

Linguistische Berichte

Reprint

Forschung Information Diskussion

Herausgegeben

**in Zusammenarbeit mit mehreren Sprachwissenschaftlichen
Instituten und Seminaren anderer Disziplinen
von Professor Dr. Peter Hartmann, Konstanz**

Redaktion: Dr. Arnim von Stechow

**Anschrift der Redaktion: 775 Konstanz, Universität,
Fachbereich Sprachwissenschaft**

» vieweg

Aus dem Inhalt der erschienenen Hefte

Heft 1: Hans-Martin Gauger: Die Semantik in der Sprachtheorie der transformationellen Grammatik

Klaus Brockhaus: **Subjekt und Prädikat in Grammatik und Logik**

R. G. van de Velde: **General linguistics and monolingual research**

Ernst-Friedrich Suhr: **Germanistik 69 - Ein Vorschlag zur Güte**

Harald Weinrich: **Sprachlehre an der Universität**

Johannes Faensen: **Zur linguistischen Ausbildung der Sprachlehrer**

Heft 2: D. Kastowsky: Wortbildung und Nullmorphem

Johannes Bechert: **Indogermanistik und generative Phonologie**

Jürgen Sternsdorff: **Neurosemantische Triebe und extensionale Orientierung**

Das Linguistische Unbewußte

Siegfried J. Schmidt: **Sprachliches und soziales Handeln
Überlegungen zu einer Handlungstheorie der Sprache**

Harald Weinrich: **Überlegungen zu einem Studienmodell der Linguistik**

Wolfgang Iser: **Überlegungen zu einem literaturwissenschaftlichen Studienmodell**

Heft 3: Hansjakob Seiler: On the Interrelation between Text, Translation and Grammar of an American Indian Language

Ilpo Tapani Piirainen: **Quantitative Methoden in der Sprachforschung**

Jens Ihwe: **Linguistik und Literaturwissenschaft: Bemerkungen zur Entwicklung einer strukturalen Literaturwissenschaft**

Bernd Dammann: **Totengräber der Germanistik - Lebendig!**

Peter Hartmann: **Linguistik und Hochschulreform**

Dieter Wunderlich: **Zur Kritik an Harald Weinrichs Studienmodell der Linguistik**

Eine Analysegrammatik

Wolfgang Klein, Saarbrücken

1

Man kann eine Grammatik als einen Formalismus auffassen, der alle Sätze einer Sprache und ihre Strukturbeschreibungen aufzählt. Eine derartige Grammatik ist kein abstraktes Sprechermodell noch gar eine Darstellung dessen, was konkret im Gehirn eines Sprechers bei der Bildung von Sätzen vor sich geht. Es ist aber kein Zufall, daß diese Verwechslung bei der Rezeption der generativen Grammatik so häufig war¹). Sie hängt — außer mit der falsch gedeuteten Bezeichnung „generativ“ — vor allem damit zusammen, wie in den bisherigen Modellen starke und schwache generative Kapazität verbunden wurden. Man könnte sich durchaus einen Grammatiktyp vorstellen, bei dem beide getrennt realisiert und erst nachträglich aufeinander bezogen werden, d.h. bei dem zwei prinzipiell unabhängige Mechanismen eine Menge von Sätzen und eine Menge von Strukturbeschreibungen charakterisieren und jedem Satz durch irgendeine Abbildung eine Strukturbeschreibung (oder mehrere, falls der Satz mehrdeutig ist) zugeordnet sind. Solche Grammatiken gibt es meines Wissens nicht. In allen Modellen, die innerhalb der vergangenen 15 Jahre im Mittelpunkt der Diskussion standen, dienen die Strukturbeschreibungen dazu, die Menge der Sätze zu charakterisieren; die mit starker und schwacher generativer Kapazität bezeichneten Aufgaben der Grammatik werden gemeinsam gelöst. Die Menge der Sätze ist die Menge der Endketten der Strukturbeschreibungen²). Wenn CHOMSKY schreibt. „Unter einer generativen Grammatik verstehe ich einfach ein Regelsystem, das ... Sätzen Strukturbeschreibungen zuordnet“³), dann ist das natürlich nicht so zu verstehen als sei dieses Regelsystem eine Abbildung aus der Menge der Sätze in die Menge der Strukturbeschreibungen⁴). Vielmehr erzeugt das Regelsystem Strukturbeschreibungen, und diese Strukturbeschreibungen führen auf die Sätze⁵). Eben dies ist auch ein Grund für die geringe Bedeutung, die das Studium der schwachen generativen Kapazität für die Linguistik besitzt⁶).

Aus dieser sehr gut begründeten Anlage der üblichen generativen Grammatiken erklärt sich ihre merkwürdige Affinität zu Sprechermodellen. Ein einfaches Kommunikationsmodell, auch wenn es so grob ist wie das folgende, macht das noch deutlicher:

(1)	Sprecherseite	Kanal	Hörerseite
Redeintention			„Verständnis“
Regelwerk	Derivation	Satz. Satz	Dekodierung Regelwerk
Sprecherstrategien			Hörerstrategien

Dieses Verfahren, die Sätze mithilfe der Derivationen (oder ihnen entsprechenden Strukturbeschreibungen anderer Art) zu beschreiben, ist natürlich nur *eine* technische Möglichkeit. Mit CHOMSKYs Worten: „Wenn wir davon sprechen, daß eine Grammatik einen Satz erzeugt zusammen mit einer bestimmten strukturellen Beschreibung, so meinen wir einfach, daß die Grammatik diese strukturelle Beschreibung dem Satz zuschreibt“⁷).

Aus dieser Art, die Grammatik anzulegen, ergibt sich eine doppelte Funktion der Derivation, die, wie mir scheint, in der Entwicklung der Transformationsgrammatik eine wichtige Rolle gespielt hat. Die Derivationen haben nämlich einerseits die Aufgabe, die Sätze bzw. Endketten zu beschreiben, und zweitens legen sie zumindest die Grundlage zur Konstitution der Bedeutung. Eine syntaktische Komponente, deren Derivationen nur die Menge der Endketten charakterisieren sollen, wie es innerhalb der Theorie der formalen Sprachen üblich ist, braucht z.B. nicht anzunehmen, daß ihren Kategorien substantiell etwas entspricht; statt NP, VP, AUX usw. kann sie — von der Willkürlichkeit der *Namen* ganz abgesehen — auch Zahlen o.ä. nehmen; die Kategorien markieren lediglich Stellen im rekursiven Regelsystem; damit entfällt weitgehend die Möglichkeit eines eigenständigen Bewertungsmaßstabes für die starke generative Kapazität, denn die syntaktischen Strukturen haben ja keinen Sinn an sich, sondern sie dienen lediglich dazu, alle Sätze mit möglichst wenig Aufwand aufzuzählen und alle Nichtsätze auszuschließen. Diese etwas kurz-sichtige Vorstellung, die noch die Argumentation in den „Syntactic Structures“ weitgehend bestimmt und sich z.B. in der bisweilen grotesk anmutenden Diskussion über die kürzeste Notation äußert, liegt auch den hartnäckigen Vorschlägen linguistisch nicht allzu versierter Logiker zugrunde, natürliche Sprachen durch logische Kalküle + Homomorphismen („Übersetzungsregeln“) zu beschreiben.

In den „Aspects“ werden die Sätze auch noch über die syntaktischen Strukturen beschrieben, doch sind Kategorien und Funktionen nicht bloß Elemente eines rekursiven Mechanismus, sondern sie bilden, den Katz-Postal-Vorschlägen folgend, bereits in gewissem Umfang das Bindeglied zwischen der linearen kommunikativen Einheit Satz und der Bedeutung; die Einzelheiten sind bekannt. Dominant ist die letztgenannte Funktion dann in der generativen Semantik geworden. Lakoff erklärt programmatisch: "I assume that a grammar of a language is a system of rules that relates sounds in the language to their correspondent meanings"⁸). Demgegenüber ist die Produktion wohlgeformter Endketten vorerst stark vernachlässigt, wie die Strukturbeschreibungen bei LAKOFF, McCAWELY, ROSS u.a. zeigen.

All diesen Modellen gemeinsam ist, daß der Satz, formal gegeben als eine Folge von Phonemen, Graphemen, Merkmalsbündeln, den Abschluß der Derivation bildet: jede Derivation, gleich wie sie angelegt ist, spezifiziert genau einen Satz. Eine Grammatik, die dies leistet, erfüllt zweifellos die Aufgabe, genau die Menge der Sätze mit dazugehörigen Strukturbeschreibungen für die betreffende Sprache aufzuzählen. Es ist aber festzuhalten, daß dies im Prinzip zumindest nur *ein* möglicher Weg ist, diese Aufgabe zu erfüllen — wenn auch allem Anschein nach der bei weitem praktikabelste.

A sei irgendein Alphabet, z.B. $\{a, b, c\}$; mit A^* bezeichnet man die Menge der durch Verkettung bildbaren Folgen aus diesem Alphabet, z.B. $\{a, aa, aaa, aababcbacbaa, \dots\}$; eine Untermenge von A^* nennt man gewöhnlich eine Sprache. Nehmen wir etwa an, A^* sei das gewöhnliche lateinische Alphabet + einige Sonderzeichen wie Punkt, Komma, Blank. Die unendliche Menge D der deutschen Sätze (in geschriebener Form) bildet eine Untermenge von A^* . Die Aufgabe einer deutschen Grammatik besteht dann u.a. darin, D im Rahmen von A^* zu bestimmen. Man kann dazu alle Elemente von D („Sätze“) und kein Element von $A^* - D$ („Nichtsätze“) durch eine Menge von Regeln aufzählen lassen; zweckmäßigerweise wird man die Regeln so wählen, daß dadurch den aufgezählten Elementen zugleich eine linguistisch interessante, d.h. für die Konstitution der Bedeutung aufschlußreiche Strukturbeschreibung zugeordnet wird. Man kann aber auch umgekehrt ein Verfahren zu entwickeln versuchen, das imstande ist, bei einem beliebigen gegebenen Element aus A^* festzustellen, ob es zu D oder zu $A^* - D$ gehört. Ein derartiges Verfahren will ich als „Analysegrammatik“ bezeichnen. Sie stellt ebenso eine Beschreibung der linguistischen Kompetenz dar wie eine gewöhnliche generative Grammatik. Unsere linguistische Kompetenz sagt uns z.B., daß die Folge „Illargia irten da“ ein Element von $A^* - D$ ist, die Folge „Der Mond ist aufgegangen“ hingegen ein Element von D . Selbstverständlich gibt es auch hier Grenzfragen, abweichende Sätze usw. Sieht man von diesen Schwierigkeiten ab, die ja nichts mit dem speziellen Grammatiktyp zu tun haben, sondern aus der linguistischen Kompetenz selbst rühren, dann ist ein solches Entscheidungsverfahren ebenso geeignet, die Satzmenge einer natürlichen Sprache zu charakterisieren wie eine generative Grammatik.

Eine Analysegrammatik untersucht also, ob eine gegebene Graphemfolge⁹⁾ bestimmten Bedingungen genügt, die erfüllt sein müssen, wenn die Graphemfolge zu D gehören soll. Leider kann man nicht sagen, alle Graphemfolgen, die zwischen 54 und 87 Graphemen lang sind, oder die mit „die“ anfangen und mit „hat“ enden, seien Sätze. Man muß auch in einer Analysegrammatik einen Teil vorsehen, der — entsprechend der starken generativen Kapazität — angibt, welche strukturellen Eigenschaften eine Graphemfolge aufweisen muß, um als Satz zu gelten. Mit der Forderung nach *strukturellen* Eigenschaften wird also die Annahme akzeptiert, daß die Angabe statistischer Verteilungen, der Übergangswahrscheinlichkeiten und dergleichen für die schwache generative Kapazität nicht genügt.

Eine Analysegrammatik besteht also in jedem Fall aus zwei Teilen:

- (2) (a) Einem *Spezifikationsteil*, d. h. einer Grammatik im engeren Sinn, in der alle möglichen Strukturbeschreibungen angegeben werden; man kann z.B. die Dudengrammatik¹⁰⁾ als Spezifikationsteil einer Analysegrammatik ansehen; das geschieht ja auch in der Praxis, wenn etwa jemand nachsieht, ob eine bestimmte Konstruktion „korrektes Deutsch“ darstellt, d.h. einen deutschen Satz ergibt.
- (b) Einem *Identifikationsteil*, der jeder gegebenen Graphemfolge aus einem Alphabet A eine (eindeutige Satz), mehrere (mehrdeutige Satz) oder

keine (Elemente von A*—D) Strukturbeschreibungen zuordnet. Beim Nachschlagen in der Dudengrammatik besteht dieser Identifikationsteil weitgehend aus der Intelligenz des Benutzers, seiner Fähigkeit, im Register nachzuschlagen, Analogien zu einem vorhegenden Paradigma zu bilden usw.

Dementsprechend hängt der Wert einer Analysegrammatik im wesentlichen von zwei Faktoren ab:

- (3) (a) Von der linguistischen Adäquatheit des Spezifikationsteils, also des zugrundelegenden Modells. Der Spezifikationsteil muß z.B. explizit sein, er muß nicht nur die erforderlichen strukturellen Eigenschaften zur Charakterisierung der Satzmenge angeben, sondern erklären, wie oder jedenfalls auf welcher Grundlage sich die Bedeutung konstituiert usw.
- (b) Von der Möglichkeit, einen Algorithmus zu konstruieren, der unabhängig von der Intuition eines Sprecher-Hörers gegebene Graphemfolgen „identifiziert“, d.h. ihnen — je nach Modell — syntaktische oder syntaktisch-semantische Strukturbeschreibungen zuordnet.

Analysegrammatiken haben — wie schon gesagt — in der theoretischen Linguistik bislang eine sehr geringe Rolle gespielt. Von eminenten Bedeutung hingegen waren sie im Bereich der linguistischen Datenverarbeitung, wo man es normalerweise mit größeren Textmengen zu tun hat, d.h. mit Folgen meist zusammenhängender Sätze, die für die verschiedensten Zwecke syntaktisch analysiert werden müssen: maschinelle Übersetzung, Abstracting, statistische Untersuchung syntaktischer Erscheinungen, automatische Lemmatisierung, stilistische Textanalyse setzen, wenn sie wirksam betrieben werden sollen, syntaktische Analyse voraus; zumindest wird ihre Effizienz dadurch bedeutend erhöht.

Einen Algorithmus, der eine Rechenanlage in die Lage setzt, beliebige Sätze einer Sprache gemäß einer bestimmten Grammatik syntaktisch zu analysieren, bezeichnet man als „Parser“. Ein Parser ist also eine technische Realisierung des Identifikationsteils einer Analysegrammatik. Seine Brauchbarkeit für linguistische Zwecke hängt nach (3) ab

- (4) (a) vom zugrundegelegten linguistischen Modell, also vom jeweiligen Spezifikationsteil und
- (b) ob und wie es gelingt, Strategien der Identifikation zu entwickeln und zu programmieren. Die Programmierung ist dabei, im Gegensatz zu anderen Problemen der linguistischen Datenverarbeitung (Statistik), vielfach sehr schwierig, einmal weil es kaum brauchbare Programmiersprachen für diese Zwecke gibt, dann aber vor allem, weil linguistische Analysestrategien und Programmorganisation meist kombiniert werden¹¹⁾.

Mit der Konstruktion von Parsern hat man vor etwa 12 Jahren begonnen, als die maschinelle Sprachübersetzung von der lexikalischen in die syntaktische Phase trat¹²). Die meisten Verfahren basieren auf kontextfreien Phrasenstrukturgrammatiken (PSG)¹³), teilweise auf PSG mit komplexen Symbolen¹⁴), auf Abhängigkeitsgrammatiken nach LECERF-HAYS¹⁵), die sich ja in eine spezielle Klasse von PSG überführen lassen¹⁶), auch auf gemischten PSG-Abhängigkeitssystemen¹⁷). Seit einigen Jahren versucht man sich auch an transformationellen Parsern; das bekannteste und meines Wissens bislang beste Verfahren stammt von S. PETRICK¹⁸); es entspricht teilweise schon in den allerdings äußerst einfachen Mustergrammatiken, die zugrunde gelegt werden, dem Stand der „Aspects“, arbeitet mit Formativketten, nicht mit Graphemfolgen und ist bislang über die Analyse einiger nicht sehr komplizierter Fälle nicht hinausgekommen.

Das Hauptproblem bei der Konstruktion transformationeller Parser — übrigens nicht nur transformationeller — liegt in der „lokalen Mehrdeutigkeit“. Es gibt relativ wenig Sätze, die in ihrer Gesamtheit mehrdeutig sind. Sehr häufig aber kann eine bestimmte vorliegende Teilkette auf verschiedene Regeln zurückzuführen sein, insbesondere wenn auch Tilgungstransformationen zugelassen sind. Man muß dann notgedrungen alle möglichen Wege verfolgen, bis man auf einen Widerspruch stößt oder aber den ganzen Satz analysiert hat. Bei mehrfacher Mehrdeutigkeit im Satz — das ist im übrigen der Normalfall — wachsen jedoch die kombinatorischen Möglichkeiten derart rasch, daß ein bloßes „trial and error“ illusorisch ist¹⁹). Man muß daher verschiedene zusätzliche Maßnahmen ausarbeiten, auf die ich hier nicht weiter eingehe²⁰). Sofern das zugrundegelegte Modell eine entscheidbare Sprache liefert, wie es ja bei den PSG der Fall ist, muß es selbstverständlich für jeden Satz eine Analysemöglichkeit geben; und da auch bei einer Analysegrammatik das Kriterium für „Satz einer Sprache“ ist, ob die untersuchte Graphemfolge bestimmte strukturelle Eigenschaften aufweist, muß sich theoretisch jedem Satz eine Strukturbeschreibung zuordnen lassen. Ein Parser, der für einen Satz 100 000 Jahre benötigt, ist aber nicht so recht befriedigend, selbst dann nicht, wenn man weiß, daß er ihn spätestens nach dieser Zeit mit absoluter Sicherheit gelöst haben wird. Erwähnt sei schließlich auch noch, daß ein brauchbarer Parser auch Sätze analysieren muß, die in geringem Maße abweichend sind, die aber von jedem Sprecher „analog“ verstanden werden. Für diesen Fall sind reine „trial-and-error“-Verfahren, die ja auf die Korrektheit und Vollständigkeit der Grammatik spekulieren, vollends unbrauchbar. Ich will aber auf dieses Problem, dem auch mit anderen Methoden äußerst schwer beizukommen ist, trotz seiner großen praktischen und theoretischen Bedeutung nicht weiter eingehen.

4

Seit etwa 1965 wurde am Germanistischen Institut der Universität des Saarlandes von der Arbeitsgruppe „Elektronische Sprachanalyse“, die von Hans EGGERS geleitet und von der DFG finanziell unterstützt wird, ein Parser fürs Deutsche entwickelt. Soviel ich weiß, handelt es sich um den ersten, der in der Tat jeden belie-

bigen deutschen Satz analysiert, ohne daß dazu irgendeine Vorbereitung erforderlich wäre; allerdings arbeitet er derzeit noch auf einer Grundlage, die nicht gerade dem neuesten Stand der linguistischen Forschung entspricht — einer in mancher Hinsicht präzisierten, den spezifischen Bedürfnissen einer programmierbaren Analysegrammatik angepaßten Schulgrammatik. Gravierend ist jedoch weniger die Unzulänglichkeit des Modells an sich, das man mit einiger Boshaftigkeit als „Dampfgrammatik“ bezeichnen könnte, als vielmehr der Umstand, daß seine Eigenschaften nicht hinlänglich genau untersucht und in eine präzise, kohärente Theorie eingebracht sind. Man kann, da ja die sprachlichen Fakten trivialerweise nun einmal so sind wie sie sind, davon ausgehen, daß eine gute einigermaßen vollständige Schulgrammatik fast alle Informationen enthält, die zur Beschreibung der betreffenden Sprache erforderlich sind. Unzulänglich ist bloß die Art und Weise ihrer Darstellung²¹). Ähnliches gilt, wenigstens bis jetzt, für den Saarbrücker Parser. Es kommt weniger darauf an, das Verfahren selbst zu ändern als vielmehr darauf, die aufgefundenen Ergebnisse — die Aussagen über die Struktur der analysierten Sätze — in eine geeignete Form der Repräsentation zu bringen. Was also hier zu leisten ist, entspricht im Bereich der Analysegrammatik etwa dem, was CHOMSKY in Kap. 2, §§ 1–3 der „Aspects“ für die generative Transformationsgrammatik unternimmt²²).

Ich will nun den Saarbrücker Parser kurz zu skizzieren versuchen. Eine ziemlich ausführliche Darstellung findet sich in dem Buch „Elektronische Syntaxanalyse der deutschen Gegenwartssprache“ von Hans EGGERS und Mitarbeitern²³). Außerdem liegen zahlreiche Arbeiten zu Einzelproblemen vor, die größtenteils in den „Linguistischen Arbeiten des Germanistischen Instituts und des Instituts für Angewandte Mathematik der Universität des Saarlandes“ (1965 ff., 1-8, hektographiert) veröffentlicht wurden.

5

Der „Spezifikationsteil“ des Parsers in seiner bisherigen Form²⁴) läßt sich am besten als ein Modell mit 5 Analyseebenen oder „Reduktionsstufen“ veranschaulichen²⁵). Wir haben sie als Wortformenebene, Wortklassenebene, Gruppenebene, Subsatzebene, Satzebene bezeichnet; die drei letzten entsprechen etwa den Termini „phrase level, clause level, sentence level“, die beiden erstgenannten sind klar.

5.1. Wortformenebene

Als einen *möglichen* Satz - d.h. als Element von A^* - betrachten wir jede Graphemfolge zwischen Punkten (oder vergleichbaren „satzschließenden“ Zeichen). Eine derartige Graphemfolge enthält in der Regel eine Reihe von „blanks“ (Leerzeichen, Zwischenräumen). Die Graphemfolgen zwischen diesen „blanks“ nennen wir „Wortformen“. Sie bilden den Ausgangspunkt der Analyse, weil zu vermuten ist, daß sie für die Strukturbeschreibung des Satzes eine besondere Relevanz besitzen. Prinzipiell könnte man den „Satz“ auch in Folgen à 7 Grapheme + Rest zerlegen; es ist aber nicht eben anzunehmen, daß diese Einheiten in den Wortformen vergleichbarer Weise syntaktische Eigenschaften des „Satzes“ reflektieren würden. Im

ersten, noch ganz trivialen Analyseschritt wird also der gesamte zu analysierende Text in Graphemfolgen zwischen „sentence boundaries“, also in mutmaßliche Sätze zerlegt; die Sätze, wie ich in Zukunft der Einfachheit halber statt „mögliche Sätze“ sagen will, werden ihrerseits in Wortformen segmentiert.

5.2. Wortklassenebene

Anschließend werden die einzelnen Wortformen mit Hilfe eines sogenannten „syntaktischen Wörterbuchs“ bestimmten Wortklassen zugewiesen, z.B.²⁶⁾ VATER, VATERS, PFERDE, TRANSPORTARBEITERINTERNATIONALE der Wortklasse „Substantiv“, HOLTE, GIBT, MOGELE der Klasse „flektiertes Verb“ usw.

Ich will diese Hauptaufgabe des zweiten Analyseschritts nun etwas präziser formulieren. Es wird angenommen, daß der Spezifikationsteil eine Liste von Paaren (W_1, I_1) , (W_2, I_2) , ..., (W_n, I_n) enthält; die gesamte Liste nennen wir syntaktisches Wörterbuch; ein beliebiges Paar (W_i, I_i) daraus besteht aus einer Wortform W_i — genauer: einem Wortformen-Type sowie einer Menge syntaktischer Merkmale oder Informationen I_i . Die wichtigste dieser Informationen ist eine Wortklassenangabe (insgesamt werden 18 Wortklassen unterschieden); weitere beziehen sich auf Genus, Numerus, Kasus, Transitivity, nominale Rektion usw.²⁷⁾. Ähnlich wie die syntaktischen Merkmale bei den Lexikoneintragen einer Transformationsgrammatik bestimmte Gegebenheiten der Ableitungsgeschichte reflektieren, können bei der Analysegrammatik umgekehrt aus I_i Rückschlüsse auf die zu erwartende Struktur gezogen werden.

Technisch gesehen ist das Wörterbuch alphabetisch nach den Wortformen geordnet auf Magnetband gespeichert. Es umfaßt derzeit ungefähr 40 000 Eintragungen. Zur Analyse werden die im Satz aufgefundenen Wortformen mit den Wörterbucheintragen verglichen und bei Übereinstimmung mit dem Wortformen-Typen durch die jeweiligen Informationen ersetzt. Bei Wortformen, die noch nicht im Wörterbuch stehen, wird angenommen, daß sie zu einer der drei Wortklassen „Substantiv (Name), Adjektiv, flektiertes Verb“ gehören, da hier am ehesten mit einer Unvollständigkeit des Wörterbuchs zu rechnen ist²⁸⁾.

Das größte Problem auf dieser Analysestufe, vielleicht sogar der maschinellen syntaktischen Analyse überhaupt, stellt die Wortklassenmehrdeutigkeit oder *Homographie* dar. Es kommt sehr häufig vor, daß eine Wortform, die im Text auftritt, zu mehreren Klassen gehören könnte. Beispiele sind etwa OEDEM als Substantiv Neutrum vs. OEDEM als Dativ Sg. des Adjektivs OED oder DIE, das Artikel und Relativpronomen sein könnte. Wir haben diese Fälle inventarisiert und zu etwa 50 Homographenklassen angeordnet²⁹⁾. Bei den entsprechenden Wortformen steht im Wörterbuch statt der üblichen grammatischen Informationen ein Hinweis auf die betreffende Homographenklasse, z.B. (DIE, HO 43), wobei HO(mographenklasse) 43 bedeutet: Artikel *oder* Relativpronomen. Dieser Hinweis „HO 43“ gibt gleichzeitig an, welcher Weg zur Lösung der betreffenden Homographie im Programm einzuschlagen ist. Ursprünglich waren für alle Homographentypen eigene Homographen-Unterprogramme vorgesehen; inzwischen sind sie jedoch in ein höchst komplexes

System integriert, das nicht nur ökonomischer, sondern auch viel wirksamer arbeitet. Es operiert vor allem mit der Stellung, den Informationen der umgebenden Wortklassen und einer differenzierten Hierarchie von „Inventarforderungen“, z.B. darf bei einem Verb nur *ein* (nichtkoordiniertes) Substantiv im Nominativ stehen, wenn es nicht ausdrücklich anders gekennzeichnet ist (wie etwa bei **WERDEN, BLEIBEN, SEIN** usw.), auf eine Präposition folgt immer eine nominale Wortklasse usw

Von den Schwierigkeiten der Homographie kann man sich erst einen gewissen Begriff machen, wenn man berücksichtigt, daß Homographen normalerweise gehäuft auftreten. Nach unserer Wortklasseneinteilung ist fast jedes zweite Wort im Text (als „token“) homograph (genau: 43 % im Durchschnitt). Deshalb kann man auch nicht einen Homographen so einfach um den anderen lösen, sondern muß mit einem sophistischen, kombinierten System arbeiten, das abwechselnd an verschiedenen Stellen operiert. Etwa 94 % der Homographen werden richtig gelöst³⁰⁾. Im zweiten Analyseschritt werden also die Wortformen, in die der Satz segmentiert wurde, durch die korrespondierenden syntaktischen Informationen, insbesondere durch die Wortklassenangabe, ersetzt. Aus diesen Angaben muß nun die weitere syntaktische Struktur erschlossen werden.

5.3. Gruppenebene

Wir unterscheiden drei Arten von Gruppen oder „Phrasen“: nominale, verbale, „sonstige“ — das sind im wesentlichen adverbiale, daneben auch Interjektionen u.a. Sie sind intern noch einmal stark differenziert; auf Einzelheiten kann ich hier nicht eingehen³¹⁾. Definiert sind sie jeweils als Folgen bestimmter Wortklassensymbole. Bei der Analyse wird demnach versucht, Folgen von Wortklassenangaben zu Gruppen zusammenzufassen; so wird etwa die Folge „Artikel-Adjektiv-Substantiv“ als nominale Gruppe Typ xyz erkannt, sofern bestimmte Bedingungen hinsichtlich der übrigen syntaktischen Informationen gegeben sind (etwa Übereinstimmung in Genus, Numerus und Kasus). Auch hier können vielfach Mehrdeutigkeiten auftreten (das Wort WAGEN z.B. kann, von der in diesem Stadium bereits gelösten Mehrdeutigkeit Verb-Substantiv abgesehen, als Substantiv alle Kasus und Numeri außer Genitiv Singular sein). Es sind also auch hier besondere Analysetechniken erforderlich, die man aber schon seit langem beherrscht³²⁾ und die auch von uns angewendet werden³³⁾. Weitaus gravierender ist das Problem der „Diskontinuität“ von Gruppen³⁴⁾. Eine Folge von Wortklassenangaben, die zu einer Gruppe zusammenfaßbar sind, kann durch eine andere Folge unterbrochen sein, die gleichfalls eine Gruppe oder gar einen Subsatz ergibt. Im Deutschen sind diese Fälle überaus häufig:

- (5) (a) WATZLAFF HAT SEINE FRAU BETROGEN.
(b) WATZLAFF HAT, MEHERE WOCHEN NACHDEM ER VON EINER ZWEIMONATIGEN ERHOLUNGSREISE NACH KLONDIKE ZURÜCKGEKEHRT WAR, SEINE FRAU UM IHRE GESAMTHABE, NÄMLICH 40 000 LEWONZEN, DIE IHR VOR KURZEM EIN UNGENANNTER WOHLTÄTER AUS NICHT SEHR DURCHSICHTIGEN GRÜNDEN ÜBERLASSEN HATTE, BETROGEN.
- (6) ER IST, WIE MIR SCHEINT, BESCHRÄNKT.

- (7) EIN, SO HÖRTE ICH EINEN MIR BEFREUNDETEN KRITIKER,
DEN ICH ALLERDINGS FÜR ETWAS BESCHEUERT HALTE,
SAGEN, SEHR DICKES BUCH ...
- (8) EIN AUSWÄRTS TÄTIGER MONTEUR ...
- (9) DIE BEISPIELE, DIE ICH ANFÜHRE, SIND ALBERN.

Beispiel (5) (b) ist nur eine Erweiterung von (5) (a); es ist klar, daß man sie unendlich weit fortspinnen könnte; eben dieser Umstand, nicht die Diskontinuität an sich wie in der Phrasenstrukturgrammatik, wo ohne komplexe Symbole weder (5) (a) noch (5) (b) zu erfassen sind, stellt für die Analyse ein Problem dar.

Von den beiden Fällen (6) und (7), also Einbettung parenthetischer Subsätze in verbale bzw. nominale Gruppen (die in sich natürlich wiederum diskontinuierlich sein können), ist in unserem Verfahren nur (6) immer lösbar, (7) vorerst nur, wenn der Subsatz eine bestimmte Struktur aufweist. Derartige Einbettungen sind auch theoretisch äußerst problematisch, da sie praktisch an jeder Stelle im Satz auftreten können und mit den üblichen Einbettungstransformationen nicht befriedigend zu erklären sind³⁵). Fall (8), also adverbiale Gruppe innerhalb einer nominalen, ist ziemlich unproblematisch; in Wirklichkeit ist das Adverb natürlich zunächst einmal von dem Verbal TÄTIG abhängig; das ist aber ein Problem des Spezifikationsteils und nicht der Identifikation.

Beispiel (9) geht über den engeren Zusammenhang hinaus, denn hier ist ein Subsatz in einen anderen Subsatz eingebettet; der Fall gehört mithin auf die nächste Analysestufe. Ich führe ihn an, um noch deutlicher zu machen, daß bei der Behandlung der Diskontinuität scharf zwischen dem linguistischen Phänomen und seiner angemessenen Beschreibung einerseits und der Behandlung im Identifikationsteil der Analysegrammatik andererseits zu trennen ist. Der Relativsatz ist Teil der nominalen Gruppe; insofern liegt bei (9) überhaupt keine Diskontinuität vor; doch unterbricht der Relativsatz in der Hierarchie der Analyse eine Einheit, die mit ihm auf gleicher Ebene steht und die wegen dieser Unterbrechung nicht sofort zusammenzufassen ist; der Subsatz, der mit DIE BEISPIELE begonnen hat, muß irgendwo "jenseits des Kommas" fortgeführt werden; bei (9) ist leicht zu erkennen, wo er weitergeht, weil nur noch ein passender Teil innerhalb des Satzes vorhanden ist: SIND ALBERN. Mit welchen Schwierigkeiten hier die Identifikation, weniger die linguistische Beschreibung, zu rechnen hat, illustrieren die folgenden Beispiele³⁶):

- (10) (a) ICH MÖCHTE EINEN ELEFANTEN.
(b) ICH MÖCHTE EINEN ELEFANTEN VERKAUFEN.
(c) ICH MÖCHTE EINEN ELEFANTEN, DEN ICH GESTOHLLEN HABE, VERKAUFEN.
- (11) ER IST, WIE SIE ANGEKÜNDIGT HATTE, GEKOMMEN, WAR ABER, WENN DER ARZT, DEN ICH GEFRAGT HABE, RECHT HAT, VERLETZT.

Sie zeigen, daß das Problem, um das es hier geht, nur teilweise mit dem bekannten Phänomen der Diskontinuität zusammenhängt; man könnte es als „Fragmentation“ bezeichnen: ein Satz „zerfällt“ in eine Reihe kleinerer Einheiten, die sich in bestimmter Weise zu Gruppen und Subsätzen ergänzen lassen. Grenzanzeiger für diese Einheiten, die wir als „Analyseeinheiten“ bezeichnen, sind Kommata, bestimmte Konjunktionen u.a. Auf die Einzelheiten der Abgrenzung und ihre Motivierung im einzelnen Fall kann ich hier nicht eingehen³⁷). Die Analyseeinheiten lassen sich nach der Art der darin enthaltenen Elemente und ihrer Stellung zu Typen klassifizieren, die sich nach bestimmten Regeln zusammenfügen lassen. In (10) (c) sind diese Einheiten

- (12) (a) ICH MÖCHTE EINEN ELEFANTEN
 (b) DEN ICH GESTOHLEN HABE
 (c) VERKAUFEN

Typ (12) (a) und (b) sind potentiell bereits vollständig, wobei (b) jedoch im Gegensatz zu (a) keinen vollständigen Satz bilden kann; Typ (12) (c) ist nicht vollständig, muß also irgendwo angeschlossen werden, damit sich sowohl eine korrekte Gruppe (MÖCHTE VERKAUFEN) wie ein korrekter Subsatz (ICH MÖCHTE EINEN ELEFANTEN VERKAUFEN) ergeben. Viel komplizierter ist (11); die Wichtigkeit der Fragmentation für die Analyse zeigt sich, wenn man in (11) die Kommata, die sie anzeigen, wegläßt:

- (11!) ER IST WIE SIE ANGEKÜNDIGT HATTE GEKOMMEN WAR ABER
 WENN DER ARZT DEN ICH GEFRAGT HABE RECHT HAT VER-
 LETZT.

Die Zerlegung in „Analyseeinheiten“ und ihre Klassifikation ermöglicht es hingegen, die vorliegenden Gruppen und Subsätze zusammenzufügen; dabei ist, wenn mehrere Einheiten zusammenfaßbar sind, eine bestimmte Hierarchie einzuhalten³⁸).

Ich habe betont, daß die Fragmentation innerhalb unseres Analyseverfahrens lediglich eine technische Funktion hat; sie wird bislang jedenfalls nicht als der Diskontinuität vergleichbares linguistisches Phänomen betrachtet. Es ist aber durchaus nicht auszuschließen, daß sie auch eine theoretische Relevanz besitzt. Mit Hilfe der Langackerschen "command-Relation"³⁹) kann man etwa die folgende Hypothese formulieren: ein Satz S weist immer dann eine Fragmentation auf, wenn (a) S eine Gruppe A und einen Subsatz S' dominiert und (b) A "commands" S'. Ich habe bislang keine Gegenbeispiele finden können; Genauereres kann dazu aber erst gesagt werden, wenn der Spezifikationsteil unseres Parsers entsprechend umgearbeitet ist.

5.4. Subsatzebene

Über die Identifikationsprobleme, die bei der Zusammenfassung der einzelnen Gruppen zu Subsätzen auftreten, wurde bereits einiges gesagt. Wir nehmen wie üb-

lich an, daß ein Subsatz in der Regel aus einer verbalen Gruppe sowie 1–3 nominalen Gruppen, deren Anzahl und Kasus der Parser partiell den Informationen des Verbs entnehmen kann, besteht. Das Problem im allgemeinen ist klar; die Schwierigkeiten bei der Identifikation im einzelnen haben vor allem drei Ursachen:

- (13) (a) Fragmentation; sie wurde bereits erörtert.
- (b) Mehrdeutigkeit ganzer Gruppen; sie kann mit der sogenannten Rektion zusammenhängen, etwa in
- (i) ER BRACHTE DIE BLUMEN DER FRAU.
Erstens wichtiger und zweitens nur in bestimmten Fällen lösbar sind die „präpositionalen Gruppen“ in Sätzen wie
- (ii) ER BETRACHTET DIE KINDER MIT DEM FERNGLAS.
(iii) ER BETRACHTET DIE KINDER MIT DEN ROTEN MÜTZEN.
Zumindest (iii) kann nur in einer Weise interpretiert werden; es ist aber sehr schwierig — und uns bisher nicht gelungen — einen Parser mit irgendwelchen Informationen zu versehen, die ihm diese und nur diese Interpretation zu finden erlauben.
- (c) Tilgungen von Gruppen, etwa verblose Sätze (sogenannte „Setzungen“), absoluter Gebrauch transitiver Verben usw.; alle dementsprechenden Forderungen an das Inventar einer Subgruppe müssen im Parser als fakultativ angesehen werden, anders gesagt: alle Spezifikationen sind negativ; so ist z.B. SEHEN nicht als [+ NP (Akk.)]⁴⁰, sondern als [- NP (Dat.), — NP (Gen.)] spezifiziert. Weitaus einschneidender und demnach auch schwieriger zu behandeln ist z.B. eine Verbelimination; daß dieser Fall vorliegt — er ist im übrigen alles andere als selten⁴¹ —, wird erst angenommen, wenn bei der Auflösung von Mehrdeutigkeiten alle anderen Lösungsmöglichkeiten ausscheiden.

5.5. Satzebene

Die Subsätze bestehen also aus bestimmten Gruppen in bestimmter Anordnung; zu ihrer näheren Charakterisierung kann u.a. die Typologie der Analyseeinheiten benutzt werden, da ja ein Subsatz aus einer oder mehreren Analyseeinheiten besteht. Ein Satz (sentence) baut sich aus Subsätzen verschiedener Art auf — im Grenzfall „Einfachsatz“ aus einem. Zur Charakterisierung dieser einzelnen Teile des Satzes verwenden wir traditionelle Termini wie „Hauptsatz“ (ein S also, das nicht von einem andern S dominiert wird), Temporalsatz, Finalsatz, Objektsatz, Prädikativsatz, Relativsatz usw. Die Position dieser Einheiten im Satzkomplex ist sehr unterschiedlich, und die Typisierung der einzelnen Subsätze mit diesen tradierten Termini reflektiert zwar diese Position innerhalb der kategorialen Komponente der Grammatik (bzw. jenem Teil, der der kategorialen Komponente entspricht), aber doch sehr unzulänglich. Hier muß der Spezifikationsteil der Analysegrammatik — nicht der Parser an sich — noch bedeutend umgearbeitet werden. Für die Resultate der Analyse, die diese Terminologie eher verschleiert als klar kennzeichnet, muß eine

geeignete Form der Repräsentation im Rahmen einer formalen Grammatik gefunden werden; dem formalen Identifikationsteil muß ein formaler Spezifikationsteil entsprechen.

6

Einige technische Nachbemerkungen: Das Programm läuft derzeit auf einer Elektronica X1, einer relativ kleinen Wortmaschine (20 480 Worte zu 27 bit im Kernspeicher) der zweiten Generation. An Externspeichern sind 3 Magnetbandeinheiten vorhanden⁴²⁾. Als Programmiersprache steht lediglich Algol zur Verfügung, das sich für linguistische Zwecke nicht sonderlich eignet. Das gesamte Programm ist daher im Assembler geschrieben; es umfaßt etwa 16 000 Befehle; zur Analyse eines Satzes benötigt es (bei einer Zykluszeit von 85 MikroSekunden) etwa 3 Sekunden (ohne Ausgabe auf Schnelldrucker)⁴³⁾.

Derzeit wird eine ausführliche Dokumentation in Flußdiagrammform vorbereitet; auf dieser Basis soll das Programm mit einer Reihe von Verbesserungen auf eine größere Rechanlage umgestellt werden, die das Rechenzentrum der Universität des Saarlandes im Laufe dieses Jahres erhalten wird.

Anmerkungen

- 1) Vgl. etwa H. Glinz, *Deutsche Syntax*, Stuttgart 1965 u. ö., S. 96 oder G. Herdan, *Götzendämmerung at MIT*, ZPhSK 21, 1968, S. 216-232; weiterhin Chomsky, *Aspekte der Syntax-Theorie*, Frankfurt 1969, S. 20 f. (ich zitiere nach der deutschen Ausgabe, rede aber im Text von "Aspects", weil diese Bezeichnung inzwischen zum Begriff geworden ist).
- 2) Es ist für diese Diskussion ziemlich unwichtig, ob man unter Satz die Endkette der syntaktischen Komponente oder deren phonologische bzw. graphematische Interpretation versteht; ich gehe daher auf diesen Punkt nicht weiter ein.
- 3) Aspekte S. 19.
- 4) Vgl. etwa auch Aspekte S. 48, (14) (iv) und die dazugehörige Anmerkung 18, S. 251/2; diese Anmerkung ist zweifellos zutreffend, aber irreführend, wenn man sich nicht die Form der Grammatik vor Augen hält.
- 5) Es wäre also sinnvoller zu sagen, daß jeder Strukturbeschreibung genau ein Satz zugeordnet wird; nur dann kann ja technisch gesehen von einer Funktion die Rede sein.
- 6) Vgl. Aspekte S. 84/85, *Formal Properties of Grammars* (in Luce-Bush-Galanter, *Handbook of Mathematical Psychology II*, New York 1963, S. 323-418), S. 325/6.
- 7) Aspekte S. 20.
- 8) G. Lakoff, *On Generative Semantics*, S. 1 (1969; zit. nach der Paperversion des Indiana University Linguistic Circle).
- 9) oder Phomenfolge; ich diskutiere hier der Einfachheit halber immer Graphemfolgen.
- 10) Mannheim 1966.

- ¹¹⁾ Neuerdings versucht man allerdings, Parsing-Systeme mit List-Struktur zu entwickeln, bei denen die Analysestrategien vom Programm selbst unabhängig sind und - ähnlich wie die zu analysierenden Sätze von Fall zu Fall eingegeben werden können; das hat vor allem den Vorteil, daß man nicht an einen Spezifikationsteil gebunden ist, sondern innerhalb eines bestimmten Grammatiktyps variieren kann; vgl. z. B. R. Banerji, *Some Studies in Syntax-Directed Parsing*, in P. Garvin, B. Spolsky, eds., *Computation in Linguistics*, Bloomington 1966, S. 76-123.
- ¹²⁾ Vgl. dazu E. D. Pendergraft, *Translating languages*, In H. Borko, ed., *Automated Language Processing*, New York 1967, S. 292-323 sowie im gleichen Band S. 217-251 den informativen, wenn auch etwas einseitigen Überblick von D. Bobrow, *Syntactical Theories in Computer Implementations*.
- ¹³⁾ Ein Vergleich der verschiedenen Verfahren findet sich in T. V. Griffith, S. R. Petrick, *On the Relative Efficiencies of Context-free Grammar Recognizers*, in Comm. ACM 8, 1965, S. 289-300.
- ¹⁴⁾ So z. B. der Parser-Teil bei W. Tosh, *Syntactic Translation*, Den Haag 1965; siehe auch V. H. Yngve, *MT at M.I. T.* 1965, in A. D. Booth, ed., *Machine Translation*, Amsterdam 1967, S. 451-523.
- ¹⁵⁾ Am systematischsten nachzulesen in D. Hays, *Dependency Theory: a Formalism and some Observations*, in Language 40, 1964, S. 511-525.
- ¹⁶⁾ Dafür gibt es zwei Beweise: Ch. Gaifman, *Dependency Systems and Phrase Structure Systems*, in *Information and Control* 8, 1965, S. 304-337, und - allerdings nur auf die schwache generative Kapazität bezogen - M. Gross, *On the equivalence of models of language used in the field of mechanical translation and information retrieval*, in *Information Storage and Retrieval* 2, 1964, S. 43-57.
- ¹⁷⁾ Z. B. bei O. S. Kulagina, I. A. Mel'čuk, *Automatic Translation: some theoretical aspects and the design of a translation system*, in A. D. Booth, ed., *Machine Translation*, Amsterdam 1967, S. 137-171.
- ¹⁸⁾ S. R. Petrick, *A Recognition Procedure for Transformational Grammars*, M. I. T. Dissertation 1965 (185 S.).
- ¹⁹⁾ So hat Petrick ausgerechnet, daß bei einer äußerst einfachen Grammatik die Analyse eines 15-Wort-Satzes ohne zusätzliche Analysestrategien auf einer IBM 7090 einige Milliarden Jahre dauern würde.
- ²⁰⁾ Speziell mit diesem Problem befaßt sich D. Varga, *Problems of improving the efficiency of parsing methods*, Int. Conf. Ling. Stockholm 1969, Preprint Nr. 61, 30 S.
- ²¹⁾ Diese Behauptungen sind etwas vereinfachend; selbstverständlich gibt es auch Gesetzmäßigkeiten, die in den üblichen Schulgrammatiken à la Grevisse, Zandvoort, Duden nicht nur nicht adäquat, sondern überhaupt nicht erfaßt werden.
- ²²⁾ Aspekte S. 88/89, bes. S. 89 unten.
- ²³⁾ Tübingen 1969. Die Mitarbeiter sind: R. Dietrich, R. Rath, A. Rothkegel, H. J. Weber, H. Zimmermann und der Verfasser dieses Artikels. Von den zahlreichen Mitgliedern dieser Gruppe war außerdem V. Wolf direkt an der Entwicklung des Parsers (in einem früheren Stadium) beteiligt. - Die „Elektronische Syntaxanalyse“ wird künftig mit ES + Seitenzahl zitiert.
- ²⁴⁾ In der theoretischen Einleitung zu ES - S. 20-37, bes. S. 28-37 - habe ich die Ansicht vertreten, es handle sich bei diesem Modell um eine taxonomische Grammatik; das stimmt höchstens in einem sehr allgemeinen Sinn; um eine PSG, die ja in der Regel als Präzisierung einer taxonomischen Grammatik genommen wird, handelt es sich keineswegs. Ich kann leider hier nicht näher auf diesen Punkt eingehen.
- ²⁵⁾ Ich sage mit Absicht „veranschaulichen“; dieses Modell ist im Augenblick nur eine Orientierungshilfe.

- 26) Da bei unserem Parser aus technischen Gründen der Unterschied zwischen Groß- und Kleinschreibung nicht berücksichtigt werden kann, verwende ich auch in den folgenden Beispielen nur Majuskeln.
- 27) Siehe dazu im einzelnen ES 56-61,179.
- 28) Für das auch theoretisch sehr interessante Problem der „unbekannten“ Wörter (welche Namen gehören zum Lexikon einer Sprache?) wird derzeit von H. Zimmermann ein lernendes Verfahren entwickelt; es betrifft natürlich nur die sogenannten „produktiven Wortklassen“ (vgl. etwa BARZELN, AUFMÜPFEN), vor allem aber auch Komposita, zu deren mechanischer Zerlegung eigene Regeln auszuarbeiten sind.
- 29) ES 62-89.
- 30) Stand Februar 1969. Nicht berücksichtigt sind Korrekturen, die in späteren Teilen des Programms erfolgen. Inzwischen hat H. J. Weber eine Reihe von Verbesserungen eingeführt, deren Auswirkung aber noch nicht an einem größeren Sample überprüft wurde.
- 31) Vgl. dazu R. Rath, *Zur syntaktischen Analyse nominaler Gruppen der deutschen Gegenwartssprache*, in BLI 9, 1966, S. 7-28 (stellt eine frühere Version dar) sowie ES 99-110 (nominale Gruppen), 110-115 (verbale Gruppen); die linguistische Beschreibung der verbalen Gruppen steckt erst in den Anfängen.
- 32) Siehe etwa H. Schnelle, *Neue Aspekte in der Theorie des Übersetzens*, in STZ 23,1967, S. 239-248, Schnelle-Engelien, *Maschinelle Sprachübersetzung*, in R. Gunzenhäuser, ed., *Nicht-numerische Informationsverarbeitung*, Wien 1968, S. 314-330.
- 33) ES 101,133-142.
- 34) ES 90-99, 116-132 sowie H. Eggers, W. Klein, R. Rath, A. Rothkegel, H. J. Weber, H. Zimmermann, *Die automatische Behandlung diskontinuierlicher Konstituenten im Deutschen*, in Muttersprache 79, 1969, S. 260-266.
- 35) Ross hat jüngst in "A first Look at Parentheticals" (unveröffentlichter Vortrag, Paris, April 1970; ich kenne nur Ross' Abstract) argumentiert, diese parenthetischen Sätze seien umgekehrt die eigentlichen Matrixsätze („Mir scheint, daß er etwas beschränkt ist“). Das scheint mir nicht sonderlich befriedigend, weil man dann offenbar auch ER IST ANSCHEINEND ETWAS BESCHRÄNKT, ER IST ALLEM ANSCHEIN NACH ETWAS BESCHRÄNKT - überhaupt alle „kommentierenden“ Adverbialkonstruktionen - analog behandeln müßte - und nicht nur diese.
- 36) Vgl. dazu auch ES 35, Diskontinuierliche Konstituenten (s. Anm. 34) S. 263.
- 37) ES 91-99.
- 38) ES 130 - 131, Diskontinuierliche Konstituenten S. 265.
- 39) R. Langacker, *On Pronominalization and the Chain of Command*, in: Reibel-Schane, eds., *Modern Studies in English*, Englewood Cliffs 1969, S. 160-186.
- 40) Die Stellung eines Objektes in der Oberflächenstruktur kann natürlich nicht in einer Lexikoneintragung angegeben werden.
- 41) In einem Korpus von 11 000 Sätzen, an dem unser Parser ausgearbeitet wurde, fehlt bei ca. 2,5 % der Sätze ein Verb.
- 42) Die Einrichtungen des Saarbrücker Rechenzentrums konnten wir mit der freundlichen Erlaubnis und der Unterstützung seines Direktors Prof. Dr. J. Dörr benutzen.
- 43) Das gesamte Programm stammt von Mitarbeitern unserer Gruppe, vor allem von Herrn H. Zimmermann. Wir danken den Mitarbeitern des Rechenzentrums für ihre Beratung.
- 44) Dieser Aufsatz wurde in einer vorläufigen Fassung von der Saarbrücker Arbeitsgruppe diskutiert Ich danke allen Mitgliedern, insbesondere R. Dietrich, R. Rath und H. Zimmermann für Ratschläge und Kritik.

Aus dem Inhalt der erschienenen Hefte

Heft 4: Werner Abraham und Robert I. Binnick: **Syntax oder Semantik als erzeugende Komponenten eines Grammatikmodells?**

Alfred Schopf: Die sogenannten Verben ohne Verlaufsform im Englischen

Klaus-Henning Schroeder: **Rumänische Linguistik heute**

Herwig Krenn und Klaus Müllner: **Linguistik und Datenverarbeitung**

Manfred W. Hellmann: **Außenstelle Bonn des Instituts für deutsche Sprache**

Henning Wode: **Freiburger Sprachpathologie**

R. G. van de Velde: Zur Linguistisierung der Philologie

Wolfgang Dressler: Theorie und Empirie in der Sprachwissenschaft

Frank Achtenhagen: Didaktik des fremdsprachlichen Unterrichts und Linguistik - Überlegungen zu den Möglichkeiten und der Notwendigkeit einer Kooperation

Heft 5: Ernst Wolfgang Orth: **Grundlagen- und Methodenprobleme der Sprachwissenschaft vom Standpunkt der Transzendental-Phänomenologie**

Stanko Žepić: **Zum Verhältnis von Komposition und Derivation**

Hans-Jürgen Heringer: **On Mr. Thümmel's zwei und drei ist vier**

Dieter Messner: **Kritische Bemerkungen zur Lusitanistik**

R. O. U. Strauch: Linguistik an der Universität Nottingham

Peter Bischoff: **Gedanken zur Rolle der Sprachwissenschaft für die Literaturforschung**

Peter Hartmann: Germanistik aus linguistischer Sicht

Alex Ströbl: Komponenten - Inhalt - Ausdruck

Karl-Dieter Bunting: **Wissenschaftliche und pädagogische Grammatik (Sprachwissenschaft und Sprachlehre)**