

# AI meets DA

---

**Wilhelm Grießhaber**



**sprachen ●● zentrum**

Münster – PALM 02/1999

© W. Grießhaber

## §0 AI meets DA

Vom Nutzen der empirischen Diskursanalyse für die automatische Spracherkennung.

1. Dialogmodellierung in der Planungsphase
  - 1.1. Dialog- und Diskursverständnis der KI
  - 1.2. Sprachanalyse der KI
2. Dialogmodellierung in der Projektlaufzeit
  - 2.1. Szenarioforschung und Datenerhebung
  - 2.2. Dialogmodellierung der KI
3. Neue Perspektiven mit der Modellierung mittels Petri-Netzen
4. Literaturverzeichnis

### Abstract

Nutzen der DA für AI:

- (a) aus einem zweckbezogenen Verständnis sprachlichen Handelns berücksichtigt die DA über die unmittelbar sichtbaren sprachlichen Handlungen auch die gesellschaftlichen Voraussetzungen und Rahmenbedingungen des Handelns und die erschließbaren mentalen Dimensionen ('Einbettung', 'Motivierung', '**Authentizität**' usw.)
- (b) aus einer funktional-zweckbezogenen Sicht auf die sprachlichen Mittel und ihre Verwendung beim sprachlichen Handeln folgt eine Modellierung des Handelns im **Handlungsmuster**, das eine adäquatere Diskursmodellierung und Klassifizierung sprachlicher Mittel erlaubt als die in der KI verwendeten flachen Modellierungsverfahren
- (c) durch Verwendung computergestützter Werkzeuge, z.B. von **Petri-Netzen**, werden dynamische Modellierungen möglich, die die statischen Musterdiagramme einerseits abstrakt modellieren, andererseits mit eingegebenen Daten konkrete Musterdurchläufe simulieren; damit werden auch die Grenzen zwischen qualitativen und quantitativen Forschungsmethoden überwunden.

### 1. Dialogmodellierung in der Planungsphase

Wesentliche linguistische Vorgaben im Forschungsverbundvorhaben VERBMOBIL wurden bereits in der vorbereitenden Studie (Verbmobil 1991) definiert, in der die wissenschaftliche Grundlage gelegt und das ambitionierte Vorhaben für eine mehrjährige Realisierung aufbereitet wird. Die Umsetzung des Konzepts in Netzpläne (Wahlster & Engelkamp 1992) operationalisiert das Vorhaben für eine

---

Ausschreibung der insgesamt 16 Arbeitspakete. Während die Netzpläne im wesentlichen den Vorgaben der Studie folgen, sind in den eingereichten Anträgen (z.B. Anträge Hamburg 1992) eigene Akzentsetzungen zu beobachten.

Die Entwicklung eines computergestützten Dolmetschsystems für einen bestimmten Diskurstyp stellt hohe Anforderungen: die sprecherunabhängige Erkennung gesprochener Spontansprache, die zuverlässige Verarbeitung der erkannten Einheiten und die Umsetzung der ausgangssprachlichen Äußerungen in zielsprachliche mit anschließender akustischer Ausgabe. Da jeder der genannten drei Bereiche für sich bisher nur ansatzweise gelöst ist, wird die Herausforderung deutlich, die die integrierte Lösung für ein Dolmetschgerät darstellt. Das Zentrum des disziplin- und fächerübergreifend angelegten Projekts liegt ungeachtet der Beteiligung von Sozial- und Sprachwissenschaftlern im Bereich der maschinellen Sprachverarbeitung, d.h. der Informatik oder der Künstlichen Intelligenz, bzw. der Artificial Intelligence.

Die Studie zeigt deutlich, daß die Ausrichtung auf die Entwicklung von Hard- und Softwarekomponenten dazu führte, daß die Informatik aus ihren sprachtechnologischen Interessen heraus die Beiträge der Linguistik zum Gesamtprojekt bestimmte. Linguistische Themen werden somit durch den Filter der Projektentwicklung betrachtet. Linguistische Forschungsergebnisse werden im Zusammenhang mit geeigneten Tools zu ihrer Implementierung in der vorgesehenen Systemarchitektur bewertet und berücksichtigt. Für die technische Realisierung spielen algorithmisierte Verfahren die Hauptrolle, auch wenn sie für den Anwender unsichtbar hinter deklarativen Teilen sind und im Projekt die Kombination von herkömmlichen und konnektionistischen Ansätzen ausgelotet werden soll. Insgesamt werden linguistische Ansätze mit strikt formalisierten Beschreibungsverfahren präferiert, z. B. HPSG-Theorien (Prouidian, D. & Pollard, C. J. 1985, Balari Ravera, S. 1990), während weniger formalisierte Ansätze an die Peripherie rücken.

### **1.1. Dialog- und Diskursverständnis der KI**

Die konkrete technologische Zielsetzung bestimmt die Interessen an linguistischen Kenntnissen. Das Vorhaben zielt in sehr konkreter Weise auf die Entwicklung eines computergestützten Dolmetschsystems für internationale Besprechungen im kommerziellen Bereich und nicht auf ein universell einsetzbares Dolmetschgerät für irgendwelche Besprechungen. Für das Verbundvorhaben VERB-MOBIL wurde das Einsatzgebiet noch weiter eingeschränkt auf "Firmenvorstellung mit Produktpräsentation" (Studie IX). Für die erste Phase wurde der Diskursbereich erheblich verengt auf "Terminplanung anhand eines Kalenders (appoint-

ment scheduling task)“ (Netzpläne, 11). Die erste technische Realisierung, der Demonstrator, sollte einen beispielhaften TA-Dialog erkennen und verarbeiten können. Das fertig entwickelte System soll also eine Menge mündlich gesprochener Äußerungen, die dem Beispieldialog entsprechen, verarbeiten können. Dabei sollte die Erkennung sprecherunabhängig sein. Das Entwicklungsziel ist also durch eine Menge von Äußerungen bestimmt, die im Beispieldialog fallen (können). Durch eine Art Drehbuch für die Datenaufnahme wird eine gewisse Homogenität der erhobenen Daten angestrebt (vgl. Beitrag xx): so sollen die Sprecher gut artikulieren, ohne Dialekt und grammatisch korrekt sprechen und zur Terminabsprache z.B. nur Tageszeiten, aber keine Uhrzeiten verwenden. Die so in vielen Simulationen mit vorgegebenem Kalender gewonnenen Daten bilden für die Entwicklung das relevante Korpus. Die Daten definieren damit faktisch auch, was unter ‘Diskurs, ‘Dialog’ oder ‘Terminabsprache’ zu verstehen ist, eine weitergehende begriffliche Bestimmung ist für die Demonstratorentwicklung nicht notwendig.

Es überrascht also nicht, daß in der Studie die Begriffe 'Dialog' oder 'Diskurs' ohne klare Definition verwendet werden. Selbst in einem Satz werden beide Begriffe verwendet (s.u.). In Kapitel 2 der Studie werden die zu berücksichtigenden Äußerungen im Top-down-Verfahren von den Kommunikationsfeldern aus bis hinunter zu den in ihnen jeweils enthaltenen Sprechakten definiert:

“In dieser Studie verstehen wir unter einem **Kommunikationsfeld** eine Menge von **Kommunikationssituationen** (gelegentlich “settings” genannt), die ihrerseits wieder durch einen jeweils spezifischen Satz von **(Sprech-)Handlungen** gekennzeichnet sind. Ein **Szenario** ist eine Konkretisierung einer Kommunikationssituation mit einem gegebenen Sprechertyp, Sprechakten und deren Sequenzen, stilistischen Eigenschaften usw. Ein Korpus für ein Szenario stellt dann eine Menge von **Äußerungen** aus einer Instantiierung eines Szenarios zusammen.” (Studie, 10; Hervorhebungen d. Verfassers)

Von ‘unten’ - den Äußerungen - her betrachtet, stellt die “(Sprech-)Handlung” die entscheidende Bestimmungsgröße dar. (Sprech-)Handlungen werden in Äußerungen realisiert und bilden, in spezifischen 'Sets' zusammengefaßt, eine Kommunikationssituation. Szenarien als Konkretisierung von Kommunikationssituationen sind jeweils durch Sprechertypen, Sprechakte und deren Sequenzen sowie stilistische Eigenschaften charakterisiert. In Kapitel 2.2.3 der Studie werden als Beispiele für Sprechakttypen, die in Szenarien auftreten und von der Maschine erkannt und analysiert werden müssen, Repräsentativa, Direktiva usw. genannt. Für die Programmentwicklung ist schließlich das Korpus als Menge von Äußerungen die relevante Bezugsgröße. Das Korpus repräsentiert die Menge sprachlicher Daten, die vom Programm bearbeitet werden muß. Die weiteren Überlegungen richten sich darauf, in den zunächst nur sehr vage strukturierten Daten regelhafte Abfolgesequenzen und Muster zu identifizieren, die programmtechnisch verarbeitet werden können.

In Hinblick auf die Ausprägung und Sequenzierung der Äußerungen in einem gegebenen Szenario werden im Zusammenhang mit pragmatischen Aspekten in Kapitel 3.2.2.1 der Studie drei Bereiche unterschieden: die Ebene des Diskurstyps, die kommunikativen Intentionen und Diskurspläne der Beteiligten und die thematische Ebene:

- “• der **Diskurstyp** und damit eine zumindest teilweise erwartbare **Diskursstruktur**: Der Diskurstyp “Firmenvorstellung” ... läßt gewisse Erwartungen darüber zu, daß der Dialog mit einer Begrüßungs-/Eröffnungssequenz beginnt und mit einer Schluß-/Verabschiedungssequenz endet. Der Beginn der entsprechenden Dialogsegmente wird meist durch relativ standardisierte bzw. konventionell determinierte Äußerungsparts signalisiert.
- die **kommunikative Rolle der Dialogpartner** sowie ihre **Intentionen** und **Diskurspläne** und **-ziele**: Diese erlauben begrenzte Voraussagen über mögliche Sprechakte, die die Kommunikationspartner mit ihren Äußerungen realisieren. Auch Phänomene der Diskurspragmatik wie Präsuppositionen und konversationale Implikaturen, lassen sich unter Einbeziehung von Sprecherintentionen leichter interpretieren ... .
- die **Domäne** und das **Thema** des Diskurses, die vor allem Einfluß auf seine Fachlichkeit haben.” (Studie 51; Hervorhebungen d. Verfassers)

Auf der ersten Strukturierungsebene wird der Begriff “Diskurstyp” neu eingeführt, der ungefähr dem Szenario entspricht. Ihm entspricht jeweils eine “Diskursstruktur”, z.B. Eröffnungs- und Schlußsequenzen, die mit konventionell determinierten Äußerungen eröffnet werden. Auf der zweiten Strukturierungsebene werden Sprechakte mit Intentionen und Diskursplänen und -zielen der Dialogpartner in Verbindung gebracht. Von den Diskursplänen versprechen sich die Autorinnen begrenzte Voraussagen über die Abfolge von Sprechakten. Die letzte Strukturierungsebene der “Domäne” und des “Thema(s)” bestimmen fachliche Strukturen.

Die Parameter sollen vor allem dazu dienen, die Analysetiefe für die Übersetzung zu steuern. Während z.B. die syntaktische Realisierung einer Begrüßungsfolskel weniger übersetzungsrelevant ist als ihre Stil- und Registerwahl, wird für den eigentlichen thematischen Dialogteil erst eine genauere syntaktische, semantische und pragmatische Analyse die Repräsentation aller übersetzungsrelevanten Informationen gewährleisten. Ziel ist eine möglichst flache Verarbeitung, um rechen- und zeitaufwendige Inferenzmechanismen in Hörer-/Sprechermodell, Diskursplänen oder Domänenmodell zu vermeiden. Im Ergebnis stellt sich somit heraus, daß computerlinguistisch betrachtet, genauere linguistische Modellierungen nach Möglichkeit zu vermeiden sind, da sie im Vergleich zu flachen, schnellen Verarbeitungstechniken ‘teuer’ und weniger effizient sind. Für das Projekt sind also von vornherein vor allem solche diskurslinguistischen Ergebnisse interessant, die Äußerungen mit geringem Rechenaufwand, ohne Rückgriff auf aufwendige Modellierung, anhand von Oberflächenmerkmalen der Äußerungen in verarbeitungsrelevante Einheiten zerlegen. Vereinfacht gesprochen geht es darum, mittels

einer Korpusanalyse typische Abfolgemuster sprachlicher Elemente zu gewinnen, die über 'einfache' Vergleichsoperationen mit prototypischen Musteräußerungen eine ausreichend genaue Zuordnung zu Diskursfunktionen erlauben.

Vor dem Hintergrund des verarbeitungsorientierten Interesses der KI an Diskursstrukturen wird verständlich, daß die KI nicht nur die von der Pragmatik entwickelten funktional bestimmten Kategorien der Diskursanalyse ignoriert, sondern auch die Ansätze der Konversationsanalyse. Selbst die oberflächennahen Ansätze von Kallmeyer & Schütze 1976 kennen zusätzlich zu Eröffnung und Beendigung für den mittleren Teil ein den Interaktionspartnern als eine Art Normalform erwartetes typisches Gerüst von Handlungsschritten(n) und Handlungsschemata, die die Interaktionspartner unter Zugzwänge setzt. Mit diesem Instrumentarium hätten sich einige spezifische Charakteristika der Ablaufstrukturen der TA ermitteln lassen. Einen Schritt weiter geht Dittmann 1979, der die themenzentrierte Interaktion durch Ablaufmuster-Hypothesen strukturiert sieht, die sich aus dem Gesprächszweck ergeben und mit einer Optimalform verbunden sind. Allerdings verfolgen die Beteiligten im sprachlichen Handeln nicht das analytisch ermittelte Ablaufmuster, sondern verfolgen bestimmte Zwecke, die eine bestimmte Struktur nahelegen. Mit dem Handlungszweck führt Dittmann eine zentrale Kategorie des sprachlichen Handelns ein, die jedoch erst in der funktionalen Pragmatik systematisch zur Entwicklung des Musterbegriffs weiterentwickelt ist (s.o. Beitrag xx). Allerdings ist die KI - wie wir oben gesehen haben - an sich gar nicht an einer möglichst genauen Kenntnis der Diskursstruktur interessiert.

Zusammenfassend läßt sich das Diskursverständnis der KI, wie es sich in der Studie zeigt, folgendermaßen bestimmen: Dialoge werden als partiell geordnete Abfolge von Sprechhandlungen betrachtet, die von den Dialogpartnern abwechselnd geäußert werden, Sprecher und Hörer sind wechselseitig austauschbar; der Diskurstyp "Firmenvorstellung" und das Szenario "Terminabsprache" hat eine dreigliedrige Diskursstruktur mit Begrüßung/Eröffnung, thematischen Dialogteil und Schluß/Verabschiedung; Sprechhandlungen werden durch Sprechakte realisiert; kommunikative Rolle, Sprecherintentionen und Diskurspläne erlauben begrenzte Voraussagen über mögliche Sprechakte.

Gegenüber dem linearen Diskursmodell der KI geht die FP von einem zweckbestimmten, musterhaft strukturierten Handeln aus (vgl. das oben vorgestellte Handlungsmuster der Terminabsprache, Beitrag xx). Die Unterschiede zeigen sich u.a. schon in der Betrachtung des Beginns: aufgrund des Musterzwecks hat die TA z.B. eine Präphase zur Realisierungusterspezifischer Funktionen und nicht einfach eine unspezifische Begrüßungs-/Eröffnungsphase. Das Verhältnis von Musterposition und Sukzession im Diskurs ist nicht deckungsgleich, so daß

statistisch ermittelte Äußerungssequenzen erst durch den Bezug auf den Muzterzweck wirklich interpretierbar sind. Auch im reziproken Muster der TA tut der Hörer etwas Hörer-spezifisches, während der Sprecher spricht: er perzipiert die Äußerungen und bildet den Sprecherplan nach, gleichzeitig bildet er einen eigenen Plan aus, er hört nicht einfach passiv zu, sondern steuert mit nonverbalen (Mimik, Gestik) und paraverbalen (Interjektionen) Mitteln den Sprecher während des Sprechens, und kann den Sprecher auch während der Verbalisierung zu einer Änderung des ursprünglichen Sprecherplans veranlassen; Sprecher und Hörer sind nicht wechselseitig austauschbar. Da die KI in der Folge der Oberflächenorientierung Äußerungen direkt mit Sprechakten identifiziert, wird weder in Betracht gezogen, daß ein Turn mehrere SAe enthalten kann, noch daß ein SA systematisch von Sprecher und Hörer in mehreren Turns kooperativ realisiert wird (s.o. xx).

## 1.2. Sprachanalyse der KI

Große Aufmerksamkeit wird in der Studie besonderen Merkmalen der mündlichen Sprache gewidmet, die unabhängig vom Diskurstyp sind. Folgende Probleme erscheinen als problematisch für die maschinelle Sprachverarbeitung: die sprecherunabhängige Erkennung der akustischen Eingabe, die syntaktische Segmentierung der erkannten Eingabe und die Behandlung typischer Erscheinungen der gesprochenen Sprache. Zu den sog. sprechsprachlichen Erscheinungen wird in der Studie 1991 bei der Erläuterung der Variationsparameter der Szenarios ausgeführt:

"Der Anteil speziell sprechsprachlicher Elemente, wie Gliederungssignale, Satzabbrüche, Selbstkorrekturen, pragmatische Floskeln, Metadialoge usw. erhöht die Verarbeitungsschwierigkeiten erheblich. Je spontaner und dialogischer die Situation angelegt ist, um so häufiger muß man mit sprechsprachlichen Eigenschaften rechnen." (Studie 1991, 27)

In den genannten Erscheinungen unterscheidet sich mündliche Sprache markant von wohlgeformten schriftlichen Äußerungen, wie man sie in redigierten Texten findet: die Äußerungen liegen in maschinenlesbarer Form vor, die Äußerungen sind durch Abstände in Wörter segmentiert, Groß- und Kleinschreibung sowie Interpunktion können als verlässliche syntaktische Gliederungsmittel verwendet werden, die Äußerungen selbst weisen normalerweise eine syntaktische Wohlgeformtheit auf. Außerdem sind die Äußerungen ohne Bezug auf den situativen Kontext verfaßt und rezipierbar (vgl. Ehlich ##). Ursprünglich sollten in VERBMOBIL auch über eine visuelle Dateneingabe Gestik und Mimik zur Verarbeitung einbezogen werden, doch wurde dies zunächst zurückgestellt.

In der Studie werden einige syntaktische Merkmale der gesprochenen Sprache bei der Diskussion des situativen Kontexts und der Interaktion der beiden Partner in einer funktionalen Perspektive erwähnt:

"Die Präsenz des situativen Kontexts und die Interaktionsorientiertheit gesprochener Diskurse hat entscheidenden Einfluß auf die syntaktischen Eigenschaften gesprochener Sprache und muß daher auch bei deren syntaktischer Analyse berücksichtigt werden. So sind syntaktische Elemente mit kommunikativ-pragmatischer Funktion wesentlich häufiger in gesprochener als in geschriebener Sprache: **Modal- und Abtönungspartikel** z.B. dienen dazu, mit Äußerungen verknüpfte Illokutionen sowie die Einstellungen, die ein Sprecher zu seiner Äußerung hat, zu verdeutlichen. Andere syntaktische Phänomene wie **Drehsätze**, **Abbruch** und **Neuanfang** von Sätzen sowie **Nachträge** zu syntaktisch bereits abgeschlossenen Sätzen, ..., sind erklärbar als Mittel zur Verständnissicherung. **Ellipsen** sind ein häufig benutztes syntaktisches Mittel zur Steigerung der Kommunikationseffizienz. Sie werden gewöhnlich aus dem situativen Kontext oder aus dem Wissen, das die Sprecher voneinander haben, erklärt oder ergänzt." (Studie 1991, 53; Hervorhebungen d. Verfassers)

Die genannten Phänomene werden in einer funktionalen Perspektive mit dem kommunikativen Handeln der Sprecher in Verbindung gebracht: Modal- und Abtönungspartikeln dienen dem Ausdruck der Illokution und zur Verdeutlichung der Einstellung des Sprechers, syntaktische Brüche wie Drehsätze usw. werden im Zusammenhang mit der Verständnissicherung gesehen und Ellipsen werden als effizienzsteigernde Verfahren betrachtet. Ebenso bemerkenswert ist die Berücksichtigung des Wissens der beiden Sprecher. Mentale Tätigkeiten beim sprachlichen Handeln spielen sonst keine Rolle. Allerdings wird der funktionale Ansatz nicht weiterverfolgt. Insbesondere fehlt der Bezug der sprachlichen Mittel zu spezifischen Funktionen beim sprachlichen Handeln (vgl. etwa die funktionale Analyse von *denn* und *da* von Redder 1990).

Die wiederholte Erwähnung von Problemen bei der Verarbeitung gesprochener Sprache in den Netzplänen reduziert sich im konkreten Fall auf die Berücksichtigung spontansprachlicher Daten im Gegensatz zu kontrollierten, z.B. vorgelesenen, Daten. So wird bemerkt, daß die Wortfehlerrate beim Erkennen von vorgelesener zu spontaner Sprache drastisch ansteigt. Ein großes Problem stellt die Einschränkung des Suchraums möglicher Worte bei der Erkennung spontansprachlicher Daten dar. Gerade hier könnten auf der Grundlage von statistischen Regelmäßigkeiten Rechenzeit sparende Modelle zur Einschränkung des Suchraums herangezogen werden, indem dem Erkennungsmodul mögliche Folgewörter vorgeschlagen werden. Während die zuletzt genannten Probleme nicht der Mündlichkeit an sich, sondern der begrenzten Leistungsfähigkeit der verwendeten Hard- und Software zuzuschreiben sind, werden auch Fragen angesprochen, die auf unterschiedliche Verwendungsweisen im mündlichen/dialektalen und schriftlichen Gebrauch basieren, wie z.B. Reaktionsunterschieden.

Auch bei der Bestimmung der Besonderheiten gesprochener Sprache dominieren sprachtechnologische Zwänge. Neben kleinen Ansätzen zu einer funktionalen Bestimmung werden die Äußerungen als Menge zu verarbeitender Symbole betrachtet, die mit möglichst geringem Rechenaufwand mit geeigneten Formalismen zu bearbeiten sind. Die Verbindung diskursiver Funktionen und sprachlich grammatischer Funktionen der verwendeten Sprachmittel gerät nicht in das Blickfeld.

## **2. Dialogmodellierung in der Projektlaufzeit**

### **2.1. Szenarioforschung und Datenerhebung**

Mit Projektbeginn setzte die Konstruktion lauffähiger Systeme die Prioritäten und diktierten die 'harten' Entwicklungsarbeiten an der Spracherkennung und Sprachverarbeitung Arbeitstempo und -richtung. Die zu verarbeitenden Sprachdaten bilden die Grundlage der Systementwicklung und setzen die Maßstäbe für die Entwicklung. Dadurch erhält das Szenario eine wichtige Rolle im Verbundprojekt, da durch das Szenario das Korpus der Sprachdaten definiert wird. Die Szenarioforschung wird dementsprechend in den Netzplänen gewürdigt:

“Die Auswahl eines geeigneten Szenarios ist eine der wichtigsten Voraussetzungen für das Gelingen des Projekts. Zu anspruchsvolle Szenarien blockieren die Arbeit wegen zu großer Komplexität, während zu 'arme' Szenarien mangelnde Plausibilität für ein öffentlich gefördertes Großprojekt zeigen. ...

Als Szenario ist ein Verhandlungsdialo g festgelegt worden.” (Netzpläne, 143)

Im Zitat kommt die Bedeutung der Szenariowahl für die Systementwicklung zum Ausdruck: ein realitätsnahes, 'anspruchsvolles' Szenario liefert zwar realistische Daten, birgt aber die Gefahr in sich, die Entwicklung durch schwer zu lösende Probleme zu hemmen; ein entwicklungsgerecht einfaches Szenario - im Extremfall z.B. vorkonstruierte, von verschiedenen Sprechern vorgelesene Dialoge - liefern unrealistische Daten, so daß das System im Einsatz versagt. Die Lösung des Dilemmas liegt in einer systematischen Erforschung des Szenarios und den daraus resultierenden Anforderungen an die Sprachverarbeitung. An dieser Stelle ist der genuine Beitrag der Diskursanalyse zur Systementwicklung anzusetzen.

Eigentlich hätte angesichts der Bedeutung der Szenarioforschung vor der eigentlichen Programmentwicklung eine Phase der empirischen Szenarioforschung vorge-schaltet sein müssen, damit die eigentliche Entwicklung auf optimal abgestimmte Daten aufbauen kann (vgl. Bade u.a. 1994). Dieser Vorlauf wurde der Szenario-

forschung nicht gewährt. Stattdessen startete das Projekt mit weitgehend festgelegten Szenarioparametern. In den Netzplänen wurden auf den Erfahrungen vorausgegangener Projekte aufbauend konkrete Szenariofestlegungen (vgl. Netzpläne, 11) getroffen, die die Szenarioforschung eigentlich weitgehend erübrigten.

Schon im Zitat oben erfolgt nach der Betonung der Szenariowahl ohne vorausgegangene Szenarioforschung die Festlegung auf Verhandlungsdialoge. Ebenfalls ohne vorherige Szenarioforschung wird am Anfang der Netzpläne unter den funktionalen Spezifikationen des Demonstrators die "Terminplanung anhand eines Kalenders" (Netzpläne, 11) als Szenario vorgegeben und mit konkreten Spezifikationen präzisiert: so sollen zwar beliebige Termine, aber keine Uhrzeiten verwendet werden, es soll keine komplexe Argumentation vorkommen, es werden Maßnahmen zur Einschränkung von Ortsnamen getroffen, und es ist vorgesehen, daß die Äußerungen grammatikalisch korrekt sind, aus mindestens einem Wort, maximal einem Satz bestehen, auf einfache Sätze mit einem finiten Verb eingeschränkt sind und vollständige Sinneinheiten darstellen.

Für die eigentlichen Datenaufnahmen nach diesen Vorgaben war ein eigenes Teilprojekt zuständig. TP 14 sollte die erforderlichen umfangreichen spontanen sprachlichen Äußerungen (ca. 50.000 satzwertige Äußerungen von Dutzenden von Sprechern) aufnehmen, transkribieren und für das Training des Systems aufbereiten. Dazu wurden die Szenario-Vorgaben für Simulationen operationalisiert. Die Simulierung von TA sollte jedoch nicht nur bei der Datenaufnahme für die Systementwicklung, sondern auch für die Erforschung des Szenarios gelten. Im Arbeitspaket 13.6 "Kommunikationsfeld", nach den Szenario-Vorgaben für die "genaue Untersuchung und Beschreibung einer konkreten Verhandlungssituation und des generischen Typs der Verhandlungsdialoge" (Netzpläne 1992, 150) zuständig, sollte nach den Netzplänen

"...eine konkrete Situation mit den dazugehörigen Texten ausgewählt werden und deren Beschreibung auf allen Ebenen erfolgen. Daraus wiederum müssen als ein sehr früher Schritt Sprachdaten **simuliert**, aufgenommen und gelabelt werden." (Netzpläne, 143; Hervorhebung d. Verfassers)

KI-Vertreter vertraten die Ansicht, daß simulierte Daten ausreichen, da das Szenario die relevanten Parameter so festlege, daß die entscheidenden (spontan-)sprachlichen Merkmale in Lexik und Syntax auch bei der Simulation auftraten. Außerdem sollte eine Überforderung der Entwicklung durch zu chaotische natürliche Daten verhindert werden (vgl. Memo 50/94 zur Szenarioentwicklung; Bade u.a. 1994). Nur mit Mühe war es den Antragstellern zu AP 13.6 gelungen, im Antrag auch die Aufnahme einzelner Terminabsprachen im natürlichen Kontext unterzubringen. Sie verwiesen darauf, daß die Untersuchung von authentischen und simulierten Bewerbungsgesprächen um eine Lehrstelle

---

im Einzelhandel systematische Unterschiede zwischen dem zitierenden Handeln und dem authentischen Handeln ergeben hatte (s. Griebhaber 1987) und daß auch bei der Terminabsprache mit systematischen Differenzen zwischen authentischen und simulierten Gesprächen zu rechnen sei. Die ausschließliche Berücksichtigung simulierter Daten muß aus diskursanalytischer Sicht dazu führen, daß das System mit falschen Daten trainiert wird und im Ernstfall versagt.

Tatsächlich zeigen Aufnahmen authentischer TA gravierende Unterschiede zu simulierten. Schon im ersten Arbeitsbericht weisen Kameyama & Maleck 1993 darauf hin, daß die Rollenspiele das authentische Handeln nur begrenzt simulieren können. Insbesondere bemerken sie, daß beim simulierten Einstieg in die TA eine wichtige Phase der authentischen TA abgeschnitten wird. In Memo 23/94 wird auf der Basis von 91 authentischen TA die Terminabsprache als sprachliches Handlungsmuster modelliert. Die Eingangsphase des Musters wird detailliert beschrieben und in ihrer Funktionalität für die Herstellung der Bereitschaft zur Terminabsprache beim Hörer herausgearbeitet. Diese Phase stellt sich in den Szenario-Daten aufgrund der Vorgaben anders dar. Dadurch wird das System mit Daten trainiert, die in der Realität des sprachlichen Handelns so nicht auftreten. Vor diesem Hintergrund ist es deshalb nicht überraschend, daß der Demonstrator die Eingabe "Guten Morgen, mein Name ist Rüttgers" als "Montag" verstand.

Das KI-Interesse an homogenen Trainingsdaten führt in Verbindung mit dem oben vorgestellten 'flachen' Dialogmodell zu einer weiteren inhärenten Beschränkung. Das Trainingskorpus bestand aufgrund des vorab festgelegten Szenarios, der Aufnahmekonstellationen und Rollenanweisungen aus einer umfangreichen Äußerungsmenge, die der gewünschten Homogenität weitgehend entsprach. Die Homogenität der simulierten Daten hat für die erste Phase der Programmentwicklung einen weiteren Vorzug: die erhobenen Daten repräsentieren ein oberflächennah realisiertes Handlungsmuster. Die Handlungspartner vereinbaren im Rollenspiel ja nicht wirklich einen Termin für ein gemeinsames späteres Treffen, sondern sie führen mit Hilfe der vorgegebenen Kalender mit unvertrauten Termineinträgen eine Art Spiel durch, bei dem ein gemeinsamer freier Terminslot zu finden ist. Die tatsächlich durchgeführte Handlung entspricht dem bekannten Spiel 'Schiffe-versenken'. Wie beim 'Schiffe-versenken' wird abgecheckt, ob ein vom Partner genannter Termin in ein freies 'Quadrat' im eigenen Kalender trifft. Umgekehrt wird für den in der Aufgabe vorgegebenen Terminrahmen ein freier Slot im Kalender gesucht, der dem Partner vorgeschlagen wird. Dies wiederholt sich solange, bis ein gemeinsamer freier Terminslot gefunden ist.

Auf der Ebene der sprachlichen Äußerungen schlägt sich die Terminslotsuche in häufigen Äußerungen *bei mir geht Termin x (nicht) nieder*. Mit *bei mir/Ihnen*

bezieht sich der Sprecher auf einen Terminslot im Kalender. *Bei Ihnen* ist in allen Beispieläußerungen zur Klassifizierung von Sprechhandlungen enthalten, in denen der Hörer gefragt wird, ob ihm ein vorgeschlagener Termin paßt (Aufforderung\_Vorschlag in Maier 1994 Memo 31/94, 18, Aufforderung\_Stellung in Schmitz & Jekat-Rommel 1994, Report 28/24, 10; sowie Jekat u.a. 1995 TechDok 26/95, 5). In (B1), das zwei Ausschnitte aus einem simulierten Dialog zeigt, der nach den Szenariovorgaben in München aufgenommen wurde, ist dieser Gebrauch klar zu sehen. (B1) zeigt, daß *bei mir* vom Sprecher zur Bestätigung eines vorgeschlagenen Termins verwendet wird.

(B1) Simulierter Dialog: M090D

- (43) BAR018 wann würde das passen **bei Ihnen** ?  
 (44) PET019 <A> <ähm> morgen wär' prinzipiell nicht schlecht,  
 (45) PET019 wie sieht's da **bei Ihnen** aus?  
 (46) BAR020 ja , da hab' ich den ganzen Tag Zeit .  
 (47) PET021 gut<Z> ,  
 (48) PET021 dann machen wir morgen ,  
 (49) PET021 <äh> welche Uhrzeit wär' Ihnen denn recht ?  
 (50) BAR022 ach , da ist mir auch alles recht ,  
 (51) BAR022 haben Sie Termine vorher oder ist auch alles frei ?  
 (52) PET023 **bei mir** ist alles frei ,

(B2) Authentische Terminabsprache HH\_TA\_LB\_032

- (44) LÜD Nein, das/ ähm das kriegen wir schon hin, diese Woche irgendwann.  
 (45) LÜD Wann würd es Ihnen denn am • besten passen?  
 (46) KEP Ich bin also die Woche noch • relati:v frei.  
 (47) KEP Relativ frei.  
 (48) KEP Also morgen früh geht s ja **bei mir** nicht.  
 (49) KEP Morgen Nachmittag wär s noch möglich.

In der langen authentischen TA in (B2) enthält Segment (48) das einzige Vorkommen von *bei mir/Ihnen*. Der Gesprächsverlauf nach der Frage von LÜD in Segment (45) zeigt, daß KEP in einem mentalen Prozeß seine zeitliche Verfügbarkeit durchgeht und über den Terminrahmen der aktuellen Woche den folgenden Tag durchgeht, zunächst den Vormittag ausschließt, bevor er den Nachmittag als möglichen Termin nennt. Die mentalen Planungstätigkeiten des Deliberierens zeigen sich an der sprachlichen Oberfläche an Verzögerungsphänomenen in Segment (46), der Wiederholung des letzten Teils von Segment (46) in Segment (47) und dem vorgeschalteten *ja* in Segment (47). Bei authentischen TA antizipiert

der Sprecher sowohl sein eigenes Terminbudget wie das des Partners, um im Hinblick auf die gemeinsame künftige Handlung, die eine spezifische Vorbereitung während einer bestimmten Vorbereitungszeit erfordert, einen geeigneten Terminvorschlag zu finden. Dagegen beziehen sich die Sprecher in den simulierten Gesprächen auf sichtbar vor ihnen liegende Kalender mit unvertrauten Termineinträgen, die erst zu überprüfen sind. Typisch für diese Prozedur ist die Sequenz (51) bis (52), in der BAR den Partner fragt, ob bei ihm wie bei ihr *auch alles frei ist*. BAR verbalisiert die Checkfrage, ob sie mit dem vorgeschlagenen Termin einen bei beiden freien Terminslot getroffen hat. Unter den Simulationsbedingungen werden also hochgradig artifizielle Daten erhoben, die lediglich in einigen Aspekten authentischen TA entsprechen.

Die Artifizialität der nach den Szenariovorgaben erhobenen Daten prädesitiniert sie jedoch für die Analyse mit den KI-Werkzeugen. Die weitgehend homogenisierten Daten werden auf wiederkehrende Muster abgesucht, die von der anderen Richtung bei der Erkennung vom System verarbeitet werden können. Für diese Art der Rasteranalyse bietet gerade die Datenhomogenität eine gute Grundlage. Die 'flachen' Simulationsdaten erlauben bis zu einem gewissen Grad das partiell erfolgreiche Arbeiten mit einem flachen 'Dialogmodell', da die Daten im Trainings- und Testset strukturell gleich sind. Allerdings kann die in der Studie angesprochene Vermeidung einer Überanpassung durch ein zu geringes Datenvolumen kann mit den simulierten Daten kaum erreicht werden, da durch eine Ausweitung der Datenmenge lediglich mehr Daten der gleichen Struktur erzeugt werden. Lediglich (mehr) Daten eines authentischen Bedingungen angenäherten Datensets könnten einer Überanpassung entgegenwirken. Entsprechende Hinweise der Linguistik stießen bei der KI jedoch auf wenig Interesse, da mit Blick auf die gegebenen Bedingungen Unrealistisches verlangt werde, statt das Machbare zu fördern. Auf längere Sicht wird jedoch die Orientierung an authentischeren Daten unvermeidlich, da die mit großem Aufwand programmierten und trainierten Systeme außerhalb der unrealistischen Trainingsszenarios versagen müssen, da sie zu eng auf die artifiziellen Datensets ausgerichtet sind.

## 2.2. Dialogmodellierung der KI

Das KI-Verständnis von Dialogen als partiell geordnete von Folge von Äußerungen (s.o. §2.1.) und das Bemühen um eine möglichst 'flache' Analyse zeigte sich auch im oberflächenverhafteten Herangehen an die konkrete Analyse und Modellierung von TA. Insbesondere die Klassifizierung von Äußerungen nach Sprechhandlungen stellt die maschinelle Sprachverarbeitung vor große Probleme. Deshalb wurde dem Problem von Anfang an mit technischen Maßnahmen zu Leibe gerückt. Die Vorgaben für VERBMOBIL sehen vor, daß die Gesprächspartner

---

dem System durch einen Knopfdruck signalisieren, daß die bei gedrücktem Knopf gemachten Äußerungen zu erkennen und zu übersetzen sind. Außerdem sollten die Sprecher pro Turn nach den Vorgaben für den Demonstrator nur maximal einen Satz äußern (s.o. §3.1.). Tatsächlich jedoch war die letzte Vorgabe nicht realisierbar. In der Praxis der simulierten Aufnahmen entspricht eine Knopfdruck-Einheit oft einem Turn mit mehreren Äußerungseinheiten. Zur Lösung des Problems wurden zunächst - vor allem aus den TP zur akustischen Aufnahme - Äußerungen mit Turns gleichgesetzt. Da sich das Vorgehen in der Verarbeitung als nicht haltbar erwies, wurde für die Dialogmodellierung schließlich aus dem Begriff des Sprechaktes der Begriff des 'Dialogaktes' abgeleitet (s. Schmitz & Quantz 1996). Ein Dialogakt kombiniert illokutionäre und propositionale Information für bestimmte Szenarien und Domänen. Anders als die nicht weiter spezifizierten Sprechakte nach Austin und Searle sind für die Dialogakte normalerweise domänenspezifische Referententypen und Prädikationen anzunehmen. Danach ist der von Geschäftsleuten vollzogene Dialogakt "einen Termin vorschlagen" sehr viel spezifischer als der von irgendjemand realisierte Sprechakt "irgendetwas vorschlagen". Für die automatische Sprachverarbeitung sind sodann im nächsten Schritt die im Grunde unendlichen Dialogakte durch empirische Korpusanalyse auf die für die TA charakteristischen einzuschränken. Auf diesem Weg wurde ein Set von Dialogakten für die Dialogmodellierung ausgearbeitet mit leicht wechselnden Bezeichnungen und leicht variierendem Umfang (vgl. Maier 1994, Schmitz & Jekat-Rommel 1994, Mast 1995 und Jekat u.a. 1995).

Auch mit dem Konzept der Dialogakte bleibt das Problem der Untergliederung von Turns bestehen. Dazu führen Schmitz & Jekat-Rommel 1994 aus, daß zwar ein Turn eine oder mehrere Äußerungen enthalten kann, daß aber durch mindestens eine dieser Äußerungen eine zentrale Sprechhandlung vollzogen wird. Deshalb versuchen sie, die Äußerungsgröße 'pragmatisch' dadurch zu determinieren, daß sie den Teil eines Turns, dem optional eine Sprechhandlung zuzuordnen ist, als Äußerung ansehen. Die Konzentration mehrerer Äußerungselemente auf einen zentralen Dialogakt ist auch im Zusammenhang mit der MÜ zu sehen. Schmitz & Quantz 1996 zitieren ein Beispiel, in dem ein Humandolmetscher die in mehreren Äußerungen realisierte Zurückweisung eines Terminvorschlags in der zielsprachlichen Version auf eine Äußerung reduziert. Programmtechnisch wird dies - soweit aus dem Beispiel ersichtlich - dadurch erreicht, daß sukzessiv aufeinander folgende gleiche Dialogakte mit dem gleichen propositionalen TA-Objekt durch einen Dialogakt ersetzt werden.

Für die Dialogmodellierung werden die Dialogakte unter Bezug auf die Korpora in eine geordnete Reihenfolge gebracht. Es wird festgestellt, mit welchen Dialogakten die TA beginnt, welche D-Akte in der Eröffnungs- und Schlußphase

auftreten und in welcher Reihenfolge sie auftreten. Dasselbe wird für die eigentliche TA durchgeführt. Für jeden D-Akt wird festgestellt, welche(r) D-Akt(e) vorausgeh(t)en und welche(r) D-Akt(e) auf einen gegebenen D-Akt folg(t)en. Auf diesem Weg gelangt man zu einem System mit einigen Zuständen, zu denen jeweils bestimmte D-Akte hinführen und von denen jeweils bestimmte D-Akte wegführen. Das Dialogmodell entspricht im Kern den Ausführungen in der Studie, wonach ein Dialog aus einer partiell geordneten Menge von Sprechhandlungen besteht (s.o. §3.1.). Ein systematischer Unterschied zwischen Sprecher und Hörer und dem jeweiligen Wissensstand wird nicht gemacht. Abbildung 1 zeigt eine graphische Darstellung des Dialogmodells nach Mast 1995.

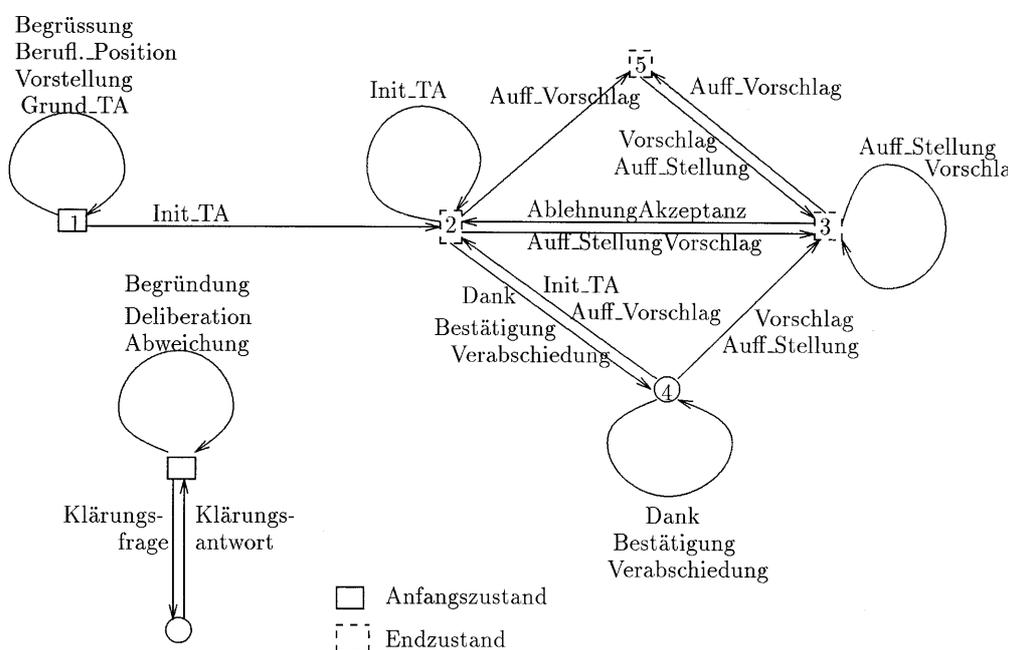


Abbildung 1: Dialogmodell nach Mast 1995, 13

Da das Modell aus simulierten Aufnahmen abgeleitet ist, fehlt im Vergleich zum Handlungsmuster der TA die Präphase, die der wechselseitigen Bereitschaft zur Vereinbarung eines künftigen Termins dient. Die Handlungen "Zwischeneintragsbestätigung", "Ergebnisfestellung" und "Bekräftigung" des Musters der authentischen TA fehlen im Dialogmodell ebenso. In funktionaler Hinsicht unterscheiden sich die beiden Modelle darin, daß das Dialogmodell die TA im Grunde nur als schwach geordnete Abfolge beliebiger D-Akte versteht, während die Pragmatik die TA als sukzessiv ablaufenden Entscheidungsprozeß der zwei Interaktanten versteht. Deshalb unterscheiden sich beispielsweise die Bestimmungen von Deliberationen in den beiden Modellen grundlegend. Während das Deliberieren im Handlungsmuster als Exothetisierung von mentalen Prozessen betrachtet wird, die einem Entscheidungsknoten zugeordnet sind, werden im

---

Dialogmodell Deliberationen gerade dadurch bestimmt, daß sie - auf der verbalen Ebene - gerade nicht dem interaktiven Vollzug der TA dienen. So schreiben Schmitz & Jekat-Rommel (1994, 18), daß sich durch die explizite Deliberation der Sprecher kurzfristig aus der TA löse, um einen möglichen Termin zu verifizieren. Im Dialogmodell können Deliberationen an jeder Stelle des Dialogs auftreten. Ihre Anordnung in einem separaten Handlungszyklus (in Abb. 1 links unten) bringt deutlich zum Ausdruck, daß sie außerhalb des Prozesses der Terminvereinbarung stehen. Dagegen sind die im Handlungsmuster nach der Vorschlagsanforderung und vor dem Vorschlag in Verbindung mit einem Entscheidungsknoten angesetzten Deliberationen integraler Bestandteil des Entscheidungsprozesses in der Terminvereinbarung.

Eine weitere Beschränkung des turnorientierten KI-Zugangs zur Handlungsanalyse liegt darin, daß ein Handlungsschritt, z.B. ein Vorschlag, nur innerhalb des Turns eines Sprechers gemacht werden kann. Rehbein u.a. (1994, 20ff.) zeigen jedoch bei der Analyse von Ankündigungen in authentischen TA, daß eine Musterposition ( Handlungsschritt) von Sprecher und Hörer in mehreren aufeinanderfolgenden Turns realisiert wird. Als allgemeine Beobachtung halten sie fest, daß ein turn häufig eine zweiteilige Binnengliederung aufweist. Im ersten Teil nimmt der den turn übernehmende Sprecher in der Rolle des vorherigen Hörers Bezug auf die Vorgängeräußerung, im zweiten Teil setzt er seinen eigenen Plan in der Sprecherrolle um. Eine weitere turn-Gliederungsdimension zeigt sich in Äußerungen mit Fragemodus. Der Sprecher spaltet die Äußerung in den ersten Teil mit dem propositionalen Gehalt, dem im zweiten Teil mit der Frageintonation die Illokution folgt. Möglicherweise bilden solche komplexen Realisierungen die Grundlage für das Vorgehen der KI, in den turns eine Äußerung zu identifizieren, die die zentrale Sprechhandlung vollzieht. Damit wird jedoch die integrale Handlungsqualität der Äußerungen aufgespalten, und lediglich ein Teil davon für das Ganze genommen. Ein derartiges Vorgehen verfestigt die Abhängigkeit der Sprachverarbeitung vom vorgegebenen Korpus, da dessen Oberflächensukzession einzig ausschlaggebendes Kriterium ist. Die Oberflächensukzession hängt jedoch von vielen Bedingungen ab und kann sehr verschiedenen Musterpositionen entsprechen. Eine tiefere Handlungsanalyse kann in diesem Sinn die Sprachverarbeitung wesentlich besser auf die Verarbeitung scheinbar chaotisch authentischer Spontansprache vorbereiten, da sie mit der Tiefenstruktur einen Bezugspunkt für funktionale Handlungsschritte besitzt, die mit jeweils bestimmten sprachlichen Mitteln prozessiert werden. Aus dieser Überlegung resultiert die Bestimmung von Realisierungsmengen (s. Beitrag xx in diesem Band).

Die Beliebigkeit der D-Aktanalyse zeigt die quantitative Verteilung von D-Akten, die Mast 1995 präsentiert. Sie listet in einer Tabelle die Sprechhandlungen für

die einzelnen Dialogzustände des Dialogmodells für 214 manuell annotierte Dialoge auf (s. Tabelle 1).

<b>Sprechhandlung</b>	<b>1</b>	<b>2,4</b>	<b>3</b>	<b>5</b>	<b>Häufigkeit</b>
Ablehnung			x		542
Abweichung	x	x	x	x	446
Akzeptanz			x		1068
Auff_Stellung		x	x	x	236
Auff_Vorschlag		x	x		268
Begründung	x	x	x	x	294
Begrüßung	x				197
Berufl_Position	x				0
Bestätigung		x			179
Dank		x			79
Deliberation	x	x	x	x	290
Grund_TA	x				203
Init_TA	x				174
Klärungsantwort					105
Klärungsfrage	x	x	x	x	141
Verabschiedung		x			251
Vorschlag		x	x	x	1948
Vorstellung	x				209

Tabelle 1: Sprechhandlungen nach Dialogzuständen nach Mast 1995, 14

Die Erkennung der D-Akte erfolgt über eine mikrostrukturelle und eine makrostrukturelle Ebene (vgl. Schmitz & Jekat-Rommel 1994). Auf der mikrostrukturellen Ebene dienen linguistische Indikatoren der Äußerungen selbst ohne weitere Informationen als Bestimmungsindikatoren. Genannt werden syntaktische Informationen über Numerus, Satztyp oder Verbmodus, semantische Informationen über bestimmte Konzepte und bestimmte Schlüsselwörter. Auf der makrostrukturellen Ebene dienen globale pragmatische Informationen als Indikatoren. Hier wird die Stelle eines turns im Zusammenhang mit der Grobgliederung der TA genannt. Die integrierte Berücksichtigung beider Ebenen wird zur D-Akt-Bestimmung benützt.

Nach dem Dialogmodell (s.o. Abb. 1) müßten eigentlich auf initiative D-Akte gleichviel reaktive D-Akte folgen. Da die nicht zur TA im engeren Sinn zählenden D-Akte wie Deliberationen oder Klärungen keinen eigenen Handlungsschritt im Modell begründen, sondern lediglich an beliebiger Stelle eingeschoben sind, sind sie dafür zu vernachlässigen. Tabelle 1 ergibt nun ein unausgewogenes Verhältnis von initiativen und reaktiven D-Akten. So liegt z.B. die Zahl der Eröffnungssequenzen (174) unter der Zahl der analysierten TA, obwohl nach dem Modell jede TA mit dem D-Akt "Init\_TA" beginnt. Die Zahl der Vorschläge (1948) liegt wesentlich höher als die Zahl der auf einen Vorschlag folgenden D-Akte. Selbst wenn die Zahl der Akzeptanzen, Bestätigungen und Ablehnungen zusammengesählt wird, fehlen fast 200 Akte. Zählt man Auff\_Stellung (236) hinzu, die auf einen Vorschlag im selben turn folgen, wird die Zahl der Vorschläge übertroffen. Die genauere Betrachtung der quantitativen Ergebnisse der 'flachen' Dialoganalyse bieten also genügend Anlaß zur Überprüfung des Vorgehens. Die Tendenz, die Äußerungen in kleine D-Akte zu zerlegen, führt dazu, daß die gebildeten Einheiten den tatsächlichen Einheiten im Handlungsprozeß nicht entsprechen und deshalb Artefakte darstellen.

Die empirische Überprüfung des Vorgehens mit Testdatensätzen aus dem Verbmobil-Simulationskorpus stellt die Dialogmodellierer noch nicht zufrieden. Z.B. konstatieren Alexandersson & Meier & Reithinger (1994, 10) eine relativ geringe Trefferquote für einen Testdatensatz von 81 TA mit 2995 D-Akten, die hinter der Quote eines Konkurrenzprojekts (Nagata & Morimoto 1993) zurückbleibt. Als Hauptgrund für die geringere Trefferquote führen die Autorinnen an, daß ihr Korpus dem konventionellen Dialogverhalten häufig nicht entspreche, da die Dialogstrukturen von Dialog zu Dialog deutlich variieren. Die genannten Variationen beziehen sich auf die simulierten Daten, die durch die Szenariovorgaben wesentlich homogener sind als die Daten der authentischen TA.

Ähnliche Ergebnisse berichtet Mast 1995 von ihrer Erprobung der D-Akterkennung. Im Vergleich von simulierten deutschen und englischen VERBMOBIL-Dialogen sind die Erkennungsrate für die englischen Dialoge um ca. 10% besser als die für die deutschen. Als einen wesentlichen Grund nennt auch sie die Variation der Dialogstrukturen, in ihrem Fall durch unterschiedliche Vorgaben. Als Variationen, die sich auf die verwendete Sprache auswirken nennt sie: die Absprache eines oder mehrerer Termine pro Dialog, rein geschäftliche oder auch private Termine. Die Absprache von einem oder mehreren Terminen pro Dialog sollte sich jedoch eigentlich nicht auf die Struktur der Absprache in der thematischen Vereinbarungsphase auswirken. Auch die Absprache privater oder geschäftlicher Termine - unter den allgemeinen Szenariovorgaben - sollte sich hinsichtlich der Dialogstruktur nicht auswirken. Die Verminderung der Erkennungsrate durch die relativ geringen Variationen zeigt die Probleme der engen

Bindung an die sprachliche Oberfläche im 'flachen' Dialogmodell. Die Erwartung, eine verbesserte Erkennungsrate durch eine größere Trainingsmenge zu erhalten, wird sich nicht erfüllen, da so lediglich mehr Daten derselben vereinheitlichten Struktur erzeugt werden. Dies zeigt gerade Masts Hoffnung, durch ein einheitlicheres Szenario besser erkennbare Daten zu bekommen. Beide Ergebnisse zeigen, daß der gewählte Ansatz aufgrund seiner Abhängigkeit von der Oberflächenstruktur der zu verarbeitenden Sprachdaten Probleme hat, Variationen der Oberflächenstruktur im Hinblick auf zugrundeliegende Musterpositionen zu verarbeiten. Ein derart von homogenisierten Daten abhängiges System kann für authentische Spontansprache nicht funktionieren. Lediglich ein der Komplexität des sprachlichen Handelns angemessener Analyse- und Modellierungsansatz kann auf längere Sicht zu wirklich besseren und stabilen Erkennungsraten für authentische Spontansprache führen.

#### **4. Neue Perspektiven mit der Modellierung mittels Petri-Netzen**

Neue Perspektiven der Diskursanalyse eröffnen KI-Werkzeuge wie die Handlungsmodellierung mit Petri-Netzen, die bislang von den Computerprojekten im Verbundprojekt nicht verwendet wurden. Mit Petri-Netzen lassen sich Pragmeme über die Darstellung in Diagrammen hinaus auch direkt mit Sprachdaten füllen, so daß einzelne TA mit dem Muster im Computer darstellbar werden (vgl. die Beiträge Benecke und Warter in diesem Band xx). Die klassische diskursanalytische Modellierung, wie sie etwa auch in Rehbein & Kameyama & Maleck 1994 verwendet wird (s.o. Beitrag xx), führt zu statischen Darstellungen möglicher Musterdurchläufe. Die Modellierung des Musters mit einem Petri-Netz führt zu einer generischen Darstellung, bei der die Positionen je nach eingegebenen Daten verschieden durchlaufen werden. Die Entwicklung mit empirisch erhobenen Daten führt zu einem stabilen Modell, das in der Lage ist, alle erhobenen Daten zu verarbeiten.

Petri-Netze bieten auch darstellungstechnisch einige Vorteile gegenüber anderen Modellierungstechniken: sie verwenden eine überschaubare Anzahl grafischer Symbole mit festgelegter Semantik. Die funktionale Modellierung bildet auch die Voraussetzung dafür, daß die Diskurse simuliert werden können. Durch die mit der Simulation gegebene Kontrolle der Modellierung wird auch Modellierung selbst genauer. Die Netze erlauben eine explizite Beschreibung von Zuständen und Aktionen, während viele andere Modellierungstechniken nur das eine oder das andere kennen oder gar keine Unterscheidung machen.

Petri-Netze eröffnen der Diskursanalyse eine Fülle neuer Modellierungsmethoden. Die generative Darstellung des Musters ermöglicht die Generierung beliebig

---

vieler Ablaufvarianten und hilft bei der Verifizierung und Verfeinerung (d.h. Konkretisierung) des linguistischen Modells. Sie schränken die Nebenläufigkeit (Parallelität) von Vorgängen nicht ein, so daß gleichzeitig ablaufende Vorgänge modelliert werden können. Dies ist beispielsweise für die Integration von interaktionalen und mentalen Tätigkeiten sehr vorteilhaft. Die Modelle können über Probeläufe so weit überprüft und verfeinert werden, daß das Modell eingegebene Diskursdaten mit ausreichender Zuverlässigkeit erkennt. Dabei läßt sich in der interaktiven Simulation auch entscheiden, was "ausreichende Zuverlässigkeit" bedeutet: 80%, 90%, ...? Die Stabilisierung der Implementierung kann unter Umständen auch entscheidende Veränderungen des Modells erfordern. Im Gegensatz zu konventionellen Methoden wird die Veränderung jedoch durch die Simulation mit Daten von den untersuchten Diskursen her motiviert und gesteuert.

Nach erfolgter Implementierung mit Petri-Netzen kann das Modell auch als Analysewerkzeug für automatisches Diskurs-Positions-Taggen verwendet werden. Bei ausreichend hoher Stabilität läßt sich eine große Zahl von Transkriptionen zum automatischen Taggen einlesen und den Segmenten Musterpositionen zuweisen. Damit könnten für die automatische Analyse - z.B. in Datenbanken - große Datenmengen aufbereitet werden. Die Behandlung wirklich großer Korpora wäre auf konventionellem Weg nur mit hohem personellem und zeitlichem Aufwand möglich und wäre ohne Werkzeuge bei humaner Rasterisierung mit unvermeidlichen Fehlern behaftet.

Besonders faszinierende Ausblicke eröffnen jedoch die mit Petri-Netzen möglichen Simulationsexperimente. Mit der werkzeuggestützten Simulation auf dem Rechner können einzelne Musterpositionen und Zustände interaktiv verfolgt und beeinflußt werden. So läßt sich z.B. durch gezielte Verletzung der Voraussetzungen zur Erreichung von Musterpositionen, bzw. durch erzwungene Transition untersuchen, was in einem bestimmten Zustand passiert, wenn eine bestimmte Transition erzwungen wird, ohne daß die Bedingungen vorliegen. Dadurch lassen sich die Bedingungen und Voraussetzungen von Handlungsschritten genauer bestimmen.

Eine weitere Einsatzart liegt in der Ermittlung von 'guten', bzw. 'problematischen' Diskursvarianten, indem Diskurse mit vielen mißlingenden Partialschleifen mit Diskursen mit wenigen mißlingenden Partialschleifen kontrastiert werden. Dadurch läßt sich die Frage behandeln, ob es Unterschiede zwischen den 'guten' und 'problematischen' Durchläufen gibt. Möglicherweise könnten die Unterschiede ja auch an der Nutzung unterschiedlicher Mittel bei der Realisierung liegen. Auch die Bestimmung der Realisierungsmengen (s.o. Beitrag Kameyama 1997) kann mit Petri-Netzen effizienter erledigt werden, indem aus 'guten' Minimaldurchläufen sich möglicherweise die für die Realisierung funktional

erforderlichen minimalen sprachlichen Mittel gewinnen lassen.

Auch die Unterschung der Funktion formaler sprachlicher Mittel, z.B. Flexionssparadigmata, kann anhand von Daten empirisch zuverlässig überprüft werden. Da für die funktionale Realisierung einer TA nicht die grammatisch voll differenzierten Zuweisungen erforderlich sind, kann eine größere Abstraktion der Zuweisung formaler Mittel einen Vergleich der Erkennungsrate mit und ohne Erkennung der voll differenzierten Mittel ermöglichen, wodurch die funktionale Leistung der grammatischen Mittel untersuchbar wird. Das interaktive Analyseverfahren verspricht auch genauere Aufschlüsse zu Funktion und Leistung von Modalität und Höflichkeit.

Eine weitere, bislang wenig berücksichtigte Dimension des sprachlichen Handelns betrifft die Abfolge des Gebrauchs von sprachlichen Mitteln. Der Gebrauch bestimmter Mittel läßt sich dadurch näher bestimmen, daß jeweils beim Auftreten des Mittels das System angehalten wird und die bis zu dem Zeitpunkt verwendeten Mittel und Diskurspositionen ermittelt und untersucht werden. Vermutlich lassen sich dadurch zur Gebrauchsfunktionen der Mittel nähere Erkenntnisse gewinnen.

#### 4. Literaturverzeichnis

Bade, U.; Eissenhauer, S.; Griebhaber, W.; Heizmann, S.; Jekat-Rommel, S.; Kameyama, S. & Krause, D.; Maleck, I.; Prahl, B.; Petzolt, S. & Preuß, W. (1994) Szenarioentwicklung im Rahmen von TP 13. Hamburg u. Hildesheim. Vm Memo 50

Balari Ravera, S. (1990) Information-based linguistics and Head-Driven Phrase Structure. In: Figueiras, M. u.a. (eds.) Natural Language Processing. Berlin: Springer, 55-101

Brinker, K. (1985) Linguistische Textanalyse. Berlin: Schmidt

Dittmann, J. (1979) Einleitung - Was ist, zu welchen Zwecken und wie treiben wir Konversationsanalyse? In: Dittmann, J. (Hg.) (1979) Arbeiten zur Konversationsanalyse. Tübingen: Niemeyer, 1-43

Hahn, W. von (1988) Analysis and Generation of Discourse. Hamburg: Fachbereich Informatik (mimeo)

Kallmeyer, W. & Schütze, F. (1976) Konversationsanalyse. In: Studium Linguistik 1/76, 1-28

Kay, M.; Gawron, J. M. & Norvig, P. (1994) Verbmobil. A Translation System for Face-to-Face Dialog. Stanford: Center for the Study of Language and Information

Mast, M. (1995) Schlüsselwörter zur Detektion von Diskontinuitäten und Sprechhandlungen. Vm Memo 57, Januar 1995. Universität Erlangen-Nürnberg

Kameyama, S. & Maleck, I. (1993) Zur Handlungskonstellation von Terminabsprachen. Arbeitsbericht 1 im Verbundprojekt Verbmobil. Hamburg: Germanisches Seminar

Kochendörfer, G. (1989) Notation, Modellbildung, Simulation: Grundlagenprobleme kognitivistischer Darstellungen in der Linguistik. Tübingen: Narr

Proudian, D. & Pollard, C. J. (1985) Parsing head-driven Phrase Structure Grammar. In: ACL Proceedings 23. Chicago, 167-171

Redder, A. (1990) Grammatiktheorie und sprachliches Handeln: "denn" und "da". Tübingen: Niemeyer

Verbmobil. Mobiles Dolmetschgerät - Studie - (1991) München

Verbundprojekt Verbmobil. Anträge der Universität Hamburg. Antragstexte. (1992) Hamburg:

FB Informatik

Wahlster, W. & Engelkamp, J. (Hrsg.) (1992) Wissenschaftliche Ziele und Netzpläne für das VERBMOBIL-Projekt. Saarbrücken: DFKI