

In: Muttersprache 9/10 1969, 79. Jahrgang, Datenverarbeitung und Linguistik, 260-265

Die automatische Behandlung diskontinuierlicher Konstituenten im Deutschen*

von der Saarbrücker Arbeitsgruppe für Elektronische Sprachforschung:

HANS EGGERS, WOLFGANG KLEIN, RAINER RATH, ANNELY ROTHKEGEL, HEINZ-JOSEF WEBER, HARALD ZIMMERMANN

I. Problemstellung

Fast alle bekannten Verfahren der syntaktischen Beschreibung gehen von der Annahme aus, dass ein Satz aus einer endlichen Anzahl von Elementen besteht, dass zwischen diesen Elementen bestimmte Relationen bestehen und dass diese Relationen irgendwie hierarchisch gestuft sind. So würde man z. B. sagen, dass in dem Satz *Er schlug ihn auf den Kopf* die Relation zwischen *ihn* und *den* hierarchisch anders ist als die Relation zwischen *den* und *Kopf*. Die letztere Relation betrachten wir als hierarchisch höher. Dieser Hierarchie entsprechend können die Elemente in der unterschiedlichsten Weise angeordnet werden. Man kann etwa die Relation als Abhängigkeiten deuten und die Elemente im Sinne eines Tesnière'schen Stemmas anordnen. Man kann - wiederum der Hierarchie gemäß - die Elemente zu Klassen zusammenfassen, die sich daraus ergebenden Einheiten erneut zusammenfassen usw. bis zu einer letzten, übergeordneten Einheit, eben der Einheit »Satz«. Die jeweils zusammengefassten Einheiten kann man als Konstituenten der sich daraus ergebenden Einheit ansehen. Nach diesem Prinzip operiert unsere Erkennungsgrammatik.

Bei einer maschinellen Erkennungsgrammatik, d. h. einem Algorithmus, der automatisch bestimmten Sätzen syntaktische Beschreibungen zuordnet, ergibt sich aus dem angedeuteten Faktum das Problem, gemäß einer bestimmten vorgegebenen Deutung der hierarchischen Verhältnisse die Stufung der einzelnen Relationen zu erkennen und die Elemente demgemäß einander zuzuordnen. Formal äußern sich die Relationen in bestimmten Nachbarschaftsbeziehungen¹. Prinzipiell stehen jene Elemente, deren Relationen hierarchisch die höchsten sind, am nächsten zusammen.

Auf diesem Prinzip kann eine Erkennungsgrammatik zunächst einmal aufbauen. Doch wird dieses Prinzip häufig durchbrochen. So würde man sagen, dass in dem Satz *Ich habe den Vater gesehen* die Elemente *habe* und *gesehen* einerseits und *den Vater* andererseits zusammenzufassen sind. Man spricht hier von »diskontinuierlichen Konstituenten« (DK). Hier kollidieren also »ordre structural« und »ordre lineare«. Es ist bekannt, dass diese Diskontinuität im Deutschen besonders häufig ist (Prinzip der Umklammerung) und gelegentlich extreme Formen annimmt, z. B. im folgenden Satz *Ich habe, kaum war ich an dem von einer großen Menge belagerten Unfallort angelangt, meinen Vater in einen erregten Wortwechsel mit einem der vielen dort stehenden Polizisten verwickelt gesehen*. Für einen Parser, der beliebige deutsche Sätze analysieren soll, ist es daher oft extrem schwierig, zusammengehörige Elemente zu erkennen.

Dieses Problem kann in verschiedener Weise angegangen werden. Zunächst bietet sich eine transformationelle Behandlung an. Die Unmöglichkeit, diskontinuierliche Konstituenten im

Rahmen einer Phrase-Structure-Grammar zu behandeln, war ja einer der Hauptanstöße zur Einführung der Transformationsregeln². Sicherlich wären Transformationsregeln die zugleich eleganteste und adäquateste Art, das Problem zu behandeln. Leider liegt bis jetzt - erst recht nicht 1963, als wir mit unseren Arbeiten begannen - keine Transformationsgrammatik des Deutschen vor, die so weit entwickelt wäre, dass sie für ein automatisches Erkennungsverfahren programmiert werden könnte³. Wir haben es daher vorgezogen, eine bereits vorliegende Art der grammatischen Beschreibung (»Schulgrammatik«) mit einer Reihe von Modifikationen so weit zu formalisieren, dass sie einen funktionsfähigen Parser zu programmieren erlaubt, einen Parser, der auch diskontinuierliche Konstituenten »bewältigt«. Das ganze Verfahren wird ausführlich in »Elektronische Syntaxanalyse«⁴ beschrieben. Wir stellen hier nur einen kurzen Ausschnitt, eben die Behandlung der diskontinuierlichen Konstituenten, dar.

II. Problematik der Erkennung von Diskontinuität bei der maschinellen Analyse

Wie schon erwähnt, stellt bereits die Erkennung von Diskontinuität ein schwieriges Problem dar. Es handelt sich hierbei um die vielfach auftretende syntaktische Mehrdeutigkeit isolierter diskontinuierlicher Konstituenten. So gibt es oftmals verschiedene Möglichkeiten, Relationen herzustellen. Als Beispiele seien genannt:

Er hat in dieser Situation zu schweigen.

Er hat in dieser Situation zu reden versäumt.

Einmal muss *hat* mit einer Partizipialform, einmal mit einem Infinitiv + *zu* verbunden werden. Sind beide Möglichkeiten im Satz vorhanden, muss die Entscheidung jeweils durch Hierarchie-Regeln getroffen werden. Dies wird durch das Prinzip erleichtert, dass zwar verschiedene Kombinationsmöglichkeiten bestehen, dass aber nicht alle DK mit allen kombiniert werden können.

Als praktisch zu lösende Aufgabe stellt sich das Problem der DK als Teilproblem einer gesamten maschinellen syntaktischen Analyse. Seine Lösung ist unumgänglich nötig, wenn man nicht auf eine erhebliche Anzahl möglicher Satzkonstruktionen und damit auf die Behandlung eines wesentlichen Teils des Sprachsystems verzichten will. Deswegen halten wir es für angebracht, das Problem der DK nicht so isoliert als vielmehr innerhalb einer syntaktischen Satzanalyse zu betrachten. Aus diesem Grunde stellen wir der speziellen Beschreibung der DK-Analyse einen Überblick über das Saarbrücker Verfahren voran.

III. Das Saarbrücker Verfahren

Die maschinelle syntaktische Analyse, wie sie in Saarbrücken entwickelt wurde, stellt eine Satzanalyse dar. Ziel der Analyse ist, möglichst alle Satzstrukturen der deutschen Gegenwartssprache zu erkennen und zu bezeichnen. Es geht also darum, alle Elemente des Satzes sowie die den Satz konstituierenden Relationen der einzelnen Elemente untereinander zu bestimmen.

Dazu werden auf mehreren Ebenen Symbolfolgen zu Gruppen zusammengefasst, die wiederum übergeordnete Gruppen bilden. Dies erfolgt so lange, bis als oberste Einheit der Satz erreicht ist. Als erste Reduktionsstufe gilt dabei die Zuordnung von Wortklassen⁵ zu den einzelnen Wortformen (Wortklassenebene). In der zweiten und dritten Reduktionsstufe werden bestimmte Wortklassen oder Wortklassenfolgen zu nominalen, verbalen und adverbialen Gruppen zusammengefasst (Gruppenebene). Auf weiteren Reduktionsstufen werden diese Gruppen zu höheren Einhei-

ten zusammengefasst, die wir >Analyseeinheiten< genannt haben. Es handelt sich hierbei um eine Kette von Gruppen der vorherigen Reduktionsstufe, die einen Satz darstellen können, die aber genauso gut als Satz noch unvollständig sein können, was mit der möglichen Diskontinuität solcher Einheiten zusammenhängt. Diese Analyseeinheiten wiederum werden zu einem Subsatz, etwa vollständiger Hauptsatz oder vollständiger Nebensatz, zusammengefasst (Subsatzebene). Die nächsthöhere und letzte Ebene stellt dann die Satzebene dar. Bereits bei der Reduktion zu nominalen und verbalen Gruppen ergibt sich von einer möglichen Diskontinuität der Elemente her eine besondere Schwierigkeit. Wenn nur kontinuierliche Teilsequenzen als Gruppen zugelassen wären, wäre eine Segmentierung des Satzes relativ einfach. Der Vergleich mit einigen Modellen könnte die Einteilung des Satzes in Gruppen und ihre anschließende Klassifizierung wesentlich erleichtern. Dies kommt aber nicht in Frage, da wir bereits auf der Ebene der nominalen und verbalen Gruppen mit Diskontinuität zu rechnen haben. Dies trifft ebenfalls für die übergeordnete Subsatzebene zu. Man könnte umgekehrt sagen, dass gerade erst in der möglichen Diskontinuität diese Ebene ihre Begründung hat. Sie ist als Vorstufe für die endgültige Satzbestimmung notwendig, da erst in einem weiteren Reduktionsschritt festgestellt werden muss, ob tatsächlich Diskontinuität vorliegt.

Aus diesen knappen Andeutungen lässt sich bereits erkennen, dass für eine Segmentierung des Satzes in Einheiten und deren Reduktion zu übergeordneten Einheiten die Diskontinuität in besonderem Maße Schwierigkeiten bietet. Doch muss betont werden, dass dieses Problem keineswegs durch die Methode der Segmentierung und Klassifizierung entsteht, sondern man kann eher umgekehrt sagen, dass das Problem der Diskontinuität besonders gut durch diese hier verwendete Methode erfasst wird. Dies wird im folgenden deutlich werden.

IV. Lösung der Diskontinuität im Saarbrücker Verfahren

Wie bereits angedeutet, kommen DK - innerhalb einer Satzebene - sowohl auf der Gruppenebene als auch auf der Subsatzebene vor. Erinnern wir uns an unsere anfangs genannten Beispiele, so stellen wir fest, dass auch dort diese Zweiteilung vorhanden war. Allerdings ist es nicht so, dass Diskontinuität von verbalen und nominalen Gruppen nur auf der Gruppenebene auftaucht. Sie kommt ebenfalls auf Subsatzebene vor. Als Unterscheidungsmerkmal dient die Tatsache, dass die zusammengehörenden Teile durch einen Satz bzw. durch eine selbständige Einheit getrennt sind im Gegensatz zu den folgenden Beispielen (a) und (c), bei denen die unterbrechenden Elemente ebenfalls in enger Beziehung zu einem DK stehen. Gliedern wir danach einige Beispielsätze, so erhalten wir:

1. Diskontinuität auf Gruppenebene

- (a) *der den Baum sehende Mann*
- (c) *er hat seinen Freund gesehen*

2. Diskontinuität auf Subsatzebene

- (b) *die, sagen wir, komplizierte Untersuchung 6*
- (d) *er hat, wie wir wissen, seinen Freund gesehen*
- (e) *die Untersuchung, die sehr lange dauerte, ist beendet*

Diese Unterscheidung besagt, dass sie ihrem Auftreten entsprechend auf einer früheren oder späteren (übergeordneten) Reduktionsstufe bearbeitet werden. Das zugrundeliegende Verfahren ist

dabei im Wesentlichen dasselbe. Aus diesem Grunde wird zunächst das für alle Typen geltende Prinzip dargestellt, im Anschluss daran einzelne Sonderprobleme.

Generell geht es darum, zu erkennen, ob und welche Elemente im Satz in diskontinuierlicher Anordnung eine Einheit bilden, d. h. in der Relation »Zusammengehörigkeit« stehen. Die in Frage kommenden Elemente müssen also in besonderer Weise hervorgehoben werden. Es handelt sich demnach um eine gezielte Segmentation. Nur bestimmte Elemente werden bezeichnet, sie müssen bestimmte Bedingungen erfüllen. Hierbei werden die beiden Bestandteile einer diskontinuierlichen Gruppe getrennt, wobei jeder Teil als zunächst eigenständige sprachliche Einheit betrachtet wird. Im folgenden wird die Berechtigung dazu deutlich werden. Die Bedingung dafür, ob eine Symbolsequenz einen DK darstellt, ist generell ihre Fähigkeit, in bestimmter Weise ergänzungsfähig zu sein. Im Einzelnen mögen die bedingenden Faktoren je nach Beschaffenheit des DK - ob es ein Element einer nominalen oder verbalen Gruppe ist oder eine Analyseeinheit - variieren. Zunächst sind gewisse Mindestanforderungen an einen DK gestellt:

1. Er muss aus einem oder mehreren Elementen des Satzes bestehen.
2. Er muss eine bestimmte Wortklassensequenz aufweisen, deren Reihenfolge zum Teil festgelegt ist.
3. Vom Inventar her muss eine Ergänzungsmöglichkeit, sei es zu einer nominalen oder verbalen Gruppe oder zu einem Satz, bestehen, er muss also in bestimmter Weise unvollständig sein.

Eine besondere Schwierigkeit ergibt sich nun daraus, dass die in Bedingung 3 genannte Ergänzung fakultativ ist. Im konkreten Satz braucht sie nicht realisiert zu sein, allein die Möglichkeit muss berücksichtigt werden.

Wir können danach verschiedene Fälle unterscheiden:

1. *Er hat das Geld erhalten.*
2. *Er hat das Geld.*
3. *Er hat das Geld, das er erhalten hat, in der Tasche.*

Bei dem ersten Typ kann eine endgültige verbale Gruppierung erst vorgenommen werden, wenn die beiden DK *hat* und *erhalten* als solche und sich gegenseitig ergänzend erkannt worden sind.

Bei dem zweiten Typ fehlt ein entsprechender Partner, die Symbolsequenz *Er hat das Geld* ist aber trotzdem eine selbständige Einheit. Dies bedeutet, dass ein Element sowohl Teil einer diskontinuierlichen Gruppe als auch eine eigenständige Gruppe sein kann. In diesen Fällen - die Ergänzung ist fakultativ - haben wir es mit potentiell diskontinuierlichen Konstituenten zu tun. Diese bezeichnen wir im Gegensatz zu Konstituenten, die in jedem Fall einer Ergänzung bedürfen, mit PDK, die anderen sind ADK (= aktualisierte DK).

Der dritte Beispielsatz zeigt Diskontinuität auf Subsatzebene. Die Symbolfolge »*Er hat das Geld*« - im zweiten Beispielsatz eine selbständige Einheit - ist hier ein PDK. Formal zu erkennen ist dies am Grenzsymbol, worin sich beide unterscheiden. Ebenfalls stellt »*in der Tasche*« ein PDK dar, während die gleiche Symbolfolge in dem Satz »*Er hat das Geld in der Tasche*« kein DK, auch kein PDK ist.

Wir erhalten ein weiteres Merkmal, wenn wir Beispiel 3 variieren:

4. *Er hat das Geld, das er in der Tasche aufbewahrt, eben erst erhalten.*

Der erste PDK ist gleich geblieben, die Ergänzung weist eine andere syntaktische Struktur auf. Daran zeigt sich, dass ein PDK durch verschiedene Arten von DK ergänzt werden kann. Andererseits lassen sich aber nicht alle Arten von DK beliebig kombinieren. Es reicht also nicht aus, festzustellen, ob im Satz eine Diskontinuität vorhanden ist. Es muss auch bekannt sein, um welche Typen (Klassen) von DK es sich handelt. Jeder DK weist Restriktionen in Bezug auf seine Kombinationsmöglichkeiten auf. So kann z. B. *hat* (finites Hilfsverb) ergänzt werden durch folgende Wortklassen (auf der Gruppenebene haben wir es mit Wortklassensequenzen zu tun):

1. *zu* + Infinitiv I
2. Partizip
3. andere unvollständige Verbalgruppen

Der Subsatz *Er hat das Geld* kann ergänzt werden durch:

1. Konstruktion mit Infinitiv + *zu*
2. Konstruktion mit Partizip
3. Präpositionales Gefüge.

Betrachtet man diese Einheiten isoliert, sind sie in Bezug auf ihre Ergänzungsmöglichkeiten syntaktisch mehrdeutig. Je nach der Ergänzung ändert sich die syntaktische Struktur. Dies bedeutet, dass die DK funktional mehrdeutig sind. Dieser Sachverhalt macht eine Klassifikation der DK nach ihren Ergänzungsmöglichkeiten notwendig. Dabei dient die Klassifikation zwei Zielen, einmal der Erkennung der DK, zum andern ihrer evtl. Zusammenführung.

Mit Hilfe dieser Klassifizierung, die auf dem Inventar basiert, lassen sich Modelle aufstellen, mit denen die einzelnen Wortklassensequenzen verglichen werden. Die Segmentierung des Satzes richtet sich nach diesen Modellen. Die Anzahl der Vergleiche lässt sich auf ein Minimum bringen, wenn als weiteres Merkmal die Grenzen der in Frage kommenden DK festgelegt sind. Auf Gruppenebene sind dies bestimmte Wortklassenfolgen, auf Satzebene Wortklassenfolgen und Interpunktionszeichen. Auf diese hin muss zunächst der ganze Satz untersucht werden, bevor sich eine Klassifikation anschließt. Da die Grenzen ein nicht nur für DK ausschließliches Merkmal sind, kann es vorkommen, dass eine Segmentierung durchgeführt wird, die kein DK bezeichnet. Erst der Vergleich mit den angesetzten Modellen wird erkennen lassen, ob es sich um ein PDK oder ADK handelt. Da, wie bereits erwähnt, bei manchen Klassen eine Ergänzung fakultativ ist, ergibt sich auch die Möglichkeit, dass laut Definition eine als PDK klassifizierte Sequenz im konkreten Satz kein DK darstellt. Auf diese Tatsache bereits bei der Klassifizierung einzugehen, erscheint zweckmäßig. Es gibt also eine Unterklasse von DK, die ebenfalls selbständige, von einer weiteren DK unabhängige Einheiten sein können und solche, die von ihrer Struktur her in jedem Fall eines diskontinuierlich angeordneten Partners bedürfen.

Die Vergleiche werden auf den verschiedenen Ebenen unterschiedlich durchgeführt, zumal sich auch die Grenzmarkierungen nach den Ebenen unterscheiden.

Die Erfassung der DK bei nominalen Gruppen auf Gruppenebene (Verkettung) setzt bereits eine Vorgruppierung voraus. *Der das Haus sehende Mann* ist in vorangehenden Reduktionsschritten bereits in drei Subgruppen unterteilt worden:

Der
das Haus
sehende Mann.

Da bei Diskontinuität von nominalen Gruppen auch bestimmte Elemente der nominalen Wortklassen⁷ als Grenze fungieren können, ist eine solche Vorgruppierung erforderlich. Bei sonstigen Nominalgruppen stellt zumeist ein Element einer nicht-nominalen Gruppe die Grenze dar. Anstelle eines Modells mit Wortklassenfolgen tritt bei der Verkettung eine Variation des Prinzips ein: die erste und dritte Subgruppe müssen im Kasus, Numerus und Genus übereinstimmen⁸.

Für die Erfassung von verbalen DK auf Gruppenebene wie auch für alle DK auf Satzebene wird als Modell eine bestimmte Wortklassenfolge angesetzt; beide unterscheiden sich nur in der Art der Grenzmarkierungen.

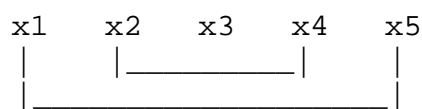
1. Verbale DK auf Gruppenebene (*Er hat das Geld erhalten*). Als Inventar der DK-Klassen für verbale DK sind Wortklassen zugelassen, die an der Bildung einer Verbalgruppe beteiligt werden können: Finites Verb, Infinitiv, Infinitiv + zu, Partizip, Verbzusatz.

2. DK auf Satzebene.

Auch hierbei wird das Modell-Verfahren angewendet. Die den Modellen zugrundeliegenden Strukturen sind gegenüber den DK auf Gruppenebene weitaus komplizierter. Inventar, Reihenfolge des Inventars und Grenzmarkierungen gelten als Kriterien für die Klassifizierung.

Nach der Erläuterung der allgemeinen Grundzüge des Saarbrücker Verfahrens zur Lösung der DK sei noch andeutungsweise auf zwei Probleme hingewiesen (die an anderer Stelle ausführlich behandelt sind⁹).

a) Außer der hier zugrundegelegten einfachen Verschachtelung wird auch eine mehrfache berücksichtigt:



x1 und x5 sowie x2 und x4 gehören jeweils zusammen. Bei mehr als zwei möglichen DK im Satz muss dieses Prinzip berücksichtigt werden. Dies geschieht in der Weise, dass zunächst die innere Klammer, dann die äußere Klammer bestimmt wird, d. h. der zweite DK hat Vorrang in der Zuordnung des nächsten DK. Hierbei tritt natürlich auch die Schwierigkeit auf, dass von z. B. vier DK zwei jeweils eine selbständige Einheit bilden. Dabei gäbe es theoretisch acht Möglichkeiten der Kombination, sofern alle DK untereinander als Partner zugelassen wären, was unwahrscheinlich ist. Doch bietet die Entscheidung zwischen zwei Möglichkeiten noch genügend Schwierigkeiten. Ein weiteres Merkmal muss also noch hinzugefügt werden. Dies kann z. B. die Stellung oder die Art der Grenzen sein. Auch können sich bei der Klassifizierung Probleme zeigen, vor allem was die Art der Klassen angeht. So erscheint es z. B. vom Verfahren her zweckmäßig, die Apposition auch als PDK zu kennzeichnen, da sie ihrer Struktur nach auch ein Teil

eines vollständigen Satzes sein kann. Dies hat nur zur Folge, dass die Fälle, die eine eindeutige Apposition darstellen, komplizierter werden als es für den konkreten Satz erforderlich ist.

b) Ebenfalls problematisch ist der Fall, wenn ein unvollständiger Satz diskontinuierlich an einen vollständigen, für sich abgeschlossenen Satz angeschlossen wird. Es erhebt sich hier die Frage, inwieweit dies als eine Variante von DK anzusehen ist. Dies trifft zu in dem folgenden Satz:

Die Kinder spielen im Garten, der sehr groß ist, und bauen dort ein Häuschen.

Auf diese Einzelheiten wollen wir hier nicht eingehen.

Zum Abschluss sei dem geschilderten Verfahren eine andere, praktisch erprobte Methode gegenübergestellt. Im Saarbrücker Verfahren liegt der Schwerpunkt auf den individuellen DK, ihrer Erkennung und ihren Kombinationen. Die einzelnen DK werden als selbständige linguistische Einheiten behandelt. Anders verfährt die z. B. in Georgetown (USA) entwickelte Analyse des Russischen von P. Toma¹⁰: Behandelt werden hier nur DK auf Subsatzebene, was von der analysierten Sprache her zu erklären ist. Hierbei liegt aber der Schwerpunkt auf der durch die einzelnen DK zusammengefassten Einheit. Als Grundlage für die Erkennung gelten ebenfalls Modelle, die jeweils mit zwei zusammengehörigen DK verglichen werden.

Diese Lösung erscheint wenig glücklich, da einschränkende Voraussetzungen gemacht werden. So ist z. B. nur eine Einheit zwischen den beiden DK erlaubt. Außerdem steht die gesamte DK-Erfassung in Abhängigkeit von der Subjekt- und Prädikatbestimmung. Es werden nur solche Einheiten als DK erkannt, die sich gegenseitig im Hinblick auf Subjekt oder Prädikat, gegebenenfalls noch auf ein Objekt ergänzen. Der Typ *Im Garten, der groß ist, stehen Bäume* würde nicht erfasst werden. Eine Erweiterung der Modelle könnte hier Abhilfe schaffen.

Gegenüber dem Saarbrücker Verfahren ergäbe sich dann aber immer noch der Nachteil, dass eine weitaus größere Anzahl von Modellen notwendig wäre. Insofern erscheint uns eine Segmentierung in die beiden Bestandteile einer diskontinuierlichen Figur zweckmäßiger, da so eine größere Flexibilität erreicht werden kann. Der Gesichtspunkt der Zweckmäßigkeit spricht dabei nicht nur für die pragmatische Ausrichtung des hier verwendeten Verfahrens, sondern überhaupt für die Forderung nach Ökonomie eines jeden Systems.

Anmerkungen:

* Vortrag, gehalten auf der Internationalen Konferenz für »Computational Linguistics«, Stockholm, 1.-4. September 1969.

1 Vgl. Tesnière »ordre lineaire« und »ordre structural«, Elements de Syntaxe Structurale, Paris 1959, S. 16-21.

2 Vgl. dazu insbesondere P. Postal, Constituent Structure, Den Haag 1964, S. 67-70.

3 Einen allerdings sehr unvollständigen Überblick über nichttransformationelle Parser, die DK erfassen, gibt D. G. Bobrow in Abschnitt 7.2.5 seines Aufsatzes »Syntactic Theories in Computer Implementations«, in Harold Borko, Hrsg., Automated Language Processing, New York 1967 S. 215-251, ebd. S. 38-241. Die beiden bekanntesten Verfahren stammen von V. H. Yngve (A Model and a Hypothesis for Language Structure, in Proc. Am. Philos. Soc. 104, 5, 1960, S. 444 bis 466) und G. H. Harman (Generative Grammars without Transformational Rules: A Defense of

Phrase Structure, in *Language* 39, 1963, S. 597-616). Beide sind Ausweitungen von Phrasenstrukturgrammatiken (PSG). Harman ersetzt die einfachen Kategoriensymbole der PSG durch komplexe Symbole, in denen zusätzlich Angaben über die Verwendung der jeweiligen Kategorie im Satz gemacht werden. Yngve führt Regeln der Form $A \rightarrow B \dots C$ ein, wobei durch die Art, wie die Regel eingeführt wird, festgelegt ist, was anstelle von »...« stehen darf. Die Einführung solcher Regeln ändert jedoch nichts an der schwachen generativen Kapazität, wie G. H. Matthews in seinem Aufsatz >Discontinuity and Asymmetry in Phrase Structure Grammars< in *Information and Control* 6, 1963, S. 137-146 bewiesen hat.

In seinem Aufsatz >MT at M.I.T.< 1965, in A. D. Booth, Hrsg., *Machine Translation*, Amsterdam 1967, S. 451-523, beschreibt Yngve S. 486-490 einen in COMIT geschriebenen Parser von Fabry, der, auf Yngve und Harman aufbauend, auch diskontinuierliche Konstituenten einbezieht. Erwähnt sei hier noch die von E. Zierer entwickelte Technik, diskontinuierliche Konstituenten darzustellen. Vgl. dazu seinen Aufsatz >Ein Modell zur Darstellung von diskontinuierlichen Konstituenten<, in: *Beiträge zur Sprachkunde und Informationsverarbeitung*, 4, 1964, S. 33-37.

4 Z. Z. im Druck: Tübingen 1969; vgl. zum Thema besonders die Abschnitte 2.13, 6.3, 6.5, 6.6.

5 Vgl. Erklärungen in >Elektronische Syntaxanalyse<.

6 Bisher noch nicht praktisch gelöst.

7 Dazu gehören u. a. Demonstrativwort und Substantiv.

8 Aufgrund der Komplexität nominaler Gruppen reicht eine einfache lineare Reduktion bestimmter Wortklassensequenzen nicht aus.

9 *Elektronische Syntaxanalyse*.

10 Vgl. dazu in Einzelnen: *Linguistische Arbeiten des Germanistischen Instituts...*, Nr. 7, 1969: Dokumentation eines Verfahrens zur Analyse russischer Sätze