# Linguistische Arbeiten 405

Herausgegeben von Hans Altmann, Peter Blumenthal, Herbert E. Brekle, Gerhard Helbig, Hans Jürgen Heringer, Heinz Vater und Richard Wiese

*Norbert Bröker*

# Eine Dependenzgrammatik zur Kopplung heterogener Wissensquellen

Max Niemeyer Verlag
Tübingen 1999

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

**Bröker, Norbert:** Eine Dependenzgrammatik zur Kopplung heterogener Wissensquellen / Norbert Bröker. – Tübingen : Niemeyer, 1999
(Linguistische Arbeiten ; 405)

ISBN 3-484-30405-7 ISSN 0344-6727

D 25

© Max Niemeyer Verlag GmbH, Tübingen 1999
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Printed in Germany.
Gedruckt auf alterungsbeständigem Papier.
Druck: Weihert-Druck GmbH, Darmstadt
Buchbinder: Nädele Verlags- und Industriebuchbinderei, Nehren

# Inhaltsverzeichnis

# Abbildungsverzeichnis

# Abkürzungs- und Symbolverzeichnis

| | |
|---|---|
| $\Phi_n$ | Projektion eines Tupels auf sein $n$-tes Element |
| $\prec$ | Präzedenzrelation (über Wörtern bzw. Dependenzen) (Kapitel 5.1.7) |
| $\ll$ | Präzedenzrelation über Wortstellungsdomänen (Kapitel 4.1.3.3) |
| $\mathcal{S}$ | Menge der Konzepte (Kapitel 5.1.6) |
| $isa_{\mathcal{S}}$ | Subklassenrelation auf $\mathcal{S}$ |
| $\mathcal{O}$ | Menge der Referenzobjekte (Kapitel 5.1.6) |
| $\mathcal{R}$ | Menge der Rollen (Kapitel 5.1.6) |
| $\mathcal{C}$ | Menge der Wortklassen (Kapitel 5.1.2) |
| $isa_{\mathcal{C}}$ | Subklassenrelation auf $\mathcal{C}$ |
| $\mathcal{L}$ | Menge der Lexeme (Kapitel 5.1.4) |
| $isa_{\mathcal{L}}$ | Subklassenrelation auf $\mathcal{L}$ |
| CG | *Categorial Grammar* (Kapitel 7.2.1) |
| DCSD | *dutch cross-serial dependencies* |
| DG | Grammatik auf der Basis von Dependenzrelationen |
| DUG | *Dependency Unification Grammar* (Kapitel 7.1.2) |
| ECPO | *Exhaustive Constant Partial Order* (GPSG) |
| FO-TAG | *Free Order* TAG |
| GB | *Government and Binding Theory*, auch *Principles and Parameters Approach* |
| GPSG | *Generalized Phrase Structure Grammar* |
| HPSG | *Head-driven Phrase Structure Grammar* (Kapitel 7.2.3) |
| LFG | *Lexical-Functional Grammar* |
| LTAG | *Lexicalized* TAG |
| MNCC | *mother-node constructing categories* (Kapitel 4.3.1) |
| MTT | *Meaning-Text Theory* (Kapitel 7.1.7) |
| PSG | Grammatik auf der Basis von Konstituenzbeziehungen |
| RG | *Relational Grammar* |
| STAG | *Synchronous* TAG |
| TAG | *Tree-Adjoining Grammar* (Kapitel 7.2.2) |
| TWE | Textwissensextraktion |
| UCFG | *unordered context-free gramar* (Kapitel 5.2.4) |
| V-TAG | *Vector* TAG |
| WG | *Word Grammar* (Kapitel 7.1.3) |

# Übersicht

Diese Arbeit stellt eine Grammatikkonzeption vor, die im Kontext eines maschinellen Textverstehenssystems entstanden ist. Primäres Ziel waren daher nicht Untersuchungen zur Syntax, sondern Fragen der Abbildung von Texten auf Inhaltsrepräsentationen. Grundlage der Grammatik ist die Dependenzrelation, die die vollständige Lexikalisierung des Analyseverfahrens erlaubt, die inkrementelle konzeptuelle Interpretation erleichtert sowie die Reduzierung von Mehrdeutigkeiten möglich macht.

Hauptziel war die Integration aller Wissenssysteme in einem wortweise inkrementellen Analyseverfahren, um Mehrdeutigkeiten in einzelnen Beschreibungsdimensionen frühzeitig zu reduzieren. Dazu wird die syntaktische Relationierung zweier Wörter mit Bedingungen an die Wörter in mehreren Wissenssystemen verknüpft. Auf diese Weise lassen sich sowohl die klassischen Beschreibungsebenen Syntax, Semantik und Diskurs koppeln als auch die Syntax in Subsysteme aufgliedern (Kategorien, Merkmale, Stellung). Wegen dieser Kopplung unterschiedlicher Beschreibungsdimensionen heißt der vorgestellte Ansatz DACHS (***Dependency Approach to Coupling Heterogeneous Knowledge Sources***).

Die Kopplung von Dominanz und Präzedenz ist in Dependenzgrammatiken nicht so eng wie in Phrasenstrukturgrammatiken; dennoch sind bisherige Beschreibungen der Wortstellung im dependentiellen Paradigma formal und/oder empirisch unbefriedigend. Zur Lösung dieser Probleme definiert DACHS Wortstellungsdomänen, die die Linearisierung von der hierarchischen Analyse trennen. Wortstellungsdomänen erlauben die widerspruchsfreie Integration von Präzedenzrestriktionen und erfassen Diskontinuitäten als Verallgemeinerung der kontinuierlichen Anbindung.

Außerdem definiere ich eine Beschreibungssprache mit modelltheoretischer Semantik für die Dependenzstrukturen. Diese auf der Modallogik beruhende Sprache fügt sich in das Paradigma der *model-theoretic syntax* ein und beruht auf dependenzgrammatisch motivierten Modellen. Damit wird erstmals eine konsistente Formalisierung dependenzgrammatischer Begriffe erreicht.

# Abstract

This dissertation presents a grammar architecture developed in the context of text knowledge extraction. The primary motivation for this architecture therefore is not syntax in itself, but rather the mapping of texts onto formal representations of their content. It is based on the primitive syntactic relation of dependency, which allows us to extend the lexicalisation to the parsing process itself, to incrementally interpret words within a terminological knowledge base, and to reduce intermediate ambiguities.

The main characteristic of the architecture is the integration of several knowledge systems within an incremental parsing scheme. Thereby, ambiguities local to any one knowledge system are eliminated as early as possible. The integration is achieved by linking syntactic relations with multi-dimensional constraints over the words to be linked. In this way, the classical linguistic systems of syntax, semantics, and discourse can be integrated, and syntax itself is split into the subsystems of categorial, morpho-syntactic, and ordering information. The name DACHS (*Dependency Approach to Coupling Heterogeneous Knowledge Sources*) derives from this coupling of different descriptional dimensions of a linguistic sign.

Dominance and precedence are not as tightly coupled in dependency grammar as they are in phrase-structure grammars; previous proposals to treat word order in dependency grammar are nevertheless formally and/or empirically inadequate. Within DACHS, so-called word order domains separate the linear from the hierarchical ordering. Precedence constraints can be consistently accomodated within this approach, which views discontinuities as a generalization of continuous attachment.

I also define a logical description language for dependency structures, which is given a model-theoretic semantics. The precise semantics as well as the linguistic motivation of the underlying models makes the description of lexical entries or syntactic analyses particularly straightforward. Previous presentations of dependency grammar did not have a precision comparable to this logic.

# Dankeschön

Diese Arbeit wäre nicht entstanden, wenn nicht viele Menschen mit mir daran gearbeitet hätten. Zuallererst sind hier meine Eltern Gisela und Paul zu nennen, die mir die notwendige Neugier sowie das Vertrauen mitgaben, mich in ein solches Abenteuer zu stürzen.

Peter Hellwig habe ich für die Einführung in die Linguistik zu danken; man wird seinen Einfluß problemlos entdecken. Udo Hahn hat mir nicht nur das Artikelschreiben beigebracht, sondern auch den Rahmen dieser Arbeit definiert und ihr viele Anstöße und fachliche Anregungen gegeben. Aus der Arbeitsgruppe CLIF an der Universität Freiburg, in der diese Arbeit entstanden ist, sind in diesem Zusammenhang insbesondere zu nennen: Susanne Schacht, der ich den schönen Namen 'Vakanz' verdanke, und Peter Neuhaus, der mir (viel zu?) viele LaTeX-Tips gegeben hat. An der Lesbarmachung dieser Arbeit waren maßgeblich Steffen Staab, Susanne Schacht und Udo Hahn beteiligt – vielen Dank!

Astrid hat wohl die größte Last getragen; sie hat mit kaum zu erschütternder Geduld dieses Projekt mit vorangetrieben. Und Lars ist am besten davongekommen – ihm bleiben immerhin die restlichen Schmierzettel zum Spielen.

Das vorliegende Buch ist eine überarbeitete Fassung der Dissertation, die den Stand der Literatur bis Mitte 1998 widerspiegelt. Für die Aufnahme in die Linguistischen Arbeiten danke ich Peter Blumenthal und Hans Jürgen Heringer.

# 1 Einleitung

> Man kann sich auch ohne komplexe Sätze verständlich machen. (Engel, 1988:240)

Die vorliegende Dissertation beschreibt eine Grammatikkonzeption, die aus Arbeiten zur Textwissensextraktion entstanden ist. Zielsetzung des Gesamtprojekts ist die Analyse von authentischen Sachtexten aus den Bereichen Informationstechnik und Medizin sowie die formale, sprachunabhängige Repräsentation des Textinhalts zum Zwecke der Weiterverarbeitung (Recherche, Zusammenfassung, Übersetzung etc.). Im Mittelpunkt des Interesses stehen dabei Fragen der fehlertoleranten Verarbeitung authentischer Texte und der Integration verschiedener Wissenssysteme, die hier Syntax, anaphorische Prozesse, Semantik, Domänenwissen sowie Lernverfahren umfassen.

Aus den Untersuchungen zur Textwissensextraktion und zum menschlichen Sprachverstehen ergibt sich die folgende These, die Ausgangspunkt sowohl für das Gesamtprojekt und als auch den Entwurf einer syntaktischen Beschreibung war.

*These 1*
*Textwissensextraktion erfordert die Integration mehrerer heterogener Wissensquellen.*

Diese These wird von psycholinguistischen Untersuchungen gestützt, die den verständnisleitenden Effekt von konzeptuellen (Haberlandt & Bingham, 1982; Granger et al., 1983; Seifert et al., 1986) und diskursiven Faktoren (Crain & Steedman, 1985; Altmann & Steedman, 1988; Konieczny et al., 1991) belegen konnten. Das daraus resultierende Problem wird deutlich, wenn man die technischen Realisierungen der einzelnen Module betrachtet, die isoliert entwickelt wurden und stark differierende Grundannahmen aufweisen. Diese manifestieren sich in unterschiedlichen Datentypen (Merkmalsgraphen, Klassenhierarchien, Syntaxbäumen, semantischen Netzen etc.) sowie darauf definierten Operationen (Unifikation, Subsumtion, Musterabgleich, Substitution etc.). Zwei Wege führen zur Integration: Erstens der Entwurf einer umfassenden Repräsentationssprache, die als Obermenge der Einzellösungen die Formulierung aller Regularitäten in einem gemeinsamen Rahmen erlaubt, oder zweitens die Kopplung bereits bestehender Einzellösungen, die durch geeignete Kommunikationsmodelle die erforderliche Interaktion unterstützt. Die erste Lösung wird im Rahmen der Unifikationsgrammatiken bereits erforscht (Pollard & Sag, 1994), führt jedoch zu einem formal aufwendigen Modell mit einer großen Zahl von Datentypen und entsprechend verallgemeinerten Operationen, das außerdem bis heute systematische Lücken bei der Integration des Domänenwissens und der Textmodellierung aufweist. Zur Inkorporation entsprechender Restriktionen ist man weiterhin auf die Kopplung angewiesen (Görz, 1992). Diese Problematik verschärft sich noch, wenn neben den statischen Wohlgeformtheitsbedingungen auch Verfahrenswissen über Verstehensstrategien integriert werden muß.

Aus diesem Grund schlagen wir eine Architektur vor, die verschiedene Repräsentationsebenen in lexikalisch motivierten Prozessen bündelt, die jeweils eine Lesart einer Wortform repräsentieren. Die lexikalischen Prozesse kommunizieren miteinander und bestimmen so die globale relationale Struktur eines Textes, die sich aus syntaktischen, textuellen und kon-

zeptuellen Relationen zusammensetzt. Diese objektorientierte Beschreibungsweise erlaubt die Integration verschiedenster Wissensquellen sowie die gemeinsame Beschreibung von deklarativem Kompetenzwissen und prozeduralem Performanzwissen durch die Definition entsprechenden Kommunikationsverhaltens.

Diese Architektur wirft die Frage auf, wie die verschiedenen Wissenssysteme gekoppelt werden können. In dieser Arbeit wird untersucht, inwieweit sich die Dependenzrelation als syntaktisches Basisprimitiv eignet und wie die verschiedenen Wissenssysteme unter Bezug auf die Dependenz gekoppelt werden können. Die Hauptthese dieser Arbeit läßt sich wie folgt formulieren.

*These 2*
*Die empirisch beobachtbaren Kombinationsrestriktionen (bzgl. Morphosyntax, Stellung, Semantik etc.) lassen sich als voneinander unabhängige Vorbedingungen einer grundlegenden Dependenzrelation zwischen Wörtern erfassen.*

Die Annahme einer primären Dependenzbeziehung zwischen Wörtern hat eine Reihe von Auswirkungen auf das Grammatikmodell. Die Abwertung des Phrasenbegriffs (der zwar rekonstruierbar, aber nicht mehr primär ist) erfordert, daß die Interpretation einer Äußerung nicht aus phrasalen Konfigurationen, sondern aus binären Dependenzrelationen abgeleitet werden. Weiterhin ergibt sich daraus die Notwendigkeit, die Wortstellungsbeschreibung mehr als in Phrasenstrukturgrammatiken von der hierarchischen Struktur abzukoppeln, da bedeutungstragende Dependenzbeziehungen nicht projektiv sind und eine hierarchische Analysestruktur nicht als Basis der Linearisierung dienen kann. Außerdem muß die Berechnung textueller Kohärenzrelationen in den Prozeß der Dependenzetablierung eingebunden werden.

Neben diesen empirischen Fragen der adäquaten Erfassung sprachlicher Phänomene gehen wir auf der methodologischen Ebene der Frage nach, wie ein angemessener formaler Rahmen für die DG aussehen kann, der an Präzision den konkurrierenden Entwürfen im phrasenstrukturellen Paradigma nicht nachsteht. Ich entwickle eine modallogische Beschreibungssprache, deren formale Primitive direkt durch linguistische Begriffe interpretierbar sind.

## 1.1 Warum ein dependentieller Ansatz?

In der Linguistik konkurrieren zwei grundlegende Beschreibungsverfahren, der phrasenstrukturelle und der dependentielle Ansatz. Phrasenstrukturelle Beschreibungen beruhen auf der Annahme nicht-lexikalischer Kategorien (den Konstituenten), die durch eine hierarchische Teil-von-Beziehung geordnet sind. Typischerweise ist diese Teil-von-Beziehung auch mit der linearen Folge gekoppelt, indem die Projektivität der Hierarchie gefordert wird. Dependentielle Beschreibungen beschränken sich demgegenüber auf lexikalische Kategorien (die Wörter), die in Abhängigkeitsrelationen zueinander stehen. Mit dieser grundlegenden Orientierung geht der Ausbau entweder des kategoriellen oder des relationalen Beschreibungsapparates einher; Phrasenstrukturgrammatiken verfügen über elaborierte Kategoriensysteme (z.B. das $\overline{X}$-Schema in GB (Haegeman, 1994) oder komplexe Kategorien in der CG (Bach, 1988)), während DGen üblicherweise verschiedene, teilweise strukturierte Relationen definieren (etwa die Relationenhierarchie in WG (Hudson, 1990)).

Die Wahl des dependentiellen Ansatzes spiegelt daher die Auffassung wider, daß natürliche Sprache sich eher als relationales Netzwerk denn als Konfiguration abstrakter Kategorien fassen läßt. Die folgenden Punkte motivieren diese Wahl und leiten daraus die Ziele dieser Arbeit ab.

### 1.1.1 Interpretation

Testverstehen muß auf Domänenwissen aufbauen, das sowohl analyseleitendes Hintergrundwissen darstellt als auch den Bezugsrahmen für die Repräsentation des Textinhalts darstellt. Hierfür ist die aus *frame*-Systemen (Minsky, 1975) erwachsene terminologische Logik besonders geeignet. Diese Beschreibungsweise beruht auf der Differenzierung von Objekten durch sog. Rollen, die die Objekte miteinander in Beziehung setzen. Es handelt sich um eine relationale, objektzentrierte Beschreibung, die im Gegensatz zu semantischen Netzwerken auf formal präzisen Beschreibungsprimitiven beruht.

Die Nähe der DG zu *frame*-Repräsentationen ist offensichtlich und wurde mehrfach bemerkt (schon bei Fillmore (1968:87f), auch in Somers (1987), Giachin & Rullent (1988), Lesmo & Lombardo (1992)). Teilweise lassen sich die Primitive der DG direkt auf Primitive der terminologischen Logik abbilden (referentielle Wörter auf konzeptuelle Objekte, Komplementdependenzen auf Argumentrelationen). Diese einfache Abbildbarkeit liegt darin begründet, daß Dependenzrelationen vielfach semantisch begründet werden (Heringer, 1993:316), und hat zu verschiedenen Entwürfen geführt, die eine PSG zur Syntaxanalyse mit einer DG zur Semantikanalyse koppeln (Fillmore, 1968; Baumgärtner, 1970). Wir nutzen die semantische Motivation der Dependenz, indem wir die Dependenzsyntax direkt mit der Wissensrepräsentation koppeln, und so zusätzliche vermittelnde Ebenen eliminieren.

### 1.1.2 Wortstellung

Im Gegensatz zur originalen Formulierung der Konstituenzbeziehung ist die Dependenzrelation zunächst unabhängig von der textuellen Präzedenz. Bedeutungstragende Dependenzen sind nicht projektiv, wie es die Formalisierung von Gaifman (1965) suggerieren will. Eine Reihe von Mechanismen zur Kopplung von Dominanz und Präzedenz wurden entwickelt, die allerdings anders als in PSGen die ursprüngliche Formulierung des Basisprimitivs nicht aufweichen, sondern es nur mit zusätzlichen Restriktionen genauer fassen. Der Blick auf die konkrete Ausgestaltung der Wortstellungskomponente verschiedener Dependenzgrammatiken offenbart jedoch erhebliche Mängel, die von der Beschränkung auf implementative Lösungen bis hin zu formal inkonsistenten Beschreibungsweisen reichen (vgl. Kapitel 7).

Ein Teil der vorliegenden Arbeit entwickelt daher eine Wortstellungskomponente, die die hierarchische Struktur von der linearen entkoppelt. Dabei zeigt sich, daß sowohl die Fixierung der linearen Folge als auch die Beschreibung möglicher Diskontinuitäten auf den unabhängig motivierten Dependenzrelationen aufgebaut werden kann. Damit wird die häufig geäußerte Meinung expliziert, daß auch die Wortstellung ein Mittel der Markierung syntaktischer Relationen sei. Zusätzliche Faktoren (wie Diskurseinflüsse, Phrasengröße u.a.) werden in dieser

Arbeit nicht betrachtet, lassen sich aber durch weitere Restriktionen berücksichtigen, da die Wortstellung durch die Dependenzsyntax nur partiell fixiert wird.

### 1.1.3 Lexikalisierung

Ein gemeinsamer Trend der modernen Grammatiktheorien ist die Assoziation grammatischer Aussagen mit Wörtern bzw. Klassen von Wörtern. Diese Lexikalisierung vermeidet die Probleme globaler, interagierender Regeln. Vererbungsmechanismen zur Strukturierung großer Datenmengen ermöglichen gleichzeitig die redundanzfreie und konsistente Beschreibung von Generalisierungen, so daß das Lexikon von einer Sammlung der Idiosynkrasien zu einer Sammlung der Generalisierungen wird.

Die lokale Natur der Dependenzbeziehungen unterstützt diese wortlokale Spezifikation der Kombinationsmöglichkeiten. Aufgrund der traditionellen Betrachtung von Dependenzrelationen als ausschließlich durch das regierende Element bestimmt ergeben sich jedoch Probleme bei den Adjunkten. Diese werden hier bereits im Lexikon von den Komplementaten differenziert, was die Unterscheidung von Funktor und Argument auch dependenzgrammatisch faßbar macht.

### 1.1.4 Lokalität

Die Lokalität kann als eine Verschärfung der Lexikalisierung angesehen werden. Das unserer Architektur zugrundeliegende verteilte Parsingverfahren (Hahn et al., 1994; Bröker et al., 1996; Neuhaus & Hahn, 1996b) definiert lexikalisch motivierte Prozesse, die durch Kommunikation syntaktische, textuelle und konzeptuelle Relationen etablieren.

Diese Kommunikation kann im dependentiellen Paradigma auf zwei Objekte, das Regens und das Dependens, beschränkt werden, zwischen denen eine direkte Relation etabliert werden soll. Die Lokalität erhöht außerdem die Fehlertoleranz der linguistischen Beschreibung, da nicht mehr vollständige Phrasen in einem Schritt konstruiert werden, sondern jeweils nur zwei Wörter kombiniert werden. Ob weitere Modifikatoren korrekt realisiert und analysiert werden, ist für die Anbindung eines einzelnen Modifikators unerheblich.

### 1.1.5 Ambiguitätsreduktion

Jedes maschinelle Verfahren der Syntaxanalyse muß die mannigfaltigen Mehrdeutigkeiten auf allen Ebenen der Sprache behandeln. Dies kann nicht allein durch Strategien des Analyseverfahrens (z.B. *head-corner*, *beam search*, Heuristiken etc.) geschehen, sondern muß von seiten des Grammatikmodells durch Vorkehrungen zur Reduktion der Alternativen unterstützt werden. Unsere Erfahrung zeigt, daß in PSGen systematische Ambiguitäten auftreten, die in DGen vermeidbar sind. Hierzu zählen insbesondere Wortstellungsvarianten, aber auch alternative syntaktische Realisierungen von Komplementen.

Unser Ziel ist hier, solche Dimensionen zu finden und zu koppeln, die diese Alternativen maximal kapseln (vgl. These 2). Für die Analyse bedeutet das zweierlei: Erstens wird der

Gesamtaufwand für die Repräsentation reduziert, da Alternativen lokal in einzelnen Dimensionen (z.B. bzgl. Wortstellung und Interpretation) repräsentiert werden und nicht expandiert werden müssen; die Zahl der strukturellen Mehrdeutigkeiten sinkt. Zweitens bedeutet es, daß die Analyse weniger Wahlmöglichkeiten hat bzw. kein Auswahlzwang entsteht. Das ist z.B. der Fall, wenn alle drei Verbstellungen in der Stellungsdimension einer Wortbeschreibung repräsentiert werden und für das Analyseverfahren so wenig sichtbar sind wie Merkmalsdisjunktionen. Dieser Ansatz kontrastiert mit dem deterministischen Parsing insofern, als er die Existenz von Alternativen explizit berücksichtigt und zum Anlaß nimmt, durch repräsentationale Vorkehrungen die Alternativen weitgehend zu isolieren und zu reduzieren.

### 1.1.6 Eliminierung der Konstituenten

Neben der Ausnutzung dieser Vorteile verfolgt die vorliegende Arbeit mit der Wahl der DG als Beschreibungsparadigma auch zwei methodische Ziele, die Eliminierung der nichtlexikalischen Konstituenten sowie die formale Fundierung dependenzgrammatischer Ideen.

Varianten des Dependenzbegriffs erfahren zur Zeit erneute Aufmerksamkeit. Das gilt für die linguistische Beschreibung wie für das Analyseverfahren. In verschiedenen Grammatiktheorien sind dependentielle Begriffe inzwischen tief verankert; man vergleiche etwa $\theta$-Rollen, Funktor-Argument-Differenzierungen, elaborierte Subkategorisierungsmechanismen sowie den Begriff der Lokalität in TAG. Die momentan wohl populärste Grammatikkonzeption, die HPSG, ähnelt in ihren Spezifikationen vielfach dependentiellen Ansätzen (mit der Ausnahme der Wortstellungsbeschreibung). Der Rückgriff auf dependentielle Begriffe ist ebenfalls sehr deutlich in TAGen. Rambow (1994:39f) definiert die 'derivationelle generative Kapazität' als die Fähigkeit, syntaktische Dependenzen zu generieren und begründet mit der für das Deutsche nicht ausreichenden derivationellen generativen Kapazität von TAGen eine Theorieerweiterung zu V-TAGen. Im Bereich der Analyseverfahren wird die Etablierung binärer Relationen wieder stärker betont, so z.B. im *Licensing Structure Parser* von (Abney & Cole, 1986), im Automatenmodell für V-TAGen (Rambow, 1994; Rambow & Joshi, 1994) oder in statistischen Ansätzen (Eisner, 1997).

Diese Beobachtungen geben Anlaß, den Status phrasaler Kategorien erneut zu überdenken. Die Spezifikation von Komplementen und Adjunkten, d.h., der hierarchischen Struktur, erfolgt in vielen modernen PSGen ohne Referenz auf phrasale Kategorien. Die in dieser Arbeit ausgearbeitete Wortstellungsbeschreibung zeigt, daß auch die Wortstellung einschließlich Diskontinuitäten unter ausschließlichem Bezug auf binäre Relationen formuliert werden kann. Unsere Konzeption eliminiert somit "*the nonobservable linguistic construct that enjoys the widest acceptance*" (Pollard & Sag, 1994:9) und führt zu einer stärker beschränkten Syntaxtheorie.

### 1.1.7 Formale Fundierung der DG

Gerade im Licht des erneuten Interesses an DG ist es notwendig, eine formal hinreichende Basis für das dependentielle Paradigma bereitzustellen. Bei der Betrachtung der existierenden Varianten von DG stellt man nämlich fest, daß ihre mathematische Basis durchweg unzurei-

chend ist. Unter diesen Umständen ist die präzise Darstellung von Analysen und noch mehr der Vergleich mit anderen Ansätzen kaum möglich. Wesentlich für die weitere Verbreitung der grundlegenden Ideen der DG ist eine dem heutigen Standard entsprechende formale Aufbereitung.

Die zweite methodische Leistung der vorliegenden Arbeit ist daher die Definition einer Beschreibungslogik,[1] die die grammatischen Ausdrucksmittel und Operationen formal fixiert. Sie basiert auf einer modalen Aussagenlogik und weist mit der modell-theoretischen Interpretation eine Präzision auf, die dependentielle Beschreibungen bisher vermissen ließen.

## 1.2 Überblick

In den beiden nächsten Kapitel werden die Rahmenbedingungen dieser Arbeit weiter ausgeführt. Kapitel 2 erläutert die Zielsetzung des Gesamtprojekts sowie die Grundprinzipien der anderen Systemkomponenten, mit denen die Grammatik interagiert. Kapitel 3 diskutiert die Grundlagen des dependentiellen Paradigmas und leitet daraus die Grundannahmen für die vorliegende Arbeit ab.

Danach folgen drei Kapitel, die die Grammatikkonzeption ausführen. Kapitel 4 betrachtet die Beschreibungsdimensionen und motiviert insbesondere die Wortklasse als Repräsentant der Distribution und die Stellungsbeschreibung durch Wortstellungsdomänen. Kapitel 5.1 gibt die Beschreibungslogik für diese Dimensionen an, so daß eine Spezifikationssprache für Lexikon und syntaktische Analysen bereitsteht. Kapitel 5.2 enthält den Nachweis, daß DGen mit linguistisch ausreichend flexibler Stellungsbeschreibung exponentielle Analysekomplexität aufweisen. Das Kapitel 6 illustriert die Grammatikkonzeption an einer Reihe verschiedener Phänomene des Deutschen.

Das Kapitel 7 vergleicht unseren Ansatz mit anderen Konzeptionen aus dem dependentiellen und dem phrasenstrukturellen Paradigma. Es wird deutlich, daß frühere Konzeptionen der DG unzureichende Wortstellungsbeschreibungen und unzulängliche (bzw. gar keine) Formalisierungen aufweisen. Auch zeigt es sich, daß das Konzept der Dependenz unter verschiedenen Namen auch in phrasenstrukturelle Ansätze Eingang gefunden hat. Im Kapitel 8 nehme ich einige offene Fragen auf und skizziere, welche Anschlußarbeiten sich daraus unmittelbar ergeben.

Zum Ende der Einleitung noch einige Bemerkungen zur Terminologie. Ich werde im folgenden die Kürzel DG und PSG als Sammelbegriffe für die beiden Paradigmen benutzen, wenn eine exakte Differenzierung zwischen einzelnen ihrer Ansätze nicht notwendig ist. Anstelle der traditionellen dependenzgrammatischen Termini 'Regens' (für das regierende von zwei dependentiell relationierten Wörtern) und 'Dependens' (für das subordinierte Wort) ver-

[1] Ich verwende den Begriff 'Beschreibungslogik', obwohl er (und seine englische Variante *description logic*) eigentlich das umfaßt, was in dieser Arbeit (dem Gebrauch in Nebel (1990) folgend) terminologische Logik heißt. Für diesen Gebrauch gibt es zwei Gründe. Erstens handelt es sich um eine logische Sprache, die Beschreibungen für linguistische Objekte zu formulieren erlaubt. Zweitens besteht neben der begrifflichen auch eine sachliche Ähnlichkeit, da die hier vorgestellte Beschreibungslogik für linguistische Objekte eine Modallogik ist, deren Äquivalenz zu den 'echten' Beschreibungslogiken Schild (1991) gezeigt hat.