

URSULA KLENK (Hrsg.)

STRUKTUREN UND VERFAHREN
IN DER
MASCHINELLEN SPRACHVERARBEITUNG

Copyright 1985
AQ-Verlag
Beim Weisenstein 6
D-6602 Dudweiler

Druck: Maro-Druck
D-8900 Augsburg

ISBN 3-922441-46-7
ISSN 0724-3103

AQ-VERLAG

BLACKBOARDMODELLE FÜR DIE ANALYSE NATÜRLICHER SPRACHEN
(am Beispiel der Analyse von Relativsätzen)

Heinz-Dirk Luckhardt

Stratifikation von Analyseprozessen gilt manchem (Computer-) Linguisten als Beweis für die Rückständigkeit von Parsern. Dabei denken die meisten an Systeme, in denen die Analyse in sequentiell ablaufende Analyseschritte (= Pass) aufgeteilt ist, die in ihrer Abfolge festliegen, voneinander nicht unmittelbar Notiz nehmen, und evtl. voreilige Entscheidungen fällen, weil sie nicht genügend Informationen haben. Diese Strategie entspricht dem Stand der Wissenschaft zu der Zeit, als diese Systeme in die Entwicklung gingen, und ermöglicht einen gewissen Prozentsatz an korrekten Analysen (je nach Gegebenheiten bis zu 90%), der für bestimmte Zwecke ausreicht (aus der Sicht der Anwender) bzw. zu einer gewissen Fehlerquote führt, die es zu verringern gilt (aus der Sicht der Wissenschaft).

Aus dem Bestreben, diese Fehlerquote zu verringern, sowie der Notwendigkeit, Parser für neue Aufgaben im Rahmen der KI zu entwickeln, zieht die Computerlinguistik einen großen Teil ihrer Daseinsberechtigung. Dabei gibt es bezüglich der Haltung gegenüber älteren Systemen zwei grundsätzlich falsche Standpunkte:

- a. ignorieren
- b. studieren und immer genau das Gegenteil tun.

Eine falsche Konsequenz aus b. wäre es nun, z.B. die Stratifikation von Prozessen zu verdammen, weil Prozesse in älteren Systemen stratifiziert sind. Daß Analyseprozesse grundsätzlich stratifiziert sein müssen, will dieser Aufsatz zeigen.

Dieses Vorhaben muß zu keinerlei Kontroversen führen, wenn man bedenkt, was Stratifikation bedeutet. Laut WINOGRAD (1983: 16ff) sind Sprachmodelle für den Computer grundsätzlich stratifiziert:

"... stratification is a natural organisation for any complex mapping process, since it allows the mapping to be organized into stages ... Computational models are also based on the idea that a complex process

can be decomposed into a collection of simpler processes, each operating to some extent independently of the others ... By operating in levels, the system has the advantage of modularity. Modularity is extremely important for a system to be flexible and expandable, ... A system organized in sequential stages is 'brittle'. ... It is possible to maintain the advantage of stratification while increasing the flexibility of a system if we separate the processing sequence from the structural levels. We can think of ... assigned structures as all being written on a **blackboard** and the component processes (often called knowledge sources) as reading information from that blackboard and writing new information onto it. The order in which things are added to the blackboard can be determined by the availability of information to any one of the processes, and a process need not be limited to reading only what was written by its upper neighbour."

Ich will Winograds Argumente kurz zusammenfassen:

1. Stratifikation ist eine natürliche Organisationsform für komplexe Abbildungsprozesse.
2. Komplexe Aufgaben werden in einfache unterteilt.
3. Die verschiedenen Prozesse arbeiten in gewissem Sinne unabhängig voneinander.
4. Modularität bringt Flexibilität und Erweiterbarkeit.
5. Ein strikt sequentielles System ist "spröde", sollte also vermieden werden.
6. Prozesse und Strukturen sollten voneinander getrennt sein.
7. Die Reihenfolge, in der Strukturen ermittelt werden, hängt davon ab, welche Informationen den Prozessen zur Verfügung stehen.
8. Ein Prozeß muß nicht auf die Bearbeitung dessen beschränkt bleiben, was sein Vorgänger produziert hat.

Wenn wir dies als allgemeine Beschreibung des Inhalts von Stratifikation akzeptieren, dann ist Stratifikation im Saarbrücker SUSY-II-Parser optimal verwirklicht, da alle 8 Aspekte berücksichtigt sind (vgl. MAAS 1984). In seinem Vorgänger SUSY fehlt zur Verwirklichung eines optimalen Konzepts die Realisierung der Punkte 5. und 8., da die Hauptprozesse strikt sequentiell ablaufen und jeder Prozeß nur auf solche nicht vom direkten Vorgänger produzierten Daten zugreifen kann, die der direkte Vorgänger auch wirklich weitergibt (vgl. LUCKHARDT 1983, LUCKHARDT/MAAS/THIEL 1985).

Im ganzen gilt das von Winograd skizzierte blackboard-Modell ohne Einschränkung für SUSY-II, mit gewissen Einschränkungen aber auch für SUSY. Warum ich dieser Meinung bin und wie sich die Stratifikation in SUSY und SUSY-II äußert, will ich am Beispiel der Analyse von Relativsätzen darstellen.

1. Relativsatzanalyse in SUSY als blackboard-Modell

Als wichtigster und gleichzeitig trivialster Grund für Stratifikation kann gelten, daß ein Prozeß nur die Informationen verarbeiten kann, die ihm zur Verfügung stehen. Eine Konsequenz aus dieser Einsicht könnte die Verwendung eines ATN-Parsers sein, der sich die Informationen, die er benötigt, sofort zu beschaffen sucht, indem er in das betreffende Netzwerk hinabsteigt. Gerade dies führt jedoch zuweilen zu ad hoc-Entscheidungen, da der Parser nicht abwarten kann, bis er genügend Informationen hat. ATN-Parser verfügen zwar über gewisse backtrack-Möglichkeiten, aber um effektiv arbeiten zu können, müßte jeder beliebige backtrack zu jedem beliebigen Knoten zu jedem beliebigen Zeitpunkt möglich sein. Im Ganzen gesehen ist das hypothesengesteuerte Parsing mit ATNs weniger flexibel, da der Linguist bei der Entwicklung seiner Analysestrategie insofern nicht frei ist, als das Funktionieren des Parsers dauernd vom Treffen irgendwelcher Entscheidungen abhängt. So begibt er sich der eigentlichen Kontrollmöglichkeiten, die es erfordern würden, daß Entscheidungen freiwillig und nicht gezwungenermaßen getroffen werden. Bei der Entwicklung eines ATN-Parsers muß der Linguist fortwährend natürlichsprachliche Strukturen antizipieren und ein passendes Korsett von Netzwerken entwerfen, anstatt dem Parser eine lockere und dennoch effiziente Kontrollstruktur anzulegen, in der er sich frei, d.h. wie es der Eingabetext erfordert, bewegen kann (wie dies z.B. bei SUSY-II der Fall ist, vgl. MAAS 1984). Systeme mit ATNs und Parser vom Typ SUSY-II haben eines gemeinsam: die Reihenfolge der Grammatiken wird vom Linguisten festgelegt. Ein ATN-Parser arbeitet jedoch in der Regel von links nach rechts, wobei er bei jedem Schritt Mutmaßungen über das nächste zu analysierende Element anstellt, wohingegen ein Parser wie SUSY-II ohne Richtungsbeschränkung die von seinen Grammatiken erwarteten Strukturen im Satz sucht. Dabei ist die Bewegung des Parsers durch den Satz linguistisch motiviert. Eine optimale "Kontrolle" bedeutet also nicht, daß der Linguist jeden Schritt des Parsers vorher festlegt, sondern daß er ihm genau die Freiheit läßt, die die Analyse unbekannter, d.h. noch nicht analysierter und

nicht ausgesuchter natürlichsprachlicher Texte erfordert.

Noch extremer äußern sich mangelnde Kontrollmöglichkeiten bei Produktionensystemen. Sie sind nicht stratifiziert im oben geschilderten Sinne, obwohl sie von den acht genannten Gesichtspunkten einige berücksichtigen. Produktionensysteme sind - als Problemlösungsverfahren gesehen - 'Zustand-Raum- Modelle' (vgl. BARR/FEIGENBAUM 1981 (I), 32):

A state-space representation of a problem employs two kinds of entities: **states**, which are data structures giving 'snapshots' of the condition of the problem at each stage of its solution, and **operators**, which are means for transforming the problem from one state to another.

Stratifikationale Modelle entsprechen eher der "Problemreduktionsrepräsentation":

In the problem-reduction approach, the principal data structures are the problem descriptions or **goals**. An initial problem description is given; it is solved by a sequence of transformations that ultimately change it into a set of subproblems whose solutions are immediate.

(BARR/FEIGENBAUM 1981 (I), 36)

Produktionensysteme scheinen für komplexe linguistische Probleme weniger geeignet, weil die Aufgliederung von Problemen in 'subproblems' zumindest schwierig ist. Dies schließt auch eine wirksame Kontrolle der Produktionen aus. Produktionensysteme haben zwar einige Vorteile, diese scheinen jedoch an Bedeutung zu verlieren, sobald die Systeme größer werden.

Modularität im Sinne von Winograd liegt, z.B., bei Produktionensystemen grundsätzlich vor:

... the individual productions in the rule base can be added, deleted, or changed independently. They behave much like independent **pieces of knowledge**. This relative **modularity** is important in building the large rule bases of current AI systems ... There are indications, however, that modularity is harder to maintain as one moves to larger systems."

(BARR/FEIGENBAUM 1981 (I), 193)

Neben den Vorteilen der Uniformität der Regeln und der Natürlichkeit der Darstellung von Wissen nennen Barr/Feigenbaum zwei wesentliche Nachteile:

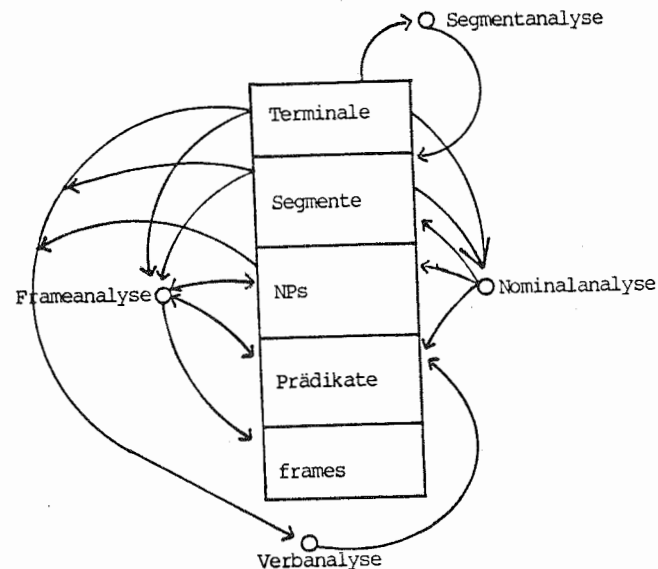
Inefficiency: ... The strong modularity and uniformity of the productions result in high overhead in their use in problem solving. ... it is difficult to make them efficiently responsive to predetermined sequences of situations or to take larger steps in their reasoning when the situation demands it ...

Opacity: ... it is hard to follow the flow of control in problem solving ... algorithmic knowledge is not expressed naturally ..."
(BARR/FEIGENBAUM 1981 (I), 194)

Ich gehe im folgenden davon aus, daß Probleme im Sinne der Stratifikation in kleinere Probleme aufgelöst werden müssen. Eng verknüpft damit ist die Frage, welches Wissen zur Lösung der einzelnen Probleme nötig ist. Wenn wir von der trivialen Einsicht ausgehen, daß ein Prozeß nur die Informationen verarbeiten kann, die ihm zur Verfügung stehen, so wurde für SUSY eine andere Konsequenz gezogen als, z. B., für ATNs. Die Konzeption von SUSY sieht vor, zunächst größere Einheiten (Satzsegmente) zu ermitteln, da dafür die Informationen auf der terminalen Ebene ausreichend schienen. Dies hat sich im Rahmen des SUSY-Gesamtkonzepts bestätigt, und die Satzsegmente haben sich als ideales Operationsgebiet für die Nominal-, Verbal- und Frame-Analyse erwiesen. Daß SUSY dennoch an gewisse Grenzen stieß, lag nicht an der Strategie, sondern an der Datenstruktur.

Wir wollen jedoch im folgenden nicht die globale Prozeßstrategie betrachten, sondern die Strategie zur Bewältigung eines bestimmten Problems, nämlich der Relativsatzbehandlung. Diese erfolgt nach dem blackboard-Modell, und zwar in der folgenden Weise:

Jeder Prozeß tut soviel, wie er kann, und schreibt seine Ergebnisse an die blackboard, wobei man sich bei SUSY mehrere parallele blackboards vorstellen muß, da jeder Prozeß für jedes alternative Ergebnis eine neue blackboard anlegt und beschreibt; jede dieser blackboards wird folgendermaßen gelesen und beschrieben:



o-> = Prozeß schreibt

->o = Prozeß liest

Bei dieser Darstellung ist die Flexions- und Homographenanalyse unberücksichtigt, in die Frame-Analyse gehen syntaktische und semantische Analyse ein.

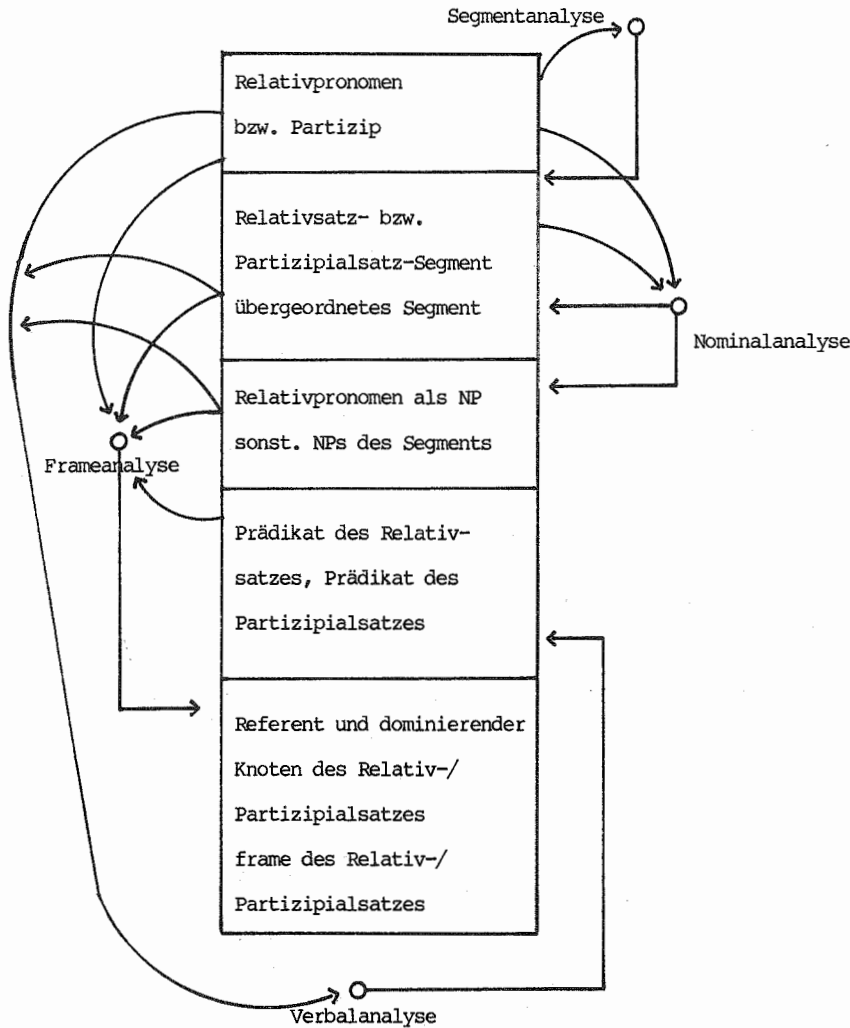
Segment-: liest die terminalen Symbole und ihre Dekorationen und schreibt Satzsegmente an die Tafel.

Nominal-: liest Terminale und Segmente und schreibt die Analyse NPs dazu, u.U. auch andere Segmentierungen. Dazu kommen die Attribute, Adjektive und Partizipien als Prädikate.

Verbal-: liest alle vorherigen Strukturen und schreibt die Analyse VG als Prädikate hinzu, dazu neue Informationen zu den Segmenten.

Frame-: liest alle vorherigen Strukturen und schreibt die Analyse gefüllten Frames hinzu, ändert auch nominale Strukturen und Prädikate.

Wie sieht nun eine blackboard für die Relativsatzanalyse aus?



Diese Darstellung erfordert einige Erläuterungen:

a. Identische Darstellung von Partizipial- und Relativsätzen:

Es liegt nicht allein in der Natur von SUSY als multilingualem System

begründet, daß Partizipialkonstruktionen und Relativsätze auf die gleiche tiefsyntaktische Darstellungsform abgebildet werden, wobei gewisse Informationen der Oberflächenstruktur erhalten bleiben. Dies ist die einzig mögliche konsistente Art der Darstellung solcher Strukturen für die maschinelle Übersetzung, die die Erzeugung adäquater zielsprachlicher Strukturen erlaubt.

b. Alternative Bearbeitung von Partizipial-/Relativsätzen durch die gleichen Prozesse:

Nachdem die einzelnen Prozesse von den in den einzelnen Sprachen vorliegenden Oberflächenstrukturen abstrahiert haben, bearbeiten sie Relativ-/Partizipialsätze nach den gleichen Regeln. Diese Bearbeitung wird durch eine außerhalb der Prozesse liegende linguistische Charakteristik für jede Sprache (vgl. THIEL (1980)) gesteuert, die z.B. Aussagen über die Form possessiver Relativsätze ('of which'- bzw. 'deren'-Typ) macht.

Es folgen einige Probleme, Vor- und Nachteile dieses Analyseansatzes.

1.1. Ambiguität des Relativpronomens

Die Ambiguität der Relativpronomina berührt die Analyse von Relativsätzen nicht, da jede der Deutigkeiten auf einer anderen blackboard steht, z.B.:

- "der": Relativpronomen
Artikel
Personalpronomen
- "welcher": Relativpronomen
Fragepronomen

1.2. Ermittlung der Relativ-/Partizipial-Satzsegmente

Hier wollen wir uns einigen besonderen Problemen bei der Erkennung und Abgrenzung von Relativ-/Partizipial-Satzsegmenten zuwenden:

B1: Er war der Feldherr, der kam, sah und siegte.

Hier besteht die besondere Aufgabe des Segmentierungsprozesses darin, die Segmente 'sah' und 'siegte' als Relativsätze zu erkennen und sie dem Segment 'der kam' als angelehnt zuzuordnen, was mithilfe der SEGMENT-Grammatik, einer augmented PS-grammar (vgl. LUCKHARDT (1984)), geschieht. Es besteht auch die Aufgabe, falsche Relativsatzsegmente zu eliminieren, vgl. die Folge von Terminalen "der Sprachen" in B2 und B3:

B2: Dies ist ein Computer, der Sprachen, aber auch

Mathematik beherrscht.

B3: Er ist ein Freund der Literatur, **der Sprachen** und der Musik.

Für B2 wird die Grammatik eine korrekte Ableitung liefern, die einen Relativsatz enthält, für B3 nicht.

1.3. Analyse der NPs

Die Analyse von NPs bereitet in unserem Zusammenhang keine besondere Schwierigkeit.

1.4. Analyse der Prädikate des Relativ-/Partizipialsatzes

Auch hier bestehen keine besonderen Probleme, da lediglich auf Segmentebene nach den Regeln der Verbalanalyse die Verbalgruppen ermittelt und zum Prädikat gemacht und auch die die Partizipialsätze einleitenden Elemente zu Prädikaten erhoben werden.

1.5. Frame-Analyse

Das wichtigste und auch linguistisch interessanteste Problem ist die Füllung der slots im Frame des Prädikats des Relativsatzes. Dabei gilt es, mehrere Dinge zu beachten.

1.5.1 Mehrdeutigkeit

Die Relativ-NP ist vor der Frame-Analyse u.U. noch stark mehrdeutig, und zwar in Numerus, Genus und Kasus, so daß zunächst der Frame des Prädikats vollständig gefüllt werden muß. Erst dann kann man die Kasus-, Genus- und Numerusvereindeutigung vornehmen, und erst dann kann der Antezedent ermittelt werden:

B4: ... die Lösung der Probleme, **die** uns bedrücken.

B5: ... die Entscheidung des Senats, **der** sich der Präsident unterwarf.

Die Relativpronomina können sich vor der Frame-Analyse in beiden Fällen auf die beiden möglichen Antezedenten beziehen, nach der Vereindeutigung mithilfe der Frames nur noch auf einen.

1.5.2 Voraussetzungen

Der Frame-Analyse stehen auf der blackboard zur Verfügung

- eine Folge von s Segmenten S
- zu jedem Segment Si eine Folge von n Nominalphrasen N, ein komplexes

Prädikat PK, bestehend aus p einfachen Prädikaten P mit dem dominierenden Prädikat OP, und einer Folge von t Terminalen T.

- zu jeder Nominalphrase Nj eine Folge von e Terminalen E, inclusive ein Governor K
- zu jedem einfachen Prädikat P ein Governor G

Segmente		S1	,		S2
Terminale	T1.1	...	T1.t ₁	T2.1	... T2.t ₂
NPs	N(K)1.1	...	N(K)1.n ₁	N(K)2.1	... N(K)2.n ₂
Terminale	E1.1.1	...	E1.n ₁ .e ₁	E2.1.1	... E2.n ₂ .e ₂
Präd.kompl.	PK1(OP(G)(P(G)1.1		PK2(...)		
	... (P(G)1.p))				

Aus der Darstellung geht hervor, daß die Terminalen Tj.1 bis Tj.tj dem Segment Sj die notwendigen Informationen für die Segmentanalyse liefern und daß die nachfolgenden Prozesse nur auf dem Segment Sj zu operieren brauchen, um die Nominalphrasen Nj.1 bis Nj.nj bzw. den Prädikatkomplex PKj innerhalb des Segments Sj zu analysieren. Jedes S, N, OP, P, T (bzw. E) verfügt über eine Dekoration mit inhärenten und relationalen Informationen. Die Aufgaben der Frame-Analyse sind folgende:

1. Die slots im Frame F jedes Governors G mit den im betreffenden Segment Si vorliegenden n NPs füllen.
2. Für jedes Segment Si, das als Relativsatz markiert ist, aus der Menge der NPs des übergeordneten Segments Sn die NP Nj ermitteln, die als Antezedent der relationalen NP anzusehen ist.
3. Im Frame des Segments Si die relative NP durch ihren Antezedenten zu ersetzen.

Hierbei interessiert in unserem Zusammenhang die 2. Aufgabe.

1.5.3 Die syntaktische Ermittlung des Antezedenten

Formal ist dies in SUSY die Suche der am weitesten rechts stehenden NP Nj aus Sn, die mit der relativen NP kongruent ist. Die Technik der Kongruenzprüfung bezüglich Genus und Numerus ist in LUCKHARDT (1982: 14) beschrieben und an dieser Stelle von keinem besonderen Interesse. Diese Kongruenzprüfung wird dazu führen, daß für B5 Nj abgelehnt und Nj-1 als Antezedent ausgewählt wird, für B4 wird Nj akzeptiert.

Ein Vergleich mit der 'psychological reality' zeigt, daß das stereotype Vorgehen (= die vorhergehende Kasusvereindeutigung der Relativ-NP und dadurch ihre Numerus- und Genus-Vereindeutigung durch die Frame-Analyse des Relativsatzes) der Maschine gegenüber dem Menschen einen gewissen Vorteil verschafft, da dieser, z.B., bei B5 auf einen 'garden path' (= Holzweg) geraten kann. Der kognitive Analyseprozeß arbeitet den Satz normalerweise von links nach rechts ab, so daß die Mehrdeutigkeit von 'der' bestehen bleibt, bis das Prädikat des Relativsatzes erfaßt ist. So ist es möglich, daß die Lesart 'Nom. Mask. Sing.' zunächst eine höhere Präferenz erhält, d.h. bis 'der Präsident' bzw. 'unterwirft' gelesen wird. Auch der kognitive Prozeß verläuft in dem Sinne stratifiziert, daß gewisse Entscheidungen erst dann gefällt werden können, wenn bestimmte Informationen vorliegen. Letztere können auch zu einem 'backtrack' (d.h. = 'noch mal lesen') führen, wie etwa im oben erwähnten Fall, häufiger wohl bei längeren Sätzen.

1.5.4. Keine semantische Ermittlung des Antezedenten

Die vorstehenden Überlegungen betreffen Beziehungen, die durch morpho-syntaktische Kongruenzprüfungen vereindeutigt werden können, auch in SUSY. Die Auswahl des korrekten Antezedenten aus einer Menge **syntaktisch** möglicher stellt sich wesentlich schwieriger dar oder ist sogar unmöglich, vgl.:

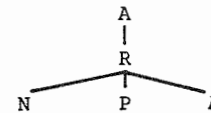
- B6: ... die Beschreibung der Umgebung, die mit dem MÜ-System nichts zu tun hat...
- B7: ... die Verwirklichung dieser Vorstellung, die erst mit der Entwicklung der Computer aufkam...
- B8: ... Vorstellungen über die Durchführung bestimmter Anweisungen durch die Computerlinguisten, die automatisch von einem Text zu dessen Übersetzung führen sollten.
- B9: ... die formalen Beschreibungen von Grammatiken mithilfe von kontextfreien Regeln, die in drei Teilgrammatiken aufgeteilt werden konnten.

Eine semantische Kongruenzprüfung findet in SUSY nicht statt, bei mehreren zur Verfügung stehenden Antezedenten wird der dem Relativsatz nächststehende gewählt. Dieser Ansatz ist sprachwissenschaftlich unbefriedigend, es ist jedoch keine praktikable Lösung in Sicht. Diese Behauptung mag angesichts der im Bereich der KI erarbeiteten Theorien und Methoden vermessen erschei-

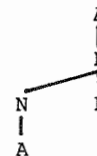
nen, sie trifft jedoch für Beispiele vom Typ B6-B9 zu: ich sehe keine Möglichkeit, für diese Beispiele die Ermittlung der korrekten Antezedenten formal darzustellen.

1.5.5. Darstellung von Relativsätzen in SUSY

Relativsätze R mit Relativ-Nominalgruppe N werden ihrem Antezedenten A untergeordnet. Gleichzeitig nimmt A in R eine Stelle im Valenzrahmen des Prädikats R(P) ein:



Die gleiche Darstellung gilt für attributive Adjektive/Partizipien und partizipiale Relativsätze, wobei allerdings N fehlt. R steht also für Relativsatz, Adjektiv- oder Partizipgruppe. Für possessive Relativsätze sieht die Darstellung folgendermaßen aus:



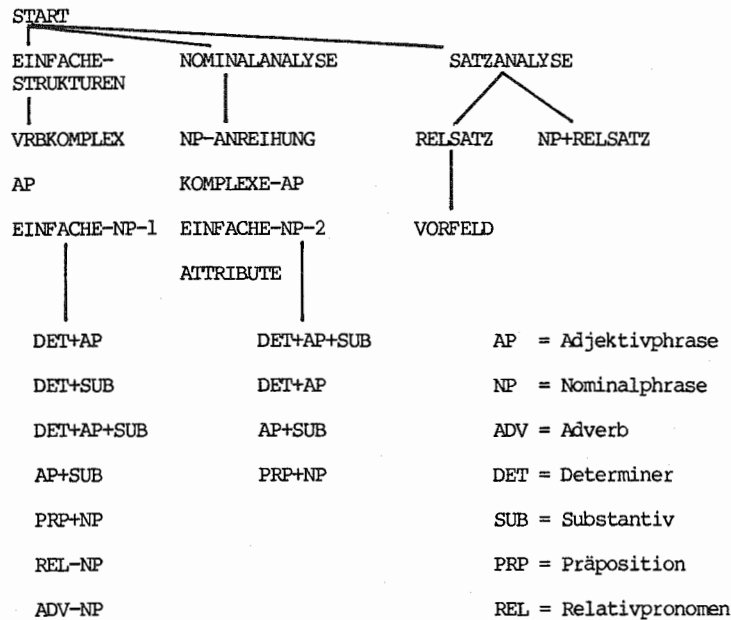
1.5.6. Zusammenfassung der Relativsatzanalyse in SUSY

- Der stratifikationale Ansatz hat sich insofern bewährt, als zu jedem Zeitpunkt der Analyse genau definierte Einheiten und wohlabgegrenzte Operationsgebiete für die Grammatiken vorliegen, was von keinem anderen Parser gesagt werden kann.
- Die Zuordnung von Relativsätzen zu einem falschen Antezedenten wirkt sich bei der deutsch-englischen Übersetzung so gut wie garnicht aus, im Gegensatz zur englisch-deutschen Übersetzung.
- Der stratifizierte Ansatz von SUSY kann sich nur dann negativ auswirken, wenn in der Homographenanalyse die Lesart "Relativpronomen" eines Wortes fälschlicherweise ausgeschlossen wird. Dieser Gefahr kann man dadurch begegnen, daß man den Determinismus der Homographenanalyse abschwächt, indem man etwa statt zwei (für Deutsch) zwölf Lesartenketten zuläßt.
- Die Darstellung von Relativsätzen reicht für die Übersetzung aus jeder

der in SUSY implementierten Sprachen in jede andere voll aus. Mehr noch: eine Verbesserung der Darstellung scheint nicht erforderlich zu sein.

2. Relativsatzanalyse in SUSY-II

Wir wollen auch hier den blackboard-Ansatz zugrundelegen, mit dem Unterschied (zu SUSY), daß die auf die Tafel zugreifenden Prozesse nicht strikt sequentiell ablaufen. Außerdem gibt es nur eine blackboard, nämlich den S-Graphen (eine Chart), der alle wohlgeformten Strukturen enthält. Zur Einführung in SUSY-II und den S-Graphen vgl. MAAS (1984). Die folgende Abbildung gibt einen Überblick über die wichtigsten direkt oder indirekt an der Relativsatzanalyse beteiligten Prozesse und gleichzeitig über die von diesen erzeugten Strukturen, d.h. die Namen der Prozesse (und Grammatiken) in der Darstellung erlauben Rückschlüsse auf die Strukturen.



Hier ist ein Grammatiktyp repräsentiert, den wir CAP genannt haben: controlled active procedures:

Die terminalen Knoten dieses Baumes sind cf-Regeln, die mit **Prozeduren** erweitert sind. Die Regeln werden von Grammatiken **aktiviert**, die wiederum von Prozessen **kontrolliert** werden.

Die CAP-Grammatik steht in der Tradition der TG und der Abhängigkeitsgrammatik und hat gewisse Ähnlichkeiten mit der lexical functional grammar. Einige ihrer Eigenschaften sind im folgenden zusammengefaßt (ausführlicher in LUCKHARDT 1985b).

a. kontextfreie Regeln

- Einfache Nominal- und Verbalphrasen werden durch Kontituentenstrukturregeln erfaßt.
- Komplexe Strukturen werden durch **inkrementelle** Regeln erfaßt, z.B. die Struktur PRED+NP+NP+NP durch iterative Anwendung der Regel PRED NP => PRED, d.h. die PRED-Struktur wird inkrementiert.

b. CAP-Regeln sind erweitert (s. unten).

c. CAP basiert auf einer starken morphologischen Komponente und verfügt über umfangreiche lexikalische Informationen.

d. Die Repräsentation ist im wesentlichen dependentiell mit den folgenden Charakteristika:

- die Oberflächenordnung bleibt soweit wie möglich erhalten
- die Knoten enthalten eine umfangreiche Beschreibung (Dekoration)
- drei Arten von Funktionen der Konstituenten bzw. Relationen zwischen ihnen werden dargestellt: oberflächensyntaktische und tiefsyntaktische Funktionen sowie semantische Relationen

e. Grundlage der Inkrementation ist im wesentlichen der Valenzrahmen der Verben, Substantive und Adjektive.

f. Die Vererbung von Merkmalen kann wesentlich stärker gesteuert werden, als dies zur Zeit bei der LFG praktiziert wird.

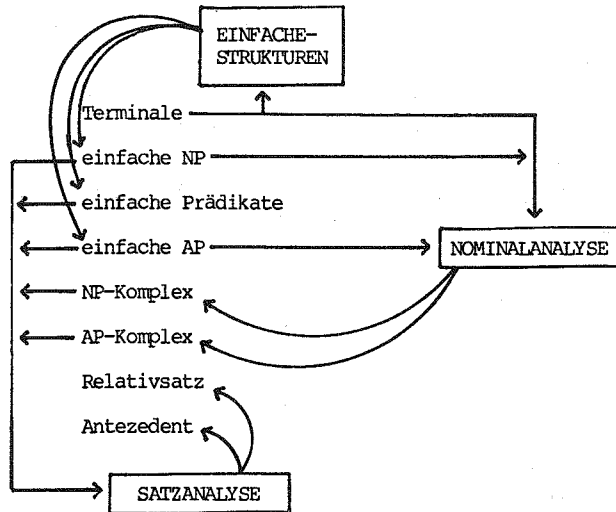
g. CAP ist aufgrund der Inkrementation und der kontrollierten Vererbung besonders für die Analyse des Deutschen geeignet. Dies zeigt sich besonders bei der Lösung einiger linguistischer Probleme, die für reine cf-Grammatiken im allgemeinen schwer zu bewältigen sind: Koordination, Diskontinuität und freie Wortstellung

h. Die Anwendung von CAP-Regeln wird auf zwei Arten gesteuert:

- durch Prozesse, die die Strategie der Regelanwendung bestimmen
- durch Grammatiken, die die Kante bzw. Folge von Kanten (Pfad) in der Datenstruktur (dem S-Graphen, einer Chart) auswählen, auf die die Regeln angewandt werden sollen

Es wird deutlich, daß Regeln, die zu einem frühen Zeitpunkt der Analyse noch nicht anwendbar sind, später auf die veränderte Datenbasis noch einmal angewandt werden können, in der Hoffnung, daß dann die erwarteten Strukturen vorliegen. So kann DET+AP+SUB (also als einfache Struktur) zunächst auf "die an der Analyse beteiligten Prozesse" nicht angewandt werden (im Gegensatz zu "die beteiligten Prozesse"). Dies geht erst, wenn in AP-KOMPLEX die einfache AP "beteiligten" durch die NP "an der Analyse" erweitert worden ist. Die Regel DET+SUB z.B. taucht unter EINFACHE-NP-2 nicht noch einmal auf, weil sie ihren Zweck schon beim ersten Aufruf erfüllt hat, d.h. die folgenden Regeln verändern die Datenstruktur nicht in der Weise, daß DET+SUB noch einmal anwendbar wird.

Der Idealtyp der Stratifikation nach Winograd bleibt auch deswegen gewahrt, weil keiner der Prozesse etwas zerstört, was einer der nachfolgenden Prozesse noch benötigt. Im groben sieht das blackboard-Modell für die Relativsatzanalyse in SUSY-II so aus:



Im folgenden die Darstellung der einzelnen Prozesse:

```

process START:
do EINFACHE-STRUKTUREN NOMINALANALYSE SATZANALYSE;
end

process EINFACHE-STRUKTUREN:
do VRBKOMPLEX AP EINFACHE-NP-1 ;
end
  
```

```

process NOMINALANALYSE:
do NP-ANREIHUNG KOMPLEXE-AP
  EINFACHE-NP-2 ATTRIBUTE;
end

process SATZANALYSE:
do RELSATZ NP+RELSATZ;
end

process EINFACHE-NP-1:
do DET+AP DET+SUB DET+AP+SUB
  AP+SUB PRP+NP REL-NP ADV-NP ;
end

process EINFACHE-NP-2:
do DET+AP+SUB DET+AP
  AP+SUB PRP+NP;
end

grammar REL-NP:
1 X : MS = PER and PROSYV = RELATIV ;
use REL-NP;
end
  
```

Die Aufgabe dieser Grammatik ist es, aus der Datenstruktur alle Kanten X mit der morphosyntaktischen Kategorie PERSONALPRONOMEN und der syntaktischen Verwendung RELATIV herauszusuchen und darauf die Regel REL-NP anzuwenden.

```

rule REL-NP:
geometry
X = Y ( X )
conditions
MS of X = PRONOMEN and PROSYV of X = RELATIV;
assignments
MS := NP & SF of X' := GOVERNOR &
INVENTAR := PRONOMEN & GENUS := PROGENUS of X &
SYVNOG := RELATIV &
if SINGULAR of X =/ null then NUMERUS := SG fi &
if PLURAL of X =/ null then NUMERUS := NUMERUS sum {PL}fi &
SINGULAR := SINGULAR of X &
PLURAL := PLURAL of X &
end
  
```

Hierdurch wird im S-Graphen eine neue NP-Kante erzeugt und mit der erforderlichen Dekoration versehen. Im folgenden einige Erläuterungen zu den Variablen und Werten:

- | | |
|-----------------------------------|--|
| MS = morphosyntaktische Kategorie | PROSYV = syntaktische Verwendung eines Pronomens |
| SF = synt. Funktion | PROGENUS = Genus e. Pronomens |
| SYVNOG = synt. Verw. e. NP | SINGULAR = Deklinationsmaske |
| PRED = Prädikat | PLURAL = Pluralmaske |
| FIV = finites Verb | HSNS = Hauptsatz-/Nebensatzstellung |
| KOM = Komma | GED = Gedankenstrich |

Sachverhalte notwendig. Zwei recht verschiedene Beispiele für die Stratifikation von Prozessen wurden mit SUSY und SUSY-II vorgeführt, wodurch die Vermutung gerechtfertigt ist, daß es für Stratifikation kein universelles Konzept gibt. Wir können lediglich sagen, daß stratifikationsmodelle die acht von Winograd genannten Aspekte in unterschiedlichem Maße berücksichtigen. Ich will abschließend darzustellen versuchen, welche Unterschiede es in verschiedenen Modellen bzw. Systemen bezüglich der Stratifikation gibt.

3.1. Stratifikation als problem-reduction-method

In SUSY, SUSY-II und EUROTRA werden komplexe Probleme in einfache aufgelöst. Dies ist z. B. bei unkontrollierten Produktionssystemen oder Inselparsern nicht der Fall.

3.2. Unabhängigkeit von Prozessen

Dies gilt für SUSY in wesentlich strikterem Maße als für SUSY-II oder EUROTRA. Dagegen kann man bei ATNs wohl nicht davon sprechen, daß in ihnen Prozesse unabhängig voneinander ablaufen.

3.3. Modularität

Modularität liegt wohl in allen Parsern vor, nur ist sie im einen Fall größer (Produktionssysteme), im anderen geringer (SUSY). Es muß jedoch gesagt werden, daß in der Regel Ineffizienz der Preis hoher Modularität ist.

3.4. Trennung von Prozessen und Strukturen

Hier müssen wir einen Unterschied machen zwischen **Regeln für Strukturen** und **erkannten Strukturen**. **Erkannte Strukturen** sind bei Produktionssystemen, Chartparsern jeglicher Art und SUSY vom Verarbeitungsprozeß getrennt. **Strukturregeln** sind bei Produktionssystemen und SUSY-II von den Prozessen getrennt, bei SUSY teilweise auch. Bei ATNs sind die Strukturregeln in die Netzwerke integriert.

3.5. Ermittlung der Strukturen

Auch beim einfachsten Parser hängt die Ermittlung der Strukturen davon ab, welche Informationen zur Verfügung stehen. Der entscheidende Aspekt dabei ist, wie die Ermittlung der Strukturen gesteuert wird: vom System (Produktionssysteme, ineffizient) oder vom Linguisten (SUSY-II, effizient).

3.6. Keine Beschränkung der verarbeitbaren Daten

Hier können mehrere Arten von Stratifikationen unterschieden werden. Bearbeitet werden.

- a. nur Daten des Vorgängerprozesses oder
- b. Daten aller Vorgängerprozesse; es ist keine Iteration oder parallele Abarbeitung von Prozessen möglich; oder
- c. Daten aller Vorgängerprozesse; Prozesse können (außer sequentiell) iterativ und/oder parallel laufen; oder
- d. Daten aller Vorgängerprozesse; die Daten können beliebig eingeschränkt werden: benutzte/unbenutzte oder neue/alte oder alle Daten

3.7. Realisierung von Stratifikation durch blackboard-Modelle

Die Definition eines blackboard-Modells schließt nach Winograd im wesentlichen die Punkte 4. - 6. ein. Die ermittelten Strukturen werden also von den beteiligten Prozessen von der blackboard gelesen bzw. auf die blackboard geschrieben. Dabei ist die Reihenfolge, in der Informationen auf die blackboard geschrieben werden, von der Verfügbarkeit von Informationen abhängig, und ein Prozeß muß nicht auf die Daten des Vorgängerprozesses beschränkt sein. Wie wir gesehen haben, gibt es verschiedene Möglichkeiten, solche Modelle zu realisieren. Ich will sie abschließend durch kurze Beschreibungen verschiedener Systeme noch einmal zeigen:

a. SUSY

SUSY verfügt über modulare Prozesse, die sequentiell ablaufen, normalerweise auf Daten des direkten Vorgängerprozesses und nur bei unvollständigen Analysen auf alle Vorgängerprozesse zugreifen, extern durch die linguistische Charakteristik der analysierten Sprache gesteuert werden, die ermittelten Strukturen auf die blackboard schreiben und bei Mehrdeutigkeiten mehrere blackboards anlegen. Die blackboard selbst ist insofern heterogen, als sie verschiedene Arten von Datenstrukturen enthält, je eine für Terminale, Nominalstrukturen, Segmente und Prädikate.

b. SUSY-II

Die Prozesse von SUSY-II können sequentiell, parallel oder iterativ ablaufen, greifen auf die Daten aller Vorgängerprozesse zu, schreiben ihre Ergebnisse in eine einheitliche Datenstruktur (den S-Graphen), laufen nach einer vom Linguisten definierten Kontrollstruktur ab, suchen sich ihr Operationsgebiet selbst und können beliebig modular angelegt werden.

c. Produktionensysteme

Produktionensysteme verfügen über ein Höchstmaß an Modularität bei den Regeln, ihre blackboard ist eine Datenbasis, in die alle Ergebnisse geschrieben bzw. von der die zu behandelnden Strukturen gelesen werden, sie bestimmen ihre Analysestrategie selbst und sind in der Regel nicht stratifiziert.

d. ATNs

ATNs verfügen normalerweise über eine Chart als blackboard, in die sie ihre Ergebnisstrukturen eintragen, in der sie aber nicht nach Belieben navigieren können; ihre Analysestrategie wird vom Linguisten strikt vorgeschrieben; sie sind in dem Sinne modular, daß es für verschiedene Aufgaben verschiedene Unternetze gibt; ihre Analyseregeln sind in die Analysestrategie integriert; sie müssen auch dann Entscheidungen treffen, wenn nicht alle nötigen Informationen verfügbar sind; man kann bei ATNs nicht von Prozessen reden, also auch nicht von "Vorgängerprozessen", es sind grundsätzlich alle Daten auf der blackboard sichtbar.

Literatur

BARR, A., FEIGENBAUM, E.A. (1981): The Handbook of Artificial Intelligence. London: Pitman

LUCKHARDT, H.-D. (1982): SATAN-Test. Beschreibung der Vorgehensweise und der Ergebnisse von Tests der deutschen Komponente des Saarbrücker Textanalyseystems SATAN. Linguistische Arbeiten des SFB 100 Neue Folge, Heft 6. Saarbrücken: Universität des Saarlandes

- (1983): Einführung in das Saarbrücker MÜ-System SUSY. Ms., Vortrag auf der DGfS-Jahrestagung 1983, Passau

- (1984): SEGMENT - die praktische Anwendung einer PS-Grammatik in einem MÜ-System und ihre Relation zu GPSG und LFG. Erscheint in: U. Klenk (Hrsgb.): Kontextfreie Syntaxen und verwandte Systeme. Linguistische Arbeiten, Tübingen: Niemeyer

- (1985): Parsing mit SUSY und SUSY-II. Strategien, Software und linguistisches Wissen. Linguistische Arbeiten des SFB 100 Neue Folge, Heft 12. Saarbrücken: Universität des Saarlandes

- (1985b): Parsing with Controlled Active Procedures. Ms., Arbeitspapier SFB 100 / TP A2. Saarbrücken: Universität des Saarlandes

LUCKHARDT, H.-D., MAAS, H.-D., THIEL, M. (1985). The SUSY-E Machine Translation System. Erscheint in: J. SLOCUM (Hrsgb.): Sonderausgabe des American Journal of Computational Linguistics über MT

MAAS, H.-D. (1983): SUSY und SUSY-II. Verschiedene Analysestrategien in der Maschinellen Übersetzung. Sprache und Datenverarbeitung 1-2/1981: 9-15

- (1984): Struktur und Steuerung der linguistischen Prozesse in SUSY-II. Erscheint in: U. KLENK (Hrsgb., 1985). Kontextfreie Syntaxen und verwandte Systeme. Linguistische Arbeiten. Tübingen: Niemeyer

- (1985, in Vorb.): SUSY-II-Handbuch. Linguistische Arbeiten des SFB 100 Neue Folge, Heft 14. Saarbrücken: Universität des Saarlandes

ROSNER, M. (1983): Production Systems. In: M. KING (Hrsgb., 1983): Parsing Natural Language. London: Academic

THIEL, M. (1980): LICA - Sprach- und texttypspezifische Parametersätze. Dokumentation A2/6. Saarbrücken: Univ. d. Saarlandes: Saarbrücken

WINOGRAD, T. (1983): Language as a Cognitive Process. Reading, Mass.: Addison-Wesley