In: Beiträge zur Linguistik und Informationsverarbeitung Heft 21. München: Oldenbourg-Verlag, 1971, S. 36-49

Zur Auflösung von Mehrdeutigkeiten bei einer maschinellen Analyse des Deutschen

Von HARALD H. ZIMMERMANN, Saarbrücken

1. Überblick

Bei einer maschinellen Analyse natürlicher Sprachen wie etwa des Englischen oder des Deutschen (wobei es im Grunde keine Rolle spielt, ob die Analyse Teil eines Übersetzungsverfahrens oder Selbstzweck ist) sind häufig Strukturen festzustellen, die je nach Satzzusammenhang mehrere unterschiedliche, einander ausschließende Funktionen ausüben können. Von diesen Mehrdeutigkeiten können drei Typen aufgrund der Häufigkeit ihres Vorkommens als die wichtigsten angesehen werden:

1. die Mehrdeutigkeit einer Wortform im Hinblick auf ihre Wortklassenzugehörigkeit (Homographie)
2. die Mehrdeutigkeit einer Wortform oder Wortgruppe im Hinblick auf Kasus und Numerus bei Nomina (bzw. auf Person und Numerus bei finiten Verben)
3. die semantische Mehrdeutigkeit

Aufgabe des Analysealgorithmus ist es, die Mehrdeutigkeiten so aufzulösen, dass der vom Autor beabsichtigte kommunikative Effekt erkannt wird. Erschwert ist die Reduktion durch das gehäufte Auftreten von Mehrdeutigkeiten in den zu analysierenden Einheiten. Daher werden geeignete Strategien benötigt, die auch diesem Phänomen gerecht werden. In diesem Referat wird das in Saarbrücken praktizierte Verfahren zur Auflösung von Homographen und zur Reduktion von Kasusmehrdeutigkeiten vorgetragen und zur Diskussion gestellt. Für den Teilbereich der semantischen Mehrdeutigkeit wird die noematische Analyse von Georg F. Meier [1] als Möglichkeit zur Vereindeutigung kurz beschrieben. Eine Alternative zum Saarbrücker Verfahren bildet die Prädiktive Analyse, eine Methode von Rhodes [2] und Oettinger [3], die abschließend skizziert wird.

2. Zur Adäquatheit des Analysealgorithmus

Wir wollen mit unseren Überlegungen ansetzen bei der Forderung Chomsky's nach einer (deskriptiv) adäquaten Grammatik. Sie könnte - im Hinblick auf die maschinelle Analyse einer Sprache formuliert - etwa lauten: Ein Analysealgorithmus muss in der Lage sein, solche und nur solche Strukturbeschreibungen - etwa von Sätzen - zu erstellen, die der linguistischen Intuition des Autors entsprechen. Obwohl diese Forderung in der Transformationsgrammatik selbst als *praktisch* niemals völlig erfüllbar angesehen wird, ist sie zumindest in der angewandten Sprachwissenschaft als Maßstab für die Adäquatheit einer 'operativen' Grammatik heranzuziehen; also ein Ziel, dem man sich asymptotisch nähern sollte. Die Schwierigkeiten, denen man bei der Formulierung der Grammatikregeln begegnet - dies gilt auch für die Lexikon-Komponente - lassen sich u. a. darauf zurückführen, dass das Sprachsystem nur in der Performanz erschlossen werden kann, sei es mittels der oft recht unzulänglichen intuitiven Erfahrung des Native-Speaker(-Linguisten) oder der ebenfalls oft recht mangelhaften statistischen Erfassung großer Textmengen.

3. Mehrdeutigkeit sprachlicher Zeichen

Eine weitere Schwierigkeit ist in der Mehrdeutigkeit (natürlicher) sprachlicher Zeichen begründet. Zunächst sei auf einige Mehrdeutigkeitstypen exemplarisch hingewiesen. Ich stütze mich dabei auf die entsprechenden Abschnitte des Berichts 'Elektronische Syntaxanalyse' der Saarbrücker Arbeitsgruppe für linguistische Datenverarbeitung [4] sowie auf die ausführliche, wenn auch nach anderen Gesichtspunkten gegliederte Behandlung der syntaktischen und semantischen Mehrdeutigkeit durch Erhard Agricola [5].

Zu den Typen:

1. syntaktisch mehrdeutige Wortformen (Homographen)
    a. SONDERN: Wir *sondern* (VRB) uns ab / nicht er, *sondern* (KON) ich
    b. b) ZU: *zu* (PRP) ihm / *zu* (ADV) sehr / das *zu* (ZADJ) singende Lied / um es *zu* (ZINF) sehen / er hört mir *zu* (VZS) / nach Osten *zu* (POP)
    c. c) BILLIGEN: die *billigen* (ADJ) Schuhe / wir *billigen* (VRB) es
2. semantische Wortformenmehrdeutigkeit
   DAME: die *Dame* des Hauses / die *Dame* im Kartenspiel
3. syntaktisch mehrdeutige, durch syntaktischen Kontext lösbare Wortgruppen
   Diesen ... Schimpansen (Sing. AKK) sollte man füttern / *Diesen ... Schimpansen* (Plur. DAT) sollte man Futter geben (Agricola S. 70)
4. syntaktisch mehrdeutige, nur durch semantischen Kontext lösbare Gruppen
   24 Tote (NOM/AKK) haben *Überschwemmungen* (NOM/AKK) gefordert (Agricola S. 71)
5. syntaktisch und semantisch mehrdeutige Strukturen
   Von ihren Eltern erstickt aufgefunden wurde ... die 14 Wochen alte Ute G. (Agricola S. 73)

Im Saarbrücker Verfahren zur automatischen syntaktischen Analyse von Sätzen der deutschen Gegenwartssprache wurden Strategien entwickelt zur Auflösung von Mehrdeutigkeiten des Typs 1 (Homographen) und des Typs 3 (syntaktisch lösbare Kasusmehrdeutigkeit). Beide Typen werden innerhalb des gesamten Analyseverfahrens in getrennten Analyseschritten behandelt. Die *Auflösung der Homographie* erfolgt unmittelbar nach der Satzeingabe, wobei jeder Wortform bereits die dem maschinellen Lexikon entnommenen Angaben beigefügt sind. Die weiteren Schritte der Analyse bauen auf den Ergebnissen der Homographenreduktion auf. Es sind im wesentlichen:

die *Abgrenzung von „Analyseeinheiten*", die in der Regel den Subsätzen (= „clauses", Haupt- oder Nebensätze) oder bei diskontinuierlich aufgebauten Subsätzen, also solchen Haupt- oder Nebensätzen, in die andere Subsätze eingebettet sind, den durch die Einbettung entstandenen Subsatz-Teilen entsprechen, z. B.:

DER MANN, DER DORT GEHT, IST MEIN FREUND,
AE 1 / AE 2 / AE 3

die *Nominale Gruppierung*, z. B.:
(1) ER
(2) DER VATER
(3) DER IHN SEHENDE MANN

die *Klassifikation der Verbalgruppe*, z. B.:

(1) SINGT
(2) HAT ... GESEHEN
(3) GESEHEN WORDEN SEIN WIRD

die *Inventarisierung*, d. h. die Klassifizierung der Analyseeinheiten aufgrund ihres grammatischen Inhalts, anschließend die - möglicherweise erst durch Zusammenfassung von Analyseeinheiten durchzuführende - endgültige *Klassifizierung* von Subsätzen. Bereits im Zusammenhang mit der nominalen Gruppierung sind Kasusreduktionen (aufgrund der Kongruenz) möglich, die anschließend noch mehrdeutigen nominalen Gruppen werden im bisher letzten Schritt der Satzanalyse, der *Kasusreduktion*, aufzulösen versucht.

Ein Vorteil dieses Verfahrens liegt in der systematischen Trennung von Strukturerkennungsprozeduren. Auf jeder Stufe der Analyse ist - ähnlich der Fulcrum-Research-Methode Garvins [6] - nur eine bestimmte Kategorie von Informationen zu befragen oder zu erarbeiten, was wesentlich zur Übersichtlichkeit des Verfahrens beiträgt. Diese Überschaubarkeit der einzelnen grammatischen Fragen selbst noch im Maschinenprogramm führt zu einer großen Flexibilität und relativ leichten Modifizierbarkeit der „Annäherungsgrammatik", die das Analyseprogramm im Grunde darstellt.

4. Auflösung von Homographen

Die einzige Ausnahme in diesem System stellt der Algorithmus zur Auflösung der Homographie dar. Die Wortklassenmehrdeutigkeit ist nur anhand von kontext-sensitiven Regeln aufzulösen. Abgefragt und berücksichtigt werden im wesentlichen:

1. die den Homographen umgebenden Wortformen und Wortklassen (Kontaktwortklassen)
2. die Flexionsangaben (Kongruenz)
3. die Wortstellung (Satzanfang, Satzende ...)
4. das satztypisierende funktionale Inventar (Satzzeichen, Konjunktionen...)

In den seltensten Fällen führt dabei eine Abfrage zur Auflösung der Mehrdeutigkeit, sondern die gefundenen Informationen werden kombiniert.

Die Reduktionsprogramme sind auf den Homographentyp zugeschnitten, anders ausgedrückt: für jeden Homographentyp ist ein spezielles Lösungsprogramm zuständig. Es würde zu weit führen, die bisher zugrundegelegten mehr als 50 Typen anzuführen, zumal die Anzahl von der Wortklasseneinteilung und dem Adäquatheitsgrad des Analysealgorithmus mitbestimmt ist (etwa wird bisher die Mehrdeutigkeit Substantiv/Eigenname noch nicht aufgelöst, andererseits wird zwischen finitem Verb und Infinitiv unterschieden). Festzuhalten ist, dass nach unseren Untersuchungen und auf der Basis unserer Homographenklassifikation (d. h. etwa unter Vernachlässigung der Großschreibung von Substantiven als homographieminderndes Merkmal) mehr als 40% aller laufenden Wortformen eines deutschen Textes der Wortklasse nach mehrdeutig sind. Es steht also zu erwarten, dass in einem zu analysierenden Satz mehr als ein Homograph auftritt. Dabei sind folgende Lösungsmöglichkeiten gegeben, die bei der Analysestrategie zu berücksichtigen sind:

1. Der Homograph kann unabhängig von - auch noch nicht aufgelösten - weiteren Homographen im Satz bestimmt werden

a) aufgrund von Stellungsregeln
b) aufgrund der Tatsache, dass alle Lösungsmöglichkeiten eines oder mehrerer anderer Homographen die Reduktion des bearbeiteten Homographen in derselben Richtung beeinflussen

2. Bei der Abfrage von Kontaktwortklassen genügt bisweilen die Kenntnis, dass ein anderer, noch nicht aufgelöster Homograph einer bestimmten Wortklasse nicht angehören kann

3. Es besteht die Möglichkeit, dass die Lösungsrichtung des Homographen von der Auflösung eines anderen, noch nicht gelösten Homographen abhängig ist.

Im letzten Fall sind wiederum zwei Möglichkeiten denkbar:

1. Der 'andere' Homograph kann unabhängig von dem ersteren aufgelöst werden
2. Die Lösungen sind gegenseitig voneinander abhängig.

Allen diesen Möglichkeiten muss das Homographenauflösungsverfahren gerecht werden. Grundsätzlich beginnt die Auflösung der Homographen beim ersten Wort des Satzes und endet beim letzten Wort. Falls jedoch die Möglichkeit 3 zutrifft, wird diese Strategie durchbrochen. Das Auflösungsprogramm für den betreffenden Homographen wird ggf. ohne Ergebnis verlassen, und der nächste Homograph im Satz wird bearbeitet. Ist das Satzende erreicht und sind noch unaufgelöste Homographen vorhanden, wird bei dem ersten nicht gelösten wieder begonnen. Auf diese Weise können alle Homographen, die nicht gegenseitig voneinander abhängig sind, bestimmt werden. Falls der vorher nicht lösbare Homograph von der Auflösung einer anderen Mehrdeutigkeit abhängig war, die aufgrund der Möglichkeiten 1. - Stellung - oder 2. - negativer Ausschluss - aufgelöst werden konnte, lässt sich seine Lösung im nächsten Satzdurchlauf` durchführen.

Bei gegenseitiger Abhängigkeit von Homographenauflösungen ist dieses Verfahren nicht mehr anwendbar. Es ist jedoch möglich, aufgrund der Art der Abhängigkeit auf die Lösungsrichtung eines oder mehrerer Homographen zu schließen. Wir haben daher versucht, die notwendigen Abfragen der Kontaktwortklassen so zu formulieren, dass in diesen Fällen die abgefragte Wortklasse als Indikator für die Auflösungsrichtung dieses Homographen angesehen werden kann.

Um ein Beispiel zu wählen: Handelt es sich bei dem zu lösenden Homographen um den Typ Präposition/Verbzusatz (HO 15, Beispiel: MIT), so lautet eine Programminstruktion „FOLGT EIN DEMONSTRATIVWORT?". Folgt als nächstes Wort ein Homograph vom Typ Demonstrativwort/Relativwort (HO 43, Beispiel: DER), so findet sich dort die Instruktion: „VORHER EINE PRÄPOSITION?". Die Abfrage in HO 15 korrespondiert also mit der in HO 43. Nachdem die Lösungsmöglichkeiten 1. und 2. nicht gegeben sind, wird schließlich für den links im Satz stehenden Homographen (hier also HO 15) die Lösungsprozedur in Abhängigkeit von der abgefragten Lösungsmöglichkeit (von HO 43 also) fortgesetzt und gelangt so zu dem gleichen Ergebnis, wie wenn statt HO 43 bereits die Information Demonstrativwort gestanden hätte. HO 43 wird später bei der entsprechenden Abfrage ein Ergebnis vorfinden und damit jetzt nach Strategie 1. gelöst.

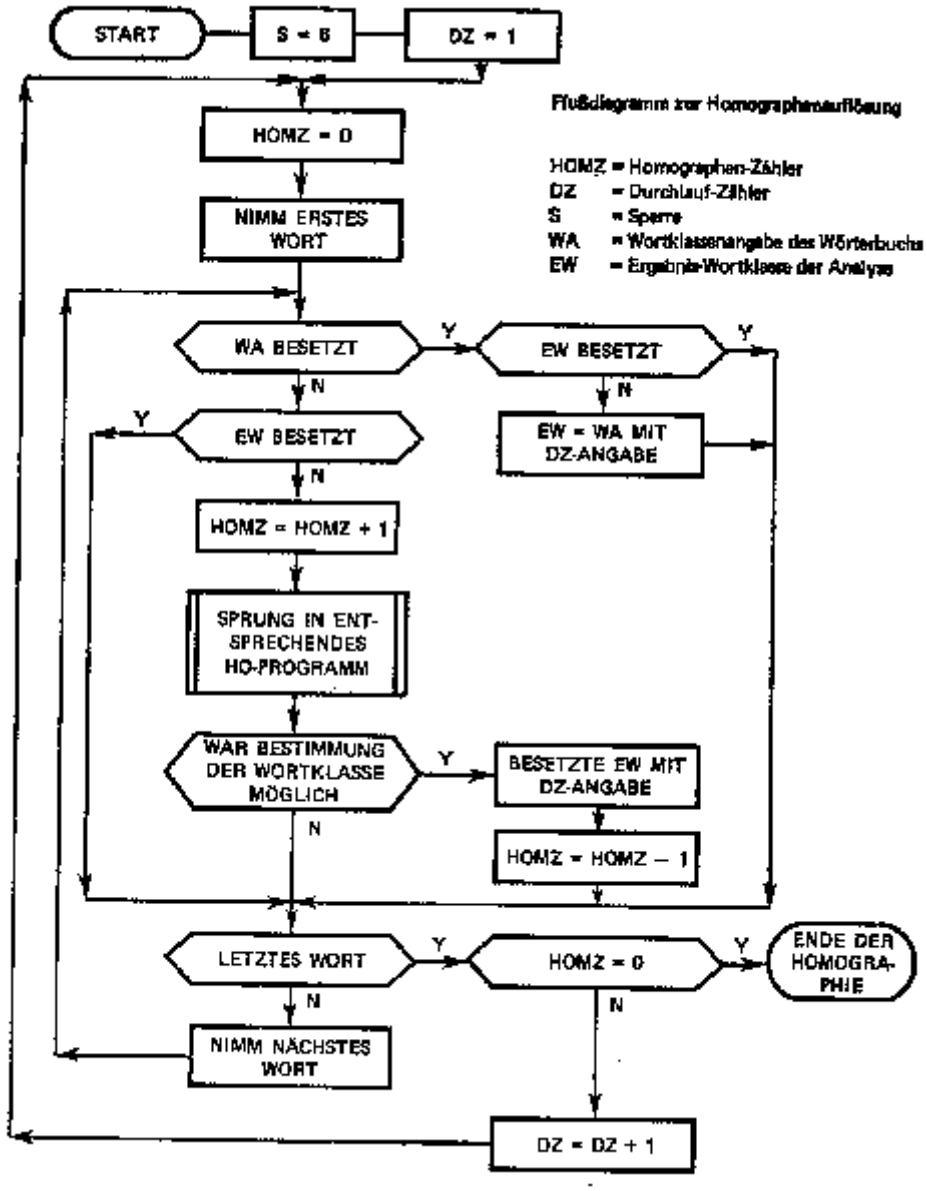

Abb. 1

Die technische Vorgehensweise zeigt das Flussdiagramm zu den Homographen-Durchläufen. Unter einem Durchlauf wird das einmalige systematische Vorgehen von links nach rechts über den gesamten Satz verstanden. Ein Durchlauf ist beendet, wenn das letzte Wort des Satzes bearbeitet wurde. Drei Indikatoren geben während der Homographenanalyse Auskunft über die strategische Phase: Ein Homographen-Zähler HOMZ notiert die Anzahl der in einem Durchlauf angetroffenen, noch nicht aufgelösten Mehrdeutigkeiten. Ist das letzte Wort bearbeitet, wird - bei HOMZ = 0 - der sog. Durchlaufzähler DZ um 1 erhöht, und der Versuch der Homographenauflösung beginnt wieder mit dem ersten noch nicht reduzierten Homographen im Satz. Daneben ist eine Sperre S errichtet, die bewirkt, dass in den Durchläufen 1 bis 5 die Strategie 3. nicht angewendet werden kann. In dieser Phase wird das spezielle Homographenprogramm verlassen, falls bei der Abfrage einer Wortklasse ein noch nicht gelöster Homograph betroffen ist, der eine relevante Mehrdeutigkeit aufweist. Ist die Sperre erreicht (S = DZ), wird sie für die betreffende Lösung außer Kraft gesetzt und gleichzeitig um 1 erhöht (dies ist in dem Flussdiagramm nicht zu erkennen, da diese Routine Teil des nicht explizit dargestellten Wortklassenabfrage-Programms ist). Dadurch wird bei der Ausgabe des Analyseergebnisses deutlich, welche Homographen im gleichen Durchlauf - dies bedeutet bei DZ >

5: in Abhängigkeit voneinander - gelöst worden sind. Auf diese Weise sind (etwa beim Testen der Programme) eventuelle Fehlerursachen leichter zu ermitteln (vgl. auch das in der Anlage beigefügte Satzbeispiel 20076, wo unter der Spalte D beim 4. und 5. bzw. 8. und 9. Wort durch die Zahlen 6 bzw. 7 entsprechende Abhängigkeitslösungen angezeigt werden).

Zum Abschluss dieses Abschnitts noch einige Bemerkungen zur Brauchbarkeit unserer Methode: Die bisher erreichte Quote an richtig reduzierten Homographen, ermittelt an mehr als 1400 analysierten Sätzen unterschiedlicher Länge, liegt etwa bei 90%, wobei vorausgesetzt ist, dass die Wortformen einschließlich der notwendigen grammatischen Angaben bereits im maschinellen Wörterbuch verzeichnet sind[1]. Die einzelnen Homographenprogramme ließen sich noch bedeutend verfeinern, so dass im Grunde noch nicht genau abzusehen ist, wo die Grenzen des Verfahrens liegen. Dennoch ist der Aufwand bereits beträchtlich, er beträgt heute - gemessen anhand der Anzahl der Programmierinstruktionen - mehr als ein Drittel (ca. 5-6000) der gesamten Analyseinstruktionen. Dies ist mit darin begründet, dass teilweise Analyseoperationen der 'nächsthöheren' Stufen vorweggenommen werden (etwa Kongruenzabfragen, verbale Wortstellung, ja sogar in gewissem Maß die Subsatzabgrenzung bzw. -Inventarisierung).

Die Lösungsquote verhält sich annähernd proportional zur Komplexität (Akzeptabilität) der Satzstrukturen und der Satzlänge; bei komplizierten Sätzen sinkt also der Anteil der richtigen Lösungen. Dem ist teilweise dadurch Rechnung getragen, dass im späteren Verlauf der Analyse für einige Homographentypen Fehlererkennungsroutinen ablaufen und ggf. automatisch Wortklassenkorrekturen durchgeführt werden. Unter Berücksichtigung der dabei ermittelten Funktion wird die gesamte Analyse (einschließlich der übrigen Homographenauflösungen) wiederholt, um Kettenreaktionen von Fehlern („Fehlerfortpflanzung") weitgehend auszuschalten.

5. Auflösung von Kasusmehrdeutigkeiten

Die durch die nominale Gruppierung noch nicht auf einen Kasus reduzierten Wörter oder Nominalgruppen werden im bisher letzten Schritt der Analyse behandelt. Folgende „natürliche" Mehrdeutigkeiten treten auf:

1 NOM/AKK	die Frau
2 NOM/GEN	der Wagen
3 GEN/DAT	der Frau
4 DAT/AKK	den Wagen
5 NOM/GEN/AKK	Häuser
6 NOM/DAT/AKK	Gott
7 NOM/GEN/DAT/AKK	Blumen

Die Voraussetzungen, auf denen die Reduktionsprogramme aufbauen, sind:

a) Die Anzahl der Nominalgruppen im Subsatz ist bekannt (Ausnahme: Gruppen mit der Mehrdeutigkeit GEN/DAT können erst hier u. U. als genitivische Subgruppe - also als Genitivattribut - bestimmt werden).

b) Der Rektionsträger (das Wort der verbalen Gruppe, das die Valenzen für den Subsatz nach sich zieht) ist - falls vorhanden - ermittelt.

c) Das finite Verb ist - falls vorhanden - bekannt.

d) Evtl. zugehörige Gliedsätze (Subjekt-/Objektsätze) sind dem Subsatz als Pseudo-Nominalgruppen mit der Mehrdeutigkeit NOM/AKK zu geordnet.

Nach folgenden Grundregeln werden die Kasusmehrdeutigkeiten zu reduzieren versucht:

1. Im Satz darf nur eine Nominalgruppe (NOG) im NOM stehen, falls der Rektionsträger keine Gleichsetzung verlangt.
2. Es dürfen zwei NOG im NOM auftreten und es darf keine NOG im AKK vorhanden sein, wenn der Rektionsträger Gleichsetzung verlangt.
3. Es dürfen zwei NOG im AKK auftreten, wenn der Rektionsträger doppelten AKK zulässt.
4. Es darf nur dann eine und nur eine NOG im Dativ auftreten, wenn der Rektionsträger eine NOG im Dativ zulässt.
5. Es darf nur dann eine und nur eine NOG im AKK auftreten, wenn der Rektionsträger eine NOG im AKK zulässt.
6. Es darf nur dann eine und nur eine NOG im GEN auftreten, wenn der Rektionsträger eine NOG im GEN zulässt (vgl. aber Voraussetzung a)).
7. Falls nach Verwendung der Regeln 1-6 nicht alle Mehrdeutigkeiten aufgelöst werden können, wird vorläufig nach Wahrscheinlichkeitsgesichtspunkten (Stellung) entschieden.

Zur Auflösung der Mehrdeutigkeiten wird eine Matrix aufgebaut, die Felder für alle möglichen eindeutigen und mehrdeutigen (künstlichen oder natürlichen) Mehrdeutigkeiten enthält (Auszug):

Kasus	RM1	RM2	RM3	RM4	RB1	RB2	RB3	RB4
NOM								
GEN								
DAT								
AKK								
NOM / GEN								
NOM / DAT								
.../ .../ .../								
NOM/GEN/DAT/AKK								

Die im Subsatz auftretenden NOG werden entsprechend ihrer Deutigkeit eingeordnet, wobei in RM (1-4) die Deklinationsmarke und in RB (1-4) die Wortnummer des ersten Wortes der Gruppe mitgegeben werden.

In Abhängigkeit von vorhandenen eindeutigen NOG werden die mehrdeutigen den Grundregeln entsprechend zu reduzieren versucht, wobei eine reduzierte NOG in das für sie neu zutreffende Feld umgespeichert wird. Die Kasusreduktion ist abgeschlossen, wenn keine mehrdeutigen Gruppen mehr in dem entsprechenden Teil der Matrix vorhanden sind.

6. Semantische Analyse

Die hier geschilderten Auflösungsverfahren für mehrdeutige Wörter und Wortgruppen entsprechen dem gegenwärtigen Stand des Saarbrücker Verfahrens. Es wäre nun verlockend,

weitere Möglichkeiten zu erproben, vor allem im Hinblick auf die Lösung semantischer Mehrdeutigkeiten (Polysemie). Einen Versuch, syntaktisch nicht weiter aufzulösende Kasusmehrdeutigkeiten anhand semantischer Kriterien zu reduzieren, schildert der Beitrag von R. Dietrich in der 'Elektronischen Syntaxanalyse'. Auch der umgekehrte Weg - die Auflösung semantischer Mehrdeutigkeiten aufgrund syntaktischer Kriterien -, sofern diese Deutigkeiten unterschiedliche syntaktische Konsequenzen nach sich ziehen, ist denkbar und - wie der Beitrag von 1. Batori [7] auf dem Linguistenkongress 1969 in Stockholm zeigte - auch praktikabel. Von großem Nutzen für die Auflösung noch verbleibender Mehrdeutigkeiten wäre auch eine über den Einzelsatz hinausgreifende, größere Texteinheiten berücksichtigende Analyse.

Das Hauptproblem einer adäquaten Analyse - vor allem im Hinblick auf eine maschinelle Übersetzung - ist jedoch die nur durch semantische Merkmale und Regeln aufzulösende semantische Mehrdeutigkeit. Bereits die Merkmalklassifikation von Lexikoneintragungen bereitet hier nahezu unüberwindliche Schwierigkeiten. Ein Beispiel dafür bietet Georg F. Meiers eingangs erwähntes „noematisches" Analyseverfahren zur Auflösung der Polysemie, in dem syntaktische und semantische Operationen und Merkmale verknüpft werden. Eine Vereindeutigung wird dabei aufgrund der Verträglichkeit bzw. Nichtverträglichkeit von Bedeutungselementen (den sog. „Noemen") der Lexeme eines Kontexts zu erreichen versucht. - Wie es scheint, ist dieses System durchaus praktikabel, doch linguistische Schwierigkeiten bei der Klassifikation der Bedeutungselemente sind nicht zu übersehen.

7. Prädiktive Analyse

Abschließend noch einmal zurück zum syntaktischen Analyseverfahren: Während man noch vor einigen Jahren in der Methode - dies trifft auch für Saarbrücken zu - abhängig war von Speicherkapazität und Geschwindigkeit des Elektronenrechners, bietet die „dritte Computergeneration" mit Großraumspeicher und Zykluszeiten im Nanosekundenbereich neue Möglichkeiten für speicher- und rechenaufwendige Analyseversuche.

Bereits 1959 hatte Ida Rhodes ein Verfahren zur Analyse von syntaktisch mehrdeutigen Sätzen und Strukturen vorgeschlagen, das der menschlichen Erkennungsprozedur vielleicht am nächsten kommt. Es ist unter den Begriffen „Predictive Analysis" bzw. „Multiple-Path-Analysis" vor allem durch seine Verwendung in A. G. Oettingers Übersetzungsprogramm in den sechziger Jahren bekannt geworden. Die dabei verfolgte Strategie lässt sich wie folgt skizzieren:

Die Analyse beginnt mit dem ersten Wort des Satzes und sagt aufgrund der bis zur gerade erreichten Stelle ermittelten Informationen voraus, welche syntaktischen Strukturen in der Folge zu erwarten (oder noch möglich) sind. Im Prinzip handelt es sich dabei um eine Trial-and-error-Methode, denn ein 'Weg' wird dann als erfolgreich angesehen, wenn die entsprechenden Voraussagen eintreffen und auf diese Weise schließlich das Satzende erreicht wird; andernfalls wird der eingeschlagene Weg abgebrochen und auf die nächste Möglichkeit zurückgegriffen. Das Verfahren kann so aufgebaut sein, dass alle möglichen Wege verfolgt werden, so dass syntaktisch mehrdeutigen Sätzen auch alle entsprechenden Strukturen zugeordnet werden.

Die Auflösung von Mehrdeutigkeiten ist hier also im Gegensatz zum Saarbrücker Verfahren in den gesamten Analyseprozess integriert. Wie hoch der bisher erreichte Genauigkeitsgrad der Auflösung ist, entzieht sich meiner Kenntnis.

Im Zusammenhang mit einer Kapazitätserweiterung des Saarbrücker Rechenzentrums (im Herbst 1970 ist eine Rechenanlage vom Typ CDC 3300 aufgestellt worden) böte sich die Gelegenheit, ein prädikatives Analyseverfahren neben dem bisher praktizierten zu entwickeln, um einen unmittelbaren Leistungsvergleich durchführen zu können. Dieses Verfahren ließe sich an dem Übersetzungsprojekt Russisch-Deutsch erproben, zu dem gegenwärtig von einigen wissenschaftlichen Mitarbeitern und Studenten unter der Leitung von Prof. Dr. Hans Eggers, der zugleich Leiter des deutschen Analyseprojekts ist, die Grundlagen und ersten Methoden entwickelt werden.

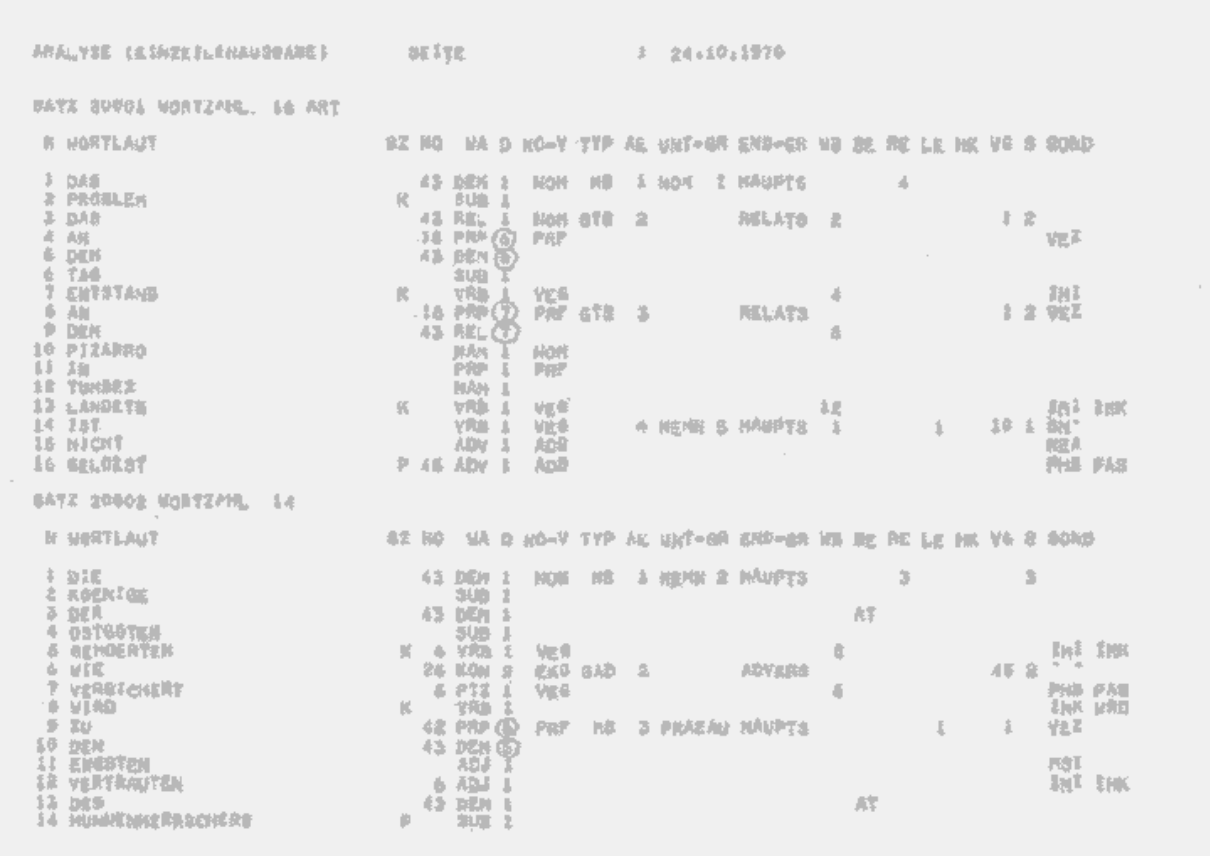

Abb. 2

Literatur:

[1] MEIER, G. F.: Noematische Analyse als Voraussetzung für die Ausschaltung der Polysemie, in: Zeichen und System der Sprache III, Berlin 1966.
[2] RHODES, I.: Syntactic Integration carried out mechanically, in: Ghizetti, Automatic Translation of Languages, Oxford-London 1966, S.205-209.
[3] OETTINGER, A. G., und SHERRY, M.: Current research an automatic translation at Harvard University and predictive syntactic analysis, in: H. P. Edmundson (ed.): Proceedings of the National Symposium an M. T. (...), Englewood Cliffs 1961, S. 173-182.
[4] EGGERS, H. u. a.: Elektronische Syntaxanalyse der deutschen Gegenwartssprache, Tübingen 1969.
[5] AGRICOLA, E.: Syntaktische Mehrdeutigkeit (Polysyntaktizität) bei der Analyse des Deutschen und des Englischen, Berlin (DDR) 1968.
[6] GARVIN, P. L.: Syntactic Retrieval, in: H. P. Edmundson (vgl. [3]), S.286-292.

[7] BATORI, I.: Disambiguating Verbs with Multiple Meaning in the MT-System of IBM Germany. Preprint Nr. 31 of the International Conference an Computational Linguistics, Stockholm 1969.

Anmerkungen:

1 Das derzeitige Wortformenbuch (Stand: April 1971) enthält ca. 40 000 Lexikoneinträge, orientiert an dem Wortschatz eines Kontrollmaterials von 11 000 Sätzen, wobei darüber hinaus die Funktionswortklassen nahezu vollständig erfasst sind und auch die nicht in den Texten belegten Deutigkeiten der Wortformen berücksichtigt wurden. Mittlerweile wurde ein Verfahren entwickelt, das es erlaubt, beliebige Sätze zu verarbeiten, d. h. auch solche, in denen Wortformen auftreten, die nicht im maschinellen Lexikon vorhanden und damit noch nicht in ihrer natürlichen Deutigkeit präklassifiziert sind. Obwohl in diesen Sätzen zu den natürlichen vorzuklassifizierende und damit funktional zumeist mehrdeutige Wortformen kommen, erhöht sich in diesem Fall die Fehlerquote nur unbedeutend, sofern der Anteil der 'unbekannten' Wörter nicht übermäßig hoch ist.