von P&P- oder minimalistischen Analysen gehören, nämlich die operativen Funktionssymbole der Baumgrammatiken $\mathbf{F}$, sowie die Hilfssymbole der regulären Kodierung: die Familie der "Kompositionssymbole" $\mathbf{C}$ und die Familie der "Projektionssymbole" Π.
Um aber die so erweiterte MSO-Sprache transparent für die Formalisierung von linguistischen Analysen nutzen zu können, muss für diese zusätzlichen Primitive eine Interpretation in der linguistischen Theorie angegeben werden. Wir zeigen, dass die "offensichtliche" Analogie über Ableitungsbäume, wobei die $c \in \mathbf{C}$ als Instanzen von MERGE, die $F \in \mathbf{F}$ als Attraktor-Merkmale und die $\pi \in \Pi$ als *checking* Merkmale interpretiert werden, zwar erstaunlich weit trägt, aber dann problematisch wird, wenn Derivationen als linguistische Entitäten ernst genommen werden.

Verarbeitungsaspekte. Im Paradigma der Constraint–basierten Grammatiktheorien wird linguistische Verarbeitung durch Theorembeweisen in der Logik modelliert. Die Entwicklung des Prototyps MoSeL (Morawietz&Cornell [37]) basiert auf dem Entscheidbarkeitsbeweis für MSO und implementiert damit zwar kein Standardverfahren, wie z.B. *Tableaux*-Methoden oder Modellkonstruktion, zum Theorembeweisen, erlaubt aber insofern automatisches Theorembeweisen für auf endliche Mengen beschränkte monadische Prädikatenlogik 2.Stufe—und damit das Testen von Formalisierungen, als aus MSO Formeln Baumautomaten generiert werden und überprüft wird, ob die von ihnen erkannte Sprache nicht leer ist.
Konkret sind die Basisrelationen der Signatur der monadischen Logik 2.Stufe in einer Datenbank als Baumautomaten repräsentiert. Entsprechend dem Formelaufbau werden diese Basisautomaten dann mit den entsprechenden Automatenkonstruktionen zu neuen Automaten verbunden. Diese neuen Automaten können dann zur Datenbasis hinzugefügt werden. Dadurch wird die Signatur der Logik faktisch erweitert, da diese Automaten nun als elementare Bestandteile in neuen Formalisierungen dienen. Durch diese inkrementelle Vorgehensweise ist es auch möglich, solche Formeln schrittweise zu kompilieren, die aufgrund ihrer Komplexität nicht in einem Schritt übersetzt werden konnten.
Weiterhin erlaubt die Implementierung, die erzeugten Automaten als Generatoren zu benutzen. Dadurch können zu testende Constraint-basierte Module und Prinzipien mit den entsprechenden erlaubten Strukturen/Bäumen visualisiert werden. Mit der Entwicklung dieses Werkzeugs ist die praktische Nutzung von MSO als Formalismus für die Computerlinguistik demonstriert worden. (Linguistische) Primitive können definiert und generiert werden, sind dann aber auch direkt und bedeutungsvoll weiter nutzbar.

Trotz der nicht elementaren Komplexität dieses Verfahrens folgt unsere Prolog–Implementation dem induktiven Entscheidbarkeitsbeweis von Rabin/Doner/Thatcher&Wright für MSO auf finiten Mengen. Möglichkeiten, die kombinatorischer Explosion in der Implementation zu beschränken, gibt es, wurden aber bisher durch uns erst zum Teil ausgeschöpft. Eine solche Möglichkeit ist die Verwendung von Knotennamen mit interner Struktur, die es erlauben, linguistisch signifikante Generalisierungen über syntaktische Merkmale und Prinzipien direkt zur Reduziering der Transitionstabellen der entsprechenden Automaten zu gebrauchen. Allerdings komplizieren derartige Verfahren die logische Formalisierung einer Grammatik—und den Nachweis ihrer Korrektheit—sehr. Ein spezielles Werkzeug für den Umgang mit dieser Komplexität ist das LISA-System, das derzeit in Saarbrücken entwickelt wird (vergl. Niehren&Podelski [38]). Es kompiliert Spezifikationen in einer merkmalsbasierten Beschreibungssprache direkt in MSO-Formeln, die Merkmalsattribute enkodieren. D. Basin und A. Podelski, die Autoren dieses Systems, das bisher hauptsächlich in der Hardware-Verifikation Verwendung gefunden hat, haben uns ihre Prototypen zur Verfügung gestellt und entwickeln eine spezielle Version zur leichteren Integration mit MoSeL.

Doch abgesehen von diesen praktischen Gesichtspunkten haben die Baumautomaten auch in theoretischer Hinsicht zum Erkenntnisgewinn beigetragen. Wie in Morawietz&Cornell [36] diskutiert, repräsentieren die Automaten sehr kompakt unser Wissen über die beschriebenen Strukturen. Die zu charakterisierende unendliche Menge von Modellen/Bäumen wird mit endlichen Mitteln dargestellt. Und die Konstruktionen, die die Automaten zu "größeren" Automaten verbinden, repräsentieren intuitiv die Art und Weise, dieses Wissen zu kombinieren. So stellt z.B. die Konjunktion der logischen Beschreibungen in der Intuition den Schnitt der charakterisierten Mengen von Bäumen dar. Genau dies wird durch die Konjunktion der Automaten erreicht. Insbesondere stellt auch die Minimisierung eine Generalisierung über die Information dar, die durch die Verbindung von Einzelautomaten entstanden ist.
Interessant ist auch der den Automaten zugrunde liegende Domänenbegriff. Es war ein Ziel dieser Projektphase, für die Stärke der P&P-Theorien, nämlich die direkte Spezifikation von nicht lokalen Beschränkungen, trotzdem einen Begriff einer syntaktischen Domäne zu entwickeln. Wie auch in MSO können die Automaten Relationen zwischen nicht benachbarten Knoten, wie z.B. *c–Kommando*, repräsentieren. Die Domänen, die die Automaten manipulieren, sind also ebenfalls *erkennbare Mengen* von Knoten. Erkennbarkeit ist gerade so definiert, daß eine Menge genau dann erkennbar ist, wenn sie von einem (Baum) Automaten erkannt wird. Dadurch können jedoch nicht beliebige Mengen als Domänen fungieren. So ist es z.B. nicht (direkt) möglich eine Menge von Bäumen zu erkennen, deren *yield* den sogenannten Kopiersprachen entspricht.

Aber auch in diesem Zusammenhang stellt sich das Problem, die Limitationen der monadischen Prädikatenlogik 2. Stufe zu überwinden. Dazu sollen u.a. Techniken der Constraint–Logikprogrammierung nutzbar gemacht werden. So gibt uns

die Definition eines Constraint-Logiksystems die Wiedereinführung eines (klassischen) Ableitungsbegriffs in Verbindung mit einer denotationellen Interpretation. Gleichzeitig stellen induktive bzw. rekursive Definitionen eine Möglichkeit zur Erweiterung der generativen Kapazität einer Logik dar.
Morawietz [35] geht einen ersten Schritt in diese Richtung. Wir betten die monadische Prädikatenlogik 2.Stufe in das von Höhfeld&Smolka [19] entwickelte Constraint-Logik Schema ein. Die entstehende relationale Extension $\mathcal{R}(\mathcal{MSO})$ hat eine semantisch widerspruchsfreie und vollständige operationale Semantik. Wir nutzen dabei die oben diskutierte kompakte Repräsentation von Lösungen von Constraints/Formeln in monadischer Prädikatenlogik 2.Stufe mit Baumautomaten. Unendliche Mengen von Lösungen werden durch die Normalform Baumautomat repräsentiert. Diese Normalform erlaubt es uns, die monadische Logik in ein Constraint-Logiksystem effizient einzubetten, da sowohl der Constraint-Speicher, als auch die mit definiten Klauseln assoziierten Constraints kompakt repräsentiert werden. Die eigentliche *Black Box* des Constraint-Lösers wird durch die Theorembeweiser wie z.B. MONA [18] oder MoSeL, vollständig spezifiziert, da sie die Desiderata der Inkrementalität und Korrektheit, die standardmässig für Constraint-Löser gefordert werden, trivialerweise erfüllen.
Wir demonstrieren die Anwendbarkeit des so definierten Systems mit Beispielen aus dem Bereich des P&P-basierten Parsings. Wir sehen den Vorteil dieses Ansatzes darin, daß Variablen nun nicht mehr auf einem Herbrand Universum von Termen interpretiert werden, sondern direkt auf Bäumen. Gerade dies gibt uns wiederum die Möglichkeit die Stärke der P&P Theorien in der Spezifikation nicht lokaler Bedingungen auf Strukturen möglichst direkt auszunutzen. Wir müssen keine Kodierung solcher Relationen vornehmen, sondern können die gewünschten Einschränkungen direkt realisieren.
Leider erlaubt die volle Extension $\mathcal{R}(\mathcal{MSO})$ die Spezifikation von unentscheidbaren Problemen. Da sie allerdings semi–entscheidbar ist, kann man sie als eine (Constraint) Logische Programmiersprache mit linguistischen Primitiven betrachten, die in ihren formalen Eigenschaften Prolog ähnelt. Die offensichtliche Einschränkung auf Individuenvariablen, anstatt von beliebigen monadischen Variablen zweiter Ordnung, in allen direkt oder indirekt rekursiven Klauseln ist entscheidbar, aber führt zu keiner Erweiterung der generativen Kapazität. Alles, was auf diese Art induktiv definiert werden kann, kann explizit in Variablen zweiter Ordnung ausgedrückt werden. Es ist ein offenes Problem, eine entscheidbare Klasse von induktiv definierten Relationen zu finden, die die generative Kapazität der Logik erweitert. Bisher haben wir mit den definierten Systemen die obere Grenze – die volle Extension – und die untere Grenze – die auf Individuenvariablen eingeschränkte Extension – aufgezeigt.

Ein interessanter Zusammenhang lässt sich zwischen dem System $\mathcal{R}(\mathcal{MSO})$ und einer um Relationen erweiterten Merkmalslogik zeigen. Die Erweiterung von Kings [23] Merkmalslogik für HPSG um die Möglichkeit, (induktive) Relationen zu definieren, war das Thema einer gemeinsame Arbeitsgruppe der Projekte A8 und B8, die innerhalb des SFB 340 über längere Zeit bestand. In dieser Logik haben die Relationen einen echten Objektstatus. Dadurch wird sowohl die unnatürliche technische Realisation derselben Relationen mit *Junkslots* vermieden, als auch ihr ontologischer Status geklärt. In diesem System, wie in $\mathcal{R}(\mathcal{MSO})$, können in der Logik selbst "Programme" geschrieben werden, die einen Teil der sonst nötigen Kodierungsarbeit direkt repräsentieren.

Eine weitere gemeinsame Arbeitsgruppe gab es mit Teilprojekt B4. In dieser Gruppe wurden gemeinsam Techniken der (Constraint-) Logikprogrammierung und ihr Nutzen für die Computerlinguistik erarbeitet. Dabei wurde ein besonderer Schwerpunkt auf Entscheidbarkeitsfragen gelegt. So führte die Analyse der Konstruktion eines exhaustiven Modells für eine Merkmalsdeskription durch eine Fixpunktkonstruktion in der Bereichstheorie zur Isolation des Problems im Entscheidbarkeitsbeweis, so daß anschließend Lösungsansätze zum Beweis der Semi-Entscheidbarkeit erarbeitet werden konnten.

4 Vergleiche mit Arbeiten außerhalb des Sonderforschungsbereichs und Reaktionen der wissenschaftlichen Öffentlichkeit auf die eigenen Arbeiten

Ein Forschungsprogramm, das sich in zentralen Bereichen der Techniken und Methoden der Logik, der algebraischen Sprachtheorie und der Bereichstheorie in einer nicht alltäglichen Kombination bedient, weist naturgemäss Berührungspunkte mit der Arbeit einer ganzen Reihe anderer Gruppen auf, die ebenfalls die formalen Grundlagen grammatischer Systeme untersuchen.

Projekt A8 hat aktiv und mit Erfolg versucht, diesen Umstand für die eigene Forschung nutzbar zu machen. So boten nicht nur die beiden durch A8 organisierten Workshops "Tag des Baumes" (Tübingen, März 1995) und "The Mathematics of Syntactic Structure" (Prag, August 1996) rege genutzte Gelegenheiten des internationalen Gedankenaustausches; auch die Tatsache, dass sich viele exponierte, an ähnlichen Fragestellungen arbeitende Forscher zu Gastaufenthalten (Blackburn, Keenan, Kracht, Rogers, Rounds) bzw. Vorträgen (Johnson, Stabler, Volger, u.a.) bei Projekt A8 aufgehalten haben, zeigt ebenso wie das Tutorium zur modelltheoretischen Syntax, das J. Rogers und T. Cornell auf Einladung der ACL bei ACL/EACL'97 halten werden, dass das Projekt sich in den letzten Jahren zu einem Kristallisationspunkt derartiger computerlinguistischer Grundlagenforschung entwickelt hat.
Eine Auswahl der Früchte der Zusammenarbeit mit an ähnlichen Fragestellungen arbeitenden Gruppen wird in dem in Vorbereitung befindlichen Buch *The Ma-*

thematics of Syntactic Structure (Hg. H.-P. Kolb und U. Mönnich) dokumentiert werden.

Besonders enge Verbindungen bestehen zu Projekten, die sich mit der Nutzbarmachung von schwachen Logiken für Syntaxtheorien der natürlichen Sprache beschäftigen. Zu nennen sind hier insbesondere das unter der Leitung von H. Volger stehende Projekt zur Entwicklung der erst-stufigen Constraint-Sprache *Clat*, die modallogischen Ansätze im Umkreis von M. Kracht bzw. P. Blackburn, und das Projekt von J. Rogers, das auf MSO aufbaut.
Die Stossrichtung dieser Ansätze ist durchaus verschieden. So ist das Forschungsinteresse des *Clat*-Projekts auf die Entwicklung einer Constraintsprache nicht nur für die Formalisierung sondern explizit auch für die Lösung des Parsingproblems von PP-Theorien gerichtet. Rogers' Interesse gilt besonders deskriptiven Komplexitätsresultaten für Syntaxtheorien natürlicher Sprache, und hier insbesondere für die PP-Theorien, für die solche Resultate notorisch schwer zu erzielen sind. Die modallogischen Ansätze wiederum fokussieren auf die Vergleichbarkeit von Grammatikmodellen und auf die Übertragbarkeit von syntaktischen Analysen via elaborierte Merkmalsstrukturen.
Entsprechend benützt die Volger-Gruppe mit *Clat* eine Logik, die sowohl spezifisch auf den Ausdruck modularer Prinzipien-Grammatiken im Sinne der *Government/Binding*-Theorie als auch auf die Erfordernisse effizienten Parsings derartiger Grammatiken zugeschnitten ist, während sowohl Rogers' Resultate als auch die der modallogischen Ansätze ihr Interesse auch gerade daraus beziehen, dass sie mit völlig generellen, unabhängig begründeten Mitteln erzielt wurden.
Was allerdings ihre Ausdrucksstärke betrifft, sind alle diese formalen Systeme beweisbar äquivalent—nähmlich stark kontextfrei, was Kracht [28] für seine modale Orientierungssprache *Olt*, die mindestens so stark ist wie die anderen vorgeschlagenen modalen Sprachen (etwa [3]), ebenso nachgewiesen hat, wie Rogers [42], basierend auf Rabin [41] u.a., für MSO und Palm [40] für *Clat.* Aus diesem Umstand ergibt sich die Notwendigkeit aller dieser Ansätze, sich mit denselben Problemen zu beschäftigen, die auch einen guten Teil der Arbeit von Projekt A8 bestimmt haben, nämlich mit der Suche nach einer linguistisch adäquaten Extension des jeweiligen Formalismus ohne Verlust von Entscheidbarkeit und anderen angenehmen formalen Eigenschaften. Dies wiederum hat zu einem regen Ideenaustausch geführt, der eine Vielzahl von Resultaten inspiriert hat. Neben Mönnichs [33] innovativer Verwendung algebraischer Methoden zur Erweiterung der Ausdrucksstärke von MSO ohne eigentliche Veränderung der *Sprache* ist hier insbesondere Rogers' [44] Behandlung generalisierter lokaler Mengen zu nennen, sowie Palms [40] Einführung von *Stacks* in *Clat*, die die Ausdruckskraft der Sprache bis zur Klasse der indizierten Sprachen steigert.

Ähnlich intensive Beziehungen bestehen zu dem Projekt einer abstrakten Syntax, in dessen Rahmen E. Keenan und E. Stabler seit einigen Jahren versuchen, mit algebraischen Mitteln den Begriff syntaktischer und semantischer Invarianten zu präzisieren. Nach dieser Theorie wird eine Sprache als eine Menge von komplexen Ausdrücken definiert, die mithilfe von generierenden Funktionen auf der Basis eines Lexikons aufgebaut werden. Die Elemente des Lexikons werden einzelnen Kategorien zugeordnet, und für die Argumente und Werte der generierenden Funktionen gilt ebenfalls, dass sie jeweils festgelegten Kategorien angehören müssen. Wenn man die Kategorien als eine Menge von Sorten auffasst und den Elementen des Lexikons Symbolkonstanten und den generierenden Funktionen Operatoren der vorgegebenen sortalen Spezifikation zuordnet, erhält man eine Signatur und die im Sinne von Keenan und Stabler definierte Sprache als homomorphes Bild der (absolut) freien Algebra der so gewonnenen Signatur. Bei dieser Charakterisierung haben wir die Tatsache ausser Acht gelassen, dass die Definitionsbereiche der generierenden Funktionen in einigen Fällen über die sortale Korrektheit hinaus noch weiteren Restriktionen gehorchen müssen, der daraus resultierende Verlust von Allgemeinheit spielt jedoch in allen Anwendungen der abstrakten Syntax keine Rolle. Vor dem Hintergrund dieser algebraischen Charakterisierung ergeben sich einige entscheidende Unterschiede, die dazu geeignet sind, die methodischen Schwerpunkte im Projekt über syntaktische Domänen von den theoretischen Voraussetzungen der abstrakten Syntax abzuheben. Die fundamentale Differenz, auf die sich alle weiteren kontrastierenden Aspekte zurückführen lassen, besteht darin, dass Keenan und Stabler zunächst weder für die Menge der generierenden Funktionen und der Lexikonelemente irgendwelche Kardinalitätsbeschränkungen vorsehen, noch die Wirkungsweise der Funktionen durch berechnungstheoretische Überlegungen einzudämmen versuchen. Es ist unter diesen Umständen nicht weiter verwunderlich, dass es ihnen ohne weiteres gelingt, verschiedene konkurrierende linguistische Theorien in ihrem eigenen Ansatz zu repräsentieren, indem sie für das Lexikon, die Sortenmenge und die generierenden Funktionen jeweils entsprechende Festlegungen verfügen. Für den Bereich der syntaktischen Domänen sind wir genau den umgekehrten Weg gegangen, auf dem ein möglichst schwaches System nur dann in seiner Ausdrucksstärke erweitert wurde, wenn empirische Phänomene einen solchen Schritt erzwangen. Signaturen waren grundsätzlich endliche Mengen, und bei den Operationen haben wir uns auf Verkettung und Substitution beschränkt und sind davon nur in den Arbeiten abgewichen, die direkt an Vorarbeiten von Stabler anknüpfen. Parallel zu dieser sparsamen Verwendung algebraischer Mittel verlief die Kontrolle durch eine schwache logische Beschreibungssprache. Eine Berücksichtigung von Ergebnissen der Theorie deskriptiver Komplexität fehlt bei den Vertretern der abstrakten Syntax völlig, und wir sehen diese Verzahnung von algebraischen und logischen Methoden als einen wesentlichen Fortschritt an. Schliesslich macht der Ausschluss von anderen Operationen als Verkettung und Substitution es überhaupt erst möglich, sich bei der Untersuchung von linguistischen Problemen auf die Ebene der Terme oder der zu ihnen isomorphen Bäume zu konzentrieren. Der Homomorphismus von der symbolischen Ebene in eine "semantische" Ebene, auf der die einzelnen generierenden Funktionen die ihrer intendierten Bedeutung entsprechenden Operationen ausführen, wird unter den skizzierten sparsamen Voraussetzungen überflüssig. Ungeachtet dieser grundsätzlichen Unterschiede haben insbesondere Stablers Arbeiten zum Minimalismus [47] zu einem äusserst fruchtbaren Dialog zwischen den Projekten zum Verhältnis zwischen repräsentationellen und derivationellen Syntaxmodellen geführt, der sich u.a. in Cornell [12, 13] und Stabler [48] niederschlägt. Auch ist zuzugeben, dass bei dem augenblicklichen Stand der Kenntnisse kein endgültiges Urteil darüber gefällt werden kann, ob ungelöste empirische und theoretische Probleme uns nicht zwingen werden, Zuflucht zu einigen der starken ontologischen Annahmen des kalifornischen Programms zu nehmen.

Ein Ansatz, dessen Relevanz für unser Projekt mit der geplanten Erweiterung auf die Behandlung auch von Derivationen noch zunehmen wird, ist die deskriptionslogische Behandlung von *Tree Adjoining Grammar*, wie sie u.a. von K. Vijay-Shanker verschiedentlich vorgeschlagen worden ist. Es ist zu erwarten, dass der rege Austausch, der sich u.a. aus der Diskussion von L. Kallmeyers [21] Behandlung von Derivationen in einer durch Äquivalenz-Constraints auf Knotennamen erweiterten Baumlogik entwickelt hat, weiter intensiviert werden wird.

Im Laufe der Arbeit hat sich erfreulicherweise herausgestellt, dass das Forschungsprogramm von Projekt A8 nicht allein durch linguistische Fragestellungen zu rechtfertigen ist, sondern auch zum Kern aktiv verfolgter Ziele in der theoretischen Informatik gehoert.

So arbeiten neben uns z.B. Gruppen in Aarhus, Freiburg, Kiel und Saarbruecken an Systemen der monadischen Logik zweiter Stufe. Am effizientesten und am weitesten entwickelt ist der automatische Beweiser MONA [18]. Seine Entwicklung und die in Aarhus angestrebte Erweiterung auf die von uns benötigten Bäume lässt darauf schließen, daß die kombinatorische Explosion in den relevanten Fällen kontrolliert werden kann. Die erreichte Kompression der Übergangstabellen der Automaten durch zum einen die Datenstruktur der *Binary Decision Diagrams* (BDD) [4] für die Repräsentation des Alphabets und zum anderen die Erweiterung auf *Guided Automata* zur Vermeidung der quadratischen Vergrößerung der Zustandstabellen bei (binären) Baumautomaten führt zu einem auch in der Informatik einsetzbaren Werkzeug. Wir hoffen, daß das auch in unseren linguistischen Anwendungen der Fall ist. Zur Zeit arbeitet eine Gruppe in Aarhus an der Extension von MONA von Zeichenketten zu den für uns relevanten Bäumen [2]. Eine Prototypimplementation auf Bäumen [49] existiert auch für das AMoRE System [29], entwickelt an der Universität Kiel.
Über die Zusammenarbeit mit D. Basin und A. Podelski in Saarbrücken bzw. Freiburg ist Projekt A8 in diesen Forschungsstrang fest eingebunden. Ihre eigenen Arbeiten zur Entwicklung der Metasprache Lisa erlaubt eine weitere Abstraktion bzw. Vereinfachung der zu spezifizierenden P&P Theorien. Lisa stellt eine Erweiterung von MONA mit Typdefinitionen und Record-artigen Strukturen dar.

Formeln im Lisa-Format können sowohl zu MONA-Formeln übersetzt, als auch direkt bewiesen werden. Diese effiziente Realisierung von Merkmalsstrukturen und Typen soll insbesondere den Vergleich und Transfer von HPSG und PP Theorien auf die gemeinsame Basis der monadischen Prädikatenlogik 2.Stufe ermöglichen – unter der Vorraussetzung, daß signifikante Teile der Theorien im Rahmen der HPSG auch in Merkmalsbäumen [38] ausgedrückt werden können. Gleichzeitig können wir von diesem Werkzeug Gebrauch machen, um eigene Formalisierungen im P&P Bereich vorzunehmen und zu testen, da besonders der erlaubte und wohldefinierte Einsatz von Typen eine Vereinfachung bzw. Präzisierung der Definitionen – und nicht nur eine Effizienzsteigerung – erwarten läßt. Ihre Kontakte zum MONA Projekt in Aarhus sichern uns den Zugang zu einem professionellen Theorembeweiser. Eine MoSeL-Schnittstelle ist in Entwicklung.

Weitere Kooperationen bestehen mit Mitarbeitern des Projekts *Construction of Computational Logics II* (ESPRIT Basic Research program) in München zur Erweiterung von Constraint–Systemen durch Amalgamierung verschiedener Constraint–Löser und des Projektes *CHORUS* des SFB 378 *Ressourcendaptive Kognitive Prozesse* in Saarbrücken zur Unifikation höherer Ordnung auf Bäumen.

Die Kooperation zur Kombination von Constraintlösern betrifft einige technische Probleme des automatischen Beweisers MoSeL. In MSO dekorieren monadische Variablen zweiter Ordnung die Knoten im Baum. Diese linguistisch nötigen Merkmale führen zu einer kombinatorischen Explosion des zu repräsentierenden Alphabets, da sie keinerlei Struktur enthalten. Eine mögliche Alternative besteht darin, komplexe Merkmalsstrukturen als Knotennamen zu verwenden. Die Probleme der Merkmalsstrukturen müssen dann separat in einem eigenen Lösungsverfahren bearbeitet werden. Dabei stellt sich offensichtlich die Frage, ob man die Lösungsverfahren und Constraint–Sprachen miteinander kombinieren kann, ohne die erwünschten Eigenschaften der MSO Logik zu verlieren. Erste Diskussionen zum Problem haben ergeben, daß die in München entwickelten Standardverfahren [1] nicht ausreichen, da sie für Constraint–Sprachen erster Ordnung definiert sind und nicht auf offensichtliche Weise für MSO adaptiert werden können. Der Stand im Moment ist, daß Existenzquantoren zweiter Stufe wohl erlaubt werden können, ohne die formalen Eigenschaften des Verfahrens zu verlieren. Das ist zwar nicht viel, schließt aber die wichtige Klasse der Σ_1^1 Formeln ein, mit denen sich unter anderem die Klasse der regulären Sprachen axiomatisieren lässt.

Die Unifikationsansätze zur Teilbaumsubstitution, die in Saarbrücken untersucht werden, haben eine enge Verbindung mit kontextfreien Baumgrammatiken und ihren denotationellen Interpretationen. Schon daher ist es interessant, den Zusammenhang zwischen einer mit der *equality-up-to* Relation [39] ergänzten Signatur erster Ordnung und MSO herzustellen. Es ist allerdings immer noch ein offenes Problem, ob equality-up-to Constraints in MSO definierbar sind.

5 Offene Fragen

Projekt A8 konzentrierte sich in der ersten Phase auf modelltheoretische Techniken für die Charakterisierung von Grammatiken. Dabei ist es gelungen, schwache Logiken zur Beschreibung syntaktischer Strukturen heranzuziehen und sie zu adäquaten Werkzeugen für die Konstruktion von Grammatikformalismen und die Analyse ihrer Modelle zu erweitern. Naturgemäss bleiben jedoch noch viele Fragen offen.

[Trans-CF:] Wir haben unserer Arbeit auf weiten Strecken wegen ihrer linguistischen Transparenz, ihrer berechnungstheoretisch vorteilhaften Eigenschaften und ihrer bereichstheoretischen Interpretierbarkeit die schwache monadische Logik zweiter Stufe mit mehreren Nachfolgerfunktionen zugrundegelegt. MSO-basierte Syntax ist jedoch strikt kontext-frei, d.h. unter ihrer natürlichen Interpretation denotiert jede Menge von MSO-Formeln eine *"erkennbare"* Menge von Bäumen (und umgekehrt). Die Blattsprache einer erkennbaren Menge von Bäumen aber ist immer kontextfrei. Zwar ist es uns durch eine Ergänzung von MSO durch Techniken der algebraischen Sprachtheorie gelungen, sprachliche Phänomene in unseren Beschreibungsrahmen einzubeziehen, die bisher ausserhalb des Anwendungsbereichs schwacher logischer Systeme zu liegen schienen, diese Erweiterung benützt jedoch formale Mittel, deren *linguistische* Adäquatheit nicht *a priori* feststeht. Ungeachtet der ersten ermutigenden Resultate wird hier noch weitere Arbeit zu leisten sein.

Gleichzeitig müssen natürlich auch andere Verfahren zur Überwindung der CF-Barriere weiter untersucht werden. Insbesondere das Problem, linguistisch adäquate induktive Definitionen auf entscheidbare Weise in das formale System zu integrieren, ist von hoher Priorität. Neben der Frage nach der Möglichkeit einer entsprechenden Erweiterung der zugrundeliegenden Logik, sind vorallem hybride Systeme bisher noch nicht ausreichend verfolgt worden. So ist die Nutzbarmachung von Techniken der Constraint–Logikprogrammierung in $\mathcal{R}(\mathcal{MSO})$ [35]—besonders auch im Hinblick auf eine Kombination von operationellen mit denotationellen Techniken—noch nicht bis zum Ende ihres Potentials vorangetrieben, und auch zweistufige Ansätze wie die Kombination beliebiger, polynomial parsebarer Generationssysteme mit MSO–Filtern auf den entstehenden Strukturen sollten untersucht werden.

[Derivationen:] Eine modelltheoretische Behandlung syntaktischer Theorien mittels MSO scheint zu einer repräsentationellen, monostratalen Sicht der Dinge zwingen. Schliesslich sind die intendierten Modelle für (Mengen von) MSO–Formeln Bäume, was die angemessene Erfassung einer derivationellen Sichtweise, die Sprachstrukturen als Ergebnis schrittweiser transformationeller Ableitungen versteht, auszuschliesen scheint. Nun hat aber die derivationelle Sicht, die seit jeher in der generativen Grammatik virulent war, in letzter Zeit durch das *minima-*

listische Programm [10] wieder stärker an Bedeutung gewonnen, so dass auch eine formale Charakterisierung von grammatischen Derivationen und die Klärung des Verhältnisses zwischen Derivationen und Repräsentationen grössere Dringlichkeit gewinnt. Zwar weisen die bekannten Möglichkeiten der denotationellen Semantik, prozedurale Begriffe mit deklarativen Mitteln zu erfassen, auf eine Möglichkeit, dieses Problem anzugehen. Auch sind natürlich die Baumautomaten, die etwa MoSeL benützt, um MSO-Theoreme zu verifizieren, in gewissem Sinne operationelle Charakterisierungen deklarativen grammatischen Wissens. Dennoch scheint die Integration von derivationellen Konzepten in ein (erweitertes) MSO-basiertes System alles andere als trivial—nicht zuletzt auch deshalb, weil die Intuition des Linguisten, wie sie sich in den angenommenen Ableitungen ausdrückt, sich nicht automatisch gleichsetzen lässt mit der operationalen Semantik einer Programmiersprache.

[Graphen:] Das Derivationenproblem zeigt, dass die Beschränkung auf Baumstrukturen, die in vieler Hinsicht sehr hilfreich ist, durchaus auch negative Seiten hat. Eine kontrollierte Erweiterung auf allgemeinere Strukturen wäre wünschenswert—und die Arbeiten etwa von Courcelle oder Engelfriet zur MSO-Theorie von Graphen, aber auch unsere eigenen algebraischen Ergänzungen der MSO, deuten darauf hin, dass sie durchaus im Bereich des Möglichen liegt.

[Bereichstheorie:] Wie oben ausgeführt gibt unsere bereichstheoretische Aufarbeitung jenen Begriffen der linguistischen Theorie eine informationstheoretische Deutung, die sich mittels MSO explizieren lassen. Die Ausdehnung dieses Ergebnisses auf die notwendigen Erweiterungen des Systems sowohl im Hinblick auf seine generative Kapazität als auch bezüglich seiner Fähigkeit, syntaktische Derivationen zu erfassen, steht notwendigerweise noch aus. Auch eine direkte deklarative Interpretation der syntaktischen Objekte—also Strukturen und Derivationen—als unabhängige Basis für Untersuchungen zu Korrektheit, Universalität der Interpretation und Effizienz der Implementation ist noch zu leisten.

[Parsing:] Der von uns vorgeschlagene Ansatz einer Constraint-logischen Erweiterung der monadischen Logik 2. Stufe erlaubt auch eine neue Sicht auf des Parsingproblem der natürlichen Sprache. Da MSO-Ausdrücke nicht auf einem Herbrand–Universum, sondern direkt auf Bäumen interpretiert werden, kann die explizite Kodierung der gewünschten Datenstrukturen in Terme vermieden werden, was es schon dem MSO-Theorembeweiser MoSeL unter anderem erlaubt, nicht lokale Domänen direkt auszudrücken und sie als neue Primitive zu betrachten. $\mathcal{R}(\mathcal{MSO})$ fügt die Möglichkeit hinzu, sie mit Restriktionen zu versehen, so dass die vorhandenen Module einer syntaktischen Theorie separat behandelt und der Kombinationsprozeß gesteuert werden kann. Ob aber die Effizienz von Baumautomaten-basierten Theorembeweisern wie MoSeL—etwa durch Minimierung der Fallunterscheidungen durch Unterspezifikation der Repräsentation der möglichen Eingabesymbole—soweit zu verbessern ist, dass diese äusserst attraktiven Eigenschaften effizient ausgenützt werden können, ist eine offene Frage. In Anbetracht der Natur der kombinatorischen Prozesse eines $\mathcal{R}(\mathcal{MSO})$-Systems bietet sich ausserdem die Erforschung von Erweiterungen in Richtung eines begrenzt parallelen, "konkurrierenden" Systems an.

6 Veröffentlichungen, Vorträge, Aktivitäten

Veröffentlichungen

Cornell, T.L.: 'Model-Theoretic Syntax.' GLOT International **2(1/2)**, 1996

Cornell, T.L.: 'A Minimalist Grammar for the Copy Language.' SFB-Bericht 79, 1996

Cornell, T.L.: 'Representational Minimalism.' SFB-Bericht 83, 1997

Grefe, C. & M. Kracht: 'Adjunction Structures and Syntactic Domains.' SFB-Bericht 80, 1996

Kim, S.-S., H.-P. Kolb, G. Müller und W. Sternefeld: 'Minimalismus in der Syntax.' SfS-Report 03-95, Univ. Tübingen, 1995

Kolb, H.-P.: 'Is I-Language a Generative Procedure?' Ms., 1995

Kolb, H.-P.: 'Segments, Categories, and other Dangerous Things.' Ms., 1995

Kolb, H.-P: 'Macro Grammars for Minimalistic Syntax.' Ms., 1997. Erscheint in: H.-P. Kolb & U. Mönnich (Hg.).

Kolb, H.-P. & U. Mönnich (Hg.): *The Mathematics of Syntactic Structure*. In Vorbereitung. Enthält Artikel von Blackburn, Cornell, Kallmeyer, Keenan, Kolb, Kracht, Mönnich, Morawietz, Palm, Rogers, Volger.

Mönnich, U.: 'Zur deklarativen Repräsentation grammatischer Datenstrukturen.' Ms., 1995

Mönnich, U.: 'On Cloning Contextfreeness.' Ms., 1997. Erscheint in: H.-P. Kolb & U. Mönnich (Hg.).

Mönnich, U.: 'Logisch Präsentierbare Bereiche für grammatische Formalismen.' Ms., 1997

Morawietz, F.: 'A Unification-Based ID/LP Parsing Schema.' *Proceedings of the International Workshop on Parsing Technologies*, ACL/SIGPARSE, 1995, pp. 162–173

Morawietz, F.: 'Domain Theory and MSO-based CLP.' Ms., 1997

Morawietz, F.: 'Monadic Second Order Logic, Tree Automata and Constraint Logic Programming.' SFB-Bericht 86, 1997. Erscheint in H.-P. Kolb & U. Mönnich (Hg.).

Morawietz, F. & T.L. Cornell: 'Compactly Representing Constraints with Automata.' Erscheint in: Proceedings of the ACL/EACL'97, 1997

Morawietz, F. & T.L. Cornell: 'On the Recognizability of Relations over a Tree Definable in a Monadic Second Order Tree Description Language.' SFB-Bericht 85, 1997

Rogers, J.: 'Capturing Linguistic Theories Model-Theoretically.' SFB-Bericht 72, 1996

Vorträge

Cornell, T.L.: 'Consistency of partial descriptions.' Workshop *Tag des Baumes*, Tübingen, 15. März 1995.

Kolb, H.-P.: 'Segment:Tree=Category:Forest' Workshop *Tag des Baumes*, Tübingen, 15. März 1995.

Mönnich, U.: 'Dynamische Eigenschaftstheorie.' Universität Passau, 19. Juni 1995

Morawietz, F.: 'A Unification-Based ID/LP Parsing Schema.' International Workshop on Parsing Technologies (IWPT'95), Prag, 21. September 1995.

Morawietz, F.: 'Deklarative Grammatikformalismen und Information Retrieval.' Universität Zürich, 24. Mai 1996

Kolb, H.-P.: 'Two Types of Linguistic Universals.' ESSLLI96-Workshop *The Mathematics of Syntactic Structure*, Prag, 12. August 1996.

Mönnich, U.: 'A Characterization of Indexed* Languages by Monadic Second Order Logic.' ESSLLI96-Workshop *The Mathematics of Syntactic Structure*, Prag, 12. August 1996.

Morawietz, F.: 'L2-Based Constraint Logic Programming.' ESSLLI96-Workshop *The Mathematics of Syntactic Structure*, Prag, 13. August 1996.

Cornell, T.L.: 'Two minimalist programs and a calculus.' ESSLLI96-Workshop *The Mathematics of Syntactic Structure*, Prag, 15. August 1996.

Kolb, H.-P.: 'Formal Universals, Substantial Universals, and the Perfect Framework.' *Interfaces*-Konferenz, Tübingen, 4. Oktober 1996.

Morawietz, F. & T.L. Cornell: 'MSO Logic, Tree Automata and Theorem Proving.' DFKI & MPI Saarbrücken, 25. November 1996.

Mönnich, U.: 'On Cloning Context-freeness.' Workshop *Morphology and the Complexity of Language*, Innovationskolleg Formale Modelle kognitiver Komplexität, Potsdam, 14./15. Dezember 1996.

Mönnich, U.: 'Über das Klonen von Kontextfreiheit.' Stuttgart-Tübingen Colloquium Theoretische Informatik, Tübingen, 6. Februar 1996

Andere Aktivitäten

Colloquia, Arbeitsgruppen, Tutorien

Tag des Baumes, 14./15. März 1995
Teilnehmer: Backofen (DFKI), Cornell, Grefe (Berlin), Hinrichs, Kapzyk, Kolb, Kracht (FU Berlin), Mönnich, Morawietz, Müller, Pafel, Rounds (UMich), Schneider (Passau), Volger (Passau).
Programm:
M. Kracht: *The modal perspective on grammatical formalism*
H. Volger: *A constraint language for attributed trees.*
R. Backofen: *A First order Axiomatisation of the Theory of Finite Trees*
T. Cornell: *Consistency of partial descriptions*
H.-P. Kolb: *Segment:Tree=Category:Forest*
C. Grefe: *Adjunction Structures*
M. Kracht: *Syntactic Coding*
Diskussion

Arbeitsgruppe "Relationen in HPSG" (mit B4, B8), 1995/96

Arbeitsgruppe "Constraint Logic Programming für Linguistik" (mit B4), 1996

Arbeitsgruppe "Minimalismus in der Syntax" (mit A3, A4, A7), seit 1995

ESSLLI '96 Workshop "The Mathematics of Syntactic Structure", Prag, 12.-16. August 1996
Programm:
H.-P. Kolb: *Two Types of Linguistic Universals*
U. Mönnich: *A Characterization of Indexed* Languages by MSO*
J. Rogers: *SωS and Generalized Recognizable Sets*
F. Morawietz: *L2-Based Constraint Logic Programming*
H. Volger: *Principle Languages and Principle Based Parsing*
A. Palm: *On CLAT*
T. Cornell: *Two minimalist programs and a calculus*
L. Kallmeyer: *Tree Description Grammars*
M. Kracht: *Nearness and Depedendency in Grammar*
Allgemeine Diskussion

Tag des Baumes II (in Zusammenarbeit mit dem Graduiertenkolleg Integriertes Linguistikstudium), Tübingen, 5. Mai 1997

ACL/EACL '97 Tutorium "Model Theoretic Syntax" (Rogers & Cornell), Madrid, 7. Juli 1997

Gastwissenschaftler

William C. Rounds 13.-17. März 1995
Thema: 'Defaults in Domain Theory'

Edward Zalta 12.-16. Juni 1995
Thema: 'Typenfreie Systeme'
Vorträge: 'The Theory of Fregean Logical Objects', 'Naturalized Platonism vs. Platonized Naturalism'

James Rogers (UPenn/UCFlorida) 6.-11. November 1995
Vorträge: 'What Does a Grammar Formalism Say About a Language?', 'On the Descriptive Complexity of Recognizable Sets: The Case of GB.'

Mark Johnson (Brown Univ.) 13.-17. November 1995
Vorträge: 'Resource Conscious Logics for LFG', 'On Constructive Negation'

Patrick Blackburn (Universität des Saarlandes, Saarbrücken) 27. November–1. Dezember 1995
Thema: 'Modal Logic and the Analysis of Natural Language Syntax'
Vorträge: 'The Basic Ingredients', 'Monostratal Grammar Formalisms.'

Marcus Kracht (FU Berlin) 18.-21. März 1996
Vorträge: 'Parikh-languages and the (non-)semi-linearity of natural language', 'Kongruenzmorphologie'

Edward Stabler (UCLA) 17. Juni 1996:
'Parsing Theories For a Derivational Minimalism'

Edward L. Keenan (UCLA) 1.-12. Juli 1996
Vorträge: 'Syntactic Invariants and Semantic Invariants', 'More than One Language'

Stephan Kepser (Universität München) 18. Dezember 1996
'Kombination von Constraint Lösern: Theoretische und praktische Aspekte'

Peter Ruhrberg (Universität des Saarlandes, Saarbrücken) 20. Dezember 1996
'On Equality Up-to Constraints over Finite Trees'

Literatur

[1] Baader, F. & K. U. Schulz (1995): 'On the Combination of Symbolic Constraints, Solution Domains, and Constraint Solvers.' In: *Proceedings of the International Conference on Constraint Programming, CP-95*, vol. 976 of *Lecture Notes in Computer Science*, Berlin: Springer.

[2] Biehl, M., N. Klarlund & T. Rauhe (1996): 'Algorithms for Guided Tree Automata.' In: *Proc. WIA '96*, LNCS, Springer-Verlag.

[3] Blackburn, P. & W. Meyer-Viol (February 1994): 'Linguistics, Logic, and Finite Trees.' Report CS-R9412, CWI, Amsterdam.

[4] Bryant, R. E. (1992): 'Symbolic Boolean manipulation with ordered binary-decision diagrams.' *ACM Computing Surveys* **24**(3): 293–318.

[5] Büchi, J. (1960): 'Weak Second-Order Arithmetic and Finite Automata.' *Z. Math. Logik Grundlag. Math.* **6**: 66–92.

[6] Chomsky, N. (1985): *Barriers*. Cambridge, MA: MIT Press.

[7] Chomsky, N. (1991): 'Linguistics and Cognitive Science: Problems and Mysteries.' In: A. Kasher (ed.), *The Chomskyan Turn*, Cambridge, MA / Oxford: Blackwells.

[8] Chomsky, N. (1991): 'Some Notes on Economy of Derivation.' In: R. Freidin (ed.), *Principles & Parameters in Comparative Grammar*, Cambridge, MA: MIT Press.

[9] Chomsky, N. (1993): 'A Minimalist Program for Linguistic Theory.' In: K. Hale & S. J. Keyser (eds.), *The View from Building 20: Essays in Honour of Sylvain Bromberger*, pp. 1–52, Cambridge, MA: MIT Press.

[10] Chomsky, N. (1995): *The Minimalist Program*. Cambridge, MA: MIT Press.

[11] Chomsky, N. & H. Lasnik (1993): 'The Theory of Principles and Parameters.' vol. 9.1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chap. 24, pp. 506–569, Berlin, New York: Walter de Gruyter.

[12] Cornell, T. L. (1996): 'A Minimalist Grammar for the Copy Language.' SFB-Bericht 79, SFB 340, Tübingen/Stuttgart.

[13] Cornell, T. L. (1997): 'Representational Minimalism.' SFB-Bericht 83, SFB 340, Tübingen/Stuttgart.

[14] Doner, J. E. (1970): 'Tree Acceptors and some of their Applications.' *J. Comput. System Sci.* **4**: 406–451.

[15] Elgot, C. C. (1961): 'Decision Problems of Finite Automata Design and Related Arithmetics.' *Trans. Amer. Math. Soc.* **98**: 21–51.

[16] Fischer, M. J. (1968): 'Grammars with Macro-like Productions.' In: *Proceedings of the 9th Annual Symposium on Switching and Automata Theory*, pp. 131–142, IEEE.

[17] Grefe, C. & M. Kracht (November 1996): 'Adjunction Structures and Syntactic Domains.' Bericht 80, Arbeitspapiere des SFB 340, Tübingen/Stuttgart.

[18] Henriksen, J. G., J. Jensen, M. Jørgensen, N. Klarlund, R. Paige, T. Rauhe & A. Sandhol (1995): 'MONA: Monadic Second-Order Logic in Practice.' In: Brinksma, Cleaveland, Larsen, Margaria & Steffen (eds.), *Tools and Algorithms for The Construction and Analysis of Systems: International Workshop*, pp. 89–110, TACAS '95, Selected Papers, Lecture Notes in Computer Science 1019, Springer.

[19] Höhfeld, M. & G. Smolka (October 1988): 'Definite Relations over Constraint Languages.' LILOG Report 53, IBM Deutschland, Stuttgart, Germany.

[20] Huybregts, M. A. (1984): 'The Weak Adequacy of Context-Free Phrase Structure Grammar.' In: G. J. de Haan, M. Trommelen & W. Zonneveld (eds.), *Van periferie naar kern*, pp. 81–99, Dordrecht: Foris.

[21] Kallmeyer, L. (Dezember 1996): 'Underspecification in Tree Description Grammar.' Bericht 81, Arbeitspapiere des SFB 340, Tübingen/Stuttgart.

[22] Kim, S.-S., H.-P. Kolb, G. Müller & W. Sternefeld (1995): 'Minimalismus in der Syntax.' SfS-Report 03-95, Seminar für Sprachwissenschaft, Tübingen University.

[23] King, P. J. (1994): 'An Expanded Logical Formalism for Head Driven Phrase Structure Grammar.' Bericht 59, Arbeitspapiere des SFB 340, Tübingen/Stuttgart.

[24] Kolb, H.-P. (1995): 'Is I-Language a Generative Procedure?' Ms., SFB 340.

[25] Kolb, H.-P. (1995): 'Segments, Categories, and other Dangerous Things.' Ms., SFB340, Tübingen University.

[26] Kolb, H.-P. (1997): 'Macro Grammars for Minimalistic Syntax.' Ms., SFB340, Tübingen University.

[27] Kracht, M. (1993): 'Mathematical Aspects of Command Relations.' In: *Proceedings of the 6th EACL*, pp. 240–249.

[28] Kracht, M. (1995): 'Syntactic Codes and Grammar Refinement.' *J. Logic, Lang. & Inform.* **4**: 41–60.

[29] Matz, O., A. Miller, A. Potthoff, W. Thomas & E. Valkema (1995): 'Report on the Program AMoRE.' Tech. Rep. 9507, Institut für Informatik und Praktische Mathematik, Christian-Albrecht-Universität Kiel.

[30] Miller, P. H. (1991): 'Scandinavian Extraction Phenomena Revisited: Weak and Strong Generative Capacity.' *Linguistics & Philosophy* **14**: 101–113.

[31] Mönnich, U. (1995): 'Zur deklarativen Repräsentation grammatischer Datenstrukturen.' Tech. rep., SFB 340, Universität Tübingen.

[32] Mönnich, U. (1997): 'Logisch Präsentierbare Bereiche für grammatische Formalismen.' Tech. rep., SFB 340, Universität Tübingen.

[33] Mönnich, U. (1997): 'On Cloning Contextfreeness.' Ms., Tübingen University.

[34] Morawietz, F. (1997): 'Domain Theory and MSO-based CLP.' Ms., Tübingen University.

[35] Morawietz, F. (1997): 'Monadic Second Order Logic, Tree Automata and Constraint Logic Programming.' SFB 340-Report 86, SFB 340, University of Tübingen.

[36] Morawietz, F. & T. L. Cornell (1997): 'Compactly Representing Constraints with Automata.' In: *Proceedings ACL/EACL '97*, WWW: `http://www.sfs.nphil.uni-tuebingen.de/ frank/papers.html`.

[37] Morawietz, F. & T. L. Cornell (1997): 'On the Recognizability of Relations over a Tree Definable in a Monadic Second Order Tree Description Language.' SFB 340-Report 85, SFB 340, University of Tübingen.

[38] Niehren, J. & A. Podelski (1992): 'Feature Automata and Recognizable Sets of Feature Trees.' In: M.-C. Gaudel & J.-P. Jouannaud (eds.), *Proccedings of the 4th International Joint Conference on Theory and Practice of Software Development*, pp. 356–375, Orsay, France: Springer Verlag, LNCS 668.

[39] Niehren, J., P. Ruhrberg & M. Pinkal (1997), 'On Equality Up-To Constraints over Finite Trees, Context Unification, and One Step Rewriting.' Submitted to CADE-97, WWW: `http://www.ps.uni-sb.de/~niehren/cv.html#papers`.

[40] Palm, A. (1997): *The Expressivity of Tree Languages*. Ph.D. thesis, Universität Passau, Fakultät für Mathematik und Informatik.

[41] Rabin, M. O. (1969): 'Decidability of Second Order Theories and Automata on Infinite Trees.' *Trans. Amer. Math. Soc.* **141**: 1–35.

[42] Rogers, J. (1994): *Studies in the Logic of Trees with Applications to Grammar Formalisms*. Ph.D. thesis, University of Delaware.

[43] Rogers, J. (Februar 1996): 'Capturing Linguistic Theories Model-Theoretically.' Bericht 72, Arbeitspapiere des SFB 340, Tübingen/Stuttgart.

[44] Rogers, J. (Februar 1997): 'The Descriptive Complexity of Generalized Local Sets.' Bericht 84, Arbeitspapiere des SFB 340, Tübingen/Stuttgart.

[45] Shieber, S. M. (1985): 'Evidence against the Context-Freeness of Natural Language.' *Linguistics & Philosophy* **8**(3): 333–344.

[46] Stabler, E. P. (1996): 'Computing Quantifier Scope.' In: A. Szabolcsi (ed.), *Ways of Scope Taking*, Boston: Kluwer.

[47] Stabler, E. P. (forthcoming): *Acquiring and Parsing Languages with Movement*. Oxford: Basil Blackwell.

[48] Stabler, E. P. (forthcoming): 'Derivational Minimalism.' In: *Proceedings of the LACL'96*.

[49] Steinmann, M. (1993): *Übersetzung von logischen Ausdrücken in Baumautomaten: Entwicklung eines Verfahrens und seine Implementierung*. Master's thesis, Universität Kiel.

[50] Thatcher, J. & J. Wright (1968): 'Generalized Finite Automata Theory with an Application to a Decision Problem of Second-Order Logic.' *Math. Systems Theory* **2**(1): 57–81.