



universität
wien

DIPLOMARBEIT

Titel der wissenschaftlichen Arbeit

„Interaktive Dialogsysteme.
Sprachwissenschaftliche Grundlagen
und
Untersuchungen zu Nutzungsmöglichkeiten“

Verfasser

Emanoil Kotsev

angestrebter akademischer Grad

Magister der Philosophie (Mag.phil.)

Wien, im August 2008

Studienkennzahl lt. Studienblatt	A 332 295
Studienrichtung lt. Studienblatt	Deutsche Philologie (Stzw) / Gewählte Fächer
Betreuer	Univ. Prof. Dr. Thomas Gloning

Vorwort

Die vorliegende Arbeit stellt das Ergebnis meiner mehrjährigen Studien in Richtung Germanistik, Computerlinguistik und Übersetzungswissenschaften dar. Als solche ist sie vom Prinzip her fächerübergreifend und widerspiegelt dadurch auch das Spektrum der Personen, die mich in meiner Studienzeit und bei dieser Arbeit fachlich und persönlich unterstützt haben.

Besonderer Dank gilt meinem Betreuer Herrn Univ. Prof. Dr. Thomas Gloning, der jederzeit beim Schreiben der gesamten Diplomarbeit mit Rat und Tat zur Seite stand und im Bereich Dialogforschung und Pragmatik hervorragende Hilfe leistete. Weiter gilt mein Dank Herrn Ao. Univ. Prof. Mag. Dr. Peter Ernst, der mich in seiner Funktion als Vorsitzender der Studienprogrammleitung fachlich betreute.

Selbstverständlich gilt mein Dank auch all jenen am Institut für Germanistik, am Institut für Medizinische Kybernetik und Artificial Intelligence und am Zentrum für Translationswissenschaft der Universität Wien, die auf verschiedenste Weise zum Erfolg meines Studiums beigetragen haben, wie auch meiner Familie für ihre moralische und persönliche Unterstützung.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Gegenstand, Methodik und Ziele	1
1.2	Inhaltlicher Überblick	5
2	Grundlagen der Kommunikaiton	7
2.1	Konvention, Bedeutung und sprachliches Handeln	8
2.2	Relevanz, Kontext und Rationalität	13
2.3	Prinzipien der Dialogorganisation	15
2.3.1	Sprachliche Handlungen und ihre Abfolge	16
2.3.2	Formen der thematischen Verknüpfung	19
2.3.3	Grammatisch-lexikalische Aspekte der Dialogorganisation	20
2.4	Dialog und Dialogforschung	24
2.4.1	Dialogspiele	25
2.4.2	Agreement, Allwissenheit und Wahrheit	27
2.4.3	Regeln	30
2.4.3.1	Dialogzug und Sprecherwechsel	31
2.4.3.2	Eingebetete Dialogsequenzen	34
2.4.3.3	Strategien und Regeln	35
2.4.3.4	Abgeleitete Regeln	35
2.4.3.5	Logische Spielregeln	39
2.4.3.6	Fehlerkorrektur	40
2.4.4	Festlegungssystem und Wissen	41
2.5	Kognitive Aspekte und KI	42
2.5.1	Pläne und sprachliche Handlungen	42
2.5.2	Nicht-monotone Logik und Sprechakttheorie	44
2.5.3	Default-Logik	46
2.5.3.1	Abduktive Argumentation	48
2.5.3.2	Transfer von Überzeugungen	48
2.5.4	Modale Logik	49

2.5.5	Die Logik rationaler Handlungen	50
3	Interaktive Dialogsysteme	53
3.1	Systemkomponenten	56
3.1.1	Automatische Spracherkennung (ASR)	59
3.1.1.1	Akustische und lexikalische Sprachverarbeitung	59
3.1.1.2	Abhängigkeit von der Stimme	62
3.1.2	Analyse gesprochener Sprache (NLU)	65
3.1.2.1	Äußerungen vs. Phrasen	65
3.1.2.2	Linguistische Regeln	65
3.1.2.3	Identifikation von Fehlern	66
3.1.3	Dialogmanager (DM)	67
3.1.4	Generierung gesprochener Sprache (NLG)	68
3.1.5	Sprachsynthese (TTS)	69
3.1.6	Fakultative Systemkomponente	70
3.2	Verarbeitung von Dialogphänomenen	71
3.2.1	Designstrategien	77
3.2.2	Initiative (wer stellt die Fragen)	85
3.2.2.1	Einwort-Dialoge	85
3.2.2.2	Systemgesteuerter Dialog	86
3.2.2.3	Gemischte Initiative	86
3.3	Evaluierung von interaktiven Dialogsystemen	88
3.3.1	Schwierigkeiten bei der Dialogsystemevaluierung	90
3.3.2	PARADISE - ein Verfahren zur Evaluierung interaktiver Dialogsysteme	94
4	Beispiele und Anwendungsmöglichkeiten	99
4.1	Plattformen	100
4.2	Assistenten	101
4.2.1	Steuerung von Soft- und Hardware	101
4.2.2	Auskunft	102
4.2.3	Organizer, E-Mail, Multimodale Systeme	103
4.2.4	Kontoverwaltung	104
4.2.5	Problemlösungen, Routenplaner	104
4.2.6	Reparaturassistenten	105
4.2.7	Unterstützung der Mensch-Mensch-Interaktion	107
4.3	Tutoring	107
4.4	Robotik	107

4.5 Übersetzungen und Dolmetschung	108
5 Schlußbetrachtung	109
Literaturverzeichnis	115
A Vorhandene Systeme	A-1
A.1 Assistenten	A-2
A.1.1 Auskunft	A-2
A.1.1.1 Das DARPA Communicator Projekt	A-9
A.1.2 E-Mail	A-12
A.1.3 Organizer	A-13
A.1.4 Reparatur	A-13
A.1.5 Problemlösungen, Routenplanung	A-14
A.1.6 Steuerung	A-15
A.1.7 Kontoverwaltung	A-17
A.1.8 Unterstützung in der Mensch-Mensch-Interaktion	A-18
A.2 Tutoring	A-19
A.3 Übersetzen und Dolmetschen	A-22
A.4 Robotik	A-22
A.5 Plattformen	A-23

Liste der Beispiele

1	direkter/indirekter Sprechakt	17
2	Referenz und Wiederaufnahme	21
3	Referenz und Wiederaufnahme (Fortsetzung 1)	21
4	Referenz und Wiederaufnahme (Fortsetzung 2)	28
5	Parameter des Dialogzugs	32
6	Optativ/Interrogativ	36
7	Entgegnung	38
8	Rekursivität von W-Fragen	40
9	Default-Logik	47
10	Notwendigkeit und Möglichkeit	50
11	explizite/implizite Bestätigungsstrategie	93
12	TRAINS-93	105
13	LARRI	106
14	VerbMobil	108

Tabellenverzeichnis

3.1 Statistik von Mensch-Mensch Konversationen aus dem Bereich der Filminformation (Glass, 1999)	76
3.3 Komplexität der Aufgaben	82
A.1 AdApt	A-2
A.2 Amities	A-2
A.3 BusLine	A-3
A.4 Bussimies / Busman	A-3
A.5 Jupiter	A-3
A.6 Bayesian Receptionist	A-4
A.7 NBA Game Update Line	A-4
A.8 NJFUN	A-4
A.9 Bildata	A-5
A.10 BirdQuest	A-5
A.11 OLGA	A-5
A.12 REA	A-6
A.13 OVIS	A-6
A.14 PEGASUS	A-6
A.15 ConQuest	A-7
A.16 Let's Go	A-7
A.17 Waxholm	A-7
A.18 Voyager	A-8
A.19 HIGGINS	A-8
A.20 How May I Help You?	A-8
A.21 DiSCoH	A-9
A.22 AT&T Communicator	A-9
A.23 CMU Communicator	A-10
A.24 MovieLine	A-10
A.25 Colorado University Communicator	A-10

A.26 Bell Labs DARPA Communicator	A-11
A.27 SRI Communicator	A-11
A.28 E-MATTER	A-12
A.29 Athosmail/DUMAS	A-12
A.30 Postimies / Mailman	A-12
A.31 Vera	A-13
A.32 LARRI	A-13
A.33 Intelligent Procedure Assistant (IPA)	A-14
A.34 TRAINS-91,93,95, 96	A-14
A.35 TRIPS	A-15
A.36 DICO	A-15
A.37 USI	A-15
A.38 WITAS	A-16
A.39 D'Homme	A-16
A.40 CTT-Bank	A-17
A.41 Clarissa	A-18
A.42 CHIL	A-18
A.43 LISTEN	A-19
A.44 NICE	A-19
A.45 MURI	A-20
A.46 ITSPOKE	A-20
A.47 The Berkeley UNIX Consultant	A-20
A.48 Ville	A-21
A.49 Verbmobil	A-22
A.50 Cogniron	A-22
A.51 Ariadne	A-23
A.52 Jaspis	A-23
A.53 DENK	A-24
A.54 Dipper	A-24
A.55 GALAXY Communicator	A-24
A.56 GoDiS	A-25
A.57 AUGUST	A-25
A.58 MUST	A-25
A.59 Siridus	A-26
A.60 RavenClaw/Olympus	A-26
A.61 SmartKom	A-27
A.62 Talk	A-27

A.63 Collagen	A-27
A.64 BOOTNAP	A-28
A.65 TRINDI	A-28
A.66 COCONUT	A-29
A.67 ATLAS	A-29

Kapitel 1

Einleitung

1.1 Gegenstand, Methodik und Ziele

Der Dialog ist das natürlichste Mittel für zwischenmenschliche Kommunikation und durch ihn wird die kommunikative Eigenschaft der Sprache am besten veranschaulicht. Der Spracherwerb und die komplexen sozialen Formen der menschlichen Existenz wären ohne eine grundlegende Dialogfähigkeit nicht denkbar. Dadurch können wir verschiedene kommunikative Situationen bewältigen, kommunikative Ziele erreichen oder uns einfach unterhalten. Mit der Entwicklung der Computertechnologie und -wissenschaft wird der Kontakt mit computergesteuerten Systemen für immer mehr Menschen möglich. Diese Systeme befinden sich bereits im Mobiltelefon, im Auto, im Bankomat oder im Fernseher und anderen Geräten des täglichen Gebrauchs. Daher kommt es zu der Anforderung, die Kommunikation mit diesen Systemen zu vereinfachen und menschenähnlicher und -freundlicher zu gestalten. Eben aus diesem Grund steht einerseits der Dialog als Kommunikationsmittel im Zentrum der vorliegenden Diplomarbeit, andererseits aber die Frage, wie ein gesprochener Dialog mit einem Computersystem möglich wäre.

Die zwei Einheiten Mensch und Maschine sind unterschiedlicher Natur schon allein deshalb, weil Menschen von Emotionen gesteuert handeln (Damasio, 1995, 2005) und der Computer nur digitale Informationen berechnen kann. Es gibt anscheinend keine Basis für den Vergleich dieser zwei Einheiten. Stellen wir also die Frage so, wie sie Turing formuliert hat: „Can a machine think?“ (Turing, 1950), dann wäre die Antwort „nein“ die richtige. Mit „denken“ bezeichnen wir die kognitive Tätigkeit des Menschen und es gibt noch keine solche Tätigkeit bei den Maschinen, die man mit dem selben Begriff bezeichnen kann. Trotzdem können durch eine geeignete Simulation dieser Denkfähigkeit des Menschen ziemlich brauchbare Ergebnisse erzielt werden. Im Hinblick auf das Thema „Dialogsysteme“ wird also die behandelte Problematik auf den Dialog als

eine kommunikative Fähigkeit des Menschen reduziert. Die Simulation dieser Fähigkeit lässt darauf schließen, dass es zwei Untersuchungsobjekte gibt, das simulierende (das interaktive Dialogsystem) und das simulierte (die Dialogfähigkeit). Dies führt zu einigen Komplikationen, denn die Antworten sind in unterschiedlichen wissenschaftlichen Bereichen zu suchen. Die kommunikative Fähigkeit des Menschen wird von der Kommunikationswissenschaft, der Kognitionswissenschaft, Psychologie, Medizin (vor allem Neurologie und Hirnforschung), den Sprachwissenschaften, der Pragmatik und der Philosophie untersucht. Die Computersysteme werden von der Informatik untersucht. Die Beiträge von Turing (1950) führten zur revolutionären Entwicklung der letzteren in der zweiten Hälfte des 20. Jh. Diese Entwicklung hat kaum einen Bereich des menschlichen Daseins unbeeinflusst gelassen. Neue Forschungsdisziplinen wurden in diesem Zusammenhang eingeführt. NLP - *natural language processing* (Natürlichsprachliche Verarbeitung) ist eine solche übergreifende Forschungsrichtung, die einerseits Teilgebiet der künstlichen Intelligenz (KI) im Rahmen der Informatik ist, andererseits die theoretischen Grundlagen der Sprachwissenschaft und sprachenbezogene Wissenschaften wie der Germanistik anwendet. Solche übergreifende wissenschaftliche Richtungen bieten auch den Sprachwissenschaftlerinnen, darunter auch Germanistinnen, neue Herausforderungen, nämlich Erkenntnisse, die im Rahmen einer wissenschaftlichen Disziplin gewonnen wurden, in einer anderen zu verwenden, in diesem Fall in der Informatik. Der Transfer der wissenschaftlichen Erkenntnisse kann nicht ohne entsprechende Anpassung an die Zielwissenschaft erfolgen, was auf die unterschiedliche Natur von Mensch und Maschine zurückzuführen ist. Eben dieser Aspekt steht im Vordergrund der vorliegenden Untersuchung. Das Ziel ist, einerseits kommunikations- und dialogtheoretische Grundlagen der zwischenmenschlichen Kommunikation zu explizieren und zu zeigen, was der Dialog ist und wie er beim Menschen funktioniert, und andererseits zu untersuchen, welche zusätzlichen Mechanismen notwendig sind, um diese für die Menschen typische Tätigkeit von Computersystemen erfolgreich nachahmen zu lassen.

Die Fähigkeit, Intentionen zu verstehen, d.h. auf das Gemeinte zu schließen (Grice, 1993; Searle, 1969), ist mit Sicherheit eine Hauptfähigkeit, Dialog zu führen, jedoch nicht die einzige. Dazu gehört auch die Fähigkeit, „Dialogzusammenhänge zu produzieren und zu verstehen“ (Fritz, 1994b, S.177), gewisse Dialogmuster zu erkennen und bestimmte Dialogprinzipien zu befolgen, wie z.B. die sequentielle Abfolge der Dialogbeiträge (aufeinander folgende Äußerungen), das Thema, worüber eben gesprochen wird, u.a. Dabei spielt die Fähigkeit des Folgerns und des rationalen Handelns eine wesentliche Rolle. Diese Aspekte der Dialogführung stellen eine Verknüpfung zu anderen Wissenschaften wie Philosophie und Kognitionswissenschaft dar, bzw. stellt sich

die Frage nach der Wahrheit und der Art und Weise wie wir mit Information und Wissen umgehen. Ein zentraler Punkt in diesem Zusammenhang ist die Bedeutung, in den beiden Aspekten: „meinen“ und „verstehen“, die in den Untersuchungen mehrerer Wissenschaftlerinnen, unter anderen Wittgenstein, Grice und Searle, eine zentrale Rolle spielen. Die Möglichkeiten auf eine bestimmte Bedeutung zu schließen, werden durch Konventionen und Regeln festgelegt. Der Dialog hat also einerseits formale Aspekte wie aufeinander folgende Beiträge der Dialogteilnehmerinnen und andererseits kognitive Aspekte wie die Fähigkeit, einzelne Äußerungen zu verstehen und in dem gesamten Dialog zuzuordnen.

Nachdem es sich herausgestellt hat, dass die Bedeutung von Äußerungen berücksichtigt werden muss, griff man auf unterschiedliche logische Modelle (Reiter (1980); Hintikka (1962)) zurück, die einen unterschiedlichen Zugang zu den Inferenzprozessen anbieten, die jede Kommunikation begleiten, um dadurch die Motivation für die Entscheidungen, die zu einer bestimmten Äußerung führen, zu erklären. Gerade die Versuche kommunikationsfähige Maschinen zu bauen, haben gezeigt, dass Menschen komplexe kognitive Prozesse in einfachsten Dialogen verwenden und dass die Formalisierung dieser Prozesse eine Herausforderung für Wissenschaft und Technik darstellt. Daher verfügen moderne Dialogsysteme über komplexe Module und Algorithmen, je nach Design, um Audio- und/oder Videoinformationen zu verarbeiten, Hypothesen und Dialogstrategien zu bilden und zu verfolgen, um eben diesen Aspekt der zwischenmenschlichen Kommunikation, die Bedeutung von Äußerungen, zu bewältigen.

Die vorliegende Untersuchung soll unter anderem zeigen, welche theoretische Ansätze es im Bezug auf zwischenmenschliche Kommunikation gibt und wie es möglich ist, diese Fähigkeit zu simulieren. Aus der Perspektive der Mensch-Mensch-Kommunikation zeichnen sich drei theoretische Richtungen ab, die besonders breite Anwendung in Dialogsystemen finden - die Sprechakt-, Dialogspieltheorie und die Theorie rationaler Handlungen. Viele Sprachwissenschaftlerinnen (Wittgenstein, Austin, Searle) betrachten die Kommunikation als eine Reihe von zusammenhängenden sprachlichen Handlungen und das Sprechen als eine Art Handlung. Dieser Ansatz hat gewisse Vorteile, denn so könnte man Eigenschaften, die zu nichtsprachlichen Handlungen gehören, auf Äußerungen übertragen und darüber gewisse Aussagen machen. Die Dialogspieltheorie sieht in der Kommunikation eine Reihe von zusammenhängenden Spielen mit dem Vorteil, dass die Erkenntnisse der Spieltheorie auf die Kommunikation übertragen werden können. Die Idee, die Sprache und in diesem Fall den Dialog als ein Spiel zu betrachten, stammt von Wittgenstein und wird von anderen Wissenschaftlerinnen (z.B. Carlson) neulich wieder aufgegriffen und weiter entwickelt. Der Vorteil ist, dass durch die Spieltheorie, die vor allem seit den 50er Jahren des 20. Jh. entwickelt wird, ein

mathematisches Modell der sprachlichen Interaktion von Menschen entwickelt werden kann, was wiederum in der KI und in der Computerlinguistik zur Abbildung von Dialogphänomenen verwendet werden kann. Daher dienen hauptsächlich diese beiden Theorien als theoretische Grundlage für die vorliegende Analyse.

Unterschiedliche Kommunikationsziele und Situationen führen zu unterschiedlichen Arten der Kommunikation bzw. zu unterschiedlichen Dialogformen und setzen verschiedene sprachliche Handlungen voraus. Manchmal geht es um eine Informationsbeschaffung (z.B. Auskunft), manchmal darum, die anderen von etwas zu überzeugen, selbst überzeugt zu werden, oder (z.B. bei Begrüßung) einfach höflich zu sein, sich also sozial korrekt zu verhalten. Die Sprache in den unterschiedlichen Dialogformen weist demzufolge bestimmte Merkmale auf, die typisch für die jeweilige Kommunikationsform sind. Die Dialogteilnehmerinnen verwenden dasselbe „Werkzeug“ (die Sprache), um unterschiedliche kommunikative Situationen zu bewältigen und unterschiedliche kommunikative Ziele zu erreichen. Die Zahl der betrachteten Dialogformen in dieser Arbeit (z.B. Auskunft, Argumentation, Begrüßung) wird dadurch eingeschränkt, dass nur jene Dialogformen in Betracht gezogen werden, die auch für eine Implementierung in Computersystemen derzeit in Frage kommen.

In den letzten Jahrzehnten wurde die Forschung auf dem Gebiet der Sprachdialogsysteme verstärkt. Dies war vor allem durch die finanzielle Unterstützung der Regierungen in der EU und in den USA möglich. Bereits in den späten 80ern wurde ein Projekt zur Entwicklung von Systemen für gesprochene Sprachen von der DARPA¹ (Walker und Hirschman, 2000) in den USA gestartet, wobei man sich vor allem auf die Erkennung der gesprochenen Sprache konzentrierte, die sich als die „harte Nuss“ der Sprachverarbeitung erwies. Die darauf folgende DARPA-Initiative, die 1998 gestartet wurde, befasste sich mit Dialogsystemen, also nicht nur mit der Spracherkennung, sondern auch mit der Sprachproduktion. In Europa gab es einige große Forschungsprogramme, die die Entwicklung von interaktiven Dialogsystemen für gesprochene Sprache vorantrieben. Das von der EU finanzierte **E**spirit **S**UNDIAL (**S**peech **U**nderstanding and **D**IALog) Projekt sponserte Forschung in vier Sprachen (Englisch, Französisch, Deutsch und Italienisch). Die Teilnehmerinnen entwickelten Informationssysteme für Bahnfahrpläne und Flugpläne (siehe. Eckert et al., 1993). Die Kooperation unter den Forscherinnen wurde forciert, denn verschiedene Gruppen mussten Komponenten für ein einziges Multistandortsystem beisteuern. Die vorliegende Arbeit konzentriert sich aus diesem Grund auf Systeme, die vor allem nach 1995 entwickelt wurden. Eine Reihe davon wird im Anhang aufgelistet. Das Ziel ist zu zeigen, welche Bereiche für die Formalisierung von Dialogwissen besonders attraktiv sind und welche Verfahren sich für welche kom-

¹einer Abteilung des US Verteidigungsministeriums

munikative Ziele am besten eignen. Ein weiterer Trend ist die Multimodalität, die die Kommunikation mit (z.T. auch die Steuerung von) unterschiedlichen Peripheriegeräten und Anwendungen, wie auch in unterschiedlicher Umgebung möglich macht. In dieser Arbeit sollen diese Untersuchungen analytisch zusammengefasst und dargestellt werden.

Das Ziel der vorliegenden Arbeit kann mit den folgenden drei Hauptfragen zusammengefasst werden: (1) was ist Dialog, (2) welche charakteristische Merkmale hat der Dialog und (3) was ist notwendig, um Dialog erfolgreich mit Hilfe von Maschinen (Computersystemen) zu simulieren. Die Arbeit wurde in der Hoffnung geschrieben, dass sie als Basis und Ausgangspunkt für weitere Forschungen in den bezeichneten wissenschaftlichen Gebieten benutzt werden kann, wie auch in der Hoffnung, dadurch die Aufmerksamkeit von Sprachwissenschaftlerinnen und Germanistinnen auf Forschungsperspektiven und Herausforderungen in diesem Gebiet zu lenken. Dabei wird auch gezeigt, wie weit die theoretischen Grundlagen, die sich auf zwischenmenschliche Dialogformen beziehen, in interaktiven Dialogsystemen implementiert werden und implementiert werden können; welche zusätzliche Mittel notwendig sind, um solche Systeme menschenähnlicher in der Dialogkommunikation zu machen, d.h. wie kann die Qualität der Mensch-Maschine-Kommunikation verbessert werden und wie können solche Systeme evaluiert werden. Im Anhang wird eine Sammlung von Dialogsystemen präsentiert, welche auch teilweise für das breite Publikum vor allem in den USA und manchen größeren europäischen Ländern zugänglich sind.

1.2 Inhaltlicher Überblick

In dem Kapitel 2 soll versucht werden, die Hauptfragen (1) „was ist der Dialog“ und (2) „welche charakteristische Merkmale hat der Dialog“ zu beantworten. Der Dialog ist eine Kommunikationsform, daher ist es auch wichtig, zuerst mehr über Kommunikation in Dialogen zu erfahren. Ich präsentiere deswegen zuerst die kommunikationstheoretischen Grundlagen im Punkt 2.1, die als Basis für die Dialogforschung dienen, um dann die oben erwähnten Hauptfragen anzugehen. Ein zentrales Thema wird dabei die Frage zur Bedeutung von Äußerungen sein, denn die Bedeutung kann nicht getrennt von den anderen Untersuchungen zur Kommunikation behandelt werden, sondern bleibt entweder im Fokus oder im Hintergrund immer präsent, wenn es um zwischenmenschliche Kommunikation geht. Im Punkt 2.2 steht die dritte Grice'sche Maxime (Relevanz) aus der Perspektive der Relevanztheorie (Sperber und Wilson, 1986) im Mittelpunkt meiner Untersuchung, ebenso die gegenseitige Abhängigkeit zwischen Relevanz, Kontext und Rationalität. In den zwei folgenden Punkten 2.3 und 2.4 werden die dialogbezoge-

nen theoretischen Ansätze (z.B. Sprechakttheorie, Dialogspieltheorie) behandelt, die die Antwort auf die ersten zwei Hauptfragen liefern sollen.

Wie es schon erwähnt wurde, unterscheiden sich Mensch und Maschine grundsätzlich voneinander, daher ist es notwendig, dies an zusätzlichen (vor allem logischen) Modellen zu erforschen, um eine erfolgreiche Simulation von Dialogfähigkeit zu gewährleisten. Unterschiedliche Ansätze in diesem Zusammenhang präsentiere ich im Punkt 2.5 „Kognitive Aspekte und KI“ auf Seite 42.

Das Kapitel 3 behandelt die Dialogsysteme in den Aspekten Struktur, Funktion und Evaluierung. Zuerst werden die minimalen Anforderungen präsentiert, die an jedes Dialogsystem gestellt werden (Punkt 3.1), wie auch die Funktion der einzelnen Komponenten. Im Punkt 3.2 werden einige Probleme und Verfahren zur Abarbeitung von dialogischer Sprache, wie auch zur Dialogteilnahme seitens der Maschinen behandelt.

Ein besonders wichtiger Punkt bei der Entwicklung von Dialogsystemen ist die Frage nach der Anwendbarkeit. Um diese gewährleisten zu können, stellt sich wiederum die Frage nach der Evaluierung von Dialogsystemen. Diese Problematik wird näher im Punkt 3.3 behandelt. Es wird auch ein Verfahren „PARADISE“ (Walker et al., 1997b) beschrieben, welches sich als „quasi“ Standard, oder Richtlinie herausgebildet hat.

Der verbreitete Einsatz von Computern und Computersystemen führt dazu, dass es immer mehr Möglichkeiten gibt, eine sprachbasierende Benutzeroberfläche für die Kommunikation mit diesen Geräten zu implementieren, dies anstatt der klassischen Konsole (Tastatur/Bildschirm). Anhand der Besonderheiten einerseits der zwischenmenschlichen Kommunikation und andererseits der Systemarchitektur von Dialogsystemen zeichnen sich bestimmte Bereiche ab, wo solche Systeme eine Anwendung finden können. Im Kapitel 4 werden die Hauptbereiche präsentiert, wo solche Systeme am häufigsten implementiert werden, wie z.B. Assistenten in der Mensch-Mensch- und Mensch-Maschine-Interaktion, bei der Ausführung unterschiedlicher Aufgaben, als Oberfläche für die Kommunikation mit Robotern in der Robotik, als Tutoren oder simultane Übersetzer, für die Auskunftsdienste allgemein, für Reservierungen, bei der Reisenavigation und bei der Planung von Reisen, wie auch bei der Steuerung von Software und Peripheriegeräten u.a. Die Beschreibungen betreffen, die in dem Anhang präsentierte Sammlung von interaktiven Dialogsystemen.

Kapitel 2

Grundlagen der Kommunikaiton

Die komplexen sozialen Bedingungen, unter denen Menschen leben, wären undenkbar ohne eine ausgeprägte Fähigkeit miteinander zu kommunizieren. Die Kommunikation läuft durch den akustischen und/oder visuellen Kanal, wobei meistens der Dialog als die kommunikative Form und die Sprache als das Kommunikationsmittel benutzt werden. Sprache und Dialogfähigkeit scheinen daher untrennbar voneinander zu sein. Die komplexen Bedingungen, unter denen ein Dialog zustande kommt, setzen eine Vielzahl von Teilfähigkeiten voraus, die je nach Situationsbedingungen in zusammenhängender Weise angewendet werden.

Die Dialogfähigkeit zeichnet sich also dadurch aus, dass die Äußerungsformen für bestimmte sprachliche Handlungen eingesetzt werden (z.B. Einladung, Bitte, Begrüßung etc.) und durch die Fähigkeit weitere Organisationsprinzipien von Dialogen anzuwenden, wie z.B. Muster zu erkennen und anzuwenden, oder linear und sequentiell die Beiträge der Dialogpartnerinnen zuzuordnen und zu analysieren, wie auch durch die Fähigkeit die Zusammenhänge zwischen den einzelnen Beiträgen zu verfolgen (vgl. Fritz, 1994b). Das bedeutet, dass die Dialogteilnehmerinnen ihre Beiträge zeitlich nacheinander machen und im Stande sind, auf die Partnerzüge in relevanter Weise zu reagieren. Das bedeutet wiederum, dass die Dialogfähigkeit durch das Gedächtnis im Wesentlichen unterstützt wird, wodurch die Dialogteilnehmerinnen im Stande sind, die abgehandelten Themen inhaltlich und chronologisch in ihrem Gedächtnis eine Zeit lang aufzubewahren, wieder zu thematisieren und in weiteren Beiträgen darauf zu referieren. Dadurch entsteht Textkohärenz, die sich durch eine logische Struktur der thematischen Entfaltung charakterisiert (vgl. Brinker, 2000).

Als Prozess besteht der Dialog in der Produktion und der Rezeption von Äußerungsformen. Beide Aspekte sind sehr komplex und nicht ganz vollständig entschlüsselt. Vor allem sind die formalen Aspekte meistens gut erforscht wie die morphosyntaktische Ebene der Sprache oder z.B. wie durch Kontraktionen der Muskulatur Laute produziert

werden und wie die Schallwellen im Ohr zu Nervenimpulse umgewandelt werden (die akustische und die phonetische Ebene der Sprache). Der Zusammenhang zwischen den Sätzen und vor allem die Funktion des Gedächtnisses und die kognitiven Aspekte (Entscheidungen, Schlussfolgerungen, usw.) stellen jedoch noch immer Herausforderungen für die Forscher dar.

Unterschiedliche Kommunikationsziele und Situationen führen zu unterschiedlichen Arten der Kommunikation bzw. zu unterschiedlichen Dialogformen. Die Sprache in den unterschiedlichen Dialogformen weist demzufolge bestimmte Merkmale auf, die typisch für die jeweilige Kommunikationsform sind. Dialogteilnehmerinnen verwenden dasselbe „Werkzeug“ (die Sprache), um unterschiedliche kommunikative Situationen zu bewältigen und unterschiedliche kommunikative Ziele zu erreichen. Die Zahl der in dieser Arbeit betrachteten Dialogformen z.B. Informationsbeschaffung (Auskunft), Argumentation, Begrüßung wird dadurch eingeschränkt, dass nur diese Kommunikationsformen und Dialogmodelle in Betracht gezogen werden, die auch für eine Implementierung in Computersystemen derzeit in Frage kommen. Dies führt auch zur Einschränkung der Zahl der theoretischen Ansätze.

In diesem Kapitel soll versucht werden, die Hauptfragen (1) „was ist Dialog“ und (2) „welche charakteristische Merkmale hat der Dialog“ zu beantworten. Der Dialog ist eine Kommunikationsform, daher ist es auch wichtig, zuerst mehr über Kommunikation in Dialogen zu erfahren. Ich präsentiere deswegen zuerst die kommunikationstheoretischen Grundlagen, die als Basis für die Dialogforschung dienen, um dann die Hauptfragen anzugehen. Die dritte Hauptfrage „welche Dialogformen ist es sinnvoll zu formalisieren und in Mensch-Maschine-Kommunikation zu verwenden“ stellt eine Anbindung an die Informatik dar und setzt voraus, neben den kommunikationstheoretischen Grundlagen auch einige Ansätze aus dem Bereich der Logik und KI (Punkt 2.5) in betracht zu ziehen, die als Basis für die Formalisierung und Simulation gesprochener Dialoge dienen und diese auch zum Teil erfolgreich und möglich machen, um dann in dem nächsten Kapitel den Aufbau und Funktion solcher Systeme näher zu untersuchen.

2.1 Konvention, Bedeutung und sprachliches Handeln

Sprachanalytische Philosophen wie z.B. Wittgenstein, Searl und Grice gehen von der Idee aus, dass die Verständigung in weiten Teilen auf sozialen Regeln (Konventionen) beruht (vgl. Searle, 1969, P.2.5, S.33-42). Diese Regeln stellen eine Brücke dar zwischen dem, was geäußert wird und dem, was in einer Sprachgemeinschaft damit gemeint bzw. verstanden werden kann. Vor allem sollte Wittgenstein erwähnt werden:

Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache. Und die *Bedeutung* eines Namens erklärt man manchmal dadurch, dass man auf seinen *Träger* zeigt. (Wittgenstein, 2003, §43)

Bremes fasst treffend diese Überlegungen Wittgensteins zusammen.

Fragt man nach der Bedeutung eines Ausdrucks, so fragt man nach den Regeln, die diesen Ausdruck in seinem Gebrauch bestimmen [...] Der Gebrauch eines Wortes [hängt] von *Lebensformen* ab, bestimmt wird durch *Regeln*, in Einklang steht mit *Familienähnlichkeiten* und geleitet wird von *Zwecken*. (Bremes, 1997, S.24)

Wie das funktioniert, kann man mit Hilfe der Äußerung „Hallo“ im Deutschen veranschaulichen. Die Äußerung „Hallo“ wird als Begrüßung verstanden, auf Grund einer sozialen Gepflogenheit im Deutschen, die folgendermaßen formuliert werden kann: „Mit der Äußerung von *Hallo* kann man im Deutschen jemanden begrüßen“. Eine solche Regelformulierung umfasst: (i) ein sprachliches Mittel (Hallo), (ii) eine Äußerungshandlung, (iii) eine Angabe, als was man eine Äußerung meinen kann (als Begrüßung), und (iv) die Angabe, dass es sich um eine Gepflogenheit handelt (kann man im Deutschen). Unterschiedliche Äußerungen können dieselbe Bedeutung haben z.B. „Guten Tag“ und „Grüß Gott“ und umgekehrt ein und dieselbe Äußerung kann situations- oder kulturbedingt (durch Konventionen) unterschiedliche Bedeutungen tragen (vgl. Searle, 1969, S.46,49).

Zu den Grundlagen der modernen Studien zur Pragmatik gehört unter anderem die Arbeit von Grice. Er wurde durch seinen Beitrag zur Erforschung der Bedeutung und vor allem der Bedeutung von Äußerungen bekannt, mit der Bemerkung, dass das Englische *meaning* sehr schwer ins Deutsche übertragen werden kann. Unter Äußerung versteht Grice allerdings auch nonverbale Äußerungen wie Gesten u.ä. neben den verbalen Äußerungen bzw. der verbalen Kommunikation. In der späteren Version seiner Theorie wird noch deutlicher und motivierter die Differenzierung zwischen Sprecher-Bedeutung (das Gemeinte) und linguistischer Bedeutung (das Gesagte).

Auch Grice sieht im Mittelpunkt der zwischenmenschlichen Kommunikation (verbale und nicht verbale) die Fähigkeit, Intentionen zu erkennen und auszudrücken und baut seine Theorie auf dieser Basis auf (Grice, 1993, S.14; S.16ff). Im Mittelpunkt seiner Untersuchungen steht die Proposition und ihre Natur:

The third constraint is that a theory of propositions should provide the basis, at least, for an adequate account of the relation between thought, action and language on the one hand, and reality on the other. For example, one

perceives the desk, walks over to sit at it, and utters sentences to mean things about it.

Since propositions are the items we specify in specifying the content of thought, perception, intention, act of meaning, and so on, an account of propositions should at least provide the basis for an account of the relation between mind and reality. (Warner und Grandy, 1986, S.29)

Die Proposition, die die Äußerung prägt, wird, so Grice, von der Intention der Sprecherin beeinflusst, so dass gerade eine Interpretation der Äußerung am wahrscheinlichsten ist. Er unterscheidet einige Arten von Bedeutung (Sprecher-, Satz-, Wort-Bedeutung) (siehe Grice, 1993, S.85ff). Im Rahmen der Pragmatik ist vor allem die Sprecher-Bedeutung - das Gemeinte vom Interesse. Nach Grice bezieht sich die Sprecher-Bedeutung auf diese Idee: „A meint irgendetwas wenn sie X sagt“, was ungefähr so viel bedeutet wie: „A äußert X mit der Absicht, eine Überzeugung zu veranlassen, dadurch dass eben diese Absicht verstanden wird“ (siehe Grice, 1993, S.2-51). Dabei verstand Grice unter dem Begriff Bedeutung zwei verschiedene Phänomene: die natürliche und die nicht-natürliche Bedeutung (Grice, 1993, S.2f, siehe auch Fritz, 1994b, S.195).

- ▷ Natürliche Bedeutung bezieht sich dabei auf das Prinzip der Kausalität (Ursache und Wirkung), wie es in folgendem Beispiel zum Ausdruck kommt: *Diese Flecken bedeuten, dass der Patient Masern hat.*
- ▷ Davon zu unterscheiden ist die nicht-natürliche Bedeutung, wie im folgenden Beispiel: *Wenn die Glocke dreimal läutet, bedeutet es, dass der Bus voll ist.*

„Etwas mit einer Äußerung meinen ist ein Fall von nicht-natürlicher Bedeutung“ (Grice, 1993, S.4f), die wiederum ein Fall von Sprecher-Bedeutung ist. In Grice's Sprachtheorie findet sich aber auch Bestandteil, der über den konventionellen Aspekt der Verständigung hinausgeht und sich auf Formen der Indirektheit, auf sog. Gesprächsimplikaturen bezieht.

Ich möchte, als Kunstbegriff, das Verbum „implizieren“ und die damit verwandten Nomina „Implikatur“ (vgl. „imply“ [andeuten]) und „Implikat“ (vgl. „what was implied“ [das Angedeutete]) einführen. (Grice, 1993, S.246)

Um zu erklären, wie nicht-wörtliche Aussagen verstanden werden können, hat Grice des weiteren ein allgemeines Kooperationsprinzip in der Konversation postuliert, ebenso wie die sogenannten Konversationsmaximen, die vom Kooperationsprinzip abgeleitet sind (Grice, 1993, S.251). Das Kooperationsprinzip ist auf die Tatsache zurückzuführen, dass die zwischenmenschliche Kommunikation mit ganz wenigen Ausnahmen

wie zum Beispiel Verhör u.ä. freiwillig stattfindet. Das bedeutet, dass eine Kommunikation, die freiwillig stattfindet, d.h. auf dem Kooperationsprinzip beruht, folgende Maximen erfüllt:

1. Maxime der Quantität (Information)

- ▷ Mache deinen Gesprächsbeitrag so informativ, wie es für den anerkannten Zweck des Gesprächs nötig ist!
- ▷ Mache deinen Beitrag nicht informativer, als es für den anerkannten Zweck des Gesprächs nötig ist!

2. Maxime der Qualität (Wahrheit)

- ▷ Versuche einen Gesprächsbeitrag zu liefern, der wahr ist!
- ▷ Sage nichts, wovon du glaubst, dass es falsch ist!
- ▷ Sage nichts, wofür du keine hinreichenden Gründe hast!

3. Maxime der Relation (Relevanz)

- ▷ Sage nur Relevantes!

4. Maxime der Modalität (Klarheit)

- ▷ Vermeide Unklarheit! („Eschew obfuscation“)
- ▷ Vermeide Mehrdeutigkeit!
- ▷ Vermeide unnötige Weitschweifigkeit!
- ▷ Vermeide Ungeordnetes!

Grice wird hier vor allem deswegen erwähnt, weil seine Arbeit einen Wendepunkt in der philosophischen Betrachtung von Bedeutung (*meaning*) in der Kommunikationsforschung darstellt.

Austin (1962) beobachtete, dass Äußerungen nicht nur Mitteilungscharakter haben, sondern dass sie eine Art Handlung repräsentieren. Durch diese Verschiebung der Perspektive des Sprachkonzepts wird Kommunikation als sprachliche Handlung betrachtet. Erst in späteren Ansätzen wurde diese Perspektive ergänzt durch den Gesichtspunkt des Handlungszusammenhangs, ein Gesichtspunkt, der bei Wittgenstein allerdings bereits prominent war. Die Sprechakththeorie, die somit begründet und von Searle (1969, 1986) bedeutend weiterentwickelt wurde, bereicherte vielseitig die Semantiktheorie und bot funktionale Klassifikation der Aussagentypologien und einen interessanten Zugang

zu der lokutionären, illokutionären und perlokutionären Bedeutung. Sie bietet auch eine Beschreibung der Bedingungen unter denen unterschiedliche illokutionäre Akte stattfinden - die s.g. „felicity conditions“ (Traum, 1999, S.2f). Die Theorie wurde auch im Rahmen der Informatik und der KI zur Abbildung von Dialogphänomenen übernommen, die in den folgenden Kapiteln näher untersucht werden.

Mit der Idee von indirekten Sprechakten wollte Searle erklären, wie es möglich ist, dass eine Sprecherin etwas sagt und etwas anderes meint. Das wäre unmöglich, oder zumindest eine nicht sehr wahrscheinliche Ausnahme, wenn in so einem Fall die Hörerin keine Chance hätte zu verstehen, was die Sprecherin meint. Die Lösung von Searle ist, dass die Hörerin fähig ist, auf die Bedeutung des indirekten Sprechakts zu schließen und dass sie einige Zeichen von sich gibt, wie das passieren kann. So argumentiert Searle, dass das beabsichtigte Ziel der Bedeutung einer Äußerung auf ihre illokutionäre und nicht so wie Grice vorschlägt auf ihre perlokutionäre Natur zurückzuführen ist.

I argued against Grice that the intended effect of meaning was *understanding* and that a non-circular account of understanding could be given in terms of the commitments of speaker in making the utterance. In brief Grice argued that meaning intentions were intentions to produce a *response* in a hearer. I argued that meaning-intentions were intentions to produce *understanding* in the hearer and that understanding consists in the knowledge of the conditions on the speech act being performed by the speaker. It is the knowledge that I call the 'illocutionary effect'. (Searle, 1986, S.210)

Demzufolge wird etwas kommuniziert, mit dem Ziel die Hörerin zu veranlassen auf eine bestimmte Bedeutung zu schließen. Dieses Ziel kann erreicht werden wie aber auch unerreicht bleiben.

It is a general feature of intentional behaviour that it can succeed or fail, and that the intentions determine the conditions of success and hence the conditions of failure. (Searle, 1986, S.215)

Abgesehen von der Tatsache, dass Intentionen nicht unbedingt kommuniziert werden, schließt Searle, dass die Sprecherin beabsichtigt die Hörerin wissen zu lassen, dass ihre Äußerung eine bestimmte Repräsentation einer Sachlage darstellt und in diesem Sinne auch verstanden werden sollte.

On the hearer's side, understanding the speaker's utterance is closely connected with recognizing his intentions. In this case of literal utterances the bridge between speaker's side and the hearer's side is provided by their common language. Here is how the bridge works:

1. Understanding a sentence is knowing its meaning.
2. The meaning of a sentence is determined by rules, and those rules specify both conditions of utterance of the sentence and also what the utterance counts as.
3. Uttering a sentence and meaning it is a matter of (a) intending (*i-1*) to get the hearer to know (recognize, be aware of) that certain states of affairs specified by certain of the rules obtain, (b) intending to get the hearer to know (recognize, be aware of) these things by meaning of getting him to recognize *i-1* and (c) intending to get him to recognize *i-1* in virtue of his knowledge of the rules for the sentence uttered.
4. The sentence then provides a conventional means of achieving the intention to produce a certain illocutionary effect in the hearer. If speaker utters the sentence and means it he will have intentions (a),(b) and (c). The hearer's understanding in the utterance will simply consist in those intentions being achieved. And the intentions will in general be achieved if the hearer understands the sentence, i.e., knows its meaning, i.e., knows the rules governing its elements. (Searle, 1969, S.48)

Es wird deutlich, dass sowohl Sprecherin als auch Hörerin, in der Abwechslung der Rollen einen wesentlichen Beitrag leisten, damit die Kommunikation erfolgreich zustande kommt (Kooperation) und dass beide dafür sorgen, dass Verständigung statt findet.

2.2 Relevanz, Kontext und Rationalität

Die zentrale Frage, wenn es um Bedeutung von Äußerungen in einem Dialog geht, ist wie Gesprächspartnerinnen das Gesagte interpretieren, so dass sie das Gemeinte verstehen, was ihnen meistens auch gelingt, und wiederum wie sie wissen, was sie in ihrem nächsten Beitrag sagen. Ein wichtiges Prinzip in diesem Zusammenhang ist die dritte Kommunikationsmaxime - die Relevanz. Die Beiträge und die Äußerungen in jedem Beitrag sind relevant zu dem Thema, das gerade besprochen wird. Grice ordnet Relevanz unter der Kategorie der Relation (Grice, 1993, S.251). Sperber und Wilson (1986, S.243) unterstreichen, dass sehr wenig zu diesem Thema im Bereich der Pragmatik gemacht wurde und schlagen eine Theorie der Relevanz (*relevance theory*) vor. Sie sehen Relevanz in einer engen Beziehung zu dem Schlussfolgerungsprozess.

[...] there is an intimate connection between inference and relevance, [...] an approach to relevance in inferential terms can yield fruitful results for pragmatic theory. (Sperber und Wilson, 1986, S.244)

Relevanz ist ihrem Vorschlag nach an erster Stelle, wie auch bei Grice, als eine Eigenschaft der Proposition zu betrachten. Die Proposition selbst, in ihrem Modell, hängt primär nicht von der Äußerung oder von dem Diskurs, sondern von dem Kontext ab. Interessant in diesem Ansatz ist noch die Berücksichtigung von dem Gedächtnis und weiteren kognitiven Fähigkeiten wie das Folgern.

We shall treat relevance in the first instance as a property of propositions (information units, combinations of sense and reference); subsidiary definitions of relevance for utterances or discourses are relatively easy to construct. Similarly we shall treat proposition as relevant in the first instance, not to a text or discourse but to a context, where context is a stock of information derived not only from preceeding discourse, but also from memory, observation and inference. (Sperber und Wilson, 1986, S.244)

Unter Relevanz wird das verstanden, was erlaubt, den neuen propositionalen Inhalt in den Kontext mit wenigst möglichen Operationen einzubetten. Das ist auf die für die meisten pragmatischen Theorien typischen Annahme zurückzuführen, dass Äußerungen in einem bestimmten Kontext gemacht werden, dass diese Äußerungen meistens aus Sätzen bestehen und dass alle Äußerungen implizite Informationen beinhalten. Die Theorie ist also ein Versuch die impliziten Schlußfolgerungen zu erklären, welche die Propositionen und die Äußerungen in einem Kontext bestimmen. Das führt dazu, dass jede Person in der Interaktion zu einer Annahme der Relevanz kommt, was eigentlich bedeutet, dass

1. implizierte propositionale Inhalte genug relevant sind, um kognitive Schritte durchzuführen, um auf diese zu schließen und dass
2. die Sprecherin bei der Kommunikation des Inhaltes genau so ökonomisch vorgeht wie die Hörerin selbst.

Sperber und Willson unterstreichen, dass ihr Konzept keineswegs die vollständige Definition des Begriffes *Relevanz* in Anspruch nimmt, sondern dass dieses Konzept einen zentralen Aspekt der natürlichen Sprache zu zeigen versucht.

It follows that information that has been deliberately communicated, unlike information from other sources, comes with a guarantee that a certain level of relevance has been attempted, if not achieved. [...] Pragmatics so conceived is about the way in which the universal search for relevance is both served and exploited in verbal communication, and the study of relevance is its foundation stone. (Sperber und Wilson, 1986, S.257)

Carlson (1983) im Kontext der Dialogspiele verweist auf den engen Zusammenhang von Relevanz und Rationalität und unterstreicht, dass Relevanz relativ ist:

First, *relevant* is a rational word: a move may be relevant (useful) for one purpose but irrelevant (useless) for another. Second, relevance is relativ to the aims of discussion - or the aims of different discussants if their aims diverge. [...] The present definitioin of relevance creates a link between the concept of relevance and rationality. (Carlson, 1983, S. 46)

Die Dialogteilnehmerinnen stellen ständig Hypothesen, so Carlson, über die mögliche Interpretation der Äußerungen auf und nehmen an, dass die folgenden Beiträge, sich auf das Gesagte beziehen, also relevant sind. Die Hörerin folgt dem kürzesten Pfad bei dem Versuch das Gesagte semantisch zuzuordnen und herauszufinden, was mit einer Äußerung gemeint ist, bis ihre Erwartungen zur Relevanz gedeckt werden. In diesem Prozess schließt die Hörerin unwahrscheinlichere Interpretationen aus (vgl. Carlson, 1983, S. 47ff). Dieser Ansatz ist einerseits dem Prinzip des minimalen Aufwands ähnlich, der auch von Sperber und Willson formuliert wurde, andererseits verwendet Carlson die epistemische Logik Hintikkas, um Schlussfolgerungen zu erklären im Kontrast zu Sperber und Willson, die eine propositionale Logik anwenden. Ich werde auf unterschiedliche logische Ansätze wie z.B. Wissen und Gedächtnis, die in jeder semantisch-orientierten Theorie eine zentrale Rolle spielen, im Punkt 2.4.4 auf Seite 41 und im Kapitel 2.5 auf Seite 42 näher eingehen.

Die Kommunikationsfähigkeit geht also aus der Fähigkeit, Intentionen zu verstehen und zu übermitteln hervor. Dabei wird Gedächtnis und Wissen über gewisse Regularitäten der Kommunikation, wie Relevanz verwendet, um die Bedeutung der von der Sprecherin geäußerten Meinung über einen bestimmten Gegenstand zu ermitteln. Es scheint, dass es die größte Herausforderung für die Wissenschaft ist, zu erklären, wie die ganz unterschiedlichen Faktoren und Gesprächsressourcen zusammenwirken, wenn Sprecherinnen Gesprächsbeiträge als in bestimmter Weise „gemeint“ verstehen. Dabei scheinen das Relevanzprinzip und die Fähigkeit zu folgern eine zentrale Rolle zu spielen.

2.3 Prinzipien der Dialogorganisation

Die Fähigkeit Intentionen zu verstehen, d.h. auf das Gemeinte zu schließen, ist mit Sicherheit eine Hauptfähigkeit Dialog zu führen, jedoch nicht die einzige. Dazu gehört auch die Fähigkeit „Dialogzusammenhänge zu produzieren und zu verstehen“,

gewisse „Dialogmuster zu erkennen“ und „bestimmte Organisationsprinzipien“ zu befolgen (Fritz, 1994b, S.177f). Betrachte man den Dialog formal, so sind einige wichtige Merkmale sofort erkennbar, wie z.B Dialoganfang und Dialogende, mehr als eine Dialogteilnehmerin, die sequentielle Abfolge der Dialogbeiträge (aufeinander folgende Äußerungen) unterschiedlicher Dialogteilnehmerinnen, ein Thema - das, worüber eben gesprochen wird.

Weitere (nicht formale) Merkmale sind, dass die Äußerungen von Dialogteilnehmerinnen nach ihrem propositionalen Gehalt beurteilt und entweder akzeptiert oder verworfen werden können. Dabei spielt die Fähigkeit zu folgern und rational zu handeln eine essentielle Rolle. Diese Aspekte der Dialogführung stellen eine Verknüpfung zu anderen Wissenschaften wie Philosophie und Kognitivistik dar, bzw. stellt sich die Frage nach der Wahrheit und der Art und Weise wie wir mit Information und Wissen umgehen. Ein zentraler Punkt in diesem Zusammenhang ist die Bedeutung, in den beiden Aspekten - „meinen“ und „verstehen“.

Beeinflusst von der Chomsky-Grammatik hat man in den 60er und 70er Jahren des 20. Jh. versucht, eine formale Dialoggrammatik zu entwickeln. Allerdings schlug dieser Versuch aus obengenannten Gründen fehl. Es stellte sich heraus, dass der Dialog nicht rein formal beschrieben werden kann. Vielmehr sollte sowohl Semantik als auch Logik miteinbezogen werden, um gewisse Aspekte des Dialogs zu erklären. Als Alternative wurde die Sprechakttheorie angeboten.

2.3.1 Sprachliche Handlungen und ihre Abfolge

In der Suche nach einem brauchbaren Modell für die Erklärung von Dialogphänomenen sehen viele Sprachphilosophen und Forscher (Wittgenstein, Austin, Searle) eine Möglichkeit, das Sprechen als eine Art Handlung zu betrachten. Das Gesagte ist nicht eine bedeutungslose Lautfolge, sondern „[eine Sprecherin] verfolgt normalerweise eine Absicht, [sie] meint etwas, und ein Hörer kann die Äußerung in einer bestimmten Weise gemeint verstehen.“ (vgl. Gloning, 1994, S.115) Dieser Ansatz hat gewisse Vorteile, denn so könnte man Eigenschaften, die zu nichtsprachlichen Handlungen gehören auf Äußerungen übertragen und darüber gewisse Aussagen machen.

Sie sind in Raum und Zeit lokalisierbar, sie werden mit bestimmten Absichten vollzogen, sie unterliegen dem Gebot der Rechtfertigbarkeit, sie haben manchmal typische Ergebnisse und Folgen, sie werden auf bestimmte Weise ausgeführt und sind befehlbar, d.h. man kann jemanden zum Vollzug sinnvoll auffordern. Auf diese Weise können naturgesetzliche Ereignisse

und nicht-intentionales Verhalten vom Handeln abgegrenzt werden. (Glo-
ning, 1994, S.115)

Die meiste theoretische Arbeit über sprachliche Handlungen (Sprechakttheorie) be-
zieht sich auf illokutionäre Akte. Die wichtigsten Aspekte hierbei sind die illokutionäre
Kraft, welche die Art der Handlung bestimmt (z.B. Frage, Antwort, Vorschlag, Warnung,
Entschuldigung, Informationbeschaffung etc.), und der propositionale Gehalt. Die illo-
kutionäre Kraft ist ein Aspekt der Bedeutung und kann nicht nach Wahrheit beurteilt
werden. Sie bestimmt u.a., was der Adressat nach der Proposition der Aussage tun
würde (z.B. glauben, versuchen) und deswegen sollte sie durch eine Handlungstheorie
und nicht durch Wahrheitstheorie erklärt werden, wie Searle (1969) deutlich zeigt.

Searle beobachtete, dass die sprachliche Handlung, die Austin nach illokutionärer
Kraft und propositionalem Inhalt unterteilt, auch unterschiedlich negiert werden kann.
Die Aussage „Ich verspreche zu kommen.“ kann z.B. propositional negiert werden „ich
verspreche, nicht zu kommen.“ Die Aussage kann aber auch auf illokutionärer Ebene
negiert werden „Ich verspreche nicht, dass ich komme.“ So unterteilt Searle eine illoku-
tionäre sprachliche Handlung in prädikative und referierende Handlung.

Ein großer Teil der Arbeit von Searle ist den indirekten illokutionären Sprechakten
gewidmet und er meinte, dass in solchen Fällen die Sprecherin mehr als das kommu-
niziert, was sie gerade sagt. Dabei verläßt sich die Sprecherin auf eine gemeinsame
Hintergrundinformation und auf die Rationalität der Hörerin und ihre Fähigkeit, darauf
Bezug zu nehmen. In diesem Zusammenhang führt Searle die Begriffe primärer und
sekundärer illokutionärer Sprechakt ein. Der primäre ist der indirekte, der nicht wört-
lich realisiert wurde. Der sekundäre ist der direkte, der mit der sprachlichen Äußerung
realisiert wurde.

(1) Sprecherin X: Wir sollten schon gehen, ansonsten verspäten
wir uns für die Vorstellung.

(2) Sprecherin Y: Ich bin noch nicht fertig.

Beispiel 1: direkter/indirekter Sprechakt

Der primäre Akt in dem Beispiel ist, dass Y den Vorschlag von X ablehnt und der
sekundäre Sprechakt ist Y's Behauptung, sie sei noch nicht fertig. Dadurch wird deut-
lich, wie wir zwei Äußerungsbedeutungen aus ein- und derselben Äußerung ableiten
können.

Betrachtet man einen Dialog als eine Reihe von sprachlichen Handlungen (z.B. Fra-
ge - Antwort, Begrüßung - Entgegnung, Vorwurf - Rechtfertigung, Abschied, etc), so
braucht man an erster Stelle eine Möglichkeit diese auch zu beschreiben.

[Viele der beschreibenden Ausdrücke] sind der Umgangssprache entnommen, wobei ihre Verwendung konkretisiert wird. [...]

Durch den *indem-Zusammenhang* kann man Zusammenhänge zwischen Handlungen beschreiben. Der *und-dann-Zusammenhang* dient zur Beschreibung von Sequenzen und Sequenzmustern. Um Handlungen und Handlungsmuster konkreter zu beschreiben, verwendet man s.g. *Spezifizierungen* „hinsichtlich der Art der Ausführung oder des Inhalts (z.B. grüßen vs. höflich grüßen). Handlungsmuster und ihre dialogischen Umgebung sind mit Spielbäumen darstellbar [...], während sich thematische Zusammenhänge mit Hilfe eines sog. thematischen Netzes übersichtlich machen lassen.“ (Gloning, 1994, S.117).

Desweiteren helfen Fragen (1) nach dem Kontext (Situation) in dem eine sprachliche Handlung realisiert wird, nach (2) dem Gegenstand der sprachlichen Handlung und des Gesprächs (was), (3) nach dem Ziel (warum), (4) nach der Art und Weise (wie) der sprachlichen Handlung. Daraus kann man folgendes schließen, ohne einen Anspruch auf Vollständigkeit zu haben.

1. Man geht davon aus, dass eine sprachliche Handlung in einer bestimmten kommunikativen Situation stattfindet.
2. Eine sprachliche Handlung wird mit einem bestimmten Ziel realisiert - z.B. behaupten, argumentieren. Aus diesem Grund lassen sich Typologien von sprachlichen Handlungen erstellen.
3. Es gibt offensichtlich solche sprachliche Handlungen, die
 - (a) dazu dienen Dialoge, zu eröffnen und abzuschließen (vgl. Schlegoff und Sacks, 1973).
 - (b) Paare bilden - die s.g. *adjacency pairs* (vgl. Jefferson, 1972), wie Frage - Antwort, Einladung - Entgegnung, Vorwurf - Rechtfertigung, etc.
4. Bestimmte implizite und explizite sprachliche Mittel können für die Vollendung einer sprachlichen Handlung eingesetzt werden, wie „ich verspreche“, „ich denke“, „also ist es ...“.

Die sprachliche Handlung, betrachtet als Grundeinheit der Kommunikation, ermöglicht, die Satzebene zu verlassen und die damit verbundenen Einschränkungen zu umgehen. In einer sprachlichen Handlung können mehrere Äußerungen erfasst werden, die wohl syntaktisch und semantisch enger zusammen hängen. Die weitere Frage, die sich

somit stellt, ist wie dieser Zusammenhang zustande kommt, so dass wir von einem Dialogthema sprechen können.

2.3.2 Formen der thematischen Verknüpfung

Jeder Dialog, betrachtet als Text, hat ein Thema, einen Textinhalt. „Textinhalt“ bezeichnet das, worüber in einem Text gesprochen wird und bezieht sich auf einen oder mehrere Gegenstände wie Personen, Sachverhalte, Ereignisse, Handlungen, Vorstellungen usw. Es gibt diesbezüglich drei verbreitete Auffassungen in der wissenschaftlichen Gemeinschaft: Propositionstheorie, Fragetheorie und Gegenstandstheorie. Die Ausgangspunkte sind von den Formulierungen her verständlich. Die Propositionstheorie geht davon aus, dass eine Proposition typischerweise ein Thema formuliert, die Fragetheorie, dass eine Frage das Thema explizit macht und die Gegenstandstheorie, dass über die Bezugnahme auf einen Gegenstand, dieser zum Thema wird. Eine andere Analyse-methode ist die Thema-Rhema-Gliederung. Der Grundgedanke dahinter ist, wie neue Informationen präsentiert werden. Ein wichtiges Prinzip der Kommunikation ist also die Fähigkeit der Dialogteilnehmerinnen, das Thema zu erkennen und relevante Beiträge zu diesem Thema zu machen.

Die zentrale Frage ist dann, nach welchen Prinzipien die Bestandteile komplexer sprachlicher Handlungen zusammenhängen. Hierzu gibt es unterschiedliche Vorschläge. Grosz und Sidner formuliert im Bezug auf Äußerungen in einem Diskurs folgendes:

Discourse structure is a composite of three interacting constituents: a linguistic structure, an intentional structure and an attentional state. Utterances - the actual saying or writing of particular sequences of phrases and clauses - are the linguistic structure's basic elements. Intentions of particular sort and a small number of relationships between them provide the basic elements of the intentional structure. Attentional state contains information about the objects, properties, relations and discourse intentions that are most salient at any given point. (Grosz und Sidner, 1986, S.3).

Das Ziel einer Diskurstheorie ist, die Textstruktur als ein Gefüge von logisch semantischen Relationen zwischen den Propositionen zusammen mit den anderen Organisationsprinzipien der jeweiligen Dialogstruktur darzustellen (siehe Brinker, 2000). Eine Dialogtheorie sollte also auch eine Diskurstheorie (siehe Carlson, 1983, Teil 2) beinhalten, um die Beziehungen zwischen den einzelnen Propositionen zu erklären. Aus der Gesamtheit der Bedeutungen der einzelnen Äußerungen, die durch diese Propositionen bestimmt werden, kann das Thema des Gesprächs formuliert werden.

Die thematisch ausgerichtete Textstrukturanalyse bezieht sich auf den semantisch-kognitiven Zusammenhang, der in dem Text zwischen den in den Sätzen ausgedrückten Sachverhalten (Propositionen) hergestellt wird und der auf unser thematisches Weltbild verweist. Grundlegend für Analyse der thematischen Textstruktur ist der Begriff des Textthemas, der innerhalb der Textlinguistik in verschiedenen Fassungen vorliegt. (Brinker, 2000, S.167f)

Auch hier werden die Logik („semantisch-kognitiven“) und der thematische Zusammenhang (Relevanz) als Zugang zur Bedeutung benutzt, welche in einem Kontext („unser thematisches Weltbild“) zu suchen ist. Ein Text enthält in der Regel aber mehrere Themen, die allerdings jeweils einen unterschiedlichen thematischen Stellenwert besitzen, so dass eine Art Themenhierarchie entsteht. Carlson (1983, S.75, 184ff) nennt das *topic hierarchy*.

Diese Tatsachen deuten darauf, dass jede Dialogteilnehmerin im Bezug auf das Hauptthema und auf die Nebenthemen im Stande ist:

1. diese zu identifizieren auch im Sinne von „verstehen“,
2. die Fakten aus Äußerungen (Aussagen) den Themen zuzuordnen und
3. sich diese nicht nur inhaltlich, sondern auch chronologisch zu merken.
4. später zu einem der bereits besprochenen Themen im Sinne von gemachten Äußerungen und einzelnen Propositionen zurückzukehren.

Solche Ereignisse wie (4) treten immerhin gar nicht so selten in Gesprächen auf.

2.3.3 Grammatisch-lexikalische Aspekte der Dialogorganisation

Texte und Gespräche zeichnen sich durch eine sehr hohe Kohärenz aus. Diese Kohärenz entsteht durch die Koreferenz zwischen Gesprächsgegenständen und die Wiederaufnahme von Gesprächsgegenständen, auf die bereits Bezug genommen wurde (siehe Brinker, 2000). Die Bezugsausdrücke sind solche sprachliche Mittel mit denen ein Gegenstand (Referenzträger), der früher thematisiert wurde, wieder im Mittelpunkt der Äußerung steht. Ein klassisches Beispiel dafür ist der bestimmte und der unbestimmte Artikel im Deutschen, Englischen und anderen Sprachen, die Pronomina oder umschreibende Ausdrücke, wie am folgenden Beispiel gezeigt.

Nach (1) ist bekannt, dass es einen Autohändler gibt und A und B zu ihm fahren und nach (2), dass es um ein neues Auto geht. In (3) referiert *diesem* offensichtlich auf das jetzige Auto von B, also ist bekannt, dass A und B mit dem Auto von B fahren. In

- (1) A: Wieso fahren wir zum Autohändler?
(2) B: Er hat ein Auto, was mich interessiert.
(3) A: Warum denn, was ist mit diesem?
(4) B: Hast du das neue Audi von Peter gesehen?

Beispiel 2: Referenz und Wiederaufnahme

(4) wird klar, dass es eine Person gibt, die Peter heißt und dass dieser Peter ein neues Auto der Marke Audi besitzt (und implizit kann man schließen, dass das der Grund ist, wieso A und B zum Autohändler fahren).

Das Wissen, das vorausgesetzt ist, wenn mit der Verwendung eines bestimmten referierenden Ausdrucks, zum Beispiel des Pronomens *sie*, ein Gegenstand oder eine Gruppe von Gegenständen identifiziert werden soll, hat man als Identifikationswissen bezeichnet. (Fritz, 1982, S.155)

So vermutet jede Sprecherin, wenn sie etwas sagt, bestimmtes Wissen bei der Hörerin, um die Verständigung gewährleisten zu können. Natürlich gelingt das nicht immer, und es werden klärende Subdialoge durchgeführt, um Mißverständnisse und Fragen auszuräumen, wie z.B. bei der möglichen Fortsetzung des oben angeführten Beispiels:

- (5) Wer ist dieser Peter von dem du erzählst?
(6) Peter, den wir gestern zu Mittag in der Stadt getroffen haben.
(7a) Ah so, ich verstehe, diesen Peter meinst du.
(8a) Ja, den Bruder von Stefan.

Beispiel 3: Referenz und Wiederaufnahme (Fortsetzung 1)

Das bedeutet wiederum, dass Verständigung genau dann stattfindet, wenn eine Äußerung genau auf einen Gegenstand, Person, etc. bezogen werden kann. Searle (1969, S.88ff) betrachtet das Referieren als Sprechakt und verweist darauf, dass die Eigennamen eher der Identifikation dienen als zu referieren und spricht dabei von einer definiten Referenz.

The point I wish to emphasize at present, however, is that it is possible for an expression to satisfy the formal requirements of the principle of identification, i.e., to be an identifying description, and still fail to be a useful referring expression. A use of such an expression may still be met with the question „who („what” or „which”) are you talking about?” and it is this question which a definite reference is designed to answer. (Searle, 1969, S.91)

Des Weiteren formuliert Searle folgende Regeln der Referenz, die richtungweisend für den Umgang mit Identifikationswissen sind:

Given that S utters an expression *R* in the presense of *H* in a context *C* then in the literal utterance of *R*, *S* successfully and nondefectively performs the speech act of singular identifying reference if and only if the following conditions 1-7 obtain:

1. *Normal input and output conditions obtain.*
2. *The utteracne of R occurs as part of the utterance of some sentence (or similar stretch of discourse) T.*
3. *The utterance of T is the (purported) performance of an illocutionary act. [...]*
4. *There exists some object X such that either R contains an identifying description of X or S is able to supplement R with an identifying description of X. [...]*
5. *S intends that the utterance of R will pick out or identify X to H.*
6. *S intends that the utterance of R will identify X to H by means of H's recongnition of S's intention to identify X and he intends this recognition to be achieved by means of H's knowledge of the rules governing R and his awareness of C.*
7. *The semantical rules governing R are such that it is correctly uttered in T in C if and only if conditions 1-6 obtain. (Searle, 1969, S.95)*

Das gemeinsame Wissen über verschiedene Gegenstände, Tatsachen und Sachverhalte wird oft in einem Weltausschnitt als Weltwissen bezeichnet. Diese Welt, in der die sprachlichen Handlungen ausgetragen werden, kann der realen Welt oder einer hypothetischen Welt entsprechen. Es gibt in der Regel viele mögliche Welten in denen Dialogpartnerinnen Aussagen über Gegenstände aus diesen Welten machen können, doch Kohärenz, so Fritz (1982, S.171ff), ist dann vorhanden, wenn Äußerungen in nur einer möglichen Welt gemacht werden, deren sich alle Dialogteilnehmerinnen bewusst sind.

Das Prinzip des Referierens und der Wiederaufnahme, so Brinker, scheint ein Grundprinzip für die Textkohärenz zu sein und es lassen sich zwei Typen der Wiederaufnahme unterscheiden - explizite und implizite.

- ▷ Die explizite Wiederaufnahme besteht in der Referenzidentität bestimmter sprachlicher Ausdrücke in aufeinanderfolgenden Sätzen eines Textes. Ein bestimmter Ausdruck (z.B. ein Lexem oder eine Lexemverbindung) wird durch einen oder mehrere Ausdrücke in den nachfolgenden Sätzen des Textes in Referenzidentität

wieder aufgenommen. Der Begriff „Referenzidentität“ besagt, dass sich der wiederaufgenommene Ausdruck und der wiederaufnehmende Ausdruck auf das gleiche außersprachliche Objekt (den sog. Referenzträger) beziehen (Brinker, 2000, S.166). Nach einer anderen Auffassung, die stärker die Aktivitäten der Sprecherin und der Hörerin in den Vordergrund stellt, nehmen Sprecher mit dem Gebrauch von Ausdrücken auf Gegenstände erneut Bezug, die bereits früher in Texten bzw. Gesprächen eingeführt wurden. Die Hörerin *identifiziert* diese Gegenstände als Thema, wenn sie über dasselbe Wissen wie die Sprecherin verfügt (vgl. Fritz, 1982, Kap.6). Die Referenzidentität besteht hier also nicht zwischen Ausdrücken, sondern zwischen Verwendungen von Ausdrücken.

- ▷ Eine implizite Wiederaufnahme ist dann gegeben, wenn zwischen dem wiederaufnehmenden Ausdruck und dem Bezugsausdruck keine Referenzidentität vorliegt. Beide Ausdrücke beziehen sich auf verschiedene Referenzträger (Brinker, 2000, S.166). Die Beziehung zwischen den beiden ist auf der semantischen Ebene - z.B. Busstation - Stadt, Krankenhaus - Chefarzt.

Was Brinker hier meint sind in gewisser Maßen ontologische Relationen, die über implizites Folgern aufgelöst werden (siehe P. 2.5.3 auf Seite 46). Es ist allerdings zu beachten, dass die grammatischen Formen der Vertextung - selbst unter Berücksichtigung von impliziten semantischen Verknüpfungen - keine zwingende Bedingung dafür liefern, dass eine Folge von Sätzen als kohärent interpretiert, d.h. als Text verstanden wird. Dialogteilnehmerinnen besitzen also Wissen, wie sie Textkohärenz über Referenzen und Wiederaufnahmen produzieren und verstehen können. Ein Beispiel dafür liefert Fritz (1982, S.207), wenn A und B über ihren Urlaub sprechen und C später dazu kommt und wissen möchte, worüber die beiden gesprochen haben. A oder B würde Ausdrücke verwenden wie „wir haben gesprochen, diskutiert, uns unterhalten über ...“.

Bei der Beschreibung der Struktur dieser Berichtssätze kann man zwei Aspekte unterscheiden: Mit dem Prädikatausdruck des Hauptsatzes (unterhalten, diskutiert, geklagt, gestritten, diskutiert, Gedanken gemacht) gibt man an, welcher Art von illokutivem Sequenzmuster man die stattgefundene Kommunikation zuordnet. Mit der präpositionalen Ergänzung (*über unseren Urlaub, darüber, daß ..., darüber, wie*) gibt man das Thema der Kommunikation an. (Fritz, 1982, S.207)

Des Weiteren kritisiert Fritz die Propositionstheorie und zeigt, dass die Gegenstands- und die Referenztheorie viel geeigneter für die Behandlung des Thema-Problems sind (siehe Fritz, 1982, S.208ff), denn das Sprechen über denselben Gegenstand oder

dieselbe Person und die Verwendung von Referenzausdrücken führt allein dazu, dass Kohärenz entsteht. Dementsprechend ist es nicht notwendig, das Thema explizit anzugeben, wenn es schon den Gesprächspartnern bekannt ist. Das führt wiederum zu Problemen, bei den Versuchen mechanische Verfahren zur Themenextraktion zu finden. Es ist vielmehr notwendig, den Text oder das, worüber gerade gesprochen wird, zu verstehen, um eine gültige Formulierung des Themas machen zu können. Daraus kann man schließen, dass (1) die thematischen Zusammenhänge zwischen den einzelnen Propositionen mit der gesamten Thema-Proposition eng verbunden sind, ohne die letztere explizieren zu müssen und dass (2) Gesprächspartnerinnen durch kognitive Schritte auf das Gesprächsthema schließen. (3) Die Verpflichtung der Dialogteilnehmerinnen gegenüber der Kommunikationsmaximen, führt dazu, dass über Referenzen und Wiederaufnahmen die besprochene Textkohärenz entsteht und wir von einem Textthema überhaupt sprechen können. Denn das Text- oder Dialogthema ist das, was als Gegenstands des Textes oder der Diskussion nach dem Verstehen der gesamten Bedeutung identifiziert wird.

2.4 Dialog und Dialogforschung

Der Dialog ist die natürlichste Form der zwischenmenschlichen Kommunikation. Die Menschen erwerben diese Fähigkeit, noch im frühen Alter zusammen mit dem Spracherwerb, so dass sie später automatisch kommunikative Situationen mit Hilfe der Dialogführung bewältigen können. Es wird daher in der wissenschaftlichen Welt nach plausiblen Wegen gesucht, die Regularitäten der Dialogkommunikation zu erklären.

Die Hauptaufgabe einer Dialogtheorie ist die Beantwortung der folgenden Frage: Worin besteht die Fähigkeit von Dialogteilnehmern, Dialogzusammenhänge zu produzieren und zu verstehen? (Fritz, 1994b, S.177)

Dialogfähigkeit, so könnte man versuchsweise sagen, besteht in der Fähigkeit, grundlegende Organisationsprinzipien von Dialogen anzuwenden, z.B. nach Sequenzmustern zu handeln, thematische Zusammenhänge zu sehen und herzustellen und kommunikative Prinzipien zu befolgen. (Fritz, 1994b, S.178)

Diese ausgeprägte Dialogfähigkeit hat zwei Aspekte: die mündliche Kommunikation (die Sprache) und die visuelle (Gesten und Mimiken). Im Rahmen dieser linguistisch orientierten Arbeit, bleiben die visuellen Aspekte außerhalb des gesetzten Rahmens. Diese Aspekte der Kommunikation gehören hier wenigstens erwähnt, denn sie spielen eine

wichtige Rolle, indem sie die Kommunikation semantisch bereichern und dadurch die Interpretation der Intentionen der Dialogteilnehmerinnen unterstützen.

In den letzten Jahren werden Dialoge auch in schriftlicher Form über Computer geführt, in s.g. Chats, Chatrooms, oder in IRC¹-Kanälen, doch auch diese Formen der Dialogführung stehen außerhab des in dieser Diplomarbeit gesetzten Rahmens, denn diese Art der Dialogführung stellt einen Kompromiss zwischen Kommunikationsbedürfnis der Menschen und dem Stand der Technik dar. Dies wird durch die rasche Verbreitung von Anwendungen für Videokonferenzen wie z.B. Skype bestätigt, die eine akustische (und visuelle) Kommunikation ermöglichen. Sobald die technische Entwicklung (Informationsverarbeitungsgeschwindigkeit) das notwendige Potential hat, kommt Wissenschaft und Technik der Nachahmung zwischenmenschlicher Kommunikation einen Schritt näher, bzw. der Möglichkeit näher der Maschine Dialogfähigkeiten zu verliehen. Im Hinblick auf das Thema „Dialogsysteme“ wird die behandelte Problematik auf den Dialog als eine kommunikative Fähigkeit des Menschen reduziert.

Neben der Sprechakttheorie zeichnet sich die Dialogspieltheorie dadurch aus, dass sie sich ausschließlich mit dem Dialog als Phänomen beschäftigt und versucht, anhand der Spieltheorie hauptsächlich durch Regeln Phänomene wie Kohärenz und Abfolgeregularitäten zu erklären. Die letzte nennenswerte Alternative bietet der Versuch durch die Prinzipien des rationalen Handelns Dialogregularitäten, Struktur und z.T. auch Regeln zu erklären (vgl. Fritz, 1994b, S.179f). In der Praxis, wie später gezeigt wird, sieht man sich gezwungen Aspekte aller dieser Theorien zu nutzen, um brauchbare (funktionierende) Ergebnisse zu erzielen. Nachdem die Sprechakttheorie schon erwähnt wurde, konzentriere ich mich auf die Dialogspieltheorie und erwähne die Theorie rationaler Handlungen kurz in dem nächsten Kapitel.

2.4.1 Dialogspiele

Die Dialogspieltheorie bietet einen sehr bemerkenswerteren Zugang zum Dialog an, indem sie die Spieltheorie auf den Dialog anwendet. Die Dialogspieltheorie hat zwei Wurzeln: Wittgensteins Sprachspielkonzept und die mathematische Spieltheorie. Wittgenstein spricht von Sprachspielen, bei denen die Sprecherinnen als Spielerinnen und die Sprache als eine Vielfältigkeit von Spielen betrachtet werden könnten.

Ich werde [...] das Ganze: der Sprache und der Tätigkeiten, mit denen sie verwoben ist, das „Sprachspiel“ nennen. (Wittgenstein, 2003, §7)

Das Wort „Sprachspiel“ soll hier hervorheben, dass das Sprechen einer

¹ engl. *internet relay channel*

Sprache ein Teil ist einer Tätigkeit, oder einer Lebensform. (Wittgenstein, 2003, §23)

Ist nicht mein Wissen, mein Begriff vom Spiel, ganz in den Erklärungen ausgedrückt, die ich geben könnte? Nämlich darin, dass ich Beispiele von Spielen verschiedener Art beschreibe; zeige, wie man nach Analogie dieser auf alle möglichen Arten andere Spiele konstruieren kann; sage, dass ich das und das wohl kaum mehr ein Spiel nennen würde; und dergleichen mehr. (Wittgenstein, 2003, §75)

Die Grundidee der mathematischen Spieltheorie ist es, Handlungsfolgen, Entscheidungen und die zugrundeliegenden Handlungsbedingungen als Kalküle zu modellieren. Nach Auffassung Carlsons verbindet eine Sprachspieltheorie die Sprache, betrachtet als ein mathematisches Kalkül, mit der Realität, die von der Sprache interpretiert, beschrieben und transformiert wird (Carlson, 1983, xiv). Es ist dabei essentiell zu unterscheiden zwischen „was gesagt wird“ (*what is said*) und „was getan wird“ (*what is done*) (siehe Carlson, 1983, Fritz, 1994a).

Ich erlaube mir hier einen kurzen Diskurs in die Grundlagen der Spieltheorie. Sie selbst ist ein Zweig der angewandten Mathematik und wird sehr oft in der Wirtschaft verwendet, um strategische Interaktionen zwischen Teilnehmerinnen, die im weiteren auch Agenten² genannt werden, zu untersuchen. In strategischen Spielen wählen die Agenten solche Strategien aus, die ihren Profit unter Berücksichtigung der Strategien der anderen Agenten und der Spielregeln maximieren würden. Das Wichtigste dabei ist, dass die Theorie eine Möglichkeit gibt, Phänomene der Interaktion zwischen intelligent agierenden Einheiten unter Berücksichtigung ihrer Umwelt zu erklären.

Die Theorie entstand 1944 durch die klassische Theorie von von Neumann und Morgenstern in ihrer Arbeit „Theory of Games and Economic Behavior“. 1950 entstand die erste Diskussion über das Gefangenendilemma³. Zu dieser Zeit entwickelte John Nash ein Kriterium zur beidseitigen Konsistenz von Spielerstrategien, welches als das Nash-Equilibrium bekannt wurde und eine breitere Anwendung auf mehrere Spielarten als nur die Theorie von von Neumann und Morgenstern fand. Das Equilibrium erlaubt eine Analyse von nicht kooperativen als auch von kooperativen Spielerinnen. Einige

²Unter Agent werden hier autonome intelligente Einheiten verstanden - Menschen oder Gruppen von Menschen. In der Informatik wird wiederum darunter vor allem Softwareagent verstanden - ein Terminus, der in der KI intelligente Systeme erfaßt, die imstande sind, selbständig zu handeln. (siehe dazu Elhadad 1992; Beskow 1997; McRoy et al. 1998; Restificar et al. 1999; McRoy und Ali 1999; Traum und Allen 1994; Sutton und Cole 1998; Kamm et al. 1999; McRoy 1998; Constantinides und Rudnicky 1999; McRoy et al. 1999; Paek 2001)

³Das Gefangenendilemma ist ein Paradox, der zuerst von zwei Mitarbeitern der RAND Corporation untersucht wurde (siehe z.B. <http://de.wikipedia.org/wiki/Gefangenendilemma>)

Nobelpreise wurden für die Forschungsergebnisse in diesem Bereich verliehen⁴.

Obwohl einige Aspekte der Spieltheorie der Entscheidungstheorie sehr ähnlich sind, unterscheidet sich die Spieltheorie von der letzteren darin, dass die Spieltheorie Entscheidungen in einem bestimmten Umfeld untersucht, in welchem die Spielerinnen miteinander interagieren. Die Entscheidungen beruhen auf den möglichen Entscheidungen der anderen Spielerinnen anhand der vorgegebenen Spielregeln, die allen Spielerinnen bewusst sind. Im Idealfall spricht man von *complete* und *perfect information*. In der Realität arbeitet man aber öfters mit *incomplete* und *imperfect information*.

Complete information ist dann gegeben, wenn alle Spielerinnen wissen, wie die Spielregeln lauten, nicht aber unbedingt Wissen über die Züge der anderen Spielerinnen besitzen. *Perfect information* ist dann gegeben, wenn alle Spielerinnen sowohl die Spielregeln als auch die Züge der anderen kennen.

Die Dialogspieltheorie (Carlson, 1983) schränkt die Spieltheorie ein, indem sie sich konkret mit dem Dialog befasst und zulässt, dass die Dialogpartnerinnen von dem nächsten Zug der anderen nichts wissen dürfen und dementsprechend ihre Wissensbasis nach jedem Zug abgleichen. In diesem Sinne ist die Dialogspieltheorie eine nicht-monotone Theorie (siehe unten „nicht-monotone Logik und Sprechakttheorie“ auf Seite 44) mit *incomplete information*. In der Dialogspieltheorie fehlen auch die Überlegungen von Profit und Verlust der Spielerinnen, die in der Spieltheorie für die Wirtschaft eine zentrale Rolle spielen. Obwohl z.B. Hintikka (1986) das Erreichen des kommunikativen Ziels als eine Art Profit betrachtet, distanziert sich Carlson von dieser Einsicht. Es geht in einem Dialog vor allem um das Gelingen des kommunikativen Ziels und Überlegungen von Profit und Verlust führen zu unnötigem Verkomplizieren des theoretischen Modells.

2.4.2 Agreement, Allwissenheit und Wahrheit

Die Dialogkommunikation zeichnet sich durch formale und kognitive Aspekte aus. Die formalen Aspekte beziehen sich auf die Fähigkeit gewisse Regularitäten des Dialogs zu erkennen und Dialogregeln anzuwenden. Die kognitiven Aspekte betreffen den Umgang mit Information und Wissen und die Denkfähigkeit des Menschen. Wenn Menschen Aussagen über etwas machen, betreffen diese Aussagen Gegenstände (Sachverhalte, Personen, Ereignissen etc.) aus und in einer bestimmten Welt. Im Sinne der aristotelischen Tradition werden manche Wissenschaftlerinnen dazu verleitet, die propositionalen Inhalte nach Wahrheit überprüfen zu wollen und mit solchen logischen

⁴Nash, Selten und Harsanyi - Nobelpreis für Wirtschaft (1994), Thomas Schelling und Robert Aumann - Nobelpreis für Wirtschaft (2005) und Roger Myerson zusammen mit Leonid Hurwicz und Eric Maskin - Nobelpreis für Wirtschaft (2007)

Modellen, die von Wahrheit handeln wie z.B. propositionale Logik oder Prädikatenlogik sprachliche Handlungen formal zu erklären. Denn es ist eine wesentliche Eigenschaft der meisten sprachlichen Handlungen, dass die propositionalen Inhalte akzeptiert oder verworfen werden können. Man kann in manchen Fällen dafür oder dagegen argumentieren. Ich komme auf das Beispiel „Referenz und Wiederaufnahme“ auf Seite 21 zurück, um diesen Gedanken zu veranschaulichen. Eine mögliche Weiterführung des Dialogs wäre z.B. 7a und 8a durch 7b und 8b zu ersetzen:

- (5) Wer ist dieser Peter von dem du erzählst?
- (6) Peter, den wir gestern zu Mittag in der Stadt getroffen haben.
- (7b) Er heißt doch nicht Peter sondern Stefan.
- (8b) Quatsch, Stefan ist sein Bruder.

Beispiel 4: Referenz und Wiederaufnahme (Fortsetzung 2)

Die Frage nach der Modellierung der Wissensvoraussetzungen und wie diese Wissensvoraussetzungen mit Fragen der Dialogorganisation zusammenhängen, stellt ein notorisches Problem in der Dialogforschung dar. Diese Frage ist sehr eng damit verbunden, wie Menschen mit Information, Gedächtnis und Wissen umgehen. Es gibt leider noch keine endgültige Erklärung für die kognitive Tätigkeit des Menschen und es wird in diese Richtung noch immer intensiv vor allem in dem Bereich der Hirnforschung und der Kognitivistik weiter geforscht. Ich präsentiere hier die Überlegungen Carlson's, die einen brauchbaren Ansatz im Rahmen der Sprachphilosophie darstellen und gehe auf unterschiedliche logische Modelle später im Punkt „Kognitive Aspekte und KI“ auf Seite 42 ein.

Nach Carlson gibt es eine ultimative Maxime und das ist **die Maxime des Agreements**. Es geht nicht um Wahrheit oder Lüge, wenn wir miteinander sprechen, sondern darum, uns darauf zu einigen, worüber wir gerade diskutieren. Die Maxime der Wahrheit ist meistens eine Voraussetzung für das Einverständnis, aber nicht unbedingt eine Notwendigkeit, denn betrachten wir einen Dialog als thematisch verknüpfte Äußerungen im Rahmen einer möglichen Welt, so ist es möglich auch fiktive Wahrheit als objektive Wahrheit in der hypothetischen Welt zu betrachten, wenn wir uns nur darauf einigen würden. Diese von Carlson postulierte Maxime scheint daher eine bessere Erklärung für die Analyse von Existenzfragen zu sein, als die von Fritz (1982, P.6.23 auf S.176ff) vorgeschlagene, die zu einem unnötigen Verkomplizieren führt. Es ist daher so, dass wir uns meistens darauf einigen, was wahr ist und was nicht, ohne die absolute Wahrheit über den betreffenden Gegenstand überhaupt zu wissen, wissen zu können, oder wissen zu müssen.

Tatsache ist, dass unterschiedliche Dialogteilnehmerinnen unterschiedliche Dinge unterschiedlich gut wissen, bzw. unterschiedlich fest an das glauben, was sie tatsäch-

lich wissen. Sie wissen nicht genau, was die anderen wissen, sondern ahnen es und versuchen im Laufe eines Gesprächs zu bestätigen, was sie eben vermuten. Daher ist der Dialog ein Spiel der unvollständigen Information (*game with incomplete information*), in dem ständig Hypothesen nach dem Gemeinten und nach der Intention der sprachlichen Handlung gestellt werden.

In vielen Fällen ist es jedoch wichtig, eine objektive Wahrheit als Basis zu haben, worauf sich alle Dialogteilnehmerinnen einigen können, z.B. im Gericht oder bei dem argumentativen Dialog. Ein zentrales Thema in jeder Dialogtheorie ist also das Problem der logischen Allwissenheit ($\frac{\vdash p}{\vdash \text{Knows that } p}$), denn dieses stellt sich immer, wenn es um die Semantik propositionaler Handlungen in möglichen Welten geht. Carlson greift auf die epistemische Logik von Hintikka (1962) zurück, um Modalität in möglichen Welten zu erklären, die bei jedem Zug in einem Dialogspiel entstehen. Die epistemische Logik ist ein Teil der modalen Logik. Hintikka schlägt vor, durch Modalität die Semantik des Wissens und nicht die Semantik der Aussagen zu beschreiben, wie es typischer Weise in der modalen Logik der Fall ist. Ich gehe hier auf die Einzelheiten aus Platzgründen nicht ein. Die theoretischen Grundlagen dazu können in Carlson (1983) und Hintikka (1962, 1986) gefunden werden. Die Schlussfolgerung ist, so Carlson, dass die Natur die einzige, wohlgerneht uninteressierte, dafür aber allwissende Spielerin in einem solchen Spiel ist. Sie erfüllt und vereinigt die Maximen der Konsistenz, der perfekten Information und der Wahrheit in einer idealen Weise. Diese drei Maximen (Agreement, Allwissenheit und Wahrheit), so Carlson (1983, S.21), können in eine einzige Maxime zusammengefasst werden: **Versuche dich mit der Natur zu einigen (*try to agree with nature*)**.

Der von Carlson vorgeschlagene Weg löst viele Probleme, die mit dem Wahrheitsgehalt von Propositionen zu tun haben und erklärt die Motivation (Intention), die hinter bestimmten sprachlichen Handlungen steht, z.B. wieso in dem oben angeführten Beispielen „Referenz und Wiederaufnahme“ auf Seite 21 genau diese Abfolge von sprachlichen Handlungen stattfindet. Dabei spielt die Tatsache, dass die besprochene Person „Peter“ oder „Stefan“ heißt, eine zentrale Rolle in diesem Subdialog und vor allem die Tatsache, dass sich beide Teilnehmerinnen darüber einigen, wie die angesprochene Person wirklich heißt. Dass diese Person „Peter“ oder „Stefan“ heißt, weiß die Natur bescheid, also wenn sich die beiden Teilnehmerinnen mit der Natur und mit sich selbst darauf einigen, dass diese Person tatsächlich z.B. „Peter“ heißt, so ist es auch wahr.

2.4.3 Regeln

Ein wichtiger Grundgedanke in Dialogspieltheorien ist es, zentrale Aspekte der Dialogorganisation mit Hilfe unterschiedlicher Dialogregeln zu modellieren im Fall, dass es keine Möglichkeit gibt eine andere Erklärung für die Phänomene zu finden. Regeln sind typisch für Spiele und beruhen auf Konventionen, ähnlich wie die Bedeutung von Äußerungen auf Konventionen beruht. Zum Beispiel merken sich Kinder grammatische Regeln ziemlich leicht. Sie haben Schwierigkeiten vor allem bei den Ausnahmen. Fragt die Großmutter ihre Enkelin, wie der weibliche Hengst genannt wird, so sagt das Kind ohne zu zögern „Hengstin“, oder im Englischen wie das Plural von „maus“ heißt, sagt das Kind „mauses“ anstatt „mice“, da die Kinder in einem bestimmten Alter eben diese Regeln leichter lernen und über Analogie auf verschiedene Regularitäten der Sprache zu schließen versuchen. Das Beherrschen des Dialogspiels setzt das Beherrschen unterschiedlicher Regeln wie Sequenzierungsregeln, Regeln für den Sprecherwechsel, Regeln wie Subdialoge funktionieren u.ä. voraus. Diese Regeln betreffen sowohl die formale Seite der Sprache („what is said“) wie auch die pragmatische Seite („what is done“).

Stellen wir uns zunächst eine tägliche Situation vor, wie z.B. wenn zwei Nachbarinnen einander vor dem Aufzug treffen. Die erste sagt „Guten Tag!“ und die zweite antwortet „Hallo!“. Beim Verlassen des Aufzugs sagt eine von den beiden „Auf wiedersehen und schönen Tag!“ und die andere antwortet „Auf wiedersehen! Gleichfalls!“ oder einfach „Danke!“. In Österreich würde die Begrüßung aber „Grüß Gott!“ heißen und trotzdem dasselbe bedeuten. Es stellt sich also die Frage wie die Dialogteilnehmerinnen wissen, was in solchen Situation zu sagen wäre und wie sie wissen, was sie als mögliche Antwort anbieten können? In dem Beispiel geht es nicht, um die wörtliche Bedeutung der Äußerungen, sondern, um die pragmatische Funktion, nämlich sich sozial und höflich zu verhalten. Solche konventionalisierte Dialogmuster können, so Carlson, mit Hilfe von Regeln beschrieben werden, weil meistens Regeln einen konventionellen Charakter haben. Durch das Anwenden der Regeln, kann die Zahl der möglichen Alternativen etwas zu verstehen, reduziert werden. Wenn Carlson also von Bedeutungskonventionen (engl.: *meaning conventions*) spricht, werden die grundlegenden Gedanken Grice's (Sprecher-Bedeutung) sichtbar, die schon früher (P. 2.1 auf Seite 8) angedeutet wurden. Anbei wird versucht, mit Hilfe der von Carlson postulierten Regeln bestimmte Dialogmerkmale hervorzuheben.

2.4.3.1 Dialogzug und Sprecherwechsel

An erster Stelle ist der Sprecherwechsel (engl. *turn-taking*, vgl. Sacks et al., 1974; Fritz, 1994b, S.184) als typisches Merkmal zu erwähnen. Das Merkmal ist zwar notwendig, jedoch nicht hinreichend, um von Dialog zu sprechen (als Beispiel kann das gemeinsame Rezitieren von Gedichten angeführt werden). Carlson kritisiert die von Sacks genannte Konvention, weil sie sich per se auf gepaarte Züge (*adjacency pairs*), wie Frage-Antwort, Einladung oder Begrüßung bezieht.

Der Begriff Dialogzug deckt dagegen nur die formale Seite des Äußerns und entspricht eher der Lokution. Carlson unterstreicht, dass eine Antwort einer Frage nur dann folgt, wenn die Antwort am besten dem Kommunikationsziel dient. Der Zusammenhang zwischen Frage und Antwort ist also auf der semantisch-pragmatischen Ebene zu suchen, auf die illokutionäre Kraft der Frage zurückzuführen und nicht allein auf die Tatsache, dass eine Äußerung eine Frage nach formalen Kriterien darstellt. Genauer betrachtet, adressiert jeder Zug einen übergeordneten Gesprächsgegenstand, der eine notwendige Voraussetzung für den vorigen Zug ist (Carlson, 1983, S. 75-77). Der Sprecherwechsel kann wiederum auf semantisch-pragmatischer Ebene relativ einfach erklärt werden.

In particular turns may be jumped quite naturally when the intervening move addresses a superordinate topic which constitutes a necessary condition for the preceeding move. (Carlson, 1983, S. 68)

Die intuitive Annahme ist, dass die Dialogteilnehmerinnen Züge wie die Spielerinnen in einem Spiel machen. Durch bestimmte Züge (*initial moves*) wird ein Bereich für mögliche Folgezüge eröffnet (vgl. Schlegoff und Sacks, 1973). Die darauffolgenden Züge werden als Gegenzüge (*counter moves*) bezeichnet. Zusätzlich, so Carlson, werden Fortsetzungszüge (*continuation moves*) definiert, mit der Restriktion, dass sie von derselben Spielerin gemacht werden und auf das schon Gesagte referieren. Die letztere Klasse von Zügen ist keine echte Dialogklasse, sondern eine monologische, weil die Züge in dieser Klasse von derselben Dialogteilnehmerin gemacht werden. Immerhin kommen solche Phänomene im Dialog sehr oft vor.

Die einfachste Einheit, die sich strukturell deutlich differenzieren läßt, ist also der Dialogzug. Jeder Dialogzug hat bestimmte formale Parameter, die bei der Umsetzung im gegebenen Kontext besetzt werden. Folgende „Parameter des Dialogzugs“ können nach Carlson (1983, S. 68) definiert werden:

1. Sprecher
2. Adressat

3. Publikum
4. Satz
5. Spielregel
6. Prämissen des Zugs
7. der Dialog, in dem der Dialogzug stattfindet

Diese Parameter werden am besten im folgenden Beispiel veranschaulicht (Carlson, 1983, S. 68).

A(1) sagt *B*(2) in der Audienz von *C*(3) den Satz *S*(4) als Antwort(5) von *S'*(6) im Dialog *D*(7).

Beispiel 5: Parameter des Dialogzugs

Interessant sind dabei Fälle, bei denen der Adressat und/oder die Audienz fehlen. Im üblichen Dialog spricht jeder in seiner eigenen Audienz und ist sich dessen bewusst, im Kontrast zum Sprechen im Schlaf zum Beispiel. Fälle, wo die Sprecherin gleichzeitig auch die Hörerin ist, treffen auch zu. In diesen Fällen wird angenommen, dass die Spielerin mit sich selbst und mit der Natur spielt. Der bewusste Ablauf dieses Spiels ist das Denken. Denken ist aber mehr als ein Dialogspiel, welches laut ausgetragen wird, denn beim Denken werden mögliche Dialogspiele anhand der epistemischen Alternativen ausgetragen. Im Dialog bestätigt der nächste Zug der Dialogpartnerin eine dieser alternativen und gibt Anlass diese zu modifizieren und weiterzuverfolgen, oder wegzuworfen. Jede epistemische Alternative besitzt eine geordnete Struktur und produziert Dialogsequenzen, die aufeinander durch die Dialogregeln Bezug nehmen.

A dialogue game (in extensive form) is an object of yet higher order of complexity, as it in turn consists of dialogue game situations ordered into a tree of possible plays. (Carlson, 1983, S. 73)

Meistens ist es so, dass je länger ein Dialog dauert, desto weniger die möglichen Alternativen werden. Zusätzlich wird dieser Prozess durch die Einschränkungen des menschlichen Gehirns reduziert, denn das menschliche Gedächtnis hat eine gewisse Kapazität. Jedoch weisen sogar die einfachsten Dialoge komplexe Strukturen der Alternativen auf. Eine schematische Darstellung präsentiert Carlson (1983, S. 74), wo am Schluss die Dialogteilnehmerinnen und die Natur in ein und derselben epistemischen Alternative operieren.

Der Dialogzug entspricht im wesentlichen einer Äußerung, die auf der semantisch-pragmatischen Ebene eine Bedeutung hat und meistens eine sprachliche Handlung

darstellt. Die allgemeine Struktureinheit, die auf dieser Ebene einen Dialog auszeichnet ist daher die Äußerung⁵:

- ▷ **Äußerung** - jede Spielerin kann eine Äußerung in beliebiger Reihenfolge machen (Carlson, 1983, S.44).

Eine Äußerung hat eine konventionelle Bedeutung, worauf die Hörerin zuerst folgen muss. Je nachdem, ob die Sprecherin an das Gesagte glaubt, oder absichtlich etwas sagt, und dabei etwas anderes meint, muss nach Ernsthaftigkeit oder Glaubwürdigkeit und Ironie unterschieden werden (Carlson, 1983, S.47f).

- ▷ **ernsthaft** - mache eine Äußerung, nur wenn du sie akzeptierst.
- ▷ **ironisch** - mache eine Äußerung, nur wenn du das Gegenteil akzeptierst.

Oft bringen Intonation, Mimiken und Gesten zusätzliche Information über die mögliche Interpretation einer Äußerung. Zum Beispiel „Ja, sicher!“ kann als Antwort, je nach begleitender Intonation, unterschiedlich interpretiert werden.

- ▷ **Vermutung**

Manchmal sagen die Dialogteilnehmer etwas was sie nicht glauben, sondern als möglich zulassen. Solche Äußerungen bringen neue Information für den mentalen Status der Gesprächspartnerinnen. Sie werden als Vermutungen bezeichnet. Die Vermutung kann propositional oder adverbial ausgedrückt werden wie z.B. „Ich glaube, denke, vermute X ist Y“ und „Vielleicht ist X Y“, mit der Bemerkung, dass die Bedeutung von Wörtern wie „vielleicht“ (engl. *perhaps*) nicht eine semantische Frage (Wahrheit), sondern eine pragmatische darstellen (Carlson, 1983, S. 49).

Eine Vermutung kann also je nach *Möglichkeit* und *Wahrscheinlichkeit* unterteilt werden.

- ▷ **möglich** - Sage nur das was du als möglich und wahr zulässt!
- ▷ **wahrscheinlich** - eine Spielerin *P* akzeptiert einen Satz in der Form:
X - wahrscheinlich - Y
nur dann, wenn *P*
X - Y
zulässt.

⁵Carlson verwendet Operatoren, die die Regeln abbilden. Der Operator für Äußerung ist daher *D.say*. Hier geht es allerdings darum, was diese Operatoren abbilden und aus diesem Grund verzichte ich auf die formale Bezeichnung.

▷ Annahmen

Ein weiterer Typ von Äußerungen sind die Annahmen (engl. *supposition*) und sie werden durch Konjunktiv II (*past subjunctive*) abgebildet:

- ▷ **Konjunktiv II** - Sage etwas in Konjunktiv II nur, wenn du es als möglich annimmst (*Put forward a sentence in past subjunctive only if you suppose it*) (Carlson, 1983, S. 53).

Die Definition dazu lautet: „A player *P* *supposes* a sentence *S* iff⁶ *S* is member of hypothetical alternative for *P*.” Hier wird unter hypothetischer Alternative eine Liste der Annahmen verstanden, die die Spielerin nach den üblichen Regeln (siehe z.B. Larchey-Wendling, 2002) zusammen stellen könnte, ohne diese in ihre epistemischen Alternativen einzuschließen.

We shall allow players to hold any number of hypothetical alternatives alongside their *bona fide* epistemic alternatives at any stage of the game. Such alternatives represent the players supposition, thoughts or plans during the conversation. (Carlson, 1983, S. 53)

Solche kontrafaktische Argumente sind schließlich der Schlüssel zu elliptischen Sätzen, die ein Problem in der Dialogtheorie darstellen wie z.B. „Ich würde das nicht versuchen“, wo aber nicht gemeint wird „ich werde es nicht versuchen“, sondern als eine Ellipse für „Ich werde es nicht versuchen, wenn ich du wäre.“ richtig verstanden wird. Diese Regel erklärt also auch den Konjunktiv II in Höflichkeitsformen.

2.4.3.2 Eingebetete Dialogsequenzen

Die eingebeteten Dialogsequenzen (Subdialoge) stellen ein weiteres Problem für jede Diskurstheorie der Konversation dar. Sie übernehmen in diesem Fall, so Carlson, die Funktion von Antworten, allerdings stellen sie relativ autonome Einheiten dar. Subdialoge sind also autonome Dialogeinheiten, die nach Carlson primär als Antworten eingesetzt werden können (siehe „Fehlerkorrektur“ auf Seite 40). Subdialoge können auch unterschiedliche Hilfsfunktionen übernehmen, die Nebenthemen behandeln mit dem Ziel z.B. eine Verständnissicherung, Referenzklärung etc. zu gewährleisten (siehe Jefferson, 1972). Die Subdialoge weisen alle Eigenschaften eines Dialogs auf, sind aber in einem anderen eingebettet und beziehen sich auf die Proposition einer sprachlichen Handlung im übergeordneten Dialog, die somit zum Thema des Subdialogs wird.

⁶iff bedeutet genau dann wenn (engl. *if and only if*)

2.4.3.3 Strategien und Regeln

Unter Strategie wird die Anwendung solcher Regeln verstanden, die den optimalen Effekt erzielen - d.h. dem kommunikativen Ziel am besten dienen. Auch hier geht Carlson von dem Kooperationsprinzip aus. Er schlägt vor, viele Phänomene durch strategische Entscheidungen anstatt durch Regeln zu erklären. Es ist sicher kein Fehler, jemanden zu fragen, was er oder sie weiß. Allerdings ist es in manchen Fällen nicht relevant (z.B. wenn jemand etwas Neues lernen möchte) in anderen aber wohl (z.B. im Gericht). Die Definition erklärt auch das Prinzip der Relevanz, denn je wahrscheinlicher wird eine Annahme, desto relevanter wird ein Beitrag zu dieser Annahme.

Regeln sollten jedoch nur dann formuliert werden, wenn sie konventionelle Tatsachen abbilden, die nicht durch eine generellere Überlegung erklärt werden können.

Rules state what is arbitrary in a game: when a choice is imposed by the rules of a language or its users. (Carlson, 1983, S. 57)

Zum Beispiel kann eine gute kommunikative Strategie durch Konventionalisierung zu einer sprachlichen Regel werden, wie z.B. manche Fragen zu Aufforderungen werden. Durch die Konventionalisierung wird der lange und komplizierte Weg von natürlicher Bedeutung zu illokutionärer Kraft über konversationelle Implikatur gekürzt und die korrekte (nichtnatürliche) Bedeutung kann dadurch verstanden werden. Viele Konventionen unterscheiden sich von Sprache zu Sprache, weil sie arbiträr entstanden sind. Aus diesem Grund gibt es keinen Weg, diesen Prozess formal zu beschreiben. Soweit bleibt ungeklärt, warum manche Aussagen stark konventionalisiert sind und warum andere mehr Freiheit zulassen und dementsprechend, wieso sich manche Sprecherinnen zugunsten bestimmter Phrasen andere sich aber für andere entscheiden.

2.4.3.4 Abgeleitete Regeln

Das Ziel der folgenden abgeleiteten Regeln ist, bestimmte konventionalisierte Dialogstrategien zu beschreiben, die in den meisten Sprachen vertreten sind. Das Vorhandensein von Begriffen wie Frage, Antwort, Behauptung, Entgegnung, Widerspruch, u.ä. sprechen für solche Konventionalisierung. Paare wie Frage-Antwort, Behauptung-Widerspruch werden als *adjacency pairs* bezeichnet und von Jefferson (1972) detaillierter untersucht.

Die von Carlson beschriebene Dialogzügen entsprechen im wesentlichen den illokutionären Handlungen, operieren aber nicht auf der semantischen Ebene. Das bedeutet, dass die Aspekte der Bedeutung nicht miteinbezogen werden. Die Regeln dienen dazu,

eine hierarchisch-funktionale Struktur über die lineare Sequenz der Dialogzüge aufzubauen. Die typischen Züge bzw. illokutionäre Handlungen und die Regeldefinitionen nach Carlson (1983, S. 59ff) sind:

▷ **Fragen**

- ▷ **sententielle** - Jede Spielerin kann jeder anderen eine sententielle (entgegengesetzte) Frage stellen.

Sententielle Fragen hängen nicht mit dem vorigen Kontext zusammen, werden also dann verwendet, wenn kein Zug gemacht wurde (z.B. Dialogeröffnung), oder wenn ein neues Thema eingeleitet wird (Diskursänderung). Alle nachfolgenden Fragen werden unter der Kategorie der kontextuellen Fragen subsumiert.

- ▷ **kontextuelle** - Wenn eine Spielerin die Präsupposition einer Frage dargelegt hat, kann sie jede andere Spielerin befragen.

Zu bemerken ist, dass kontextuelle Fragen einen Gegenzug darstellt (*countermove*), der die explizite oder implizite Anwesenheit einer Präsupposition voraussetzt.

▷ **Antworten**

Der Gegenzug für kontextuelle Fragen ist eine Antwort:

- ▷ wenn eine Dialogpartnerin eine Frage stellt, kann eine andere Dialogpartnerin diese Frage beantworten.

Die Frage- und Antwortregeln sind sehr einfach, denn die Komplexität der Fragen ist auf ihre Semantik zurückzuführen, die aber in das Feld der semantischen Spielregeln gehört. Das einzige was für die Dialogspielregeln übrigbleibt, ist die Bedeutung des optativen⁷ Operators aus den Analysen Hintikka's (Hintikka, 1986) oder, anders gesagt, die Bedeutung des interrogativen Modus zu explizieren.

(1) A: Du unterschreibst das?

(2) B: Ja.

Beispiel 6: Optativ/Interrogativ

Der interrogative Modus ist ein einfaches Frage-Antwort-Spiel und das ist im Einklang mit dem Unterschied, den Wittgenstein zwischen Frage und Behauptung sieht. Eine Frage wird nicht wie eine Behauptung akzeptiert oder verworfen, sondern beantwortet oder nicht (siehe Carlson, 1983, S. 61).

⁷Der Optativ drückt Wunsch oder Hoffnung aus.

▷ logische Züge

Vieles im Dialog geschieht auf der semantisch-pragmatischen Ebene. Zwei logische Operationen können den Schlussfolgerungsprozess - das Folgern (*infer*) und das Erklären (*explain*) beschreiben.

- ▷ **Folgern** - wenn eine Dialogteilnehmerin eine Äußerung gemacht hat, kann sie aus dieser eine weitere folgern.
- ▷ **Erklären** - wenn eine Dialogteilnehmerin eine Äußerung gemacht hat, kann sie diese im nachhinein erklären.
Erklärungen sind eine Unterkategorie der induktiven Argumentation, also argumentative Züge (Carlson, 1983, S. 62).

Die logischen Züge werden durch Wörter wie „also“, „demzufolge“, „dann“ u.ä. lexikalisch markiert. In Fällen, in welchen solche explizite Markierung fehlt, wird die Markierung durch Stimmlage oder in geschriebener Sprache durch Fragezeichen angezeigt. Die logischen Züge sind meistens Fortsetzungszüge (*continuation moves*) (Carlson, 1983, S. 62).

▷ Argumentieren

Argumentation und Argumentieren stellen einen Mittelpunkt in jeder Dialogtheorie dar und werden sogar getrennt behandelt (Toulmin, 1958). Carlson sieht in den argumentativen Zügen einer Dialogteilnehmerin den Wunsch, eine Bestätigung für eine Proposition zu bekommen und fasst zusammen:

When an interlocutor refuses to accept an answer, what one may do is *agree* for it by producing evidence for it. Such relevant evidence are of course sentences that confirm the answer. Hence argumentation is relative to some explication of the relation of confirmation. [...] a sentence is evidence for another sentence if it adds to the probability of the latter. (Carlson, 1983, S. 62f)

Die Ausgangsüberlegung Carlsons ist, dass wenn die Wahrscheinlichkeit für h (*hypothesis*), durch das Auftreten von e (*evidence*) erhöht wird $p(h|e) > p(h)$, dann ist e Argument für h . Und umgekehrt e spricht gegen die Wahrscheinlichkeit $p(h)$, wenn die letzte durch die Anwesenheit von e kleiner wird $p(h|e) < p(h)$.

Diese Wahrscheinlichkeit kann man wiederum nach der Regel der konditionellen Wahrscheinlichkeitstheorie:

$$p(h|e) = \frac{p(h \wedge e)}{p(e)}$$

berechnen, und folgendes ableiten:

$$\frac{p(h \wedge e)}{p(h)p(e)} \neq 1$$

was bedeutet, dass durch das liefern eines Arguments oder eines Gegenarguments die Hypothese gestärkt oder geschwächt wird. Argumentation wird demzufolge folgendermaßen definiert:

- ▷ wenn eine Spielerin eine Aussage gemacht hat, kann sie durch eine weitere Aussage einen Beweis liefern und die erste dadurch argumentativ unterstützen.

▷ **Entgegnen**

Deklarative Gegenzüge und deklarative Äußerungen im Unterschied zu Fragen und Antworten, beantworten eine Frage, Vorwurf u.ä. nicht direkt. Wenn eine Äußerung nicht mit „Ja“ oder „Nein“, sondern mit einer anderen Äußerung beantwortet wird, geschieht das, weil der epistemische Nutzwert der Äußerung zu niedrig ist und die Sprecherin diesen ignoriert.

- ▷ wenn eine Spielerin eine Antwort gegeben hat, kann jede andere Spielerin dieser Antwort entgegnen (Carlson, 1983, S. 64f).

Auf der semantischen Ebene, kann es aber sein, dass das Thema plötzlich gewechselt wird, dass anstatt einer Antwort ein deklarativer Satz folgt. Das bedeutet, dass die Züge in einem Dialog nicht strukturell motiviert sind, sondern sie werden durch Entscheidungen geleitet, die von den Dialogteilnehmerinnen ausgehen. Das kann im Bezug auf die oben definierte Entgegnung am folgenden Beispiel veranschaulicht werden:

A: Magst du Milch und Zucker?
B: Ich trinke ihn schwarz.

Beispiel 7: Entgegnung

Sie korrigiert indirekt oder gibt eine indirekte Antwort einer Frage, zusätzlich beantwortet sie aber die sententielle (topische) Frage. Dies passiert auf Grund von impliziten Annahmen, vor allem wenn die Dialogteilnehmerin glaubt, dass ihre Gesprächspartnerin ihren elliptischen Beitrag richtig verstehen wird. Manchmal führt das auch zu Mißverständnissen, die im nachhinein, meistens in Subdialogen (eingebetteten Dialogsequenzen), aufgeklärt werden müssen.

2.4.3.5 Logische Spielregeln

Bestimmte logische Operationen werden durch entsprechende Lexeme sprachlich angedeutet. Die semantische Bedeutung der Lexeme ist nicht mit einer logischen Operation äquivalent, sondern nur ein Hinweis. Zum Beispiel *wenn ... dann*, deutet auf die logische Implikation. Das Kopulaverb *sein* widerspiegelt die logische Äquivalenz. Wenn jemand *oder* sagt, ist es ein Zeichen dafür, dass er/sie eine andere Möglichkeit zulässt, also logische Disjunktion und *und* ist ein Zeichen für die logische Konjunktion. Eine weitere logische Operation ist die Negation, die im Deutschen auf der Oberfläche mit *nein*, *kein* und *nicht* realisiert wird. *Kein* negiert den nominalen Inhalt der Aussage, *nicht* negiert den propositionalen Inhalt und *nein* ist eine Negation für die ganze Aussage⁸. Das bedeutet, dass bestimmte Wörter die Fähigkeit besitzen, logische Operationen abzubilden. Ich füge einige Beispiele aus Carlson (1983) ein:

▷ **der bestimmte und der unbestimmte Artikel**

Die Anwesenheit vom bestimmten oder unbestimmten Artikel im Beispiel „Referenz und Wiederaufnahme“ auf Seite 21 veranlasst die Hörerin, das dazugehörige Nomen als Referenz zum damit gemeinten Gegenstand als bekannt oder unbekannt zu identifizieren und die entsprechenden Informationen zuzuordnen - „ein Auto“.

▷ **pronominale und andere Referenzen**

Die Verwendung von Pronomina und die Auflösung der Referenzen, die dadurch entstehen (z.B. „er“, „diesem“ ebd.), sprechen dafür, dass wir Texte als ganzes wahrnehmen und mit jedem Satz und jedem Wort uns ein genaueres Bild von dem Textgegenstand und der Intention der Äußerung verschaffen.

▷ **einfache Fragen**

Basierend auf der Logik von Hintikka und vor allem auf seiner Semantik und Pragmatik der Fragen sieht Carlson in einfachen Fragen wie z.B. „Regnet es gerade?“ auch eine eigene semantische Bedeutung, nämlich diese einer elementaren Frage. Einfache negative Fragen sind, so Carlson, immer negative Fragen. Der Vorteil von elementaren Fragen ist, dass sie durch eine einfache Regel erklärt werden können:

Antworte eine elementare Frage mit seiner deklarativen Form (*Answer a question by its declarative form*).

⁸Hier ist nochmals zu unterstreichen, dass die logischen Regeln auf einer bestimmten Sprache oder Sprachfamilie bezogen sind, denn es wird z.B. in manchen slavischen Sprachen eine doppelte Negationen verwendet. D.h. es wird sowohl der propositionale als auch der nominale Inhalt negiert.

Konjunktion (*und*) und Disjunktion (*oder*) helfen komplexe Fragen zu erklären.

▷ W-Fragen

Eine andere Gruppe bilden die W-Fragen (vgl. Carlson, 1983, S.84ff), die besonders oft von Kindern verbalisiert werden. Sie setzen typische Antwortmuster, sowohl inhaltlich, als auch strukturell voraus. Carlson bemerkt, dass W-Fragen rekursiv auf ihren Output anwendbar sind. Daher sollte eine Art unabhängige Charakteristik für Konklusivität definiert werden, damit Rekursivität vermieden werden kann. Das kann mit Hilfe einer Überprüfung der Schlußfolgerungen auf Konklusivität gemacht werden (vgl. Carlson, 1983, S.89ff.).

```
A1: Wer ist Peter?  
B1: Peter ist der Bruder von Stefan?  
A1: Wer ist der Bruder von Stefan?  
B1: Der Bruder von Stefan ist Peter?  
A1: Wer ist Peter?  
...  
A2: Was ist U?  
B2: U ist V?  
A2: Was ist V?  
B2: V ist X?  
A2: Was ist X?  
B2: X ist U?
```

Beispiel 8: Rekursivität von W-Fragen

2.4.3.6 Fehlerkorrektur

Fehler sind sowohl bei der Textproduktion als auch bei der Textrezeption möglich. Es ist dann möglich, dass keine Dialogteilnehmerin den Fehler der anderen bemerkt, weil z.B. die Hörerin nicht genau die Äußerung zuordnen kann und die Sprecherin denkt, sie hätte das richtige gesagt, oder die Hörerin den Fehler der Sprecherin selber reparieren kann, ohne die Sprecherin darauf aufmerksam zu machen. Wenn eine der beiden allerdings den Fehler aufdeckt und der epistemische Nutzwert hoch genug ist, kommt es meistens zu einer Korrektur der Äußerung, die in Form eines Subdialogs abläuft. Jefferson et al. (1977) untersuchen ausführlich solche Phänomene, die ich hier nicht näher eingehen werde. Nach Jefferson (1972) werden auf diese Weise die s.g. *side-sequences* eingeleitet, die von Carlson und anderen Wissenschaftlerinnen Subdialoge genannt werden (siehe oben P.2.4.3.2).

Nach der impliziten oder expliziten Fehlerkorrektur geht der Dialog weiter, an der Stelle, wo er unterbrochen wurde. Alle Dialogteilnehmerinnen, haben in ihrem Gedäch-

nis die Stelle markiert, wo ein Subdialog, bzw. eine *side-sequence* eingeleitet wurde, so dass sie danach den eigentlichen Dialog an dieser Stelle fortsetzen können.

2.4.4 Festlegungssystem und Wissen

Unter Festlegungssystem wird die Fähigkeit der Dialogteilnehmerinnen verstanden, bestimmte Verpflichtungen in den sprachlichen Handlungen einzugehen und festzuhalten. Sie hängen von dem Textverständnis und der Intention der jeweiligen Dialogteilnehmerinnen ab, sollten aber von ihren tatsächlichen Annahmen, Einstellungen und Gefühlen unterschieden werden (Fritz, 1994b, S.187f).

Dieses Phänomen ist eng damit verbunden, wie wir mit Wissen umgehen und auch mit der Rolle, die das Wissen in jedem Dialog spielt. Daher wird es auch als Dialogwissen bezeichnet. Gerd Fritz fasst folgende Grundfragen einer Theorie des Dialogwissens zusammen (nach Fritz, 1994b, S.188):

1. Welche Arten von Wissen kann man unterscheiden?
 - (a) nach dem Gegenstand des Wissens (Wissen-dass, Wissen-wie, Wissen-
wer, sog. episodisches Wissen (Ereigniswissen)),
 - (b) nach der Herkunft des Wissens (allgemeines Erfahrungswissen, Fachwissen, Wahrnehmungswissen, spezielles Wissen aus dem laufenden Dialog),
 - (c) nach der Wissenskonstellation (gemeinsames Wissen, nicht-gemeinsames Wissen, rollenspezifisches Wissen etc.),
 - (d) nach der Art der Präsenz (Wissen, das gerade im Brennpunkt des Dialoggeschehens steht vs. Hintergrundwissen)
2. Wie greift das Wissen in das sprachliche Handeln ein (z.B. als Voraussetzung für bestimmte Handlungen)?
3. Wie entsteht Wissen im Dialog? Wie wird Wissen vermittelt?
4. Welches Wissen ist jeweils bei einem bestimmten Dialogstand relevant?
5. Wie verändert sich das Wissen im Zeitverlauf, speziell: im Dialogverlauf?

Meistens ist eine Dialogform mit bestimmter Wissenskonstellation verbunden, so dass Typologien entstehen z.B. Lehr- und Lerndialoge, Informationsdialoge, bei denen zu Dialogbeginn eine Asymmetrie der Wissensverteilung in Bezug auf das thematische Wissen gegeben ist (Fritz, 1994b, S.188). Für das Ziel dieser Arbeit sind vor allem Lehr-

und Lerndialoge, Auskunftdialoge (Wissenstransfer), argumentative Dialoge (Debatte, Diskussion und Beratung) von besonderer Interesse.

Wissen ist ein immanenter Bestandteil einer jeden Kommunikationsform (unter anderem auch Dialog). Es ist daher wichtig eine Wissensrepräsentation sowohl des Wissens, welches von allen Dialogteilnehmerinnen gesichert ist, als auch ihrer Annahmen und Überzeugungen aufzubauen und intakt zu halten. Eine elegante Variante mit dem Wissen im Rahmen einer Dialogtheorie umzugehen, wurde schon mit der Maxime „*try to agree with Nature*“ vorgeschlagen (Carlson, 1983, S.21).

Dialogteilnehmerinnen sind also im Stande Tatsachen und Dialogabläufe, sprachliche Handlungen und andere relevante Elemente der Dialogkommunikation in ihrem Gedächtnis chronologisch in Evidenz zu halten und Wissen über den Gesprächsgegenstand und die Regularitäten des Dialogs anzuwenden. Im folgenden Punkt möchte ich einige logische Modelle präsentieren, die vor allem im Bereich der KI entwickelt wurden, um die logischen Prozesse für bestimmte Aufgaben abzubilden. Diese Modelle finden tatsächlich eine Anwendung in Dialogsysteme, wie es später gezeigt wird.

2.5 Kognitive Aspekte und KI

In seiner bedeutenden Arbeit „On Computable Numbers, with an Application to the Entscheidungsproblem“ (Turing, 1936-1937) hat Turing Kurt Gödel's aus dem Jahr 1931 stammende Unvollständigkeitssatz neu formuliert, und dabei Gödel's universelle arithmetische Formalsprache mit der heute sogenannten Turing-Maschine ersetzt. Er bewies theoretisch, dass so eine Maschine jedes nur erdenkliche mathematische Problem lösen könnte, solange es als Algorithmus dargestellt werden kann. Die Turing-Maschinen und der Turing-Test (Turing, 1950) bleiben bis heute ein zentrales Thema der Berechenbarkeitstheorie, Informatik und KI. Gerade die Versuche kommunikationsfähige Maschinen zu bauen, haben gezeigt, dass Menschen komplexe kognitive Prozesse in einfachsten Dialoge verwenden und dass ihre Formalisierung eine Herausforderung für Wissenschaft und Technik darstellen. Daher stellt sich die Frage, wie können diese kognitive Prozesse abgebildet werden, wenn es überhaupt möglich ist, sie abzubilden.

2.5.1 Pläne und sprachliche Handlungen

Vielleicht die beste Arbeit im Bereich KI und Sprechakttheorie wurde an der Universität von Toronto von Allen, Cohen und Perrault geleistet (Cohen und Perrault, 1979; Allen und Perrault, 1980). Diese untersuchen das Phänomen der Spracherzeugung und de-

finieren sprachliche Handlungen als Planoperatoren (*plan operators*), welche die Überzeugungen (*beliefs*) der Sprecherin und der Hörerin beeinflussen. Der Grundgedanke dahinter ist, dass es von einem Ist-Zustand ausgegangen wird, mit dem Ziel einen Soll-Zustand zu erreichen. Das Planen besteht darin, eine Sequenz von (sprachlichen) Handlungen zu finden, die zur Änderung des Ist-Zustandes in den Soll-Zustand führen kann. Sie legen fest, dass die Analyse jeder sprachlichen Handlung folgende Fragen beantworten sollte:

1. Unter welche Umstände kann eine Beobachterin glauben, dass eine Sprecherin aufrichtig und erfolgreich einen Sprechakt durchgeführt hat, indem er oder sie eine Aussage für die Hörerin produziert hat?
2. Welche Veränderung des Sprechermodells von der Hörerin und des Hörermodells von der Sprecherin hat der Sprechakt erzeugt?
3. Wie hängt die Bedeutung (Sinn/Referenz) der Aussage *X* bezogen auf die sprachliche Handlung mit den Handlungen zusammen, die beim Äußern von *X* können durchgeführt werden?

Allen, Cohen und Perrault schlagen auch vor, dass eine Theorie, die auf Pläne von sprachlichen Handlungen beruht, folgendes spezifizieren soll:

1. Planungssystem: eine formale Sprache zur Beschreibung der Zustände der betrachteten Welt, eine Sprache zur Beschreibung von Operatoren, einen Satz von Schlußfolgerungen zur Plangestaltung, eine Spezifikation von zugelassenen Planstrukturen und die Semantik der formalen Sprachen sollte auch gegeben sein.
2. Definitionen von sprachlichen Handlungen als Operatoren in dem Planungssystem.
Welche sind ihre Wirkungen? Wann können sie angewendet werden? Wie kann man sie in Worten beschreiben und verstehen?

Das Modell der mentalen Zustände (im Sinne von Wissensstand von Hörerin und Sprecherin) von Cohen und Perrault besteht aus zwei Typen von Strukturen: Überzeugungen (*beliefs*) und Wünsche (*wants*). Diese sind modale Operatoren. Überzeugungen akzeptieren zum Beispiel zwei Argumente: jemand der glaubt (*agent*) und etwas (*proposition*), was geglaubt wird. Weiters kann ein Agent glauben, dass etwas einen Wert hat, ohne zu wissen, was für einen Wert es hat oder dass ein anderer Agent weiß, ob eine Proposition wahr oder falsch ist, ohne selber zu wissen, wieso diese Proposition wahr oder falsch ist. Wünsche unterscheiden sich von Überzeugungen - sie sind

andere modale Operatoren, die die Ziele der Agenten bestimmen. Die Motivation von Plänen als Ergebnis der Intentionen zu sprechen, ist darauf zurückzuführen, dass wir Menschen automatisch als Ergebnis des Gehörten eine Reaktion produzieren, die Maschinen machen es aber nicht. Durch die Planoperatoren wird eine Art Umsetzung der Intentionen in die Oberflächenstruktur der Konversation durchgeführt.

Allen und Perrault (1980) folgen dem selben Muster allerdings mit einem anderen Zweck. Sie untersuchen Schlußfolgerungen und Pläne bei kooperativem Verhalten. Sie zeigen auch, wie wichtig die Erkennung von Plänen anderer Agenten für die Interpretation von Äußerungen ist. Es ist von großer Bedeutung die Intention zu erkennen und Hilfe anzubieten, auch wenn sie nicht gefordert wurde, wenn es darum geht, sich kooperativ zu verhalten. Eine Vertiefung in die weitere Entwicklung dieser theoretischen Ansätzen würde allerdings den Rahmen dieser Arbeit sprengen. Die formalen Modelle, mit denen diese Theorien arbeiten, beruhen auf der nicht-monotonen und modalen Logik. Die Aspekte der Umsetzung werden in dem Punkt „Designstrategien“ auf Seite 77 näher beschrieben.

2.5.2 Nicht-monotone Logik und Sprechakttheorie

Wie es schon erwähnt wurde, gibt es einen wesentlichen Unterschied zwischen Menschen und Maschinen. Menschen denken und besitzen (erwerben im frühen Alter) eine Dialogfähigkeit, die ihnen hilft unterschiedliche kommunikative Situationen zu bewältigen. Dabei werden sie durch Wissen und eine Fähigkeit rational zu handeln und zu denken wesentlich unterstützt. Maschinen im Gegenteil führen logische Operationen durch, mit einem Vorteil, dass sie das sehr schnell machen. Diese logische Operationen werden auf binären Daten ausgeführt. Dadurch entsteht eine Asymmetrie zwischen der kognitiven und sozialen Tätigkeit der Menschen, wenn sie sich an einem Dialog beteiligen, und der rechnerischen Tätigkeit einer Maschine. Nach Turing (1950) kann aber eine Maschine jede mögliche Aufgabe lösen, die in Form eines Algorithmus repräsentiert werden kann. Daher sucht man nach formalen Modellen, um Dialogsituationen, sprachliche Handlungen etc. abbilden zu können. Aus diesem Grund wird intensiv nach logischen Modellen gesucht, die zum Abbilden von kognitiven Prozessen verwendet werden können, um dadurch die formalen Kommunikationsmodelle zu untermauern. Die unten beschriebenen Modelle werden zur Modellierung der Fähigkeiten verwendet, Schlussfolgerungen zu ziehen und für den Dialogverlauf zu nutzen und spielen eine wesentliche Rolle in intelligenten Systemen, darunter auch in den Dialogsystemen. Daher ist es auch wichtig das passende logische Modell zum Abbilden von solchen Phänomenen zu finden.

Die nicht-monotone Logik ist eine formale Logik, die im Gegensatz zur monotonen Logik nicht-monotone Relationen der Schlußfolgerungen erlaubt. Die monotone Logik erlaubt keine Reduktion der Schlußfolgerung nach dem Hinzufügen einer neuen Regel (siehe unten Default-Argumentation und das Beispiel bezüglich „*Vögel fliegen normalerweise*“). Monotonie bedeutet, dass das Hinzufügen neuer Tatsachen, bzw. das Lernen neuen Wissens, das bereits Gewußte nicht reduzieren kann. Die monotone Logik kann unterschiedliche Aufgaben des logischen Denkens nicht lösen, wie Denken nach Vorgabe (*reasoning by default*)⁹, abduktives Denken (*abductive reasoning*)¹⁰, Argumentation oder Denken über Wissen (*reasoning about knowledge*)¹¹ und ebenso nicht das Revidieren von Überzeugungen (*belief revision*). Diese Aspekte sind aber aus großer Bedeutung für das Abbilden von kognitiven Prozessen bei dem Formalisieren von Dialogsituationen und verdienen eine kurze Erklärung:

▷ **Default-Argumentation**

Ein Beispiel für eine Annahme im Sinne der Default-Logik (Reiter, 1980) ist der Vogel, der fliegen kann (siehe „Default-Logik“ auf Seite 46). Daraus folgt, dass wenn man von irgend einem Tier weiß, dass es ein Vogel ist und sonst nichts über das Tier bekannt ist, davon ausgeht, dass es fliegen kann. Das muss aber zurückgenommen werden, wenn sich herausstellt, dass das Tier ein Pinguin ist. Dieses Beispiel zeigt, dass die Default-Logik nicht-monotone Phänomene untersucht. Die Logik, welche die Default-Argumentation in Formeln fasst, kann grob in zwei Kategorien unterteilt werden: einerseits in die Logik, mit der man beliebige Default-Annahmen behandeln kann (Default-Logik - *default logic*, anfechtbare Logik - *defeasible logic*, Antwortmengenlogik - *answer set programming*) und andererseits in die Logik, die die spezifische Default-Annahme beschreibt, die davon ausgeht, dass man standardmäßig die Fakten, von denen man nicht weiß, ob sie wahr sind, als falsch voraussetzt (Umschreibung - *circumscription*, Annahme zur geschlossenen Welt - *closed world assumption*).

▷ **Abduktive Argumentation**

Abduktive Argumentation bedeutet den Prozess der wahrscheinlichsten Ableitungen aus den bekannten Fakten abzuleiten. Abduktive Logik sollte nicht-monoton sein, denn die wahrscheinlichsten Ableitungen müssen nicht unbedingt die richtigen sein. Zum Beispiel ist die wahrscheinlichste Erklärung für nassen Rasen, dass es geregnet hat. Diese Schlussfolgerung muss aber zurückgezogen werden,

⁹Tatsachen können bekannt sein, nur weil es keine Beweise für das Gegenteil gibt.

¹⁰Tatsachen werden abgeleitet, weil diese die wahrscheinlichsten Erklärungen anbieten.

¹¹Das Ignorieren einer Tatsache muss zurückgezogen werden, wenn die Tatsache bekannt wird.

wenn die Tatsache bekannt wird, dass ein Rasensprenger zum Einsatz gekommen ist. Da jetzt die alte Erklärung (der Regen) wegen der hinzugekommenen neuen Information durch eine andere Erklärung (Rasensprenger) ersetzt wird, sollte diese Logik nicht-monoton sein.

▷ **Argumentation über Wissen**

Falls eine Logik Formeln beinhaltet, die mit unbekannten Variablen arbeiten, sollte diese Logik nicht-monoton sein. Vielmehr sollte für den Fall, dass vormals unbekanntes neues Wissen hinzukommt, die Formel, welche die Unbekannte ausweist, entfernt werden. Diese zweite Abänderung allerdings verletzt die monotone Bedingung (Hinzufügen einer neuen Regel führt zur Reduktion des Gewußten). Eine Logik für die Argumentation über Wissen ist daher die (auto)epistemische Logik (vgl. Hintikka, 1962).

▷ **Annahmenüberprüfung**

Annahmenüberprüfung (*belief revision*) bedeutet den Prozess der Abänderung einer Annahme, um sie einer neuen Annahme anzupassen, die mit der alten unvereinbar ist. Vorausgesetzt die neue Annahme gilt als wahr („Die Erde ist rund“), müssen im Sinne der Folgerichtigkeit einige der alten Annahmen überprüft und gegebenenfalls als wahr zurückgezogen werden („am Rand der Erde“, „am Ende der Welt“). Dieses Zurückziehen von alten Annahmen als Antwort auf das Hinzukommen von neuen Annahmen macht jede Art der Annahmenüberprüfung nicht-monoton. Dieser Ansatz der Annahmenüberprüfung ist eine Alternative zur para-konsistenten Logik, welche auch inkonsistente Annahmen zulässt, anstatt diese zu entfernen. Die letztere wird hier vollständigshalber erwähnt und wird nicht näher behandelt.

2.5.3 Default-Logik

Default-Logik ist eine nicht-monotone Logik, die von Raymond Reiter (1980) vorgeschlagen wurde, um Formeln für Argumentation mit Default-Annahmen zu finden. Default-Logik kann Tatsachen beschreiben wie: „Üblicherweise ist etwas wahr“. Im Unterschied dazu kann die übliche Logik etwas nur als wahr oder falsch beschreiben. Dies birgt Probleme, denn normalerweise beinhaltet die Argumentation Fälle, die zwar meistens, aber eben nicht immer wahr bzw. falsch sind. Ein klassisches Beispiel ist: „Vögel fliegen normalerweise“. Diese Regel kann in der üblichen Logik entweder so beschrieben werden: „Alle Vögel fliegen“ ($\forall X (Vogel(X) \Rightarrow Fliegen(X))$), was auf bestimmte Vögel wie Pinguine u.ä. nicht zutrifft, oder „Alle Vögel fliegen oder sie sind Pinguine und die Pinguine fliegen

nicht" ($\forall X((Vogel(X) \Rightarrow Fliegen(X)) \vee (Vogel(X) \wedge Pinguin(X) \Rightarrow \neg Fliegen(X)))$). Letztere Regel benötigt viele Ausnahmen, bzw. eine Umformulierung des ganzen Satzes, wie z.B. „Es existieren Vögel, die fliegen und es existieren solche, die Pinguine sind und nicht fliegen ($\exists X(Vogel(X) \Rightarrow Fliegen(X)) \wedge \exists X(Vogel(X) \wedge Pinguin(X) \Rightarrow \neg Fliegen(X))$). Die Default-Logik zielt auf das Finden von Rückschlussregeln wie die oben angeführte ab, die aber nicht alle Ausnahmen beinhalten müssen.

Eine Definition für Default-Logik könnte folgender Weise formuliert werden(Reiter, 1980): Eine Default-Theorie sei ein Paar $\langle D, W \rangle$. W ist die Menge logischer Formeln, genannt die Hintergrundtheorie (*background theory*), welche die sicher bekannten Tatsachen formalisiert. D ist die Menge von Regeln, von denen alle der Form sind, wie am Beispiel gezeigt:

$$D = \frac{Prerequisite: Justification_1, \dots, Justification_n}{Conclusion_n}$$

$$W = Condition_1, \dots, Condition_n$$

Beispiel 9: Default-Logik

Die logischen Formeln in W und alle Default-Formeln wurden ursprünglich als Formeln der Prädikatenlogik der ersten Stufe (PL1) eingestuft, sie können aber potentiell auch in einer beliebigen formalen Logik vorkommen. Der Fall, in dem sie einer propositionalen Logik angehören, gehört zu den am meisten untersuchten.

Die Default-Regel „Vögel fliegen normalerweise“ kann in folgender Weise ausgedrückt werden:

$$D = \left\{ \frac{Bird(X): Flies(X)}{Flies(X)} \right\}$$

Diese Regel bedeutet, dass wenn X ein Vogel ist und wenn Vögel fliegen können, auch X fliegen kann. Eine Hintergrundtheorie (*background theory*), die ein paar Tatsachen über Vögel beschreibt, ist z.B. folgende:

$$W = \{Bird(Condor), Bird(Penguin), \neg Flies(Penguin), Flies(Eagle)\}$$

Nach dieser Regel fliegt ein Kondor, denn die Vorbedingung $Bird(Condor)$ ist erfüllt und die Begründung $Flies(Condor)$ steht nicht im Widerspruch zu dem, was bekannt ist. Andererseits erlaubt $Bird(Penguin)$ nicht den Rückschluß $Flies(Penguin)$, selbst wenn die Vorbedingung (*prerequisite*) der Default-Regel $Bird(Penguin)$ zutrifft, so ist die Begründung $Flies(Penguin)$ inkonsistent mit dem, was bekannt ist. Aufgrund dieser Hintergrundtheorie und dieser Default-Regel kann $Bird(Eagle)$ nicht gefolgert werden, denn die Regel erlaubt nur $Flies(X)$ aus $Bird(X)$ abzuleiten, aber nicht umgekehrt.

2.5.3.1 Abduktive Argumentation

Das Bezugselement (*antecedens*) aus der Folgerung einer Rückschlussregel abzuleiten, ist das Ziel der abduktiven Argumentation. Eine übliche Default-Annahme ist es anzunehmen, dass das, was nicht als wahr bekannt ist, als falsch eingestuft wird. Das wird als Annahme zur Weltabgeschlossenheit (*Closed World Assumption*) bezeichnet und wird in der Default-Logik mit der folgenden Default-Regel für alle Fakten (F) formuliert:

$$\frac{:\neg F}{\neg F}$$

Zum Beispiel verwendet die Computersprache Prolog eine Art von Default-Regel, wenn mit einer Negation angefangen wird: „wenn ein negative atomare Einheit nicht als wahr bewiesen werden kann, so wird angenommen, dass sie falsch ist. Zu beachten ist aber, dass Prolog die sogenannte Negation als Nichterfüllung betrachtet: Wenn der Interpreter das Atom $\neg F$ auswertet, versucht es zunächst zu beweisen, dass F wahr ist und erst wenn das nicht gelingt, schließt Prolog, dass $\neg F$ wahr ist. In der Default-Logik gilt aber, dass eine Default-Regel mit $\neg F$ als Begründung nur dann gilt, wenn $\neg F$ konsistent mit dem bestehenden Wissensstand ist.

2.5.3.2 Transfer von Überzeugungen

Perrault (1990) geht davon aus, dass eine Aussage alleine nicht ausreichend ist, um den Effekt einer sprachlichen Handlung zu bestimmen. Die Deutung einer sprachlichen Handlung hängt von dem vorigen Kontext ab, einschließlich der mentalen Zustände der Agenten, wie auch von dem, was eigentlich gesagt wurde, von der Proposition. Demzufolge, so Perrault, sollte ein Axiom, das die Deutung einer deklarativen Aussage bestimmt, die Möglichkeiten berücksichtigen, dass diese Aussage eine Lüge, eine Wahrheit, Ironie oder Standardinformation ist. Perraults Zugang ist im Rahmen der „*Default-Logic*“ (Reiter, 1980), so dass der Effekt einer einfachen und üblichen Aussage direkt abgeleitet werden kann, wenn es keinen Einspruch gibt. Perrault verwendet eine einfache Axiomatisierung der Überzeugungen (*beliefs*), Intentionen und Handlungen (*actions*) zusammen mit Standardregeln inklusive Regeln zum Transfer der Überzeugungen (*belief transfer rule*). Diese Regel besagt, dass wenn ein Agent glaubt, dass ein anderer Agent etwas glaubt, der erste dasselbe glauben wird. Eine deklarativen Regel (*Declarative Rule*), besagt, dass wenn ein Agent etwas sagt, er dann an den propositionalen Inhalt dieser Aussage glauben wird. Dieses einfache Schema erlaubt erwartete Konsequenzen von deklarativen Aussagen in unterschiedlichen Kontexten abzuleiten.

Der Ansatz erklärt also, wieso Tatsachen ohne Argumentation (Beweisführung) akzeptiert werden können und wie sprachliche Handlungen motiviert durch eine bestimmte Überzeugung zustande kommen, wie auch wieso anhand von dem propositionalen Inhalt einer Äußerungen auf die dahinter liegenden Intention und Motivation gefolgert werden kann. Obwohl der Formalismus einfach und elegant ist, bleiben viele offene Fragen. Vor allem erweist sich das Revidieren von Überzeugungen als ein Mangel. Wenn zum Beispiel angenommen wird, dass die Motivation für eine sprachliche Handlung X, Y ist und es sich herausstellt, dass diese Schlussfolgerung falsch ist, dann gibt es keine Möglichkeit, diese Schlußfolgerung zurückzuziehen und gleichzeitig eine Konsistenz des Wissens zu gewährleisten.

2.5.4 Modale Logik

McRoy und Hirst (1995) verwenden in ihrer Arbeit ein uniformiertes abduktives Verfahren für das Generieren und Analysieren von sprachlichen Handlungen, von bestimmten Aspekten des Kontextes wie auch für die Analyse der Oberflächenstruktur. Dieser Zugang gibt ihnen die Möglichkeit, Fehler im Prozeß der Interpretation in einer Art und Weise zu erkennen, die auch für das Argumentieren über negierende Präsuppositionen und Implikationen geeignet ist. Um das theoretische Verkomplizieren zu vermeiden, wenden sich andere Wissenschaftler zu anderen logischen Modellen, die bessere Möglichkeiten und eine übersichtlichere Semantik anbieten. Die meisten sind Varianten der dynamischen Logik (Harel, 1979). Sie ist wiederum eine Erweiterung der modalen Logik, die hier kurz dargestellt wird.

In der formalen Logik ist eine Logik dann modal, wenn sie Modalitäten abhandeln kann, also Konzepte wie: Möglichkeit, Existenz und Notwendigkeit. Darüber hinaus werden Logiken modal genannt, wenn sie eine Zahl von anderen Ideen abhandeln können, die lexikalisch durch *eventuell*, *formal*, *können*, *könnte*, *möge*, *mag*, *muss* realisiert werden können. Das ist so, weil diese Phänomene ähnlich abgearbeitet werden.

Eine formale modale Logik repräsentiert Modalitäten mit Hilfe von monadischen modalen Operatoren. Z.B.: „Der Anschlag auf Schmidt war eine Möglichkeit“, „Schmidt wurde möglicherweise ermordet“, „Es ist möglich, dass Schmidt ermordet wurde“ beinhalten die Idee von Möglichkeit. In der modalen Logik wird das durch einen Operator *Möglichkeit* dargestellt, der dem jeweiligen Satz „Schmidt wurde ermordet“ angehängt wird.

Die Hauptoperatoren werden folgenderweise geschrieben: $\Box$ oder L für Notwendigkeit und $\Diamond$ oder M für Möglichkeit. In der klassischen modalen Logik kann jeder durch den anderen in Kombination mit einer Negation ersetzt werden.

$$\begin{aligned}\Diamond P &\leftrightarrow \neg \Box \neg P \\ \Box P &\leftrightarrow \neg \Diamond \neg P\end{aligned}$$

Beispiel 10: Notwendigkeit und Möglichkeit

Also Schmidt könnte ermordet worden sein, nur dann, wenn es nicht notwendig ist anzunehmen, dass Schmidt nicht ermordet wurde (nach Hughes und Cresswell, 1996; Garson, 2006).

2.5.5 Die Logik rationaler Handlungen

Cohen und Levesque (1990) zeigen mit Hilfe einer detaillierten Logik der Interaktion und des mentalen Zustandes, dass die illokutionäre Kraft von den generellen Prinzipien des rationalen kooperativen Verhaltens abgeleitet werden können. Sie behaupten, dass kommunikative Handlungen nicht primitiv sind und definieren, was es für einen Agenten bedeuten würde, sich kooperativ und hilfsbereit zu zeigen. Dazu geben sie eine Charakteristik der Imperative und Aufforderungen. Sie behaupten, das Erkennen der illokutionären Kraft einer Aussage sei überhaupt nicht notwendig, sondern es sei wichtig, dass der Hörer das tut, was der Sprecher will. Sie behaupten also, dass der illokutionäre Akt als ein deskriptives Modell der Handlung betrachtet werden soll und definieren alle illokutionäre Handlungen als Versuche. Aufforderungen (*Performative*) wie „Ich möchte, dass du ...“ haben eine auffordernde Seite und eine leere Stelle. Diese leere Stelle wird durch den Versuch zur und nicht durch den Erfolg der sprachlichen Handlung eröffnet.

Sadek (1992) verwendet eine ähnliche Logik rationaler Handlungen als Basis für die Semantik sprachlicher Handlungen. Die Logik unterscheidet sich von dieser von Cohen und Levesque dadurch, dass mögliche Welten als Zeitpunkte und nicht als ganze Zeitsequenzen von Punkten gesehen werden. Sadek versucht, eine Logik zu entwickeln, die nicht nur als Meta-Theorie der rationalen Handlungen von Agenten fungieren kann, sondern auch eine, die von den Agenten selber verwendet werden kann, um ihre Ziele zu argumentieren und zu erreichen. Die Logik erlaubt das Planen von Handlungen als Ergebnis einfacher Überlegungen über den eigenen mentalen Status der Agenten ohne andere logische Prozeduren (siehe Sadek, 1994; Bretier und Sadek, 1997).

Die kommunikativen Handlungen von Sadek wurden anhand der Voraussetzungen für ihre Durchführbarkeit und der perlokutionären Wirkung formuliert. Jede von ihnen kann zu einer weiteren Planung (z.B. *reasoning*) führen. Die beabsichtigten perlokutionären Ziele einer Handlung sind nicht immer mit den erreichten Zielen der durchgeführten Handlung identisch, sondern eher das Ziel, wofür die Handlung geplant wurde. Manche dieser Ziele werden andere Handlungen erfordern, bevor sie vervollständigt

werden. Diese Handlungen sind allerdings mehr Dialoghandlungen und nicht einfach sprachliche Handlungen.

Die erste Hauptfrage - „was ist Dialog“, die hier gestellt wurde, kann also folgender Weise beantwortet werden: Der Dialog ist die pragmatische Anwendung der Sprache in der zwischenmenschlichen Kommunikation. Ein großer Teil dieser Kommunikation erfüllt die vier Kommunikationsmaximen von Grice. Dabei geht es darum, ein kommunikatives Ziel und Verständigung zu erreichen - etwas mit einer Bedeutung zu äußern, so dass diese auch verstanden wird. Die zweite Frage - „welche charakteristische Merkmale hat der Dialog“ wurde mit der Sprechakttheorie und mit der Spieltheorie erklärt. Der Dialog ist demzufolge eine Reihe von zusammenhängenden sprachlichen Handlungen oder Sprachspielen, die miteinander durch ein Thema zusammenhängen. Die menschliche Dialogfähigkeit zeichnet sich durch eine ausgeprägte Fähigkeit typische Organisationsprinzipien des Dialogs und Dialogmuster identifizieren und anwenden zu können. Die Prinzipien betreffen einerseits die kognitiven Fähigkeiten des Menschen wie die Fähigkeit zu schließen, rational zu handeln, Intentionen zu verstehen, thematisch unterschiedliche Inhalte zuzuordnen etc. Andererseits zeichnet sich der Dialog durch unterschiedliche formale Aspekte aus, wie der Sprecherwechsel (Dialogzug), die Fähigkeit syntaktisch-lexikalische Mittel zu verwenden und dabei zu referieren, durch Identifikationswissen Referenzen zu identifizieren und bestimmte sprachliche Handlungen durchzuführen und die dahinter liegenden Intentionen zu verstehen. Vor allem die Sprechakttheorie und die Dialogspieltheorie geben einen brauchbaren Ansatz in der Dialoganalyse und eine Möglichkeit formal die Dialogkommunikation zu beschreiben. Die formale Beschreibung kann nur durch die Anwendung mathematisch-logischer Modelle abgebildet werden, die die kognitive Fähigkeit des Menschen formal abbilden können. Diese logischen Modelle zeichnen sich durch Nicht-Monotonie und Modalität aus, und werden mit Logiken wie die Default- und die modale Logik beschrieben.

Kapitel 3

Interaktive Dialogsysteme

Um die dritte in dieser Diplomarbeit gestellte Hauptfrage zu beantworten, nämlich „was ist notwendig, um Dialog erfolgreich mit Hilfe von Maschinen (Computersystemen) zu simulieren“, ist es sinnvoll zu untersuchen, was diese Systeme eigentlich sind und wie sie funktionieren. An erster Stelle sollte der Begriff „interaktives Dialogsystem“ geklärt werden. Dieser Begriff wird unterschiedlich interpretiert, bezieht sich aber im Allgemeinen immer auf ein interaktives System, welches in einem bestimmten Bereich (*domain*) arbeitet, wie z.B. Reservierungen, Auskunft u.ä. (siehe Kap. „Beispiele und Anwendungsmöglichkeiten“ auf Seite 99). Dass das System interaktiv ist, bedeutet, dass es im Stande ist, mit anderen Einheiten (in diesem Fall Menschen) zu interagieren. „Dialog“ als Attribut zu System stellt eine zusätzliche Bedingung, nämlich, dass diese Systeme ein gewisses Sprachverständnis benötigen. Reine Spracherwiderungssysteme (IVR - *interactive voice response*) oder Diktiersysteme (*speech typewriters*) werden daher ausgeschlossen (Bernsen et al., 1994). Die Frage nach interaktiven Dialogsystemen wird vor allem aus pragmatischen Gründen gestellt, denn mit der Entwicklung der Computertechnologie werden auch die Einsatzmöglichkeiten für Computer immer mehr. Das führt dazu, dass viel mehr Menschen, vor allem Nicht-Experten, am Computer arbeiten. Die Kommunikation mit dem Computer läuft allerdings, mit Ausnahme der Erfindung und Einführung der Maus in den 1960er/1970er, bis heute noch vor allem über die Konsole (Tastatur/Monitor). Schreiben ist aber eine sekundäre Tätigkeit für die Menschen im Vergleich zu Sprechen und weil Menschen vor allem mündlich miteinander kommunizieren, wird eben diese Anforderung gestellt, dass die Ein- und Ausgabe von Informationen und Kommandos am Computer durch gesprochene Sprache abgewickelt wird.

Ein Dialogsystem sollte also die Mensch-Mensch-Kommunikation in unterschiedlichen kommunikativen Situationen nachahmen können. Das bedeutet Aufnahme von Audio- und ggf. auch von visueller Information, Identifikation von Intentionen, Entschei-

dungen im Bezug auf die nächsten möglichen Beiträge, Generieren von Intentionen bzw. digitalen Gegenstücken zu Intentionen, ihre Verbalisierung und Audio- und Videoausgabe. Die Situationen, in denen eine Interaktion zwischen Mensch und Maschine erforderlich ist, sind allerdings anderer Natur als die typische zwischenmenschliche Kommunikation. Wir benutzen den Computer für Informationsverarbeitung, für Unterhaltung und Kommunikation mit anderen Menschen, aber nicht mit ihm selbst. Es wäre sinnlos den Computer zu fragen „Wie geht es dir heute?“, außer man möchte „menschliches“ Verhalten nachahmen. In einem anderen Aspekt können aber persönliche Fragen gestellt werden, z.B. „Wo hast du die Datei X gespeichert?“. Das bedeutet, dass gewisse Komponenten der zwischenmenschlichen Kommunikation untypisch für die Mensch-Maschine-Kommunikation sind, andere aber transferierbar sind. Es gibt keine endgültige Meinung darüber, wie weit Maschinen personifiziert werden sollen. Da die Entwicklung solcher Systeme eine zeit- und kostenintensive Sache ist, wird zuerst ein Minimum an Brauchbarkeit (*usability*) gedeckt und nachher wird jedes Modell mit Möglichkeiten erweitert, zu persönlichen Fragen Stellung zu nehmen. Dadurch werden bestimmte Dialogtypen (Streit, Beichte, Liebeserklärungen u.ä.) zunächst oder auch langfristig ausgeschlossen.

Die Komplexität der Kommunikationsformen, die in Frage kommen, führt allerdings dazu, dass die vorhandenen Systeme vor allem eine bestimmte kommunikative Situation (z.B. Auskunft und Reservierungen) in einer bestimmten Domäne (z.B.: Luftfahrt oder Bahn) verarbeiten können. Aus der Perspektive des Benutzers ist die Eingabe von Befehlen (die Steuerung) und die Ein- und Ausgabe von Informationen typisch für die Arbeit am Computer. Auf der anderen Seite aber können Maschinen viele monotone Tätigkeiten von Menschen übernehmen und effizienter ausführen wie es schon der Fall ist, wie z.B. Reservierung oder Informationssuche im Internet.

Eine typische Situation, die als Algorithmus dargestellt werden kann, ist, wenn jemand irgendwohin fliegen will und ein Hotelzimmer buchen und/oder Auto mieten muss. Was man üblicherweise tut, ist nach Informationen zum passenden Flug und Hotel zu suchen, ein Flugticket zu kaufen und ein Zimmer zu reservieren. Ein anderes Beispiel wäre der Kauf eines Fahrscheins. Man geht zum Ticketautomaten, wählt das gewünschte Reiseziel, erkundigt sich nach den Fahrtmöglichkeiten und kauft das Ticket durch Drucken von Tasten oder von Schaltflächen auf einem interaktiven Bildschirm. Maschinen sind für solche monotone Tätigkeiten sehr geeignet und setzen sich in vielen Bereichen durch. Die Kommunikation läuft allerdings über Tastatur oder ein Interaktives Display, was eher untypisch für die menschliche Kommunikation ist.

Die Kommunikation in solchen Fällen hat meistens zwei Teilnehmerinnen: die Benutzerin und die Maschine und zwei Aspekte: das Erwerben und Anbieten von X, wobei

X eine Dienstleistung, Information o.ä. sein kann. Das könnte formal mit den unten beschriebenen Algorithmen (Algorithmus 1, Algorithmus 2) ausgedrückt werden. Ich gehe von folgenden Überlegungen aus:

- ▷ Jede X-Bezieherin möchte das bestmögliche X beziehen (Algorithmus 1), sei es Information, Dienstleistung oder etwas anderes.
- ▷ Jede X-Anbieterin möchte der Kundin das bestmögliche X anbieten (Algorithmus 2), sei es Information, Dienstleistung oder etwas anderes, wonach die Kundin, die Benutzerin oder Dialogpartnerin sucht.

Algorithmus 1 Beziehen von X

1. Starte eine Kommunikation mit dem Ziel, X zu beziehen
 2. Suche nach X-Anbieterin
 - (a) Finde mehrere X-Anbieterinnen
 - (b) Vergleiche die Angebote aus 2 (a)
 3. Beziehe ein X von dem X-Anbieterin nach Präferenzen
-

Algorithmus 2 Anbieten von X

1. Starte eine Kommunikation mit dem Ziel, X anzubieten
 2. Erfahre was für ein X gesucht wird und zusätzliche Präferenzen
 3. Finde X-Anbieterin, die der Anfrage aus 2. entsprechen
 4. Biete ein X aus der Liste der X-Anbieterin in 3.
-

Es ist so, dass viele Dienstleistungen und Informationen schon über Internet erreichbar sind. Man braucht also ein interaktives Dialogsystem als Vermittler zwischen X-Anbieterin und X-Bezieherin. Dabei wird deutlich, dass es :

1. ein Ziel oder Aufgabe (*goal, task*) gibt,
2. eine Möglichkeit gibt, zusätzliche sprachliche und andere (z.B. eine Suche nach gewünschten Informationen) Handlungen durchzuführen, um das Erreichen des Ziels zu gewährleisten.

3. eine Möglichkeit gibt, bestimmte kognitive bzw. logische Schritte durchzuführen, also eine Art Entscheidung zu treffen, welche Ergebnisse dem Ziel am besten Entsprechen und
4. eine Möglichkeit gibt, die Aufgabe als erfüllt zu betrachten (nachdem X bezogen oder angeboten wurde).

Folgende Bereiche sind aus dieser Sicht für die Entwicklung solcher Dialogsysteme besonders interessant: Systemsteuerung (Eingabe von Kommandos), Informationsverarbeitung (Eingabe und Ausgabe von Daten), Unterstützung (Ausführung von Aufgaben, Informationssuche, Terminkalenderverwaltung etc.), Tutoren. Solche Anwendungen können auf der Seite der Benutzerin (der X-Bezieherin), auf der Seite der X-Anbieterin und als Client-Server-Applikationen implementiert werden. Besonders weit definierte Unterstützung betrifft die Möglichkeit unterschiedliche Geräte zu steuern (FAX-Geräte, Fernseher, Videogeräte etc.) und unterschiedliche Aufgaben durchzuführen. Das hat in den letzten Jahren zur Entwicklung multimodaler Systeme geführt, die unterschiedliche kommunikative Situationen bewältigen können. Andererseits werden interaktive Dialogsysteme für die Informationsverarbeitung entwickelt, um monotone Tätigkeiten von Menschen zu übernehmen, wie z.B. Informieren und einfachere Formen des Beratens z.B. in Werkstätten, wo die Handwerker Informationen über unterschiedliche Geräte vor Ort beziehen müssen, dabei aber ihre Hände für Reparaturtätigkeiten frei halten müssen.

3.1 Systemkomponenten

Die Art der zwischenmenschlichen Kommunikation führt dazu, dass jedes System zumindest folgende Komponenten haben sollte (siehe Abb.3.1), um die Dialogfähigkeit des Menschen nachahmen zu können:

- ▷ ein Modul für automatische Spracherkennung (*Automatic Speech Recognizer* - ASR, seltener *Speech-To-Text* - STT genannt), um die Umwandlung einer Sprache in Text zu bewerkstelligen,
- ▷ eine Einheit für die Analyse gesprochener Sprache (*NLU* - *Natural Language Understanding*), die dem Sprachinput eine „Intention“ zuschreibt,
- ▷ eine Art von Dialog-Manager (auch Kontroller genannt) - kurz DM, um die Interaktion mit den Benutzerinnen zu steuern,

- ▷ eine Komponente zur Informationsübertragung an die Benutzerinnen - die s.g. NLG Komponente (*Natural Language Generation*), die die Äußerungen für die Text- und/oder Sprachgenerierung bereitstellt und
- ▷ die Komponente für die Sprachausgabe, die die Äußerungen in Audiosignal umwandelt. Meistens wird dafür Sprachgenerierung verwendet, auch Sprachsynthese in Anlehnung an den Speechsynthesizer genannt, die allerdings nur eine mögliche Lösung darstellt.

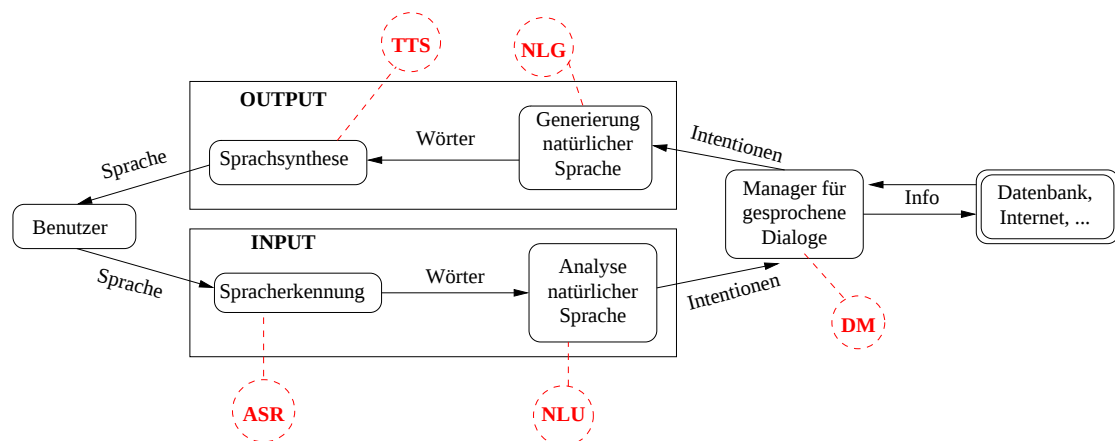


Abbildung 3.1: Grundarchitektur eines Dialogsystems (nach Cuayáhuatl, 2005, S.8)

Typischerweise wird das Sprachsignal der Benutzerinnen über ein Mikrofon oder ein Telefon gesendet. Es geht dann an die ASR Komponente, welche aus dem Sprachsignal die entsprechenden Wörter herausfindet und diese an die NLU Komponente weitergibt. Die letztere wiederum sucht „Absichten“, die mit den Wörtern bzw. Dialogbeiträgen in Verbindung stehen und leitet diese an die DM Komponente weiter. Die Hauptaufgabe des Dialogmanagers ist es, die Konversation effektiv zu steuern, indem die zu den Absichten der Benutzerinnen passenden „Systemabsichten“ aneinander gebracht werden, wobei nicht nur die jeweiligen Absichten der Benutzerinnen, sondern auch die aus der Datenbank ablesbare Information zum derzeitigen Stand und Verlauf der bisherigen Konversation benutzt werden.

Die Output-Komponenten sind das Gegenstück der Input-Komponenten. Sie empfangen einen Set von „Systemabsichten“ und geben entsprechende Sprachsignale an die Benutzerinnen weiter. Die „Systemabsichten“ gehen zur NLU Komponente, welche eine dem Kontext entsprechende Antwort sucht und die dazu notwendigen Wörter in Form von grammatisch korrekten Sätzen an die TTS Komponente weiterleitet. Diese

überträgt nun die Information an die Benutzer als Audiosignal in Form von gesprochener Sprache. Auf diese Weise werden Mensch-Maschine-Konversationen aus den Aktionen der Benutzer und denen der Maschine zusammengesetzt, bis nach einer endlichen Reihe von sich wiederholenden Schleifen das Ziel der Konversation erreicht ist, bzw. die Intention des Benutzers korrekt erkannt und abgearbeitet wird.

Der oben beschriebene Prozess ist nach wie vor eine große Herausforderung für die Wissenschaft und die Technik. Keine der oben beschriebenen Komponenten arbeitet selbst bei einfachen Aufgaben fehlerfrei. So kann es vorkommen, dass die ASR Komponente eine inkorrekte Sequenz von Wörtern wegen Kanalgeräuschen, Hintergrundgeräuschen oder unbekannten Vokabeln aufnimmt und weitergibt. Die NLU Komponente kann inkorrekte Sequenzen von Absichten erzeugen, weil eine Wortsequenz falsch erkannt worden ist oder mehrere Interpretationen zulässt. Hierbei spielt insbesondere das Problem der Verrechnung von Wissenshintergründen beim Verstehen von Äußerungen eine sehr wichtige Rolle. Die DM Komponente hat eine besonders schwierige Aufgabe, da die Input-Komponente falsche Benutzer-Absichten übermitteln kann. Die DM Komponente muss also immer mit einem Unsicherheitsfaktor arbeiten. Um Missverständnisse der Input-Komponenten auszuräumen und die besten „System-Absichten“ zu wählen, muss sie dazu immer die gesamte Information aus dem Verlauf des Gesprächs heranziehen. Letztlich bietet auch die TTS Komponente ein unnatürliches Sprachsignal, das roboterartig klingt, was die Aufmerksamkeit der Benutzer ablenken und einen eher negativen Eindruck bei den letzteren hinterlassen kann.

Obwohl die ASR Komponente außerhalb des Umfangs dieser Arbeit steht, weil die Spracherkennung ein sehr komplexes und an und für sich abgeschlossenes Thema ist, ist der Input, den sie liefert wichtig für die richtige Interpretation der Äußerungen. Wenn die NLU Komponente nicht korrekt funktioniert, kann es passieren, dass ein falscher Inhalt identifiziert und thematisiert wird. Dasselbe gilt auch für die Generierung und die Sprachsynthese. Das bedeutet, dass der Erfolg jeder Komponente von dem Erfolg der vorigen abhängt und dementsprechend die Fehlerquote exponentiell wächst. Daher ist es notwendig alle Hypothesen, die einen Sinn machen, in Betracht zu ziehen. Fehler können daher nur mit einer Art Diskursrepräsentation und einer Verifizierung der Hypothesen vermieden und korrigiert werden. Dies führt zu einer hohen Komplexität und erklärt, wieso die Komponenten gut aufeinander abgestimmt sein müssen. Zum Beispiel kann die NLU-Komponente die n -besten Hypothesen der ASR Komponente im Input bekommen und in seine Diskursrepräsentation einbinden, um vielleicht in einem späteren Zeitpunkt, wenn es sich erweist, dass die verfolgte Strategie die falsche ist, auf eine alternative Hypothese zuzugreifen, um eine neue Strategie zu definieren. Ich bespreche im Folgenden die einzelnen Arten von Systemkomponenten genauer.

3.1.1 Automatische Spracherkennung (ASR)

Automatische Spracherkennung ist der Prozess der Analyse eines Audiosignals zur Bestimmung der Wörter, die von Sprecherinnen ausgesprochen wurden. Sie stellt eine der komplexesten Herausforderungen für die Softwareentwickler dar. Erst seit ein paar Jahren gibt es taugliche kommerzielle Spracherkennungssoftware, was auch auf die gestiegene Prozessorleistung und die erhöhte Speicherkapazität der Computer zurückzuführen ist (siehe Lippmann, 1997; Loos und Zorn, 2007; Jurafsky und Martin, 2007).

Spracherkennung erscheint viel einfacher als es tatsächlich ist, denn Menschen meistern diese Aufgabe so gut und selbstverständlich. Diese Fähigkeit jedoch stammt aus Millionen von Jahren der gemeinsamen Evolution der Stimmorgane und der Biologie des Hörens. Obwohl Wissenschaftler die Verarbeitung der Hörsignale auf niedrigem Level schon recht gut erklären können, sind die Hirnfunktionen auf dem höheren Level noch nicht hinreichend erforscht. Die Algorithmen für die Spracherkennung und die ASR als Prozess, hat im Endeffekt mit dem biologischen Prozess nicht viel zu tun, jedoch ist eine kurze Beschreibung hier durchaus angebracht.

3.1.1.1 Akustische und lexikalische Sprachverarbeitung

Die folgende Abbildung zeigt einen typischen Spracherkennungsprozess, mit der Hauptunterscheidung der dargestellten Typen der akustischen und lexikalischen Verarbeitung. Die akustische Verarbeitung erzeugt ein Netzwerk möglicher phonetischer Sequenzen, die dem Sprachinput entsprechen könnten. Die Lexikalische Verarbeitung verfeinert das Netzwerk durch die Anwendung von kontextueller Information verpackt in Grammatiken.

▷ Audio-Verarbeitung

ASR beginnt mit dem Einfangen des Audiosignals, um es in eine elektrische Form umzuwandeln und es für die Computeranalyse zu digitalisieren. Das Ergebnis ist ein Strom von Zahlen - ungefähr 8000 pro Sekunde (8KHz) für Sprache, die in Telefonqualität erkannt werden soll.

Ein Sprachdetektor, auch End-Pointer genannt, identifiziert Anfang und Ende der Sprache innerhalb des Signals, was bedeutet, dass nichtsprachliche Signale wie Stille und Hintergrundgeräusche nicht verarbeitet werden.

Die gewonnenen Zahlen (nach dem Herausfiltern der gesprochenen Sprache) werden in sogenannte Frames gruppiert, von denen jeder typischer Weise ungefähr zehn Millisekunden lang ist. In Anlehnung an das menschliche Ohr wird

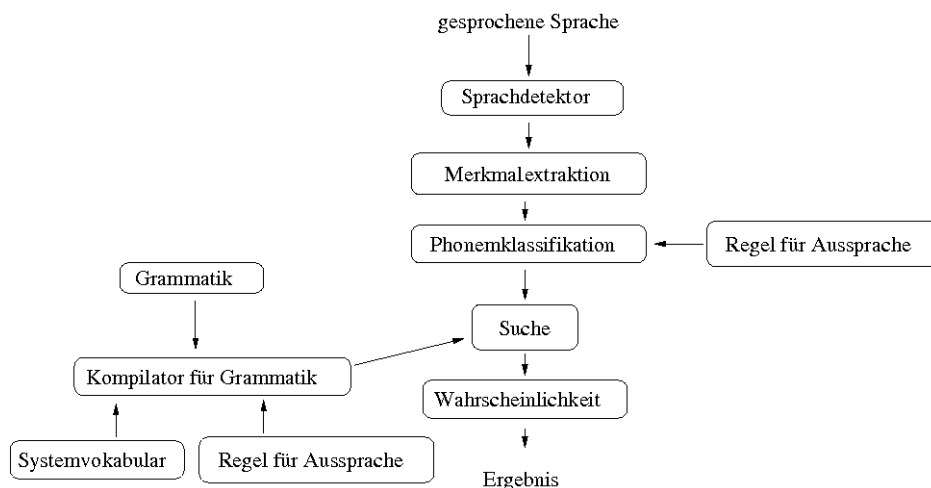


Abbildung 3.2: der Spracherkennungsprozess
(nach OpenSpeech Recognizer - ScanSoft 2007)

jeder Frame weiterverarbeitet, um auf sein Spektrum (die bei verschiedenen Frequenzen vorhandenen Energiemengen) schließen zu können. Dieses Spektrum kann noch weiter verarbeitet werden, um Eigenschaften herauszufiltern, die für die Spracherkennung leichter verwertbar sind (siehe Wahlster, 2006, S.85-107). Zum Beispiel anhand von Mustern, die für jeden Menschen einzigartig sind, kann eine Person identifiziert werden und anhand von Mustern, die für eine Sprache typisch sind, kann Spracherkennung durchgeführt werden.

Gesprochene Sprache ist aus Sequenzen von Wörtern zusammengesetzt, und gesprochene Wörter bestehen bekanntlich aus Phonemen. Weil Phoneme die kleinsten unterscheidbaren Einheiten darstellen, bedeutet das Abändern eines einzigen Phonems die Veränderung eines Wortes in ein anderes oder in ein sinnloses. So wird durch das Ändern von „g“ in „k“ aus dem Wort „Gabel“ das Wort „Kabel“. Um das zu vermeiden, werden statistisch-mathematische Modelle, die s.g. stochastischen Verfahren, angewendet. Solche sind z.B. „Viterbi Algorithmen“ (Wahlster, 2006, S.86) und „Hidden Markov Modelle“ (Jurafsky und Martin, 2007), die die Phonemsequenz in dem durch End-Pointer begrenzten Sprachsignale durch das Berechnen der höchstwahrscheinlichen Phonemsequenz, die den Sprach-Input am besten beschreibt, identifizieren. Diese Wahrscheinlichkeiten werden über den Gebrauch von mathematischen Akustik-Modellen ermittelt, die archetypische Phoneme repräsentieren. Diese akustischen Modelle werden aus Millionen von aufgenommenen Äußerungen (die s.g. Korpora) herausgefiltert und die essentiellen Eigenschaften daraus ermittelt. Eine riesige Anzahl von Daten ist notwendig, um die Vielfalt all der unterschiedlichen Artikulationen zu meistern. Diese Vielfalt ergibt sich aus Unterschieden im Zusammenhang mit der

Physiologie der Sprecherinnen und der Sprechgeschwindigkeit sowie auch aus den Hintergrundgeräuschen, den Kanaleigenschaften, dem linguistischen Kontext und aus vielen anderen Faktoren. Die Kunst der Spracherkennung liegt in der Auswahl des geeigneten Maschinentrainings, das sicherstellen soll, dass die resultierenden Modelle die höchstmögliche Genauigkeit bei der Spracherkennung auch in der Praxis erreichen (Anttila, 2001).

Sogar Menschen verwechseln ähnlich klingende Wörter oder Phrasen. Das ist besonders häufig bei ungewohntem Kontext oder in lauter Umgebung und wenn Menschen z.B. den Liedtexten der Sängerinnen zuhören. In ähnlicher Weise versucht ASR nicht eine spezifische Phrase zu bestimmen, sondern schlägt eine Reihe *n*-bester ähnlich klingender Phrasen vor, die nach Wahrscheinlichkeit geordnet sind. So entsteht der s.g. *word hypothesis graph*, der weiter an die NLU Komponente übermittelt wird (Wahlster, 2006, S.88).

▷ **lexikalische Verarbeitung**

Bei diesem Verarbeitungsschritt muss die Spracherkennungskomponente das Netzwerk möglicher Phonemsequenzen durchsuchen, um die wahrscheinlichste Phrase, die gesprochen wurde zu finden. Die meisten der möglichen Phrasen sind für die Anwendung bedeutungslos, weil sie außerhalb des Kontextes stehen. Der Kontext der Anwendung wird über eine Erkennungsgrammatik ermittelt. Grammatik bedeutet hier einen Set von zu einem bestimmten Zeitpunkt erlaubten Phrasen. Anders als die Schulgrammatik sollen ASR-Grammatiken gar nicht alle Phrasen in einer Sprache repräsentieren. Statt dessen kann jede Eingabe einer Sprachanwendung eine einzigartige Grammatik aufrufen, welche die assoziierte zu erkennende Sprache auf einen bestimmten Bereich eingrenzt. Sobald sich der Kontext der Anwendung ändert, werden andere Grammatiken geladen (siehe Jurafsky und Martin, 2007).

Grammatiken werden erstellt, indem man die erlaubten Phrasen auflistet, wobei bei häufigen Gebilden Abkürzungen verwendet werden. Diese Abkürzungen können auch anzeigen, dass ein Wort aus einer Phrase an verschiedenen Stellen erscheinen kann, was in bewusstem Unterschied zur Anforderung einer expliziten Liste aller vollständigen Phrasen steht.

So ist zum Beispiel:

[Ich möchte] [Johann Mayer|David Baum|Marianne Forster] [sprechen]

kompakter als:

[Ich möchte Johann Mayer sprechen]

[Ich möchte David Baum sprechen]

[Ich möchte Marianne Forster sprechen]

Pfade in der Grammatik können Gewichtungen anzeigen, um eine höhere Wahrscheinlichkeit einer bestimmten Wortsequenz anzuzeigen. Zum Beispiel ist die Sequenz Wien - Österreich wahrscheinlicher als die Sequenz Wien - Deutschland. Die Grammatik kann Wörtern und Sequenzen auch einen semantischen Wert zuordnen, um verschiedene Wörter als das gleiche Ergebnis darstellen zu können (ja, jawohl, bestimmt, sicher usw.).

Alle Wörter müssen in Phoneme umgewandelt werden, wie sie in der Spracherkennung verwendet werden. Jedes Wort wird zunächst in einem Aussprachewörterbuch gesucht, welches die entsprechenden Phoneme bereitstellt. Angepasste Aussprachewörterbücher können verwendet werden, um die Aussprache bei speziellen Anwendungen zu berücksichtigen. Wenn ein Wort nicht in einem Aussprachewörterbuch gefunden wird, werden Ausspracheregeln angewendet, um eine Antwort zu finden, die der einer natürlichen Person nahekommt. Das Resultat dieses Umwandlungsprozesses ist ein Netzwerk von Phonemketten, welche die von der Grammatik erlaubten Phrasen repräsentieren.

Der letzte Schritt der ASR ist das Zusammenführen der Netzwerke von durch Phonem-Label gekennzeichneten akustischen Segmenten mit den entsprechenden von der Grammatik erlaubten Phonemketten. Eine gelungene ASR verfügt über Techniken für diese komplexe Suche, die so effizient sind, dass dies ohne bemerkbare Verzögerung geschieht (Wahlster, 2006, S.129). Diese Beschreibung der ASR umfasst nicht alle Details, die einen wesentlichen Einfluss auf die Usability, die Genauigkeit oder die Performance eines Systems haben. Näheres ist z.B. in Wahlster (2006, S.85-107) zu lesen.

3.1.1.2 Abhängigkeit von der Stimme

Grundsätzlich gibt es zwei Ansätze zur Implementierung von ASR Systemen nach Abhängigkeit von der Stimme der Sprecherin:

- ▷ **abhängig von der Stimme** (*speaker independent*), wenn die Spracherkennung keine Besonderheiten der Prosodie der Sprecherin in Betracht zieht, z.B. in herkömmlichen mobilen Telefongeräten.

Es hat sich erwiesen, dass diese Art von Spracherkennung sehr gut bei Eingabe von wenigen Wörtern funktioniert und sich somit für die Eingabe von Kommandos oder Namen eignet (siehe unten „Einwort-Dialoge“ auf Seite 85), z.B. in der

Mobiltelefonie „Wähle Johann Mayer!“. Meistens sind es kleine Vokabulare, die die meisten Wörter im Bereich Telefonieren und die Eigennamen einer Sprache abdecken. So lassen sich Sprachanwendungen in diesem Bereich kostengünstig umsetzen.

- ▷ **unabhängig von der Stimme** (*speaker dependent*) wird bei wesentlich komplexeren Eingabesequenzen eingesetzt, besonders wenn es sich um semantische Analyse des Gesagten im Nachhinein handelt (Wahlster, 2006, P.4 auf S.93,95).

Die ASR Komponente ist essentiell für die Performanz des ganzen Systems und es zeichnen sich folgende Schwerpunkte ab, die jedes System berücksichtigen soll:

- ▷ **Kanalbedingungen**

Die akustischen Inputbedingungen hängen natürlicherweise von Einstellungen der jeweiligen Systeme ab. In den letzten Jahren ging man weg von geräuschdämpfenden Nahfeldmikrophonen und jetzt verwendet man normale Telefone. Diese ermöglichen einem viel größeren Benutzerkreis den Zugang aber die geringe Bandbreite, das geringe Signal-Rausch-Verhältnis (S/N - *signal/noise level*) und die verschiedenen Bauweisen der Hörer bergen viele Probleme in der akustischen Umsetzung. Mobiltelefone müssen noch dazu mit zusätzlichen Störgeräuschen und Aussetzern fertig werden (Wahlster, 2006, P.3.2 auf S.92).

- ▷ **Variationen in der Sprechweise**

Als die Systeme in der Öffentlichkeit getestet und Daten gesammelt wurden, stellte es sich heraus, dass die verschiedenen Dialekte und die Aussprache von Nicht-muttersprachigen eine sehr umfassende Datensammlung notwendig machten. Je nach Stärke des Dialekts oder des Akzents werden die Sprecher-unabhängigen Erkennungsmethoden gefordert. Auch die Kindersprache stellt einen eigenen Typus dar. Die Systeme müssen speziell entwickelt werden, auch deshalb, weil die Kanalbandbreite einen beträchtlichen Anteil der Sprache von Kindern eliminiert. Die Kinder erwiesen sich erstaunlicher Weise als sehr ausdauernde Testpersonen, tendierten aber dazu, sich zu wiederholen anstatt ihre Äußerungen neu zu formulieren, falls sie nicht oder falsch verstanden wurden (Wahlster, 2006, S.89f).

- ▷ **Adaption**

Adaption wird benutzt, um die Leistung eines Systems für individuelle Benutzerinnen zu optimieren. Die traditionellen Adaptionmechanismen benötigen eine große Datenmenge und sind nicht geeignet, dort eingesetzt zu werden, wo es nur

eine geringe Anzahl an Fragen gibt und wo keine Benutzer-Identifikation benötigt wird. Deshalb muss man Technologien finden, die Adaptionen in kurzer Zeit bewerkstelligen können. Anwendungen mit bekannter Benutzer-Identität könnten auf eine Art Benutzer-Profil zurückgreifen, in dem akustisch-phonetische Eigenheiten, Vokabular, Dialekt und auch die Region der Benutzerin gespeichert werden, was die Adaption wesentlich beschleunigen könnte (Wahlster, 2006, P.4 auf S.93ff).

▷ **Hinzufügen neuer Wörter**

In Systemen mit uneingeschränktem Input ist das Vokabular nicht spezifisch auf die Benutzerin abgestimmt, daher können diese auch Wörter verwenden, die dem System nicht bekannt sind. Diese Wörter werden OOV (*out-of-vocabulary*) bezeichnet. Bei JUPITER (Zue et al., 2000) und ähnlichen Systemen, die nicht alle Städte der Welt kennen, versuchen die Benutzer oft, sich durch Buchstabieren der Wörter oder durch Betonen einzelner Silben der Wörter dem System verständlich zu machen, was aber die Sache eher noch verkompliziert. Wenn also unbekannte Wörter als solche erkannt werden, wäre es sehr nützlich, daher sollte das System über Möglichkeiten verfügen, die Korpora um dieses neue Vokabular dynamisch zu erweitern oder alternative Methoden anzuwenden. Eine Möglichkeit die im SMARTKOM System verwendet wird, ist das akustische Modell des Wortes zu speichern, anstatt das Wort mit einem ähnlichen mit niedriger Wahrscheinlichkeit aus dem Vokabular zu ersetzen wie es andere Systeme machen (Wahlster, 2006, P.6 auf S.102ff).

▷ **Unterbrechen - (*Barge-In*)**

Nur wenige Systeme unterstützen das s.g. *barge-in* oder Unterbrechen beim Sprechen. Dieses Phänomen kommt sehr oft in Mensch-Mensch-Dialogen vor und ist etwas Selbstverständliches für die zwischenmenschliche Kommunikation, wenn dies auch nicht immer als höflich bezeichnet werden kann. Die Abarbeitung des Phänomens überschreitet die Grenzen der ASR und betrifft alle Komponenten des Systems. Wenn das System eine Äußerung macht und die Benutzerin anfängt zu sprechen, dann ist es meistens so, dass sie auf den Ausgang der Äußerung schließen kann oder z.B. die Äußerung bestätigen oder bezweifeln will (uhu, ja, ok, gut, uhm) ohne diese unterbrechen zu wollen. Oft passiert es aber, dass sie sich irrt und ihre Äußerung auf falschen Präsuppositionen beruht. Das System sollte also im Stande sein, eine Unterbrechung zu registrieren, die Äußerung der Sprecherin auf Kompatibilität zu überprüfen, die eigene Äußerung zu unterbrechen, zu korrigieren oder einfach fortzusetzen (Wahlster, 2006, P.3.1, S.90).

3.1.2 Analyse gesprochener Sprache (NLU)

Sobald ein Dialogsystem über den Input isolierter Wörter hinausgeht, wird es notwendig, dass es Wortsequenzen verarbeiten kann, um diese entsprechend weiterverwenden zu können. Hierzu wurde eine große Anzahl von Methoden erforscht um herauszufinden, wie linguistische Regeln formuliert werden können und wie diese in die Suchfunktion integriert werden können.

3.1.2.1 Äußerungen vs. Phrasen

Manche Systeme betreiben einen Recognizer¹ mit einer formalen Grammatik. Dieser Ansatz hat sich aber als potentiell zu restriktiv erwiesen. Flexiblere Modelle versuchen nicht Äußerungen als Ganzes zu analysieren, sondern machen Phrasen und Schlüsselwörter ausfindig. Dieser Weg war sehr erfolgreich im DARPA ATIS (*AirTraffic Information System*) Programm (Moore et al., 1995) und wurde auch ebenso erfolgreich in kommerziellen Anwendungen eingesetzt. MIT hat unter anderem einen gemischten Ansatz gewählt, bei dem zuerst eine komplette Analyse der Äußerung stattfindet und wenn diese erfolglos ist, mit Hilfe eines Parsers zur anderen Methode übergegangen wird.

In allen Fällen hat sich stochastisches Parsen als sehr nützlich erwiesen und es ist klar geworden, dass eine vollständige linguistische Analyse in manchen Fällen überflüssig und nicht notwendig ist. Für einfache Anwendungsgebiete wie bei JUPITER (Zue et al., 2000) genügt das Auffinden von Phrasen und Schlüsselwörtern offensichtlich vollkommen. Für komplexere Aufgaben reicht das aber nicht aus. Je komplexer der Aufgabenbereich eines Systems, desto höher auch die Komplexität der abgebildeten Welt. In moderneren Systemen wie SMARTKOM (Wahlster, 2006, Kap.3, S.74ff) und dem Nachfolger SMARTWEB werden Ontologien angebunden, um Wissensrepräsentation im Dialog zu unterstützen. Dieses Verfahren hat gezeigt, dass die Identifikation des Diskurses bzw. des Textthemas, wie auch die Rationalität der Entscheidungen bezüglich Interpretation und Generierung von Äußerungen dadurch deutlich verbessert werden.

3.1.2.2 Linguistische Regeln

Viel Anstrengung ist in die Richtung gegangen, wie man linguistische Regeln in eine Art grammatisch-syntaktische Formalismen packen kann, um diese in die Suchfunktion von entsprechenden Mustern einbauen kann. Für die Suche wird meistens eine

¹in der Informatik ein bestimmtes abstraktes Maschinenmodell

Datenbank in Form eines Korpus bzw. einer Ontologie zur Verfügung gestellt und eine Menge von Regeln. Die einfachste Strategie dafür ist, die zwei Prozesse ASR und NLU völlig zu trennen, wobei die automatische Spracherkennung den wahrscheinlichen Top-Choice Satz ermittelt und diesen an die NLU Einheit für weitere Analyse zusammen mit den n -besten Hypothesen weitergibt. Derzeit können viele Systeme einen Set von Sätzen n -bester Wahrscheinlichkeit oder Wort-Graphen ermitteln, die von der NLU Einheit verarbeitet werden. Zwar hat man auch mit enger gekoppelten Methoden experimentiert, stieß aber auf hohen Rechenaufwand bei komplexeren Grammatiken. Allgemein versucht man so früh wie möglich linguistische Regeln in die Suchfunktion einzubauen, ohne aber Suchfehler oder Rechenzeit länger als *realtime* zu provozieren (Glass, 1999).

In dem SMARTKOM System wurde eine neue Methode entwickelt (SPIN), um das Entwerfen von Regeln und ihre Anwendung zu vereinfachen und eine höhere Übersichtlichkeit zu erreichen (Wahlster, 2006, S.195f). Die Regeln repräsentieren vollständige oder unvollständige Äußerungen, bei denen manche Wörter mit bedingten Regeln im Bezug auf die Ontologie des Systems ersetzt werden. Die Regeln stellen Templates dar, die auf das Arbeitsgedächtnis (WM - *working memory*) angewendet werden. Das Arbeitsgedächtnis ist der Ort, wo Tatsachen, Festlegungen und anderen Dialoginformationen gespeichert werden, um bearbeitet zu werden. Jede Regel kann Elemente aus dem WM entfernen, auswählen oder neue Elemente hinzufügen (Wahlster, 2006, S.197f).

3.1.2.3 Identifikation von Fehlern

Ein notorisches Problem stellt die Tatsache dar, dass der Input nicht immer korrekt ist - entweder wegen Fehler bei der Spracherkennung (7,9%-21,1% laut Wahlster, 2006, S.195) oder weil sich die Sprecherin versprochen hat. Weitere Ursachen sind die akustische Fehlinterpretation, der nicht verstandene Sprechstil, die Unterbrechungen, Begriffe, die nicht im System gespeichert sind, oder Fehler beim Parsen. Nach Wahlster gehen die meisten Fehler im Deutschen entweder aus falschen Flexionen oder aus komplett falschen Wörtern hervor. Um die Fehler im ersten Fall zu reduzieren, wird der Wortstamm verwendet und im zweiten Fall durch Kombination von Hinzufügen und Entfernen von Wörtern. Die Schwierigkeit besteht nun darin, zu erkennen, dass ein Fehler überhaupt aufgetreten ist. Ist ein Fehler festgestellt worden, so kann auch die Ursache ermittelt und eine Reaktion an die Benutzer gesendet werden, um den Fehler zu beheben.

Eine Anzahl von Systemen haben eine Art Wahrscheinlichkeitsstatistik eingebaut, um problematische Inputs zu erkennen. Das System wechselt nach der Identifizierung

eines Fehlers zu einer anderen Routine, um den Benutzerinnen weiterzuhelfen oder auf einen anderen, direkteren Dialogmodus zu wechseln, der den Benutzerinnen eindeutigere Antworten abverlangt. Erfahrungen mit JUPITER (Tab.: A.5 auf Seite A-3) haben gezeigt, dass die statistische Wahrscheinlichkeit hoch ist, dass auf einen vom System nicht verstandenen Input ein weiterer nichtverstandener folgt. Einige Benutzerinnen scheinen eine gewisse Tendenz zu haben, in einen Kreislauf von Fehlern zu kommen. Ein präziseres Feedback könnte bei solchen Fehlern Abhilfe schaffen (z.B. engere Einschränkung auf einen bestimmten Bereich des neuerlichen Inputs). Das Erkennen von Wortteilen, wenn möglich, könnte auch hilfreich beim Identifizieren von nicht gespeicherten Wörtern sein.

Die NLU Komponente z.B. im SMARTKOM System verfügt über WM und eine *Agenda* und führt Operationen auf diese zwei interne Strukturen aus. In diesen Strukturen werden verschiedene mögliche Alternativen gespeichert. Es kann mehrere WM's geben, die verschiedene Alternativen repräsentieren. In der Agenda werden die weniger wahrscheinlichen Alternativen gespeichert, um in einem späteren Zeitpunkt darauf zugreifen zu können.

Die höchstwahrscheinliche Alternative innerhalb der Agenda wird übernommen, dann werden die Templates in vordefinierter Reihenfolge angewendet. Eine Art Verfeinerung der Ergebnisse wird zusätzlich durchgeführt, wobei die Alternativen wieder in der Agenda gespeichert werden. Zusätzlich wird geprüft, ob optionale Bedingungen zutreffen und wenn es notwendig ist, wird auch Hintergrundwissen (unter anderem erkannte Gesten) konsultiert. Die Ergebnisse werden mit einer Wahrscheinlichkeit versehen und ausgegeben. Abbildungen wie auch Details zu der NLU Komponente des SMARTKOM Systems können in Wahlster (2006, S.195-207) gefunden werden.

3.1.3 Dialogmanager (DM)

Die Funktion des Dialogmanagers ist es, den Verlauf der Konversation zu kontrollieren. Daher hängt er sehr eng mit der Architektur eines jeden interaktiven Dialogsystems zusammen. Seine Tätigkeit deckt:

1. die Aufnahme des Benutzerinputs in Form von „Intentionen“ (bzw. formalen Gegenstücke zu Intentionen, die man den Benutzerinnen zuschreibt),
2. das Generieren von Strategien für die Umsetzung der Intentionen und für den weiteren Dialogablauf,
3. die Auswahl von Äußerungen, die für die Benutzerin verständlich sind (unter anderem Formulieren von Vorschlägen, die Mißverständnisse klären) und

4. die Interaktion mit unterschiedlichen internen und externen Ressourcen.

Aus diesem Grund wird auch angenommen, dass die Dialogstrategie eine Subkomponente des Dialogmanagers ist, was bedeutet, dass der DM die Dialogstrategie konsultiert, um Handlungen vorzunehmen (Cuayáhuitl, 2005, S.25f). Unter Dialogstrategie wird hier das verstanden, was die besten kommunikativen Ergebnisse in einer kommunikativen Situation liefert.

At each time step the dialogue strategy will aim to provide the actions that achieve the highest reward (or lowest punishment). (Cuayáhuitl, 2005, S.25)

Das Systemdesign bestimmt das Design des Dialogmanagers, denn der letztere steuert letztendlich die Konversation seitens der Maschine. Der Dialogmanager kann demnach bestimmte Dialogtypen abarbeiten und eine oder mehrere Dialogsituationen bewältigen, wobei mit dem Fortschreiten der Forschung und Entwicklung auf diesem Gebiet die Systeme immer „intelligenter“ werden und komplexere Dialogmodelle und Aufgaben bewältigt werden können. Es ist allerdings so, dass derzeit unterschiedliche Modelle bei der Suche nach einem optimalen Konzept für die Verarbeitung von Dialogphänomenen konkurrieren. Die verbreitetsten dieser Modelle werden näher in P. 3.2 auf Seite 71 beschrieben. Man ist darauf gekommen, dass sehr oft eine Mischung aus unterschiedlichen Verfahren und Modellen bessere Ergebnisse liefern kann.

3.1.4 Generierung gesprochener Sprache (NLG)

Wurde eine Äußerung vom System einmal verstanden, wird es notwendig, den Benutzerinnen eine Reaktion, z.B. eine angeforderte Information oder ein weiteres Rezipienzsignal zu geben, so dass die Benutzerin weiß, dass die Äußerung verstanden wurde. Für Systeme mit einem Bildschirm reicht es aus, diese Information in einer Tabelle auszugeben. Die Erfahrung hat jedoch gezeigt, dass es von großem Vorteil ist, irgendein linguistisches Feedback zu geben, damit die Benutzerinnen wissen, was die Maschine verstanden hat und was sie gerade tut. Dadurch wird es leichter, einen Fehler seitens der Maschine zu orten und auf diesen zu reagieren. Das Feedback ist gewöhnlich eine knappe Zusammenfassung und ergänzt oft die Information der ausgegebenen Tabelle, wie z.B. „Ihre Suche nach X liefert folgende Ergebnisse“.

Der Aspekt der Sprachgenerierung wird wesentlich schwieriger, wenn die Maschine die gesamte Information via Sprachausgabe zu übermitteln hat. So kann zum Beispiel keine lange Liste von Flügen aufgelistet werden, ohne die Benutzerin mit Details zu überfordern. Daraus ergibt sich, dass ein nicht unerheblicher Teil der Generierung

mit Entscheidungen zu tun hat, wieviel, bzw. welche Information in welcher Portionierung und in welcher Abfolge übermittelt werden soll in Anlehnung an die Probleme die im Kap. 2 angesprochen wurden. Walker et al. (1998a); Walker (1998) schlagen eine Theorie, *centering* genannt, vor, um brauchbare Formalismen für die Kohärenz, die pragmatische Ausrichtung der Äußerungen und potentielle Referenzen zu gewinnen.

Centering is formulated as a theory that relates focus of attention, choice of referring expression, and perceived coherence of utterances, within a discourse segment. (Walker, 1998)

Beobachtungen haben gezeigt, dass die genaue Formulierung bzw. Einschränkung der Antwortmöglichkeiten (z.B. ja/nein-Fragen der Maschine) einen großen Einfluss auf den Input der Benutzerinnen hat: je ungenauer, desto größer wird die Variation der Benutzerreaktionen. Wie sehr dies erwünscht ist, hängt von der Ausrichtung des Dialogmanagers ab. Wenn es das Ziel ist, eine große Bandbreite an Antworten (benutzerseitig) zu Lernzwecken auszuwerten, dann ist eine ungenaue Formulierung passend. Allerdings würde meistens eine zielgerichtete Benutzerreaktion kurzfristig die Leistung steigern.

3.1.5 Sprachsynthese (TTS)

In Abhängigkeit von der Natur des Feedbacks müssen Systeme eine Form der Sprachsynthese für die Benutzerinnen durchführen - vor allem bei Telefonsystemen. Die Formen der Sprachsynthese unterscheiden sich deutlich: beginnend mit aufgezeichneten Phrasen über Wort- und Phrasenverkettungen bis hin zu *general purpose synthesizer*. Letztere haben die Eigenschaft, unnatürlich zu wirken, denn bei der Synthese gehen verschiedene Aspekte der natürlichen Sprache verloren. Doch gerade die Sprachsynthese, oder besser gesagt der Sprachoutput des Systems ist es, was den bleibenden Eindruck bei den Benutzerinnen - vor allem auch wegen der Unnatürlichkeit - hinterlässt. Aus diesem Grund war die Sprachgenerierung bei den meisten Systemen für gesprochene Dialoge eher statisch und es wurden festgelegte Muster zur Erwidern der Benutzeranfragen verwendet. Eine gewisse Variation der festgelegten Muster konnte den Benutzerinnen den Eindruck der Natürlichkeit vermitteln und die Systeme weniger roboterartig erscheinen lassen. Man wird sehen, ob das Einführen einer verstärkt stochastisch generierten Sprache bei den Benutzerinnen gut ankommen wird, was das eigentliche Ziel von neueren Systemen ist. Variationen der Prosodie des Maschinenoutputs wären eine gute Möglichkeit, die Monotonie der Computerstimme zu bekämpfen.

Obwohl die Prosodie sowohl das Sprachverständnis als auch die Sprachgenerierung beeinflusst, ist sie bis jetzt am meisten in TTS aufgenommen worden, um wenigstens die Natürlichkeit der generierten Sprache zu verbessern. Es gibt jedoch Versuche,

die Information aus der Prosodie sowohl für Spracherkennung als auch für Sprachverständnis nutzbar zu machen, die hier aber nicht behandelt werden können.

Im Projekt „Verbmobile“ (Tab.: A.49 auf Seite A-22) konnte gezeigt werden, dass die Verarbeitungsgeschwindigkeit rapide gesteigert wird, wenn die Information aus der Prosodie in die Suchfunktion aufgenommen und weiterverarbeitet wird (Pierrehumbert, 1993). Neuere Systeme wie SMARTKOM machen einen Gebrauch davon und verwenden ein Modul zur Abarbeitung von Prosodie sowohl am Input als auch am Output (siehe Wahlster, 2006, S.139-151). Die Kontrolle über die Generierung der prosodischen Merkmale obliegt dem Dialog-Manager. Es bleibt aber eine philosophische Frage, ob es Sinn ergibt, den Dialog-Manager zu personalisieren. Die Meinungen sind geteilt. In manchen Situationen erscheinen gewisse Reaktionen nur in zwischenmenschlicher Konversation sinnvoll (Zum Beispiel: „Es tut mir Leid“), in anderen Fällen, wie schon erwähnt („Wo hast du die Datei X gespeichert“), sind solche Fragen durchaus transferierbar. Allerdings hat eine persönliche Note der Dialogmanager noch nie eine negative Reaktion seitens der Benutzer hervorgerufen (Glass, 1999).

3.1.6 Fakultative Systemkomponente

Sogar einfache Anwendungen verlangen anspruchsvolle Komponenten für erfolgreiche Konversation unter realen Bedingungen. Robuste ASR, NLU und DM können in der Praxis die Performanz eines Systems wesentlich steigern und eine TTS, die mit Algorithmen für Prosodie ausgestattet ist, kann den Eindruck bei der Benutzerin deutlich verbessern. Komplexere Systeme werden im allgemeinen weitere, auf den DM abgestimmte Mechanismen zur Bewältigung lokaler Diskursphänomene einbauen und weiters eine Form der semantischen Repräsentation unterstützen. In neuesten Systemdesigns werden zusätzlich Ontologien benutzt, um Wissen in einem bestimmten Bereich (*domain*) einzugrenzen und Ambiguitäten und Referenzen aufzulösen (siehe Milward und Beveridge, 2003; Wahlster (2006, S.71-84)).

Zu den fakultativen Systemkomponenten zählen unter anderem Komponenten für die Erkennung von Gesten und Mimiken und ihre Verarbeitung, auf die hier nicht näher eingegangen werden soll (siehe Wahlster, 2006, S.153-193). Die dadurch gewonnene Information bereichert die Interpretation der Äußerungen und erhöht die Erkennung der Benutzerintention. Vor allem die Erkennung von Gesten kann beim Verstehen und Interpretieren von deixischen Phänomenen sehr nützlich sein, wie z.B. „Ich möchte **diesen** Film aufnehmen.“ (begleitet mit einem Zeigen auf dem am Bildschirm dargestellten Fernsehprogramm).

Sehr oft wird eine Komponente für die Abarbeitung von Diskursphänomenen hinzu-

gefügt (engl.: *discourse manager*), um die Erkennung der Benutzerintentionen zu verbessern (siehe Wahlster, 2006, S.237-299), zusätzlich wird auch oft eine Komponente für die Kontextidentifikation benutzt, um das Auflösen von Ambiguitäten zu unterstützen. In multimodalen System wird eine Komponente für die Steuerung des Betriebs in verschiedenen Modi verwendet, wobei für jeden Modus die Systemintentionen übersetzt und entsprechend ausgeführt werden. Die Wahl der fakultativen Komponenten hängt daher vom Systemdesign ab.

3.2 Verarbeitung von Dialogphänomenen

Im Unterschied zu der Mensch-Mensch-Kommunikation spielt der Aufwand (die Zeit und die Kosten/Nutzen u.ä.) für die Herstellung eines Dialogsystems eine wesentliche Rolle und das führt zu gewissen Überlegungen im Bezug auf die gesamte Komplexität des Systems - auf die Zielsetzung und auf die Umsetzungsmöglichkeit. Die wichtigste und entscheidendste Design-Frage eines Dialogsystems ist die Frage nach der Struktur des Dialogmanagers, denn diese Komponente kontrolliert den Ablauf der Konversation und koordiniert die Arbeit der restlichen Komponenten. Dabei zeichnen sich folgende Überlegungen der Systemgestaltung aus, die bei allen Systemen zumindest in Erwägung gezogen werden:

▷ die Initiative

Die an einem bestimmten Punkt des Dialoges leitende Dialogteilnehmerin übernimmt an diesem Punkt die Initiative und entscheidet, worüber als nächstes gesprochen wird. Daher auch die Frage, wie weit der Benutzerin im Sinne einer Dialogpartnerin die Möglichkeit überlassen wird, den Dialog zu führen. Ein zentrales Handlungsmuster der Dialogsteuerung sind Fragen. Im engeren Sinne betrifft es daher die Entscheidung, ob die Benutzerin Fragen stellen darf, denn die Fragen bestimmen auch das Thema oder Subthema der Konversation und fordern bestimmtes Dialogwissen, um die richtige Antwort für den nächsten Zug zu generieren. Die Statistik zeigt (vgl. Bernsen et al., 1994; Levin et al., 2005, S.4), dass je mehr die Benutzerin die Initiative übernehmen darf, desto höher auch das Fehlrisiko wird, bzw. desto komplexer die Systemarchitektur, um dieses Risiko zu minimieren. Dementsprechend wird entweder die Initiative dem System überlassen oder für die Bewältigung von bestimmten Situationen eine gemischte Initiative (*mixed initiative*) gewählt. Die Frage nach der Initiative ist eine zentrale Frage in dem Konzept eines jeden Dialogsystems und wird näher und ausführlicher im Punkt 3.2.2 auf Seite 85 behandelt.

▷ **Voraussagen** (*predictions*)

Voraussagen sind Erwartungen hinsichtlich des Nachfolgenden, von den Benutzerinnen Gesprochenen. Hier werden damit Voraussagen auf Dialogebene gemeint im Unterschied zu Voraussagen bei ASR (phonetisch-phonologische Ebene) (vgl. Blaylock und Allen, 2004). Die Voraussagen helfen, das Sub-Vokabular und die Sub-Grammatiken zu bestimmen, die je nach behandeltem Thema verwendet werden sollen. In komplexeren Systemen, die über die Fähigkeit verfügen mehrere unterschiedliche sprachliche Handlungen abzuarbeiten, ist es notwendig (meistens pro sprachliche Handlung) das Vokabular und die Grammatik einzuschränken, um die Leistung zu steigern. Manchmal (in kleinen geschlossenen Domänen, mit eher kurzen Dialogen) ist es überhaupt nicht notwendig Voraussagen zu formulieren und zu überprüfen.

Der Grundgedanke hinter *predictions* ist, dass wenn eine Äußerung phonetisch von dem ASR-Modul und inhaltlich von der NLU-Einheit erkannt wurde, diese auch klassifiziert wird, mit dem Ziel eine konversationelle Funktion im Sinne eines Sprechakts, der dem Benutzerinput entspricht, vorauszusagen und zu bestimmen. Die Systemantwort wird anhand des vorausgesagten Sprechakts generiert.

▷ **Systemfokus**

Ein anderer Aspekt ist der Systemfokus (vgl. Bernsen et al., 1994). In Aufgabenorientierten Systemen, was die meisten auch sind, ist der Systemfokus der Subset von Unteraufgaben, worauf sich die Benutzerinnen voraussichtlich in der nächsten Äußerung beziehen werden. Die vorher erwähnten Voraussagen basieren auf dem Subset der Unteraufgaben, der sich gerade im Systemfokus befindet.

▷ **Back-Channel und System-Feedback**

Forscher haben auch begonnen, die Back-Channel-Kommunikation (die Verwendung von Rezipienzsignalen) zu studieren, um die Interaktion natürlicher zu gestalten. Es scheint, als ob das *back-channelling* in einer gewissen Beziehung zur Grundfrequenz der Stimme der Sprecherin steht, wie Auswertungen der prosodischen Information gezeigt haben (Pierrehumbert, 1993). Dieser Aspekt spielt eine besondere Rolle bei der Identifikation von sprachlichen Handlungen z.B. bei deklarativen Fragen wie „Sie heißen X?“ (siehe Pierrehumbert, 1993, S.276) und bei der Auswertung der Benutzersicherheit in argumentativen oder Tutoring-Systemen. Wenn man unsicher ist, spricht man leiser, bzw. sind gewisse Abweichungen von der Grundfrequenz der Sprecherin festzustellen.

Das System-Feedback (vor allem explizit), ist eine durch das System getätigte

Wiederholung der von den Benutzerinnen bei den vorhergehenden Sprecherwechseln gegebenen Schlüsselinformation. Unmittelbares Feedback des Systems kann außerdem informativer sein als Pausen, wenn sich unvermeidbare Verzögerungen ergeben (zum Beispiel: „Warten Sie bitte, während ich den billigsten Flug nach Wien suche...“).

Das Bereitstellen von ausreichend Feedback an die Benutzerinnen ist bei Benutzer-unabhängigen Systemen besonders entscheidend, denn Missverständnisse beim Benutzer-Input sind bei diesen Systemen besonders häufig. Fortlaufendes Feedback kann reichen von einer Wiederholung (auch versteckte) bis hin zu einer expliziten Aufforderung an die Benutzerinnen, Input zu bestätigen oder zu wiederholen bzw. zu sichern. Diese Art von Feedback ermöglicht es, auf direkte und indirekte Weise, die Schlüsselinformation der von den Benutzerinnen getätigten Aussagen zu wiederholen. Sprecher-unabhängige Systeme bieten ständiges Feedback (Wiederholungen und implizite Wiederholungen) bezüglich der Entscheidungen der Benutzerinnen. Diejenigen, die das Feedback akzeptieren, müssen nicht erneut ihre Entscheidungen wiederholen, da das System mit der nächsten Unteraufgabe innerhalb der gleichen Äußerung fortsetzt. Eine ausgeklügeltere Methode, Feedback bestehend aus Wiederholungen mit solchen bestehend aus ausdrücklichen Bestätigungsaufforderungen zu verbinden, verwendet akustische Auswertungen als Entscheidungsgrundlage für die Art des Feedback, wie von Bilange (1991) vorgeschlagen. Wenn der akustische Input unter eine bestimmte Schwelle sinkt, was Unsicherheit ausdrücken kann, wird das Feedback aus einer ausdrücklichen Bestätigungsaufforderung bestehen. Akzeptieren die Benutzerinnen dieses Feedback nicht, braucht man Meta-Kommunikation.

▷ **Dialogverlauf**

In den meisten Systemen wird der Dialogverlauf (*dialogue history*) in mehreren Aspekten bis zum Ende der Konversation verfolgt und gespeichert wie auch der linguistische Dialogverlauf (*linguistic dialogue history*), der die tatsächlich gesprochenen Wörter, sowie deren Reihenfolge aufzeichnet. Der Dialogaktverlauf zeichnet die Reihenfolge der Dialogakte und deren semantische Inhalte auf. Die Aufgabenliste (*task record*) registriert die für die Aufgaben relevanten Informationen, die im Laufe des Dialoges vermittelt wurden, entweder alle von ihnen oder nur die, welche vom System bzw. nur die, welche von den Benutzerinnen kommen, wie es die Anwendung eben erfordert. Die Erfolgsquotenliste (*performance record*) berichtigt laufend ein Modell des Dialoges zur Überprüfung, ob der Dialog auch gut voranschreitet und kann die weitere Vorgehensweise des Systems beeinflussen

(siehe Bernsen et al., 1994).

▷ **Sprecher-Modelle**

In einem Mensch-Mensch-Dialog entwickeln die Dialogteilnehmer Sprecher-Modelle, um das Dialogverhalten anzupassen. Manchmal bringen die Dialogteilnehmerinnen solch ein Modell schon vorher mit. Zum Beispiel hat ein Zahnarzt bestimmte Erwartungen wie ein Dialog mit der Patientin verlaufen könnte und die Patientin ihrerseits genauso. Daher hat das Phänomen sehr früh die Aufmerksamkeit der Forschung auf sich gelenkt und man versucht Sprecher-Modelle für die Systeme zu entwickeln.

A user model is a knowledge source in a natural-language dialog system which contains explicit assumptions on all aspects of the user that may be relevant to the dialog behaviour of the system. These assumptions must be separable by the system from the rest of the system's knowledge.

A user modeling component is that part of the system whose function is to incrementally construct a user model; to store, update and delete entries; to maintain the consistency of the model; and to supply other components of the system with assumptions about the user. [...] With the definition given above we propose to restrict the term 'user model' to assumptions about aspects of the human directly interacting with the system. (Kobsa und Wahlster, 1989, P.1.2)

Interaktive Dialogsysteme brauchen solch ein Sprecher-Modell, um das Dialogverhalten anzupassen, oder müssen solch ein Modell erstellen und in Evidenz halten, wenn Interaktion mit derselben Benutzerin erforderlich ist (siehe Bernsen et al., 1994; Kobsa und Wahlster, 1989).

▷ **Unterbrechen (*barge-in*)**

Viele Systeme können z.B. Unterbrechungen verarbeiten, indem sie eine sogenannte „Barge-In“ Funktion enthalten; ist es zu einer Verzögerung gekommen, so kann der Benutzer das System übersprechen. Hat das System die Verzögerung erkannt (und das ist sowohl schwierig wie auch essentiell, um den Dialog sinnvoll weiterzuführen), kann es feststellen, wann die Unterbrechung stattgefunden hat und beginnt mit einem Update, um den Dialog wieder dort aufzunehmen, wo er unterbrochen wurde. Dann wird die erste Äußerung nach der Unterbrechung ausgewertet und der Dialog kann sinnvoll weitergeführt werden. Dafür muss das

Mikrofon die ganze Zeit ein Signal empfangen, was meistens zum Entstehen des s.g. „Echo-Effekts“ führt. Daher ist das Implementieren von einer *barge-in* Funktion immer an einer *echo cancellation* Funktion gebunden. Die letztere erlaubt das Herausfiltern des User-Inputs aus dem gesamten Audiosignal und wird als Vorstufe bei der Spracherkennung implementiert (siehe dazu Wahlster, 2006, S.90ff).

▷ **Integrieren multipler Modalitäten** (multimodale Dialogsysteme)

Systeme für gesprochene Dialoge stellen den Benutzern recht unterschiedliche Input- und Output-Modalitäten zur Verfügung. In Systemumgebungen ohne Display kann es notwendig sein, den Dialog so zu gestalten, dass die Benutzerinnen nicht von dem Angebot an Information überwältigt werden. Ist ein Display aber verfügbar, so ist es erstrebenswert, die Information in Tabellen zusammenzufassen und sie so den Benutzerinnen (und somit in verschiedenen Dialogen) gegliedert zugänglich zu machen. Sollten Input mittels Stift oder Handbewegungen möglich sein, so ist in der selben Weise vorzugehen. Welche der Modalitäten am effektivsten ist, wird von der spezifischen Verwendung und den Vorlieben der Benutzerinnen abhängen. In den letzten Jahren versucht man allerdings unterschiedliche Systemkomponenten flexibel und portierbar zu gestalten. Das erlaubt die Möglichkeit, Komponenten zwischen den Systemen auszutauschen, so dass ein und dasselbe System in unterschiedlichen Modalitäten arbeiten kann. In diesem Zusammenhang spricht man von *modality fusion* und *modality fission* (Wahlster, 2006, S.223ff;S.379ff). Die Fusion (*fusion*) stellt den Prozess der Zusammenfassung des Inputs aller dem System zur Verfügung gestellten Eingabegeräte dar und die Spaltung (*fission*) - den entgegengesetzten Prozess - die Verteilung der kommunikativen Aufgaben auf die unterschiedlichen Ausgabegeräte, die zur Verfügung stehen.

▷ **Nachahmung von Mensch-Maschine-Dialog**

Eine andere offene Frage für Dialogsystem-Designer ist es, inwieweit der Mensch-Maschine-Dialog dem beobachteten Mensch-Mensch-Dialog nachempfunden werden sollte. Das Studium des letzteren ist für sich ein aktives Feld der Forschung. Viele der Konversationsmerkmale, die im Kapitel 2 beschrieben wurden, können nicht so einfach interpretiert und verstanden werden, es sei denn, man kennt den genauen Kontext und hat das notwendige Hintergrundwissen, um auf die richtige Bedeutung zu schließen. Es hat sich herausgestellt, dass das beste Verfahren ein stochastisches auf Basis vorhandener Daten ist. Viele grosse Korpora wurden für diesen Zweck erstellt und analysiert (Allen, 1995; Carletta et al., 1997). Durch

stochastische Methoden ist es dann möglich, die Bedeutung anhand vom Kontext mit einer höheren Wahrscheinlichkeit zu bestimmen.

▷ **Kompetenz der Benutzerinnen**

Eines der schwierigeren Probleme beim Kreieren von Systemen für gesprochenen Dialog ist es, die Fähigkeiten des Systems den Benutzern zu vermitteln. Das beinhaltet sowohl die Art des Sprach-Inputs, der vom System verarbeitet werden kann, als auch den Bereich der Systemkenntnis selbst. Experten können oft recht gut zumindest mit einem Teil des Systems umgehen, auf Neulinge trifft dies allerdings nicht zu. Diese haben dann auch oft beträchtliche Schwierigkeiten, das System zu benutzen, wenn die Erwartungen an das System nicht mit den tatsächlichen Fähigkeiten desselben übereinstimmen.

Tabelle 3.1 zeigt die Statistik von aufgezeichneten Mensch-Mensch-Dialogen aus dem Bereich der Filminformation (Flammia, 1998). Man sieht, dass beinahe die Hälfte der Kundendialoge nur Rezipienzsignale sind (wie zum Beispiel: ja, gut, aha, usw.)². Die Statistik zeigt auch, dass die Kundenanfragen nicht sehr lang sind. Die Untersuchung von Flammia (1998) über 100 Stunden Kundengesprächen aus mehr als 1000 Mensch-Mensch-Dialogen im Bereich verschiedener Informationsdienste zeigte, dass über 80% der Kundenäußerungen aus weniger als 12 Worten bestand, und mehr als die Hälfte aus vier oder weniger.

	KUNDE		CALL CENTER AGENT	
Handlung	Frequenz	Wörter	Frequenz	Wörter
Rezipienzsignale	47,9	2,3	30,8	3,1
Anfrage	29,5	9,0	15,0	12,3
Bestätigung	13,1	5,3	11,3	6,4
Information	5,9	7,9	27,8	12,7
Aussage	3,4	6,9	15,0	6,7

Tabelle 3.1: Statistik von Mensch-Mensch Konversationen aus dem Bereich der Filminformation (Glass, 1999)

Ein weitere Überlegung bei der Systemumsetzung ist, ob sich das aktive Vokabular und die Bandbreite der verstandenen Antworten abhängig vom Verlauf des Dialoges ändern soll. Einige Systeme sind so strukturiert, dass die Benutzerinnen sämtliche Fragen an jeder Stelle des Dialoges stellen können, so dass das gesamte Vokabular die ganze Zeit aktiv ist. Dagegen schränken andere Systeme das Vokabular bzw. die Sprache ein, welche an einer bestimmten Stelle im Dialog verwendet werden kann. Man

²Ein durchschnittlicher Dialog bestand dabei aus über 28 Dialogzüge zwischen Kunden und Call Center Agents.

tauscht also Flexibilität (der Reaktion auf Fragen und Antworten der Benutzerin) gegen eine erhöhte Genauigkeit ein, die durch die eingeschränkten Input-Möglichkeiten der Benutzerin erreicht wird.

3.2.1 Designstrategien

Die Fähigkeit, gewöhnlich vorkommende, der menschlichen Kommunikation eigene Phänomene verarbeiten zu können, würde Dialogsysteme natürlicher erscheinen lassen und sie wären für Menschen leichter zu benutzen. Dies würde aber die Systeme komplexer und fehleranfälliger machen. Das ist auch der Grund, warum es so unterschiedliche Strategien zur Entwicklung der Systeme für gesprochene Dialoge gibt. Die Entscheidungen betreffen vor allem das Design des Dialogmanagers, beeinflussen aber auch die Art und Weise wie die Komponente miteinander arbeiten und welche Informationen noch intern ausgetauscht werden.

Für die Entwicklung von DM-Architekturen werden vor allem die Zustand-orientierte (*state-based*), die Frame-orientierte (*frame-based*), die Plan-orientierte (*plan-based*), die Aufgaben-orientierte (*task-oriented*), die Agenten-orientierte (*agent-based*) und die Informationszustand - orientierte Architektur (*information-state*) verwendet.

▷ Zustand-orientierte Architektur

In der Zustand-orientierten (*state-based*) Architektur wird die Dialogstruktur in Form eines Netzwerks dargestellt, wo jeder Knoten eine Frage darstellt und die Übergänge (*transition*) zwischen den Knoten alle möglichen Dialoge darstellen (McTear, 1998). Die Architektur entspricht einem s.g. endliche Automaten (EA) (engl.: *finit state*) und erlaubt den Benutzern nur die Möglichkeit Antworten zu geben (siehe unten P. „Systemgesteuerter Dialog“ auf Seite 86).

▷ Dialoggrammatiken und Frames

Dialoggrammatiken sind Systeme, die lokale oder globale Dialog- und Sprechaktmerkmale (Searle, 1969) der Oberflächenstruktur, wie auch die darauf folgenden Reaktionen und Antworten identifizieren und abbilden können.

Sie verwenden eine preskriptive Grammatik für Musteridentifikation von Dialogsequenzen. Die ersten entwickelten Grammatiken beschrieben die Struktur des gesamten Dialogs von Anfang bis Ende. Spätere Ansätze machen Gebrauch von den Untersuchungen von Jefferson (1972) über *adjacency pairs*. Exkurse und Fehlerkorrekturen werden daher durch eingebettete Sequenzen abgebildet.

Für die Umsetzung von Dialoggrammatiken werden Phrasenstrukturgrammatiken und einige endliche Automaten (EA) verwendet. Das SUNDIAL System, zum

Beispiel, verwendet Dialoggrammatik, um eine Reisekonversation zu bewältigen. Trotzdem wurden Dialoggrammatiken wegen der mangelnden Flexibilität kritisiert, sowohl im Bezug auf Variationen und Abweichungen im Dialog, als auch im Bezug auf die Portabilität zu anderen Domänen.

Die Frame-orientierte Architektur ist eine bemerkenswerte Erweiterung von Dialoggrammatiken und basiert auf Frames³, welche von der Benutzerin ausgefüllt werden müssen. Dabei enthält jeder Frame die s.g. Slots, die die Benutzerin durch den Dialog führen. Die Idee ist, dadurch einen ganzen Dialogakt abzudecken (Lemon et al., 2001). Bei diesem Modell hat die Benutzerin die Möglichkeit die Initiative zu übernehmen (siehe auch Goddeau et al., 1996, Chu-Carroll und Carberry, 2000), d.h. beliebige Slots in beliebiger Reihenfolge zu befüllen. Zum Beispiel kann eine Benutzerin in unterschiedlichen Dialogzügen Informationen zu ihrem Reiseziel, Abfahrt- oder Ankunftszeit, etc. übermitteln. Das Ermitteln der Informationen wird durch Fragen eingeleitet, die aber in beliebiger Reihenfolge gestellt werden können. Sobald die Informationen zu dem betreffenden Frame gesammelt werden, geht die Verarbeitung weiter. Angenommen die Benutzerin will von Berlin nach Wien am 21. September fliegen und in Wien um 10 Uhr morgens sein. Da durch dieses Verfahren der Benutzerin die Möglichkeit überlassen wird, die Konversation zu führen, wird der Dialog mit einer Standardfrage eröffnet:

0) System: Wie kann ich Ihnen helfen?

Darauf kann die Benutzerin beliebige Antworten in diesem Kontext geben, wie:

1a) User: Ich möchte nach Wien am 21. September fliegen.

1b) User: Ich möchte in Wien am 21. September um 10 Uhr morgens ankommen.

1c) User: Ich möchte von Berlin nach Wien fliegen.

Mögliche Systemfragen wären dann:

2a) System: Um wieviel Uhr möchten Sie in Wien landen?

2b) System: Von wo möchten Sie nach Wien fliegen?

2c) System: Wann möchten Sie nach Wien fliegen?

³Frames (engl. für: Rahmen) sind Konstrukte zur Wissensrepräsentation, die komplementär zur Repräsentation von Wissen mittels Logik (zum Beispiel Prädikatenlogik) sind. Als Erfinder der Frames gilt Marvin Minsky. Man kann sie sich im Wesentlichen als Objekte ohne Methoden vorstellen, d. h. Frames sind in einer natürlichen und hierarchischen Vererbungsstruktur eingebettet und besitzen Attribute, sogenannte Slots (etwa: Schubfächer). Übergeordnete Frames können dabei ihre aktuellen Slotwerte, die sogenannten Filler, an untergeordnete Frames vererben. Eng verwandt mit den nicht nicht-Logik-basierten Frames sind die semantischen Netze. (http://de.wikipedia.org/wiki/Frames_%28Wissensrepr%C3%A4sentation%29)

So wird eine bestimmte Flexibilität in der Kommunikation erreicht und dadurch wird der Dialog menschenähnlicher gestaltet. Jeder Dialogzug, bzw. jede Systemfrage zielt darauf, einen Slot zu befüllen. Im Falle, dass es zu Mißverständnissen kommt, welche nicht gelöst werden können, sind manche Systeme so konzipiert, dass sie auf Systeminitiative-Modus (siehe P.3.2.2.2 „Systemgesteuerter Dialog“ auf Seite 86) umschalten und anfangen, elementare Fragen zu stellen, die mit „ja“ oder „nein“ zu beantworten sind.

Die Agenda-orientierte (*agenda-oriented*) Architektur basiert auf der Frame-orientierten Architektur, baut allerdings komplexere Strukturen auf (durch die Anwendung dynamischer Bäume (*dynamic trees*)) und ermöglicht dadurch die Konversation in komplexeren Domänen (Xu und Rudnicky, 2000).

▷ **Plan-orientierte Architektur und kollaborative Systeme**

Die Plan-orientierte Architektur beruht auf der Idee, dass Menschen kommunizieren, indem sie Ziele (*goals*) erreichen und dass bei dieser Kommunikation sich der mentale Status (Informationszustand) der Hörerin ändert. In der Plan-orientierten Theorie der kommunikativen sprachlichen Handlungen und des Dialogs (Cohen und Perrault, 1979; Cohen und Levesque, 1990; Cohen, 1992; Allen und Perrault, 1980) werden die Äußerungen als sprachliche Handlungen (Searle, 1969) betrachtet, die dazu dienen, die vorher definierten Ziele zu erreichen (siehe P.2.5.1 „Pläne und sprachliche Handlungen“ auf Seite 42). Die Hörerin muss den primären Plan der Sprecherin erkennen und entsprechend reagieren (handeln). Zum Beispiel auf die Frage „Wo sind die Kekse aus der Werbung?“ antwortet der Verkäufer „Wieviele möchten Sie denn?“ ganz korrekt, weil er den primären Plan erkannt hat, nämlich die Kekse zu kaufen.

Die Plan-orientierte Architektur versucht diese Idee in die Praxis umzusetzen und explizit die (globalen) Ziele abzubilden. Das SMARTKOM System (Wahlster, 2006, S.302f) verwendet eine Variante der Plan-orientierten Architektur und Elemente der Aufgaben-orientierten Architektur (siehe unten). In der Hintergrundtheorie wird der Bezug zu Intentionen deutlicher:

Each application requires knowledge sources defining the set of plan operators to specify the services and the available dialogues of the application. (Wahlster, 2006, S.307)

In most cases, user utterances must be integrated into an overall plan of ongoing dialogue. The dialogue manager maintains a stack of goals that have been adopted during the dialogue. At any time one goal is

focused for the planner, the *current goal*. A plan is created and then pursued that specifies a possible course of subsequent communicative acts to reach the goal. If the user fully cooperates with the system the originally generated plan will be followed. However, frequently the user will make contributions that are not intended by the plan, or he will provide information in an order different from the planned sequence. In this case, the plan has to be revised and modified to account for changed circumstances, and to continue the dialogue. (Wahlster, 2006, S.311)

Der Plan-orientierte Ansatz, wie in SMARTKOM umgesetzt, gibt eine flexible Möglichkeit, Regeln je nach Kontext zu laden, die nächste sprachliche Handlung zu planen und durchzuführen.

▷ **Aufgaben-orientierte Architektur**

Die konversationelle Spieltheorie, die auf den Theorien von Carlson (1983) und Carletta et al. (1997) beruht, verwendet Verfahren sowohl aus dem Bereich der Diskursgrammatik als auch aus der Plan-orientierten Architektur, indem sie ein Ziel- oder Plan-orientiertes Verfahren in dem strukturellen Ansatz einbettet. Dadurch kann sie zum Abbilden der Mensch-Maschine-Konversation in einem Aufgaben-orientierten (*task-oriented*) Dialog verwendet werden. Ein Aufgaben-orientierter Dialog besteht aus ein oder mehreren Transaktionen, wo jede Transaktion eine Unteraufgabe repräsentiert. Jede Transaktion erfasst eine Zahl von Dialogspielen und jedes Spiel besteht aus einem Eröffnungszug (*initial move*) und aus einem (manchmal optionalen) schließenden Zug (*closing*). Ein Beispiel dafür ist ein Anweisungsspiel (*INSTRUCTION game*), welches aus 3 eingebetteten Spielen besteht: Erläuterungsspiel (*EXPLAINING game*), Ja-Nein-Fragespiel (*QUERY-YN game*) und Überprüfungsspiel (*CHECKING game*). Das Überprüfungsspiel kann z.B. aus einer Ja-Nein-Frage und einer Ja- oder Nein-Antwort bestehen. Das Aufgaben-orientierte Verfahren wurde in dem RavenClaw System (Tab. A.60 auf Seite A-26) implementiert und somit in der Architektur unterschiedlicher Anwendungen integriert (LARRI A.32, Vera A.31, ConQuest A.15 u.a.).

Das Verfahren eignet sich zum Abbilden von Subdialogen (*side sequences*), Erklärungen, Erläuterungen etc., indem es einem Spiel erlaubt aus anderen eingebetteten Spielen zu bestehen - eine Technik, welche sich für die Abbildung der Komplexität von Mensch-Mensch-Dialoge sehr gut eignet. Das zeigt auch deutlich, dass es keine feste Grenze zwischen diesem und dem Frame-orientierten

Ansatz gibt, denn Pläne können als Frames seit der Veröffentlichung der Plan-Skripten-Theorie von Schank (1982), abgebildet werden.

Der kollaborative Ansatz ist eine Variante des Aufgaben-orientierten Ansatzes und betrachtet den Dialog als einen kollaborativen Prozess. Beide Gesprächspartnerinnen, arbeiten gemeinsam, um ein beidseitiges Verständnis von dem Dialog zu bekommen. Die Erkenntnis, dass dieser Prozess auf beiden Seiten statt findet, gibt eine brauchbare Erklärung für Diskursphänomene wie Bestätigungen oder Erklärungen, die für die Mensch-Mensch-Kommunikation typisch sind. Der kollaborative Ansatz versucht eher die Intentionen hinter dem Dialog und die Dialogmechanismen zu erklären, anstatt sich auf die Struktur der Aufgabe zu konzentrieren. Dadurch wäre die Überzeugung (*belief*) von wenigstens zwei Dialogteilnehmerinnen explizit modelliert. Ein Ziel, welches von der(n) anderen Dialogteilnehmerin(en) akzeptiert wurde, wird zu einem Teil der gemeinsamen Überzeugung und die Dialogteilnehmerinnen werden in Kooperation arbeiten, um dieses Ziel zu erreichen.

Eine zwischen dem System und den Benutzerinnen interaktiv zu erledigende Aufgabe besteht aus einer oder mehreren Unteraufgaben. Aufgaben können in andere Aufgaben eingebettet sein und dadurch mit den Unteraufgaben in Beziehung stehen. Eine Aufgabe wird erledigt, indem die Unteraufgaben erfolgreich erledigt werden. Bei Systemen für gesprochene Dialoge beinhaltet jede Aufgabe mindestens einen Dialogzug, besteht aber oft aus mehreren Dialogzügen. Ein Dialogzug entspricht einer Äußerung, die ein oder mehrere Dialogakte, wie zum Beispiel Bestätigungen oder Fragen beinhalten. Aufgaben, die von systemgesteuerten Dialogen abgewickelt werden, haben eine stereotype Struktur, die vorschreibt, welche Information zwischen den Dialogpartnerinnen zur Aufgabenbewältigung ausgetauscht werden muss und darüber hinaus auch noch die Reihenfolge wie dies zu erfolgen hat, um ein natürliches Ergebnis zu erzielen.

In diesem Zusammenhang hat man den Terminus *ill-structured* (komplexe) Aufgaben eingeführt. Diese Aufgaben repräsentieren Probleme der realen Welt und enthalten typischerweise eine große Anzahl von optionalen Unteraufgaben und sind daher für systemgesteuerte Dialoge eher nicht geeignet. Zu wissen, dass die Benutzerinnen Reiseinformationen wollen, hilft nicht bei der Entscheidung, was und in welcher Reihenfolge es zu geschehen hat. In solchen Fällen ist es bis zu einem gewissen Grad notwendig, auf die gemischte Initiative als Ansatz zurückzugreifen, um ein akzeptables Mindestmaß von Natürlichkeit zu erlauben. Die Klasse der nicht-stereotypen Aufgaben scheint dabei ziemlich groß zu sein. Es kommt sogar vor, dass das System und die Benutzerinnen nicht einmal das

Wissen bezüglich der stereotypen Aufgabe teilen müssen, und zwar in den Fällen wo

1. dem System die Situation der Benutzerin ausreichend bekannt ist und
2. die Benutzerin beim System das bezüglich der stereotypen Aufgabe notwendige Wissen voraussetzt.

Das birgt eine beträchtliche Anzahl von Unteraufgaben-Typen im Rahmen eines systemgesteuerten Dialoges und beinhaltet auch viele Aufgaben bei denen die Benutzerin entweder neu oder nur wenig Erfahrung hat und das System zugleich Anweisungen gibt, was für die Benutzerin zu tun ist. Darunter befinden sich Aufgaben, bei denen sich die Benutzerin Informationen, Rat und Unterstützung holt oder sich einen bestimmten Teil aus dem Wissenspool des Systems zunutze machen wollen (Bernsen et al., 1994).

Je nach Umfang und Struktur der Aufgabe kann man diese grob in drei Typen einteilen, welche eine zunehmend komplexere Dialogverarbeitung zur Unterstützung der Funktionalität benötigen. In Tabelle 3.3 wird diese Einteilung dargestellt.

Derzeitige interaktive Dialogsysteme schaffen es mittlerweile, die Aufgaben aus den ersten zwei Spalten der Tabelle 3.3 auf brauchbare Weise abzuwickeln. Vor allem in den letzten Jahren konzentrierte sich die Forschung auf die Systeme, die in die dritte Spalte eingestuft werden.

Immer wenn die Technologie so weit vorangeschritten war, um einen Schritt in die nächste Spalte zu erlauben, wurde es auch möglich, dass nicht nur die neuen Aufgabenklassen hinzukamen, sondern auch die alten dahingehend weiter verbessert werden konnten.

Aufgabentyp: - kleine und einfache Aufgaben	Aufgabentyp: - umfangreiche, <i>well-structured</i> Aufgaben - eingeschränkte Domänen	Aufgabentyp: - umfangreiche, <i>ill-structured</i> Aufgaben, - eingeschränkte Domänen
Dialogtyp: - Einwort-Dialog	Dialogtyp: - Systeminitiative	Dialogtyp: - gemischte Initiative

Tabelle 3.3: Komplexität der Aufgaben

▷ Agenten-orientierte Architektur

Die Agenten-orientierte Architektur (*agent-based*) wurde von Allen et al. (2001) vorgeschlagen. In seinem Vorschlag werden die Komponenten eines Dialogsystems in drei Klassen geteilt: Interpretation (*interpretation*), Generierung (*generation*) und Verhalten (*behaviour*). Jede Einheit besteht aus einem allgemeinen Kontrollmodul und kommuniziert mit den anderen Einheiten, um Informationen und Mitteilungen auszutauschen. Der Agent, der zuständig für das Verhalten ist, ist auch für den Verlauf des Dialogs zuständig (Cuayáhuitl, 2005, S.6).

Das Konzept von Traum et al. (1996) über Konversationsagenten erweitert die BDI-Theorie (*Beliefs Desires Intentions*). In dem BDI-Modell beeinflusst jede Handlung, die in der Welt ausgeführt wird, die Überzeugungen der Agenten, die darüber Meinungen äußern und dadurch Wünsche und Intentionen formulieren können. Die Überzeugungen (*beliefs*) widerspiegeln die Welt, wie sie empfunden wird; die Wünsche (*desires*) - wie die Welt sein sollte und die Intentionen formulieren Pläne, wie diese Wünsche zu erreichen wären.

Es gibt zwei Probleme mit dem BDI-Modell. Die Weltperzeption beeinflusst nicht nur die Überzeugungen, sondern gleichzeitig auch die Wünsche und die Intentionen der Agenten. An zweiter Stelle unterstützt das Modell nur einen Agenten. Traum erweitert und implementiert das Modell in dem TRAINS-93 (Tab.: A.34 auf Seite A-14) und dem nachfolgenden TRIPS Projekt (Traum et al., 1996), wobei er gemeinsame Überzeugungen einführt (was beide Agenten glauben, dass wahr ist) und auch zulässt, dass die Perzeption der Welt nicht nur die Überzeugungen, sondern auch die Wünsche und die Intentionen der Agenten beeinflusst (Wilks et al., 2006, S.2).

Sehr viel wurde im Feld Dialogmanagement geforscht mit dem Ziel, eine flexible und robuste Interaktion mit kompakten Softwareagenten zu erzielen. Diese Ansätze erfassen das Agenda-orientierte Dialogmanagement (Constantinides und Rudnicky, 1999) und die Erweiterung RavenCraw (Tab.:A.60) (Bohus und Rudnicky, 2005) der von DARPA gesponserten *Communicators*⁴ (Anh.: A.1.1.1). Auch andere Systeme implementieren diesen Ansatz (zB. Collagen Tab.:A.63). Sehr oft bei der Umsetzung wird objekt-orientiert gearbeitet. Dass wichtige dabei ist, dass die Vererbung⁵ dafür verwendet wird, allgemeines Dialogmanagement von domain-spezifischen Interaktionen zu unterscheiden.

Der modulare Agenten-orientierte Ansatz zum Dialogmanagement ermöglicht die

⁴<http://communicator.sourceforge.net/index.shtml>

⁵im Sinne einer objektorientierten Programmiersprache

Kombination unterschiedlicher Ansätze wie den Zustand-orientierten und Frame-orientierten Ansatz. Darüber hinaus können auch unterschiedliche Strategien wie Systeminitiative und gemischte Initiative je nach Anforderungen der Situation und des Kontextes verwendet werden (siehe Levin et al., 2005; Jordan und Walker, 1996). Wenn einige Agenten für die Bewältigung ein und derselben Aufgabe verwendet werden, können auch Regel-basierte Verfahren und maschinelles Lernen kombiniert werden. Zum Beispiel in dem AthosMail (Tab.:A.29) System werden diese Verfahren eingesetzt, um das System robust für unterschiedliche Anwenderinnen zu machen (Jokinen et al., 2004). Das erste Verfahren verwendet Agenten für die pragmatischen Verarbeitung und die Markierung von möglichen Bedeutungen, während das zweite Verfahren unterschiedliche spezielle Dialogagenten verwendet, um eine multilinguale Interaktion zu ermöglichen. In der Realität wird die Agenten-orientierte Architektur mit dem Ziel verwendet eine echte gemischte Initiative im Dialog zu ermöglichen, wie auch eine multimodale Anleitung um Neueinsteigerinnen zu unterstützen.

Einige Dialogmanager verwenden autonome Agenten, basierend auf der BDI-Theorie als Alternative. Einige Implementierungen der BDI-Theorie in DM wurden unabhängig voneinander durchgeführt und es hat sich gezeigt, dass sich die Systeme durch den Vorteil auszeichnen, intentionales Verhalten und Planabbrüche in plausibler psychologischer Manier zu bewältigen (Wilks et al., 2006, S.5f).

▷ Informationszustand-orientierte Architektur

Die Informationszustand-orientierte (*information-state*) Architektur baut auf der Idee auf, dass sich der Informationszustand für alle Dialogteilnehmer im Laufe des Dialogs ändert (Georgila und Lemon, 2004). Der Zustand, der die vorhandene Information in einem bestimmten Zeitpunkt darstellt, kann dazu dienen, Dialoge voneinander zu unterscheiden, vorige Handlungen zu repräsentieren und weitere Handlungen zu motivieren (Larsson und Traum, 2000; Lemon et al., 2002; Traum und Larsson, 2003).

The term information state of a dialogue represents the information necessary to distinguish it from other dialogues, representing the cumulative additions from previous actions in the dialogue, and motivating future action. For example, statements generally add propositional information, questions generally provide motivation for others to provide specific statements. Information state is also referred to by similar names, such as „conversational score“, or „discourse context“ and „mental state“. Larsson und Traum (2000, S.1f)

Die Theorie wurde in der *dialogue move engine* TRINDI (Tab.:A.65) implementiert und folgender Weise formuliert (Larsson und Traum (2000); Traum und Larsson (2003)):

- ▷ Beschreibung der relevanten Komponenten wie Kontext, Faktoren wie Gemeinsamkeiten, Verbindlichkeiten, Überzeugungen, Intentionen, etc
- ▷ Formale Repräsentationen der aufgelisteten Komponenten (Z.B. Listen, Sätze, Strukturen, Aufnahmen, etc)
- ▷ Ein Set von Dialogzügen, die zum Update des Informationszustandes führen (auch in Korrelation mit extern ausgeführten Handlungen wie manche sprachliche Äußerungen
- ▷ Ein Set von Update-Regeln, die das Updaten des Informationszustandes und des Dialogablaufs kontrollieren und ein Set von Regeln zur Auswahl von Regeln
- ▷ Update-Strategie um zu entscheiden, welche Regel(n) zu einem bestimmten Zeitpunkt aus dem Set der zugelassenen Regeln anzuwenden sind.

Eine andere Implementierung dieses Ansatzes ist DIPPER (Tab.:A.54) (Bos et al., 2003).

3.2.2 Initiative (wer stellt die Fragen)

Die Aufgaben, die mittels auf Ein-Wort-Äußerungen beruhenden Dialogen bewältigt werden, sind kurz und einfach und brauchen nur wenige Dialogzüge. Die mit systemgesteuerten Dialogen bewältigbaren Aufgaben erfordern zwar viele Dialogzüge, sind aber gut strukturiert (*well-structured*). Mit gemischter Initiative bewältigbare Aufgaben enthalten typischerweise eine große Anzahl von Unteraufgaben von denen jede einige Sprecherwechsel beinhalten kann. Jede Unteraufgabe kann zwar gut strukturiert sein, wenn aber die Beziehungen zwischen den Unteraufgaben schlecht definiert sind, wird daher so ein Dialog auch Sets von unvorhersehbaren Unteraufgaben einschließen. Die Anzahl und die Komplexität der in einer Anwendung gebrauchten Dialogelemente hängt von dem zu unterstützenden Dialogtyp ab.

3.2.2.1 Einwort-Dialoge

Systeme mit auf Ein-Wort-Äußerungen beruhenden Dialogen sind normalerweise entweder mit großem Vokabular ausgestattet und Sprecher-abhängig oder mit kleinem

Vokabular ausgestattet und Sprecher-unabhängig. Die Initiative geht entweder nur vom System oder nur von den Benutzerinnen aus. Wenn die Benutzerinnen an der Reihe sind, verläuft die Dialogstruktur flach, was heißt, dass es zum Beispiel eine Auswahl von Wörtern für Anweisungen zum Speichern, Drucken usw. von Dateien gibt. Wenn das System an der Reihe ist, verläuft die Dialogstruktur weder flach noch hat es wirklich tiefergehende Ebenen. System-Feedback kann erforderlich sein, wenn zum Beispiel das System verstanden hat, dass die Benutzerin Geld auf ein Konto überweisen will. In diesem Fall ist es meistens notwendig, sich das von der Benutzerin vor der Durchführung noch einmal bestätigen zu lassen. Dies geschieht aber ohnehin dadurch, dass die darauf folgende Aktion zeigt, welche Auswahl der Benutzerin vom System erkannt worden ist.

In kleineren Systemen werden zum Beispiel Hypothesen (*predictions*) überhaupt nicht gebraucht. Der Systemfokus wird der jeweiligen Benutzer-Aufgabe entsprechen und in der Dialogstruktur fest verankert sein. Auch der Dialogverlauf wird nicht benötigt, da man selten Verweise auf vorhergehende Dialogteile braucht. Man kann sogar die Elemente für die explizite Meta-Kommunikation auslassen. Fehlerkorrektur erfolgt dann über das Wiederholen der Anweisung oder durch eine neue. Viele Systeme dieser Gruppe passen sich an die Sprecherin an. Der Dialog selbst wird aber nicht an die Benutzerin angepasst und ein ausgereifteres Benutzer-Modell ist daher nicht notwendig (Bernsen et al., 1994).

3.2.2.2 Systemgesteuerter Dialog

Was hier als systemgesteuerter Dialog bezeichnet wurde, ist normalerweise wesentlicher komplexer als der Einwort-Dialog. Die gesteigerte Komplexität ist hauptsächlich auf die gesteigerte Komplexität der Aufgabe selbst zurückzuführen, sowie die Länge der Äußerungen und auf den gesteigerten Umfang des Vokabulars. Bei systemgesteuerten Dialogen übernimmt, wie der Name sagt, das System die Initiative. Es übernimmt und behält die Initiative indem es alle Sprecherwechsel mit Fragen an die Benutzerinnen - mit Ausnahme des letzten - abschließt. Die Fragen zeigen implizit, dass die Initiative vom System auszugehen hat.

3.2.2.3 Gemischte Initiative

Um Dialoge mit gemischter Initiative (*mixed initiative*) zu unterstützen, muss ein System die Reihenfolge der Sprecherwechsel festlegen können. Eine Methode dafür kann darin liegen, auf Dialog-Kontext basierende Kontroll-Regeln neben einer simplen Klassifizierung der Dialogakte zu verwenden, wie durch Ferguson (1995) gezeigt. In der

Praxis gibt es eine kontinuierliche Abfolge zwischen der durch Fragen, Feststellungen oder Aufforderungen ausgeübten vollen Kontrolle des Systems einerseits und Dialogen mit gemischter Initiative andererseits, bei denen das System nur dann die Initiative ergreift, wenn es natürlich erscheint. Sogar die interaktiven Dialogsysteme für stereotype Aufgaben brauchen ein Maß an *mixed initiative* Dialog um natürlich zu wirken. Und Systeme, welche nicht-stereotype Aufgaben erfüllen, wie zum Beispiel eine größere Anzahl von Unteraufgaben, können oft zu einen systemgesteuerten Modus wechseln, sobald eine stereotype Unteraufgabe, welche die Benutzerinnen durchgeführt haben wollen, identifiziert worden ist (siehe Yang et al., 2004). Wenn ein Teil der Initiative den Benutzerinnen überlassen worden ist, können Abweichungen von der voreingestellten Aufgabe des Bereichs erwartet werden und das System muss dann imstande sein, den im Systemfokus befindlichen Set von Unteraufgaben in Laufzeit zu generieren. Dialoge mit gemischter Initiative benötigen daher einen dynamisch determinierten Set von Unteraufgaben im System Fokus. Eine Voraussage (*prediction*) über den User-Input ist schwieriger bei solchen Systemen, vor allem bei nicht-stereotypen Aufgaben. Die größte Herausforderung für das System bei Empfang einer Aussage in Dialogen mit gemischter Initiative ganz allgemein und insbesondere bei nicht-stereotypen Aufgaben ist es, die Unteraufgabe zu identifizieren, die die Benutzerinnen beabsichtigen. Phänomene des Diskurses wie zum Beispiel Anaphern kommen in Dialogen mit gemischter Initiative häufiger vor. Der linguistische Dialogverlauf wird hauptsächlich dazu verwendet, die Auflösung der Anaphern und Ellipsen zu unterstützen und muss seine Arbeit getan haben, bevor der semantische Input weiter verarbeitet wird (Andersen et al., 1999). Der Dialogaktverlauf kann dazu verwendet werden, die Korrektur des weiter zurückliegenden Inputs zu unterstützen. Wie auch beim Mensch-Mensch-Dialog ist es die bequemste Lösung, die zu korrigierende Information zu nennen. Dies benötigt unter anderem ein Aufrechterhalten der gegenseitigen Abhängigkeiten (*inter-dependencies*) zwischen den Aufgaben-Werten, einer Umsetzungsstrategie, um die vorangegangenen Abschnitte des Dialoges wieder aufsuchen zu können und einen Aufgabenbericht. Zusätzlich kann dieser Aufgabenbericht auch die noch durchzuführenden Aufgaben aufzeichnen. Das System kann die laufende Aufgabe anhalten, wenn es klar wird, dass es zum Fortfahren einige Werte braucht, die nur aus Aufgaben, die in der Dialogstruktur eigentlich vorher angesiedelt sind, bezogen werden können.

Die Adaptionsfähigkeit des Systems kann durch den Einsatz eines Benutzer-Modells weiter gesteigert werden, welches dem System bei der Beurteilung hilft, wie mit den Benutzerinnen weiter verfahren werden soll. Zum Beispiel ob es verstärkt die Benutzerinnen zum Buchstabieren veranlassen soll oder dezidierte Ja/Nein-Antworten verlangen soll oder ob vielleicht Multiple-Choice-Fragen das Gelingen der Aufgabe erleichtern.

Untersuchungen in diesem Bereich wurden von Litman und Pan (1999) und Chu-Carroll und Nickerson (2000) gemacht.

Vom System veranlasste sogenannte *graceful degradation* des Benutzer-Levels ist eine vielversprechende Methode, einen fehlgeschlagenen Dialog wieder herzustellen. Beim Anwenden von *graceful degradation* übernimmt das System die Initiative und erfragt in immer einfacher werdenden Schritten die fehlende Information. *Graceful degradation* wird solange fortgesetzt, bis das System den Input verstanden hat oder keine weitere *graceful degradation* mehr möglich ist. Wurde der Input verstanden, setzt der Dialog wieder bei dem Benutzer-Level direkt vor der *graceful degradation* fort. Da man von den durchschnittlichen Benutzern nicht erwarten kann, dass sie sich viele Schlüsselwörter merken, ist für eine verbesserte Meta-Kommunikation der Ansatz des *mixed initiative* Dialogs unabdingbar (Bernsen et al., 1994). Die im Anhang aufgelisteten Systemen verwenden mit wenigen Ausnahmen die gemischte Initiative als Verfahren.

3.3 Evaluierung von interaktiven Dialogsystemen

Eine wichtige Frage bei der Entwicklung von Dialogsystemen ist ihre Evaluierung. Daher wird gleichzeitig mit dem Systemdesign meistens ein Plan entwickelt, wie die Ergebnisse des gesamten Systems und der einzelnen Systemkomponenten überprüft werden können. Zum Beispiel können solche Systeme hinsichtlich folgender Kriterien ausgewertet werden: Rate der falschen erkannten Wörter, Rate der Falsch erkannten Intentionen, Rate der erfolgreich erledigten Aufgaben, Produktivität (bezieht sich auf die für das erfolgreiche Erledigen benötigte Zeit), und Benutzerzufriedenheit. Dialogsysteme müssen in all diesen Bereichen eine gute Leistung erbringen. Ein wichtiges Kriterium ist die Einhaltung der vier (Konversations-) Maximen von Grice: der Beitrag des Systems sollte echt und genau sein - Qualität, der Beitrag des Systems sollte kurz und frei von Zweideutigkeiten sein - Modalität, das System sollte genügend Information bieten - Quantität, und der Beitrag sollte relevant sein - Relevanz (Baum et al., 2001, S.33). Üblicherweise werden Benutzerinnen aufgefordert, unter verschiedenen Umständen bestimmte Aufgaben durchzuführen (ruhige Umgebung im Gegensatz zu belebter, vordefinierte Aufgabe im Gegensatz zu einer freien), auch der Grad des Vertrautseins der Benutzerinnen mit dem System ist dabei unterschiedlich. Die Auswertung kann qualitativ oder quantitativ sein.

Eine der Fragen, mit der die Entwickler von Systemen für gesprochene Dialoge konfrontiert werden ist, wie der Fortschritt ausgewertet werden kann, wie man erkennen kann, ob man ein brauchbares System entwickelt hat. Die Entwickler müssen Maßstäbe festlegen, anhand derer der Fortschritt getestet werden kann. Dabei sollten die

einzelnen Komponenten wie auch die gesamte Performanz geprüft werden.

Bei Systemen, die gesamte Vorgänge bewältigen müssen, kann man leicht testen, ob die Benutzerinnen die Aufgabe mit Hilfe des Systems erledigen konnten. Dabei kann man auch Statistiken führen, wie lange der Prozess gedauert hat und wieviele Dialogzüge dafür notwendig waren. Allerdings wurde festgestellt, dass solche Statistiken weniger aussagekräftig sind, als die Kundenzufriedenheit. Zum Beispiel kann ein Dialogmanager länger brauchen, um ans Ziel zu kommen als ein anderer, trotzdem kann die Alternative für die Benutzer weniger zufriedenstellend sein. Eine bessere Form der Evaluierung ist es daher zu fragen, ob die Benutzer eine echte Aufgabe erledigen konnten, ob sie das System wieder benützen würden oder auch weiterempfehlen würden.

Evaluierungen einzelner Komponenten sind sehr gebräuchlich für ASR und NLU, weniger aber für Sprachgenerierung oder Sprachsynthese, weil letztere weniger zugänglich für automatisierte Evaluierungsmethoden mit eindeutigen richtigen oder falschen Antworten sind. ASR Evaluierung ist normalerweise am eindeutigsten, aber auch hier gibt es eine Reihe von Fällen, die nicht so einfach auszuwerten sind (Unterbrechen und Dazwischenreden, Murmeln, nur Wortteile aussprechen, Hintergrundgeräusche, die Qualität der Telefonverbindung). Die NLU Einheit kann auch evaluiert werden, indem man die ermittelten Bedeutungen mit einer Referenz vergleicht. Das Problem dabei ist aber, dass es keine gemeinsame Basis für die Evaluierung der Bedeutungserfassung bei den verschiedenen Forschergruppen gibt. Beim DARPA ATIS Programm (siehe Anhang „DARPA Communicator“ auf Seite A-9) konnten sich die Teilnehmer letztendlich nur darauf einigen, Antworten mit einer gemeinsamen Datenbank zu vergleichen (siehe Walker und Hirschman, 2000).

Leider führte dies zur Notwendigkeit, ein umfangreiches Dokument zu schaffen, in dem die Prinzipien der Interpretation für alle möglichen Fragen definiert wurden (Walker und Hirschman, 2000). Um die Antworten aller Systeme konsistent zu halten, wurden die Entwickler der verschiedenen Systeme davon abgehalten, Eigeninitiativen zu ergreifen. Dies bedeutet aber eine gröbere Einschränkung der Dialogforschung.

Ein Weg den Fortschritt eines bestimmten Systems zu zeigen ist, langfristig wieder und wieder die Fähigkeiten der Spracherkennung und des Sprachverständnisses auszuwerten. Im Falle von JUPITER (Tab.:A.5) verfolgte man eine Reihe von Standard-Tests, die periodisch überarbeitet werden, damit nicht nur ein bestimmter Datenset angesprochen wird (Zue et al., 2000). Dies ist relativ einfach, denn es werden immer mehr Daten gesammelt.

Einige Systeme verwenden bereits den Kontext des Dialogs, um nützliche Einschränkungen bei der Spracherkennung zu erreichen. Es werden zum Beispiel die Teile der Äußerungen herangezogen, die das Schlüsselwort der Anfrage enthalten.

Daher muss das System imstande sein, bei neuerlichen Tests auf Verbesserung der Spracherkennungs- und Verständniseigenschaft, diese Äußerungen in den selben Kontext zu stellen, wie er im originalen Dialog mit den Benutzerinnen schon bestanden hat. Um das zu erreichen, muss der Dialogkontext zugleich mit der Datensammlung gespeichert werden und in der darauf folgenden Verarbeitung wiederverwendet werden, um zu verhindern, dass das Ergebnis verzerrt wird.

Das Problem bei solchen Auswertungen ist, dass sich Systeme voneinander stark unterscheiden und nicht auf einen gemeinsamen Nenner zu bringen sind. Daher wurde in Richtung Evaluierung von interaktiven Dialogsystemen viel geforscht mit dem Ergebnis, dass sich einige Verfahren dabei durchgesetzt haben

Traum und Andersen (1999) entwickelten eine Methode, um den Dialogstatus in Domäne- und aufgabenunabhängigen Systemen zu repräsentieren. Walker et al. (1997b) schlagen PARADISE vor. Polifroni und Seneff (2000) schlagen vor, Galaxy-II als Architektur zur Evaluierung zu verwenden (siehe auch Seneff et al., 1998, 1999). Tim Paek (2001) schlägt empirische Verfahren zur Evaluierung vor. Das deutet darauf hin, dass es kontroverse Meinungen gibt, die auf bestimmte Schwierigkeiten zurückzuführen sind.

3.3.1 Schwierigkeiten bei der Dialogsystemevaluierung

Interaktive Systeme für gesprochene Dialoge basieren auf vielen technischen Komponenten: Spracherkennung, Dialogmanagement, Sprachverständnis, Sprachgenerierung und Sprachsynthese wie auch Datenbank-Abfragesprachen u.a. Während die Messungen für die einzelnen Komponenten gut verstanden werden, ist es schwierig, standardisierte Messungen für ganze Systeme zu entwickeln, die all diese Technologien integrieren. Das Problem besteht darin, dass es einfach so viele potentielle Messungen gibt, die man für die Evaluierung eines Dialogsystems benutzen kann. So kann man zum Beispiel ein System darauf prüfen, ob es imstande ist, den Benutzerinnen beim Erreichen ihrer Ziele zu helfen, oder auf die Fähigkeit, Fehler der Spracherkennung oder des Sprachverstehens zu registrieren und diese zu beheben, oder auch auf die allgemeine Qualität der Interaktion mit den Benutzerinnen (Danieli und Gerbino, 1995). Ein weiteres Problem ergibt sich aus der Tatsache, dass sich die Auswertungen von Dialogsystemen nicht auf die reine Auswertung der Transkripte reduzieren lassen, auch nicht auf einen einfachen Vergleich mit einem als Referenz vorgefertigten Set von Antworten, weil es sehr viele potentiell richtige Dialoge als Ergebnis geben kann.

Polifroni et al. (1998) schlagen eine Methode der Evaluierung von Systemen für gesprochene Dialoge auf zwei Wegen vor:

a) Auswertung des Systemverhaltens durch manuelle Untersuchung jedes einzelnen

Frage/Antwort-Paare, und

- b)** Auswertung durch automatische Untersuchung der Leistung jeder einzelnen Komponente (Spracherkennung, Parser der natürlichen Sprache, natürliche Sprachgenerierung und Domain-Server).

Ein Vergleich zwischen der manuellen und automatischen Auswertung zeigte eine signifikante Übereinstimmung.

Im Hinblick auf Dialogmodelle ist die Simulation auf der Ebene der Intentionen die praktischste, denn die Auswirkungen der Fehler bei Spracherkennung und Verständnis können vormodelliert werden und die Feinheiten der natürlichen Sprachgenerierung können vermieden werden (Cuayáhuitl, 2005, S.13).

Levin und Pieraccini (1997) bereiteten den Weg für das Konzept der automatischen Dialogmodellierung als ein Optimierungsproblem unter Zuhilfenahme des *reinforcement learning* (RL) *framework*, welches das „*learning by trial and error*“-Paradigma anwendet (siehe Henderson et al., 2005; Lemon et al., 2006; Litman et al., 2000; Singh et al., 2002).

Die Messkriterien für Dialoge sind sowohl objektiver als auch subjektiver Natur. Die objektiven Messkriterien, welche benutzt wurden, um Dialoge in ihrer Gesamtheit auszuwerten beinhalten (nach Danieli und Gerbino, 1995; Walker et al., 1997b,c; Walker und Hirschman, 2000; Walker und Passonneau, 2001) :

- ▷ Prozentsatz der korrekten Antworten im Hinblick auf einen Set von Referenz-Antworten
- ▷ Anzahl von Sprecherwechsel und Äußerungen
- ▷ Dialogdauer oder benötigte Zeit für das Erledigen der Aufgabe
- ▷ Durchschnittliche Reaktionszeit der Benutzerinnen
- ▷ Durchschnittliche Reaktionszeit des Systems
- ▷ Häufigkeit der Diagnose-Fehlermeldungen
- ▷ Prozentsatz der nicht trivialen (länger als ein Wort) Äußerungen
- ▷ Durchschnittliche Länge der nicht trivialen (länger als ein Wort) Äußerungen

Diese objektiven Messkriterien können ohne Rückgriff auf menschliche Beurteilungen ausgewertet werden und in vielen Fällen kann das automatisch durch die Systeme für gesprochene Dialoge erfolgen. Eine mögliche Ausnahme ist die Aufgaben-orientierte

Erfolgsmessung, wie zum Beispiel das Prüfen auf Erfolg der Transaktion, ob die Aufgabe vollständig erfüllt wurde und wie qualitativ hochwertig die Lösung war.

Ob eine Messung objektiv oder subjektiv ist, wird dadurch beeinflusst, ob die Ziele der Benutzerinnen schon am Anfang des Dialoges klar definiert werden. Dies ist der Fall in kontrollierten Experimenten aber in Feldstudien benötigt man menschliche Entscheidungskraft, um festzustellen, ob eine Aufgabe erfolgreich bewerkstelligt wurde. Subjektive Messungen werden von Personen angewendet, die sich dafür entweder des Systems selbst bedienen oder auf die Ergebnisse von Versuchen mit Testpersonen stützen, um den Dialog oder die darin enthaltenen Äußerungen im Rahmen verschiedener qualitativer Bereiche zu kategorisieren. Weil diese Messungen auf menschlichen Beurteilungen basieren, müssen sie von allen Beurteilenden gleich angewendet werden, um eine mit den nach objektiven Kriterien durchgeführten Messungen vergleichbare Wiederholbarkeit zu gewährleisten.

Subjektive Messungen können durchaus quantitativer Natur sein, wenn zum Beispiel ein Verhältnis zwischen zwei subjektiven Kategorien zu berechnen ist. Sehr oft werden unter anderem folgende Methoden der subjektiven Messungen verwendet, die von Danieli und Gerbino (1995) u.a. vorgeschlagen wurden :

- ▷ Implizite Wiederherstellung (*implicit recovery*) (IR): die Fähigkeit eines Systems, aus dem Zusammenhang einen wegen nur teilweise erfolgter Spracherkennung fehlgeschlagenen Dialog wieder herzustellen.
- ▷ Explizite Wiederherstellung (*explicit recovery*) (ER): der Anteil von sowohl vom System, als auch von den Benutzerinnen zur Wiederherstellung des Dialoges getätigten expliziten Äußerungen. Systemseitige Korrektur - *system turn correction* (STC), und benutzerseitige Korrektur - *user turn correction* (UTC).
- ▷ Kontextabhängige Angemessenheit (*contextual appropriateness*) (CA): der Grad der Abstimmung der System-Äußerungen auf den Dialog-Kontext. Äußerungen können entweder passend (*appropriate* - AP), unpassend (*inappropriate* - IP) oder zweideutig (*ambiguous* - AM) sein.
- ▷ Kooperativität der System-Äußerungen: beurteilt, wie sehr das Systemverhalten die Grice'schen Konversationsmaximen (Grice, 1993) einhält.
- ▷ Korrekte und teilweise korrekte Antworten.
- ▷ Passende und unpassende Anweisungen und Analysen: das System erteilt den Benutzerinnen Anweisungen, während Analysen die Mitteilungen des Systems an die Benutzerinnen sind, mit denen es vermittelt, was den Fehler verursacht hat und warum das System keine Antwort geben kann.

- ▷ Benutzerzufriedenheit: eine Messung, welche darauf abzielt, die Einschätzung der Benutzerinnen bezüglich der Benutzerfreundlichkeit einzufangen. Dies wird üblicherweise mit multiple-choice Fragebögen durchgeführt. Die letzteren fordern Benutzerinnen auf, die Systemleistung in verschiedenen Belangen hinsichtlich der Benutzerfreundlichkeit innerhalb einer Skala zu benoten.

Sowohl die objektiven als auch die subjektiven Messungen haben sich beim Vergleichen verschiedener Systemen für gleiche Aufgaben für die Entwickler der Sprachdialogsysteme als sehr nützlich erwiesen, aber trotzdem sind diese Messungen auch begrenzt. Eine weithin anerkannte Einschränkung ist, dass das Verwenden von Referenz-Antworten unbrauchbar für den Vergleich von Systemen ist, die verschiedenen Dialogstrategien zum Ausführen der gleichen Aufgabe verwenden (vgl. Walker et al., 1997a; Farfán et al., 2003). Die Referenz-Antwort-Methode verlangt nach vorschriftsmäßigen Antworten (also einzig richtige Antworten), die für jede Benutzeräußerung vordefiniert sein müssen. Daher ist es nicht möglich, die gleichen Referenz-Antworten zu benutzen, um ein System zu evaluieren, welches zum Beispiel in einem Fall eine Zusammenfassung als Antwort bietet, eine klärende Frage in einem anderen Fall, oder mit einem Set von Datenbankwerten in einem weiteren Fall reagiert. Ebenfalls limitierend wirkt, dass verschiedene Messungen eng verbunden miteinander sind und daher redundante Information über die Leistung liefern. Solche Verbindungen zu bestimmen erfordert eine Gruppe von Messungen, die weitläufig erfolgen, um dann zu prüfen, ob diese Verbindungen mehrere verschiedene Dialogsysteme betreffen. Eine dritte Einschränkung resultiert aus der Tatsache, dass sich verschiedene Messungen nicht austauschen oder kombinieren lassen und ebensowenig Verallgemeinerungen möglich sind. Man betrachte das unten angeführte Beispiel 3.3.1 - den Vergleich zweier Auskunftssysteme für Zuginformation (nach Danieli und Gerbino, 1995), wo System A in Dialog 1 eine explizite Bestätigungsstrategie verfolgt, während System B in Dialog 2 eine implizite Bestätigungsstrategie:

- (1) User: I want to go from Torino to Milano.
Agent A: Do you want to go from Trento to Milano? Yes or No?
User: No.
- (2) User: I want to travel from Torino to Milano.
Agent B: At which time do you want to leave from Merano to Milano?
User: No, I want to leave from Torino in the evening.

Beispiel 11: explizite/implizite Bestätigungsstrategie

Danieli und Gerbino haben herausgefunden, dass System A eine höhere Erfolgsrate pro Transaktion hatte und weniger unpassende bzw. Korrektur-Äußerungen erzeugte als System B. Weiterhin fanden sie heraus, dass das System A mit seiner Strategie

ungefähr doppelt so lange Dialoge wie System B produzierte, konnten aber nicht herausfinden, ob System A mit dem besseren Transaktionserfolg oder System B mit der höheren Effizienz eine bessere Leistung erbrachte. Die Fähigkeit, die Faktoren zu identifizieren, welche die Leistung am entscheidendsten beeinflussen, ist die wichtigste Basis für systemübergreifende Generalisierungen. Es wäre nützlich zu erkennen, wie die Eindrücke der Benutzerinnen hinsichtlich der Leistung eines Systems von der jeweils angewandten Strategie abhängen und wie hier das Verhältnis der einzelnen Faktoren wie Effizienz, Geschwindigkeit oder Genauigkeit eine Rolle spielen. Neben den dem System eigenen Faktoren wie zum Beispiel die verschiedenen Dialogstrategien in Dialog 1 und 2, können auch andere Faktoren wie die Größe der Datenbank und umweltbedingte Faktoren wie zum Beispiel Hintergrundgeräusch durchaus leistungsrelevant sein.

3.3.2 PARADISE - ein Verfahren zur Evaluierung interaktiver Dialogsysteme

In dem Zusammenhang mit den beschriebenen Anforderungen, Forschungsrichtungen und Problemen der Evaluierung wurde PARADISE (Walker et al., 1997b, 1998b) als Verfahren entwickelt, mit dem Ziel als Basis für die Evaluierung interaktiver Dialogsysteme zu dienen.

The most widely embraced evaluation method proposed so far is PARADISE. This method uses a decision-theoretic framework to specify the relative contribution of various factors for the overall system performance. Performance is modelled as a weighted function of: task success (exact scenario completion), dialogue efficiency (task duration, system turns, user turns, total turns), dialogue quality (word accuracy, response latency) and user satisfaction (sum of TTS performance, ease of task, user expertise, expected behaviour, future use). (Cuayáhuitl, 2005, S.11)

Das PARADISE Verfahren bezieht sich auf die von Danieli und Gerbino (1995) vorgestellten Messungen zur Evaluierung von Dialogsystemen und basiert auf einer Struktur von Zielen. Der Einsatz der Entscheidungstheorie setzt die Spezifizierung sowohl der Ziele des Entscheidungsproblems als auch des Sets von Maßnahmen (in der Entscheidungstheorie Attribute genannt) voraus, um die Ziele verfügbar zu machen. Das Modell postuliert, dass auf der obersten Ebene die Leistung in Verbindung mit sinnvollen externen Kriterien, wie zum Beispiel Benutzerfreundlichkeit, gesetzt werden kann. Daher ist es die Aufgabe eines Dialogsystems, ein Ziel besonders anzustreben, dass

die Benutzerfreundlichkeit maximiert. Benutzerbezogene Zufriedenheitswerte sind die am meisten verbreiteten externen Indikatoren der Benutzerfreundlichkeit eines Dialogsystems. Das Modell postuliert weiters, dass zwei Faktoren potentiell relevant für die Benutzerfreundlichkeit sind, nämlich Erfolg bei der Erfüllung der Aufgabe und der Kostenfaktor (Geschwindigkeit und Leistung bezogen auf verbrauchte Ressourcen). PARADISE verwendet eine linear rückschreitende Methode, um den relativen Beitrag des Erfolges bei der Aufgabe und des Kostenfaktors zur Benutzerfreundlichkeit zu quantifizieren (vgl. Walker et al., 1997b). Die Messung des Erfolgs bei den Aufgaben baut auf die vorhergegangenen Messungen der Transaktionserfolge (Danieli und Gerbino, 1995), verwendet aber auch den Kappa-Koeffizienten. Der Kappa-Koeffizient ist eine statistische Messung, die sich von der reinen prozentuellen Auswertung unterscheidet, indem die Übereinstimmung der Beurteilenden (*inter-rater agreement*) berücksichtigt wird (Carletta et al., 1997).

Der Kostenfaktor teilt sich in zwei Bereiche: Die Effizienzmessungen stammen aus einer Liste von Leistungsmessungen früherer Arbeiten, wie oben beschrieben. Qualitative Messungen zielen auf die Qualität des Dialoges ab. Diese basieren auf objektiven und subjektiven Messungen von vorausgegangenen Tests wie zum Beispiel die Häufigkeit von Analyse- und Fehlermeldungen, die Rate von unpassenden Äußerungen und der Anteil von Äußerungen zur Aufhebung von Fehlern.

Wie schon bei früheren Ansätzen zur Auswertung, verwendet man bei Leistungsevaluierungen mit dem PARADISE Modell einen Korpus von Dialogen zwischen Benutzerinnen und Dialogagenten, bei denen die Benutzerinnen einen Set verschiedener Szenarien durchführen. Jedes Szenario hat eine entsprechende Attribut-Wert-Matrix (AVM). Jedes durchgeführte Szenario hat eine entsprechende Beispielsammlung, welche die Informationsanforderungen für das Szenario anzeigt und wo jedes Attribut mit einem aus dem Dialog ermitteltem Attributwert gepaart ist.

Walker et al. (1997b) haben das PARADISE-Modell vorgestellt, indem sie es dazu benutzt haben, die Leistung zweier hypothetischer Dialog-Agenten in einer vereinfachten Zugauskunft-Umgebung zu vergleichen, wie am Beispiel „explizite/implizite Bestätigungsstrategie“ auf Seite 93 gezeigt wurde. Man nehme an, dass in einem Szenario die Benutzerinnen einen Zug von Turin nach Mailand finden sollen, der am Abend abfährt. Der Erfolg bei Erfüllung der Aufgabe für einen ganzen Dialog (oder Subdialog) wird daran gemessen, wieviel der notwendigen Information am Ende des Dialogs (Subdialogs) zwischen Agent und Benutzerin übermittelt worden ist.

PARADISE wurde dabei eingesetzt, um eine Leistungsfunktion für diese Aufgabe abzuleiten, indem die relativen Beiträge eines Sets von potentiellen Anzeichen zur Benutzerzufriedenheit geschätzt wurden. Das Verfahren ist in folgenden Schritten durch-

zuführen:

1. Definition einer Aufgabe und eines Sets von Szenarios
2. Spezifikation einer *AVM* Repräsentation der Aufgaben
3. Experimente mit alternativen Dialogagenten für die Erfüllung der Aufgabe
4. Berechnung der Benutzerzufriedenheit anhand von Umfragen
5. Berechnung des Aufgabenerfolges
6. Berechnung der Kosten für den Dialog mit Effizienzmessungen und qualitativen Messungen
7. Schätzung einer Leistungsfunktion mittels linear rückschreitender Methode und Werte für die Benutzerzufriedenheit und die Dialogkosten
8. Vergleich mit anderen Agenten und Aufgaben, um die Faktoren zu bestimmen, die am meisten Gewicht haben, um zu einer generellen Leistungsfunktion zu gelangen
9. Verfeinerung des Leistungsmodells.

Wichtig ist, dass all diese Schritte notwendig für die Ermittlung und die Festlegung der Leistungsfunktion sind. Sind aber die Gewichtungen in der Leistungsfunktion festgelegt, braucht man keine Benutzerzufriedenheitswerte mehr zu sammeln. Anstelle können Vorhersagen zur Benutzerzufriedenheit auf Basis der Anzeichenvariablen gemacht werden, was in der Anwendung von PARADISE auf Subdialoge veranschaulicht wird (siehe Walker et al., 1997b). Den jetzigen Wissensstand vorausgesetzt, müsste man sehr viele Experimente durchführen, um zu einer generellen Leistungsfunktion zu gelangen. Leistungsfunktionseinschätzungen müssen für viele verschiedene Aufgaben und Dialogstrategien aufzählend erfolgen, um zu sehen, welche Faktoren generalisiert werden können. Auf diese Weise kann man durch Feldforschung Fortschritte beim Identifizieren der Beziehungen zwischen den verschiedenen Faktoren erzielen und somit weiter in Richtung besser vorhersagbarer Modelle für Dialogagenten gehen.

Die dritte in dieser Diplomarbeit gestellte Frage - „wie kann Dialog von Maschinen erfolgreich simuliert werden?“, kann folgender Weise beantwortet werden: Interaktive Dialogsysteme bestehen mindestens aus Komponenten, welche die Sprachrezeption, die kognitive Tätigkeit (das Verstehen und das Generieren von Intentionen) und die Sprachproduktion des Menschen abbilden. Das sind die Spracherkennung (ASR), die Identifikation von Intentionen (NLU), der Dialogmanager (DM), das Generieren von

grammatisch richtigen Sätzen (Umsetzung der Systemintentionen in der NLG Komponente) und die Sprachsynthese (TTS). Weitere Komponenten können z.B. Gesten und Mimiken verarbeiten und dadurch die Arbeitsqualität der anderen Komponenten positiv beeinflussen. ASR, NLU, DM und NLG arbeiten sehr eng zusammen, denn es geht dabei, um das Verstehen und Generieren von Äußerungen, die sich auf das Gesprächsthema beziehen. Die TTS sollte auch in Kooperation mit den anderen Komponenten arbeiten, um mögliche Fehler bei der Umsetzung der Systemintentionen oder bei der Sprachgenerierung in Betracht zu ziehen und auszubessern. Je genauer die Komponenten die Dialogfähigkeit des Menschen abbilden, desto menschenähnlicher wirkt ein System. Um diese Fähigkeiten erfolgreich abbilden zu können, bedient man sich mathematisch-logischer Modelle, statistischer Verfahren und Korpora.

Der logische Aufbau der Dialogsystemen betrifft die Entscheidung, mit was für Wissen das System umgehen wird und in was für einem Bereich (*domaine*) es tätig sein wird. Diese Entscheidungen wirken sich auf die Flexibilität und Größe der vorbereiteten Korpora, Datenbanken, Modelle etc. und auf den Umfang der zu verwendenden Algorithmen. Weitere Überlegungen betreffen einerseits die Initiative, also die Frage wer darf die Konversation führen, bzw. wer darf die Fragen stellen und andererseits den theoretischen Ansatz zur Abarbeitung von Dialogphänomenen. Einfache Systeme verwenden die Systeminitiative und unterstützen typischer Weise Ein-Wort-Dialoge mit größeren statischen Thesauri und kleineren Sets von Regeln. Sie arbeiten meistens in kleinen geschlossenen Domänen. Komplexere Systeme verwenden gemischte Initiative und unterstützen meistens mehrere kleinere dynamische Thesauri, die je nach Bedarf geladen werden können.

Viel wurde besonders bei der Architektur des Dialogmanagers (die Art und Weise wie Dialog verarbeitet wird) experimentiert. Das ist auf die Tatsache zurückzuführen, dass einige Wissenschaftlerinnen mehr durch die Sprechakttheorie beeinflusst wurden, andere von der Dialogspieltheorie, dritte aber von der Theorie der rationalen Handlungen. Daher werden manche Systeme so konstruiert, dass sie mit Zielen und Überzeugungen umgehen, andere mit Plänen, weitere mit Aufgaben usw. Die Entscheidungen über das Design werden durch das Ziel der Anwendung geleitet, das im nächsten Kapitel behandelt wird. Man kann daraus schließen, daß im Allgemeinen sich Systeme durchsetzen, die eine gemischte Initiative unterstützen und multi-modal sind. Die am häufigsten verwendeten Verfahren sind das Plan- und Aufgaben-orientierte Verfahren. Eine sehr breite Anwendung finden die s.g. Agenten, die autonome Einheiten darstellen und meistens austauschbar unter den Systemen sind.

Eine weitere wichtige Frage im Bezug auf Dialogsysteme ist ihre Evaluierung. Die unterschiedliche Natur der Systeme führt dazu, dass es praktisch unmöglich ist, Systeme

me miteinander zu vergleichen. Unter den Vorschlägen, wie das möglich wäre, zeichnet sich das PARADISE Verfahren als eine Richtlinie für die Evaluierung von interaktiven Dialogsystemen aus, um die s.g. Leistungsfunktion zu bestimmen, wodurch Systeme auf einen gemeinsamen Nenner gebracht werden können.

Kapitel 4

Beispiele und Anwendungsmöglichkeiten

Der verbreitete Einsatz von Computern und Computersystemen führt dazu, dass es immer mehr Möglichkeiten gibt, eine sprachbasierende Benutzeroberfläche für die Kommunikation mit diesen Geräten anstatt der klassischen Konsole (Tastatur/Bildschirm) zu verwenden. Anhand der oben beschriebenen Besonderheiten der menschlichen Kommunikation (Kap. „Grundlagen der Kommunikation“ auf Seite 7) und der Systemarchitektur (Kap. „Interaktive Dialogsysteme“ auf Seite 53) zeichnen sich bestimmte Bereiche aus, wo die Anwendung solcher Systeme durchaus möglich ist. Das sind vor allem:

- ▷ Assistenten in der Mensch-Mensch und Mensch-Maschine-Interaktion
 - ▷ für die Auskunft allgemein, für Reservierungen, bei der Reisenavigation und bei der Planung von Reisen
 - ▷ für die Organisation im Büro (E-Mail, Terminverwaltung)
 - ▷ bei der Reparatur von verschiedenen Geräten
 - ▷ bei der Steuerung von Software und Peripheriegeräten
 - ▷ für die Kontoverwaltung bei der Bank
- ▷ Tutoren für die Vermittlung von Wissen mit einem didaktischen Ansatz
- ▷ Robotik (Oberflächen für die Kommunikation mit Robotern)
- ▷ simultane Übersetzungsprogramme

Bei allen Systemen geht es darum, eine bestimmte Aufgabe zu erledigen, wie auch ein oder mehrere damit verbundene Ziele zu erreichen. Die meisten heutigen Systeme verwenden gemischte Initiative als Verfahren mit der Möglichkeit, unter Umständen die Initiative zu übernehmen, um schließlich durch elementare Fragen in Subdialogen bestimmte Missverständnisse zu klären. Ein weiterer Trend ist die Multimodalität, die die Kommunikation (z.T. auch die Steuerung) von unterschiedlichen Peripheriegeräten und Anwendungen in unterschiedlichen Umgebungen möglich macht. Typische Umgebungen sind zu Hause, im Auto (unterwegs) und im Büro, wie es zum Beispiel in SMARTKOM (Wahlster, 2006) der Fall ist. Eine typische Umgebung für NASA ist das Weltall (siehe das IPA System Tab.:A.33).

Man unterscheidet zwei Varianten der Kommunikation mit dem System: direkter Kontakt, oder über Telefonverbindung. Beim direkten Kontakt wird das Audiosignal der Konversation direkt zum am System gebundenen Mikrophon geleitet. Bei dem zweiten Verfahren wird berücksichtigt, dass die Qualität der Telefonverbindung proportional zur Performanz steht, wie auch die Tatsache, dass die Benutzerin sich nur einen Teil der gesamten Information bei einem Dialogzug merken kann (Besonderheiten der Modalität). Die meisten Auskunftssysteme werden daher speziell auf Kommunikation über Telefon ausgerichtet und können mit Sprache in schlechter (Telefon-) Qualität arbeiten.

Viele Systeme sind multilingual, oder können mit lokalen sprachlichen Varianten (Dialekten) umgehen. Die Vervollständigung von Korpora und Sprachmodelle für die Unterstützung dieser Funktionalitäten, die als besonders benutzerfreundlich gelten, ist keineswegs abgeschlossen und bietet viele Möglichkeiten für Sprachwissenschaftlerinnen, darunter auch Germanistinnen, zur Forschung in diesem Bereich.

4.1 Plattformen

Sehr oft bei der Entwicklung von Dialogsystemen wird eine Art Plattform benutzt, um mehrere Systeme auf dieser Basis zu entwickeln. Dadurch können Systemkomponenten für eine bestimmte Tätigkeit bzw. eine bestimmte Domäne modelliert werden.

Turunen und Hakulinen (2001) beschreiben Jaspis (Tab.: A.52) als „a general speech application development architecture“. Es ist ein adaptives System mit einer Agentenorientierten Architektur. Einige konkrete Projekte basieren auf dieser Plattform: Bussimies / Busman (Tab.: A.4), Athosmail/DUMAS (Tab.: A.29) und Mailman (Tab.: A.30).

Der Galaxy II Communicator (Tab.: A.55) ist eine andere solche Entwicklerumgebung, die im Rahmen der DARPA Initiative entstanden ist. Seneff et al. (1998) präsentieren Galaxy II als eine Hub-basierende Client-Server Architektur. Der Hub wird dafür verwendet, die Kommunikation zwischen den Servern zu ermöglichen und zu koordinie-

ren. Die Architektur erwies sich als sehr flexibel und wurde in der Implementierung von verschiedenen Dialogsystemen, meistens in der ATIS Domäne (siehe unten P.4.2.2) verwendet. Die Systeme können Flugauskünfte durchführen, Flüge und Reisen planen und buchen; manche davon können Auto und Hotelzimmer reservieren.

Das Trindi Toolkit ist eine andere Entwicklerumgebung für Sprachsysteme im Bereich der Telematik und stellt eine Implementierung der Theorie der Informationszustände (Traum und Larsson, 2003) dar. Damit wurden Siridus (Tab.: A.59) und GoDiS (Tab.: A.56) entwickelt, die wiederum zur Entwicklung anderer Systeme verwendet wurden, wie z.B. DICO (Tab.: A.36) oder Talk (Tab.: A.62). Die Architektur dieser Systeme ist multi-modal und multi-lingual.

RavenClaw/Olympus ist eine weitere Umgebung, die unbedingt erwähnt werden soll (Tab.: A.60). Die Architektur besteht aus dem RavenClaw Dialogmanager, der als „a task-independent dialogue engine based on the AGENDA dialog manager“ von Rudnicky und Xu (1999) bezeichnet wird und auf dem CMU Communicator beruht. Die Plattform wurde zur Entwicklung von unterschiedlichen Dialogsystemen benutzt, wie LARRI (Tab.: A.32), Clarissa (Tab.: A.41), VERA (Tab.: A.31), ConQuest (Tab.: A.15) etc.

Es gehört hier vielleicht noch erwähnt, dass ein Teil der Systeme eine animierte virtuelle Figur für die Kommunikation mit den Benutzerinnen verwendet. Solche sind OLGA (Tab.: A.11), AUGUST (Tab.: A.57) und Smartakus in SmartKom (Tab.: A.61) .

4.2 Assistenten

Betrachtet man die Art und Weise, wie wir unterschiedliche technische Geräte verwenden, so kann man zwei Bereiche der Anwendung definieren: Unterhaltung und Unterstützung einer produktiven Tätigkeit. Diese Tätigkeit kann sein, etwas herzustellen, Information zu finden, Termine zu verwalten usw. Die Maschinen unterstützen die Menschen bei der Ausführung unterschiedlicher Aufgaben in beiden Bereichen. Daher bieten diese Bereiche die meisten Möglichkeiten für den Einsatz von Dialogsystemen.

4.2.1 Steuerung von Soft- und Hardware

In den letzten Jahrzehnten wurden immer mehr mechanische Geräte in beiden Bereichen (Unterhaltung und produktive Tätigkeiten) durch elektronisch gesteuerte Äquivalente ersetzt, was sich auf die Effizienz allgemein positiv auswirkt. Grund dafür ist die Entwicklung der Computertechnologie. Bei der Bedienung von Maschinen, geht es um die Steuerung der letzteren. Die Steuerung durch Hebel und Knöpfe wird durch Soft-

ware und Tasten, die s.g. Benutzeroberfläche (UI - *user interface*) oder die graphische Benutzeroberfläche (GUI), ersetzt. Die Benutzeroberfläche kann über die Eingabe- und Ausgabegeräte, meistens die Konsole (Tastatur/Bildschirm), bedient werden. Die Software hilft dem Menschen effizient die Hardware zu steuern und somit die Maschine für eine oder mehrere Aufgaben zu verwenden. Aus dieser Perspektive ist klar, dass es am einfachsten wäre, diese graphische Benutzeroberfläche durch eine, die gesprochene Sprache unterstützt, entweder komplett zu ersetzen oder zu erweitern, so dass die Steuerung (die Befehleingabe) und die Systemrückmeldungen nicht mehr über Tastatur (und Maus, oder einen interaktiven Bildschirm) läuft, sondern über ein Sprachmodul, welches in dem System integriert ist wie z.B. DICO (Tab.:A.36), WITAS (Tab.:A.38), D'Homme (Tab.:A.39). Man versucht auch universelle Benutzeroberflächen für gesprochene Sprache wie USI (*universal speech interface*) (Tab.: A.37) zu entwickeln.

Betrachtet man also die Bereiche in denen Maschinen eingesetzt werden, so zeigt sich die Informationsverarbeitung in allen Aspekten besonders attraktiv: Auskunft, Dokumentenverwaltung, Organisation, Terminplanung, Kommunikation etc.

4.2.2 Auskunft

Die meisten Systeme werden als Auskunftssysteme entwickelt, vielleicht auch deswegen, weil die DARPA Initiative die Forschung in diese Richtung unterstützt hat (P. A.1.1.1). Es zeichnet sich ab, dass dies hauptsächlich Informationsdienste sind, wie Auskunft für den Flug-, Bahn-, Bus-, Schiffverkehr, aber auch für den Immobilienmarkt, für das Wetter, die Vogelkunde und einige andere Bereiche, wie hier aufgelistet:

1. Immobilien: AdApt (Tab.:A.1), REA (Tab.:A.12)
2. Allgemeine Auskunft: Amities (Tab.:A.2), NJFUN (Tab.:A.8), OLGA (Tab.:A.11), HIGGINS (Tab.:A.19)
3. Einzelne Systeme
 - (a) Film/Kinoauskunft: MovieLine (Tab.:A.24)
 - (b) Telefonistin: How May I Help You (Tab.:A.20)
 - (c) Wetterauskunft: Jupiter (Tab.:A.5)
 - (d) Auskunft an der Rezeption: Bayesian Receptionist (Tab.:A.6)
 - (e) Sportauskunft (NBA): NBA Game Update Line (Tab.:A.7)
 - (f) Gebrauchtwagenhandel: Bildata (Tab.:A.9)

(g) Vogelkunde: BirdQuest (Tab.:A.10)

4. Flugauskunft (Air Travel Information service - ATIS) und Reiseauskunft

- ▷ PEGASUS (Tab.:A.14)
- ▷ Voyager(Tab.:A.18)
- ▷ Die DARPA Initiative
 - ▷ AT&T (Tab.:A.22)
 - ▷ CMU (Tab.:A.23)
 - ▷ Colorado University (Tab.:A.25)
 - ▷ Bell Labs (Tab.:A.26)
 - ▷ SRI (Tab.:A.27)

5. Öffentliche Verkehrsmittel: BusLine (Tab.:A.3), Busman (Tab.:A.4), OVIS (Tab.:A.13), Let's Go (Tab.:A.16), WAXHOLM (Tab.:A.17)

6. Auskunft über Konferenzen: ConQuest (Tab.:A.15), DiSCoH (Tab.:A.21)

An dieser Stelle ist vielleicht auch Clarissa (Tab.: A.41) zu nennen. Es ist ein System für die Unterstützung von Astronauten bei ihren täglichen Tätigkeiten im All:

"Just try to analyze a water sample while scrolling through pages of a procedure manual displayed on a computer monitor while you and the computer float in microgravity," challenges astronaut Michael Fincke, who recently completed a six-month stay on the ISS. "To be able to speak to the system and hear the step-by-step instructions while my hands are free to complete the procedure will be like having another crew member aboard."

(<http://www.nasaspaceflight.com/content/?cid=3011>)

4.2.3 Organizer, E-Mail, Multimodale Systeme

Ein Teil der Systeme kann organisatorische Tätigkeiten durchführen, um dadurch lästige und zeitraubende Aufgaben zu übernehmen. Die Systeme E-MATTER (Tab.:A.28), Athosmail/DUMAS (Tab.:A.29), Mailman (Tab.:A.30) können am Telefon E-Mails vorlesen, aufnehmen und verschicken, sind „quasi“ interaktive Sprachoberflächen für E-Mail.

Andere Systeme können so eine Art Sekretärin sein und für uns Termine vereinbaren, indem sie z.B. beim Arzt anrufen. In einer geschäftlichen Umgebung können sie Sitzungen vereinbaren und die Teilnehmer einladen und bei eventuellen Änderungen darüber informieren, wie es am Beispiel von Vera (Tab.:A.31) gezeigt wird.

Die multi-modalen Systeme wie SMARTKOM (Tab.: A.61) können unterschiedliche Tätigkeiten ausführen, wie z.B. TV und andere Geräte des täglichen Gebrauchs bedienen und steuern, E-Mails versenden und in der Büroarbeit unterstützen oder als Navigationssystem dienen, wenn man unterwegs ist.

4.2.4 Kontoverwaltung

Das CTT-Bank System (Tab.: A.40) wurde mit dem primären Ziel entwickelt, die Authentifizierung über gesprochene Sprache an einem Banksystem zu studieren. Die Idee dahinter ist, dass man z.B. einige Banktransaktionen durchführen kann, während man mit dem Auto unterwegs ist. Man kann im Auto keine Geräte für die visuelle Kommunikation mit einem System verwenden, daher bietet sich nur die Möglichkeit an, über Telefon solche Tätigkeiten durchzuführen. Für die Entwicklung des Systems wurde die Plattform ATLAS (Tab.: A.67) verwendet (Melin, 2001).

4.2.5 Problemlösungen, Routenplaner

Eine zentrale Rolle in der Entwicklung von Dialogsystemen spielt das TRAINS Projekt (Tab.:A.34) (Ferguson et al., 1996) und sein Nachfolger TRIPS (Ferguson und Allen, 1998) (Tab.:A.35). Es ist ein interaktives kollaboratives Planungssystem. Das beinhaltet, im Unterschied zu den Auskunftssystemen, eine Fähigkeit zu Argumentieren und Vorschläge zu machen. Das System wurde hauptsächlich dafür benutzt kollaboratives Verhalten zu studieren, indem der Transport verschiedener Ladungen mit der Bahn simuliert wurde. Unten ist ein Transkript eines Dialogs <http://www.cs.rochester.edu/research/cisd/resources/93dialogs/d92a-4.1> mit TRAINS-93 (Traum und Allen, 1994, S.3), um diese Idee zu veranschaulichen.

Dialogue: d92a-4.1
 Number of utterances files: 30
 Length of dialogue: 82.862090
 Estimated number of turns: 24
 -
 utt1: s: hello <sil> can I help you
 utt2: u: <click> <sil> uh <sil> alright <sil> um I'm not sure <sil> <laughter>
 <sil> uh <sil> alright
 utt3: um <sil> I have to ship a boxcar of oranges to Bath by eight a.m.
 <sil> and it's midnight <sil> okay
 utt4: so <sil> <click> <sil> the orange warehouse is in Corning
 utt5: s: right
 utt6: u: okay so I need an <sil> engine
 utt7: um <sil> I guess <sil> <noise> <sil> I take engine <sil> two from
 <sil> Elmira
 utt8: s: mm-hm
 utt9: u: okay how long is it then from Elmira to Corning
 utt10: s: uh <sil> two hours
 utt11: u: two hours <sil> okay <sil> and how long is it from Corning to Bath
 utt12: s: uh <sil> two hours
 utt13: u: two hours
 utt14: s: yep
 utt15: u: alright so <sil> can I take an engine from Elmira <sil> to Corning
 <sil> um <sil> pick u(p)- <sil> oh okay <sil> alright how far is it
 from Corning to <sil> Dansville
 utt16: s: uh one hour
 utt17: u: okay so can I take an engine from <brth> Elmira <sil> to Corning
 <sil> to Dansville which should take three hours
 utt18: s: uh-huh
 utt19: u: uh <sil> pick up a boxcar <sil> <click> <sil> take another hour
 back to Corning
 utt20: s: + right +
 utt21: u: + which will be + four hours <sil> and fill it up <sil> and then
 bring it to Bath
 utt22: s: uh yes <sil> that'll + get you there at <sil> uh +
 utt23: u: + oh <sil> h- six + uh
 utt24: s: well seven <sil> because + it takes + an hour <sil> to load the
 oranges
 utt25: u: + seven +
 utt26: okay <sil> um
 utt27: s: + so + I <sil> think your <sil> + problem's + solved + <laughter> +
 utt28: u: + i- +
 utt29: + u- +
 utt30: + okay + <sil> great <sil> <laughter>

Beispiel 12: TRAINS-93

4.2.6 Reparaturassistenten

Interaktive Dialogsysteme wie LARRI (Tab.:A.32) oder Intelligent Procedure Assistant (IPA) (Tab.:A.33) bieten Unterstützung bei Wartungsarbeiten. Im Fall von LARRI ist es

das F18 Flugzeug und IPA sollte Astronauten bei Wartungsarbeiten auf der ISS unterstützen. Beide basieren auf RavenClaw/Olympus (Tab.: A.60). IPA wurde als Basis benutzt, um die Clarissa (Tab.: A.41) Plattform zu entwickeln, die dann tatsächlich auf der ISS installiert wurde, allerdings aber eine andere Dialogmanagerarchitektur implementiert.

Der allgemeine Ablauf der Interaktion mit LARRI ist Authentifizierung, Anleitung zur Durchführung bestimmter Aufgaben und Abmelden vom System, wie am folgenden Beispiel gezeigt (nach Bohus und Rudnicky, 2002, S.4).

```
1: L: Hi, this is LARRI, State your name please
2: U: Hi, my name is John Doe
3: L: Hello John Doe. Before you start on your scheduled work, we need
    you to finish an urgent task
4: U: OK
5: L: Let's reprioritize your schedule: first you will perform the INS Ground
    Initiated BIT Test on plane 301. You last performed this task 3 months
    ago. Is that correct?
6: U: Yes, that's right.
7: L: Then you will do the BXF radar check on plane 304. You last performed
    this task 2 weeks ago. Is that correct?
8: U: Affirmative
9: L: When you are ready to start working on the INS Bit Test, please say
    ready to work
10: U: Ready to work
    ... ..
11: L: Step A: On RDDI press and release the MENU pushbutton switch until HSI
    pushbutton switch until HSI pushbutton switch option appears
12: U: Done
13: L: Does the RDDI menu display the HSI?
14: U: Affirmative
15: L: Step B: Now press the HSI pushbutton switch.
16: U: Ok. RDDI displays the HSI
17: L: Ok. Step C: Press the DATA pushbutton switch
18: U: <long timeout>
19: L: Did you push the data button?
20: U: Yes, I just did
21: L: Does the RDDI display DATA option display with the WYPT option boxed?
22: U: Yes
23: L: OK. Step D: Select WYPT 0
24: U: Okay. WPT is 0 and the latitude is local
24: L: Good. Skipping to step J. On RDDI press and release MENU pushbutton
    switch until BIT pushbutton switch option appears.
26: ... ..
```

Beispiel 13: LARRI

4.2.7 Unterstützung der Mensch-Mensch-Interaktion

Das Ziel des CHIL (Tab.: A.42) Projekts ist eine fundamental neue Methode der Computeranwendung zu erschaffen, indem der Computer die Unterstützung der zwischenmenschlichen Kommunikation und Interaktion übernimmt. Das bedeutet, dass die Maschinen im Hintergrund bleiben und die menschlichen Tätigkeiten beobachten und nur dann aktiv werden, wenn ihre Hilfe gebraucht wird. *Computers in the Human Interaction Loop* (CHIL) bedeutet, dass die Computer in dem Kreislauf der zwischenmenschlichen Kommunikation betrachtet werden, im Gegensatz zu der jetzigen Situation, wo Menschen in einem Kreislauf von Maschinen arbeiten müssen.

Das CHIL Projekt ist ein umfangreiches Projekt, welches mehrere Teilprojekte mit vielen prominenten Teilnehmerinnen umfasst. Einige der angewendeten Technologien sind Multi-Sensoren, multi-modale Verarbeitung für robuste Identifikation und Verfolgung von menschlichen Objekten unter unterschiedlichen unvorhersagbaren Umständen (akustisch, visuell, nicht-frontal, etc.), Erkennung von Gesten und Mimiken und Körpersprache u.a. Ich verweise auf die Webseite des Projekts (Tab.: A.42), wo weitere Information zu finden ist.

4.3 Tutoring

Ein besonders interessanter Bereich ist die Vermittlung von Wissen. Meistens sind es Kinder, die lernen, doch auch Erwachsene greifen ständig auf diese Fähigkeit zurück. Daher ist die Domäne des Spracherwerbs für Kinder und Erwachsene mit Sicherheit eine vielversprechende. Mit diesem Ziel wurden LISTEN (Tab.: A.43) und NICE (Tab.: A.44) entwickelt. NICE zeichnet sich dadurch aus, dass zum ersten Mal Computerspielentwickler in solche Tätigkeiten eingebunden werden, so dass eine virtuelle Welt, gestaltet nach H. C. Andersen's Märchen, zum Lernen entsteht. ITSPOKE (Tab.: A.46) und MURI (Tab.: A.45) sind für Studierende bestimmt und „The Berkeley UNIX Consultant Project“ (Tab.: A.47) kann jemandem erklären, wie das UNIX Betriebssystem funktioniert bzw. alle mögliche Fragen in diesem Zusammenhang beantworten. Auch hier versucht man die virtuelle TutorIn mit einer 3D animierten Figur zu personifizieren (Ville, Tab.: A.48).

4.4 Robotik

Eine Idee in der Robotik ist es mit den Robotern mehr oder weniger wie mit Menschen zu kommunizieren. Das Ziel des von der EU unterstützten Projektes Cogniron (Tab.:

A.50) ist es, eine solche Oberfläche zu erschaffen, die diese Kommunikation ermöglicht. Das ist allerdings nur ein Teil der kognitiven Tätigkeit des Menschen, die ein Roboter besitzen muss, um als menschliche Begleitung Anwendung zu finden. Das Projekt befasst sich also mit all diesen Aspekten als Ganzes. Ein Beispiel dafür ist die Sprache in einem Dialog. Sie bezieht sich auf die Umwelt, also verwenden wir deiktische Ausdrücke, mit denen wir auf etwas verweisen. Um solche Ausdrücke zu verstehen, muss ein Roboter Weltwissen besitzen und eine Fähigkeit Kopf- und Händebewegung zu verfolgen und zu analysieren, sowie mehrere andere für die Menschen selbstverständliche Fähigkeiten (Burger et al., 2008) z.B. die Fähigkeit, Neues zu lernen (Otero et al., 2008).

4.5 Übersetzungen und Dolmetschung

Als letztes Beispiel möchte ich das System VerbMobil (Tab.: A.49) anführen, welches im Bereich der Dolmetschung, der Übersetzung gesprochener Spontansprache, in zwei Domänen tätig ist: die Terminvereinbarung und die Planung von Geschäftsreisen (Alexandersson, 2003). Das Interessante an dem System ist, dass der Dialogmanager keine typische Funktion erfüllt, denn das System darf sich, wenn möglich, nicht in die Konversation einmischen. Es muss aber die Äußerungen in Evidenz halten, sowie die Stellung, die dabei von den Dialogpartnerinnen eingenommen wurde. Daher spielt das VerbMobil eine Vermittlerrolle zwischen den Dialogpartnerinnen, die unterschiedliche Sprachen sprechen und hilft, sich zu verständigen. Als Beispiel wird hier ein Transkript (Alexandersson, 2003, S.77) angeführt.

```
(37) ja ich kenne mehrere Hotels in Hannover  
    I know several hotels in Hanover  
(38) good  
    can you make reservations for us there?  
(39) können Sie das bitte wiederholen?  
    could you please repeat that?  
(40) yes  
    will you please make reservations at a hotel?  
(41) ja  
    yes  
(42) welches Hotel von den dreien möchten wir nehmen?  
    which one of the three hotels should we take?  
(43) you choose .  
(44) ich schlage das Hotel Luisenhof vor  
    I suggest the hotel Luisenhof  
(45) good  
(46) I trust your choice  
(47) why don't we meet at the station on Wednesday?
```

Beispiel 14: VerbMobil

Kapitel 5

Schlußbetrachtung

In der vorliegenden Diplomarbeit wurden drei Hauptfragen behandelt: „was ist Dialog“, „welche charakteristische Merkmale hat der Dialog“ und „wie kann Dialog von Computersystemen erfolgreich simuliert werden“. Die letzte Fragestellung setzt voraus, dass es zwei Einheiten, das simulierende und das simulierte gibt, die Maschine und den Dialog. Die zwei Einheiten sind unterschiedlicher Natur, denn der Dialog bezieht sich auf Mensch-Mensch-Kommunikation und die Computer werden als technische Geräte, als Rechner verwendet. Die Frage ist trotzdem berechtigt, denn Computer sind im Stande, so Turing, jedes nur erdenkliche Problem zu lösen, wenn das Problem in Form eines Algorithmus dargestellt wird. Findet man also ein formales Modell, welches den Dialog beschreibt, so kann man auch dieses formale Modell von einem Computer ausführen lassen. Dies erweist sich allerdings als ziemlich schwierig, denn manche Aspekte der zwischenmenschlichen Kommunikation, vor allem, jene die mit der kognitiven Tätigkeit des Menschen zu tun haben, nicht vollständig untersucht sind und dementsprechend nicht formalisierbar sind. Immerhin zeigen die Versuche, solche Systeme zu gestalten, dass durch geeignete Algorithmen ziemlich brauchbare Ergebnisse erzielt werden können, was die vorliegende Arbeit durchaus berechtigt.

Die erste hier gestellte Frage kann also folgender Weise beantwortet werden: der Dialog ist eine pragmatische Anwendung der Sprache in der zwischenmenschlichen Kommunikation, wobei es darum geht, Verständigung zu erreichen - etwas mit einer Bedeutung zu äußern, so dass diese auch verstanden wird. Ein großer Teil dieser Kommunikation erfüllt die vier Kommunikationsmaximen von Grice (Quantität, Qualität, Relation, Modalität). Die Bedeutung der Äußerungen kann nicht unbedingt aus der Bedeutung der einzelnen Wörter abgeleitet werden, sondern beruht auf Konventionen und Regeln. Menschen lernen diese Bedeutung, indem sie die Sprache lernen. Damit Maschinen mit Bedeutung von Äußerungen umgehen können, werden sowohl Regeln, als auch Korpora und stochastische Verfahren eingesetzt, die meistens

auf eine sprachliche Handlung aus dem Gesagten schliessen, oder die s.g. Benutzer-Intention herauszufinden versuchen. Dasselbe gilt für den umgekehrten Prozess. Für eine System-Intention wird eine treffende Äußerung gefunden, die wiederum von der TTS Komponente verbalisiert wird.

Die zweite Frage - „wie funktioniert Dialog?“ hat zwei Aspekte - formale und kognitive und kann am besten entweder mit der Sprechakttheorie, mit der Spieltheorie oder mit der Theorie der rationalen Handlungen beantwortet werden. Der Dialog ist demzufolge eine Reihe von zusammenhängenden sprachlichen Handlungen oder Sprachspielen, die auf das Erreichen eines kommunikativen Ziels ausgerichtet sind.

Die menschliche Dialogfähigkeit zeichnet sich durch die Fähigkeit aus, typische Organisationsprinzipien und Muster zu identifizieren und anzuwenden. Die Prinzipien betreffen einerseits die kognitiven Fähigkeiten des Menschen wie die Fähigkeit zu folgern, rational zu handeln, Intentionen zu verstehen, Themen zu erkennen, unterschiedliche thematische Inhalte zuzuordnen, zu referieren, durch Identifikationswissen Referenzen zu identifizieren, die dahinter liegenden Intentionen zu verstehen und bestimmte sprachliche Handlungen durchzuführen. Die sprachlichen Handlungen können in direkte und indirekte geteilt werden; in solche, bei denen man direkt auf die Sprecher-Bedeutung schließen kann und in solche, die auf eine Bedeutung ausgerichtet sind, die aber implizit ist und zuerst aufgeschlüsselt werden muss. Dabei spielt das Prinzip der Relevanz eine wesentliche Rolle, bei der Entscheidung was, wann und wie gesagt wird, wie auch bei der Identifikation von Referenzen und Referenzträgern und bei der Formulierung des Themas (und ggf. der Subthemen). Andererseits zeichnet sich der Dialog durch unterschiedliche formale Aspekte aus, wie durch die Teilnahme mindestens zweier Teilnehmerinnen, Anfang und Ende des Dialogs, die sequentielle Abfolge der Dialogbeiträge (Sprecherwechsel oder Dialogzug genannt), die Fähigkeit syntaktisch lexikalische Mittel zu verwenden, Textkohärenz durch diese Mittel zu produzieren u.a.

Vor allem die Sprechakttheorie und die Dialogspieltheorie geben ein brauchbares Modell für die Dialoganalyse und bieten eine Möglichkeit, formal die Dialogphänomene zu beschreiben. Die Struktureinheiten nach Carlson, bzw. die sprachlichen Handlungen nach Searle können folgender Weise zusammengefasst werden. Die allgemeine Struktureinheit im Dialog ist der Dialogzug (auch Sprecherwechsel genannt), der sich durch eine oder mehrere Äußerungen seitens einer der Gesprächspartnerinnen auszeichnet. Jeder Dialogzug hat bestimmte Parameter: Sprecher, Adressat, Publikum, Prämisse etc. Eine Äußerung kann ironisch oder ernsthaft gemeint sein. Die Äußerungen besitzen meistens, so Searle, eine illokutionäre Kraft, die den Typ der sprachlichen Handlung bestimmt. Typische sprachliche Handlungen sind: Frage, Antwort, Begrüßung, Entgeg-

nung, Beschuldigung u.a. Bestimmte syntaktisch lexikalische Mittel sprechen für bestimmte kognitiv-logische Operationen wie Vermutung, Annahme, Folgern, Erklären, Argumentieren. Dialoge zeichnen sich noch dadurch aus, dass sie im Rahmen eines anderen Dialogs auftreten können, d.h. sie können eingebettet werden (Subdialoge, *side-sequences*).

Die formalen Aspekte des Dialogs, im Unterschied zu den kognitiven, sind meistens gut erforscht. Eine wesentliche Herausforderung für die Wissenschaft ist zu erklären, wie Menschen mit Information und Wissen umgehen. Die formale Beschreibung der letzteren kann nur durch die Anwendung mathematisch-logischer Modelle abgebildet werden, die die Dialogfähigkeit und die damit verbundene kognitive Tätigkeit des Menschen formal beschreiben können. Diese logischen Modelle zeichnen sich, soweit Forschung in diese Richtung betrieben wurde, durch Nicht-Monotonie und Modalität aus, und werden mit Logiken wie die Default- und die modale Logik beschrieben. Diese logischen Modelle bieten z.B. eine Möglichkeit, Schlussfolgerungen zu revidieren und zurückzuziehen, Default-Annahmen zu definieren und mit modalen Propositionen umzugehen.

Die dritte in dieser Diplomarbeit gestellte Frage - „wie kann Dialog von Maschinen erfolgreich simuliert werden?“, kann folgender Weise beantwortet werden: Interaktive Dialogsysteme bestehen mindestens aus Komponenten, die die Sprachrezeption, die kognitive Tätigkeit (das Verstehen und das Generieren von Intentionen) und die Sprachproduktion des Menschen abbilden. Diese sind die Spracherkennung (ASR), die Identifikation von Intentionen (NLU), der Dialogmanager (DM), Umsetzung der Systemintentionen (NLG), welche grammatisch richtige Sätzen anhand von Systemintentionen generiert und die Sprachsynthese (TTS). Weitere Komponenten können z.B. Gesten und Mimiken verarbeiten und dadurch die Qualität der anderen Komponenten positiv beeinflussen. ASR, NLU, DM und NLG arbeiten sehr eng zusammen, denn es geht dabei um das Verstehen und Generieren von Äußerungen, die sich auf das Gesprächsthema und auf die Interaktion mit den Benutzerinnen beziehen. Die TTS sollte auch in Kooperation mit den anderen Komponenten arbeiten, um mögliche Fehler bei der Umsetzung der Systemintentionen oder bei der Sprachgenerierung zu vermeiden, festzustellen und ggf. auszubessern. Je genauer die Komponenten die Dialogfähigkeit des Menschen abbilden, desto menschenähnlicher wirkt ein System und desto höher ist die Benutzerzufriedenheit. Um diese unterschiedlichen Dialogfähigkeiten erfolgreich abbilden zu können, bedient man sich grundsätzlich mathematisch-logischer Modelle, statistischer Verfahren und Korpora.

Der logische Aufbau der Dialogsysteme betrifft die Entscheidung, mit was für Wissen das System umgehen wird (z.B. Ort und Zeit der Zugankunft oder -abfahrt) und

in was für einem Bereich (*domaine*) es tätig sein wird (z.B. Bahnauskunft). Diese Entscheidungen wirken sich auf die Flexibilität und Größe der vorbereiteten Datenbanken und auf den Umfang der Algorithmen aus. Weitere Überlegungen betreffen die Initiative, also wer die Konversation führen, bzw. wer die Fragen stellen darf, und den theoretischen Ansatz zur Abarbeitung von Dialogphänomenen. Einfache Systeme verwenden die Systeminitiative und unterstützen typischer Weise Ein-Wort-Dialoge (Dialoge mit elementaren Fragen) mit größeren statistischen Thesauri und kleineren Sets von Regeln. Sie arbeiten meistens in kleinen geschlossenen Domänen. Komplexere Systeme verwenden gemischte Initiative und unterstützen meistens kleinere dynamische Thesauri, die je nach bedarf geladen werden können und auf welche eine größere Zahl von Regeln angewendet werden.

Viel wurde besonders bei der Architektur des Dialogmanagers, also die Art und Weise wie Dialog verarbeitet wird, experimentiert. Das ist auf die Tatsache zurückzuführen, dass einige Wissenschaftlerinnen mehr durch die Sprechakttheorie, andere von der Dialogspieltheorie, dritte aber von der Theorie der rationalen Handlungen beeinflusst wurden. Daher werden manche Systeme so konstruiert, dass sie mit Zielen und Überzeugungen umgehen, andere mit Plänen, weitere mit Aufgaben usw. Die Entscheidungen über das Design werden durch das Ziel der Anwendung geleitet. Im Allgemeinen setzen sich Systeme durch, die Aufgaben-orientiert arbeiten, die eine gemischte Initiative unterstützen und multi-modal sind. Das bedeutet, dass sowohl die Benutzerin als auch das System nach bedarf Fragen stellen und beantworten können und dass das System (multi-modal) in unterschiedlichen Umgebungen (z.B. zu Hause, im Büro, unterwegs) und mit unterschiedlichen Geräten und Anwendungen arbeiten kann (z.B. E-Mail, Terminkalender, Videorecorder, Scanner etc). Eine sehr breite Anwendung finden die s.g. Agenten, die autonome Einheiten darstellen und meistens austauschbar unter den Systemen sind.

Die Grundüberlegungen bei dem Design eines solchen Systems betreffen vor allem die Architektur des Dialogmanagers in folgenden Aspekten: Initiative, Hypothesen (Voraussagen was demnächst folgen könnte), Systemfokus (welche Information, wann und wie geliefert werden soll), Rezipienzsignale (wann wäre angebracht, die Benutzerin über die Vorgänge zu informieren). Die meisten Systeme, so wie die Menschen, verfolgen und speichern den Dialogverlauf, um ggf. auf Stellen, die schon besprochen wurden, referieren zu können. Seltener können (vor allem neuere) Systeme Unterbrechungen (*barge-in*) verarbeiten und adäquat reagieren.

Ein wichtiger Aspekt der Gestaltung von Dialogsystemen ist die Anwesenheit von Sprechermodellen. Man hat festgestellt, dass Menschen in bestimmten Situationen ein solches Modell mit sich tragen, oder ganz schnell ein solches aufbauen. Dies erweist

sich mehrfach als vorteilhaft. Einerseits können dadurch stochastische Modelle genauer formuliert werden, andererseits können bestimmte Gewohnheiten der Benutzerin, wie Defekte in der Aussprache oder korrigierte Fehler gespeichert werden.

Eine weitere wichtige Frage im Bezug auf die Gestaltung der Dialogsysteme ist ihre Evaluierung. Eine gelungene Evaluierungsmethode des Systems oder einzelner Systemkomponenten, könnte die Performanz des gesamten Systems steigern, indem bestimmte Parameter justiert werden. Die unterschiedliche Natur der Systeme führt dazu, dass es praktisch unmöglich ist, Systeme miteinander zu vergleichen. Unter den Vorschlägen, wie das möglich wäre, zeichnet sich das PARADISE Verfahren aus, als eine Richtlinie für die Evaluierung von interaktiven Dialogsystemen, um die s.g. Leistungsfunktion zu bestimmen, wodurch Systeme auf einen gemeinsamen Nenner gebracht werden können. Das Verfahren umfasst eine Definition einer Aufgabe und eines Sets von Szenarios, Spezifikation einer *Attribute-Value-Matrix* (AVM) Repräsentation der Aufgaben, Experimente mit alternativen Dialogagenten für die Erfüllung der Aufgabe, Berechnung der Benutzerzufriedenheit anhand von Umfragen, des Aufgabenerfolges, der Kosten für den Dialog mit Effizienzmessungen und qualitativen Messungen, Schätzung einer Leistungsfunktion mittels linear rückschreitender Methode, Werte für die Benutzerzufriedenheit und die Dialogkosten, Vergleich mit anderen Agenten und Aufgaben, um die Faktoren zu bestimmen, die am meisten Gewicht haben, um zu einer generellen Leistungsfunktion zu gelangen, Verfeinerung des Leistungsmodells. Die Messungen bei der Evaluierung können objektiv und subjektiv sein, denn in vielen Fällen wird menschliche Beurteilungskraft benötigt, um bestimmte Aspekte beurteilen zu können. Dies führt wiederum zu Komplikationen, denn es sollte doch letztendlich alles objektiv sein. Unterschiedliche Experimente wurden auch in diesem Bereich durchgeführt und am besten eignen sich solche Verfahren, die den Kappa-Koeffizienten verwenden, der eine statistische Messung ist. Diese unterscheidet sich von der reinen Auswertung nach Prozentsätzen, indem die Übereinstimmung der Beurteilenden berücksichtigt wird.

Die existierenden Systeme sprechen dafür, dass ein gewisses Interesse für solche Anwendungen präsent ist. Die „Evolution“ der Systeme in den letzten Jahrzehnten deutet darauf hin, dass theoretische und empirische Modelle für Implementierungen vorbereitet werden in der Hoffnung, dass sobald die Technik die notwendige Preis/Leistungs-Relation erreicht hat, diese Modelle auch vermarktet werden können. Das kommerzielle Interesse für diese Systeme ist positiv für Sprachwissenschaftlerinnen, unter anderem auch Germanistinnen, die in unterschiedlichen Bereichen des Systemdesigns und der Systementwicklung tätig sein könnten.

Wie es ersichtlich wird, werden derzeit solche Dialogsysteme in dem Bereich der Auskunft verwendet, der eigentlich mit Datenbankabfragen zu tun hat. Bei vielen dieser

Systeme ist die Fähigkeit vorhanden, auch Auto- und Hotelreservierungen durchzuführen und komplexe Routen, unter Berücksichtigung vieler Kriterien zu planen. An zweiter Stelle kommen Assistenten im Bürobereich, so etwas wie virtuelle Sekretärinnen, die Termine, elektronische Post und andere typische Tätigkeiten wie z.B. Erinnerungsfunktion (über Telefonanruf) ausführen können. Weniger verbreitet sind Systeme, die bestimmte Informationen bezüglich technischer Daten liefern können, die meistens in Umgebungen eingesetzt werden, wo die Benutzerin, ihre Hände benötigt, um Reparaturaufgaben durchzuführen, dabei aber gewisse Details nachschlagen soll. Ein weiterer Bereich der Anwendung solcher Systeme ist die Robotik, mit dem Ziel Maschinen für die Bedienung einfacher und benutzerfreundlicher zu machen. Abgesehen von den menschenähnlichen Robotern, die leider nur in den Science-Fiction-Filme zu finden sind, kann man z.B. in Prototypen mancher Autohersteller die Stereoanlage, das Licht und andere Geräte (z.B. Mobiltelefon, Spiegel, etc) über gesprochener Sprache steuern.

Ein weiterer Bereich für den Einsatz von interaktiven Dialogsystemen ist das Tutoring und der Bereich Übersetzung und Dolmetschung. Man könnte sich vorstellen, dass interaktive Dialogsysteme als Tutoren sehr günstig und effizient Wissen vermitteln können, besonders in einer Zeit, wo es immer mehr zu wissen gibt. Dasselbe gilt auch für Übersetzungen und Dolmetschungen im Hinblick auf die globalisierte Welt und der Mehrsprachigkeit der Europäischen Union. Systeme in diesen Bereichen sind für den Erfolg vorbestimmt, denn durch sie wären viele Aufgaben einheitlicher und vor allem günstiger erfüllt.

Literaturverzeichnis

- Alexandersson, Jan: *Hybrid Discourse Modeling and Summarization for a Speech-to-Speech Translation System*. Dissertation, Universität des Saarlandes 2003.
- Allen, J.; D. Byron; M. Dzikovska; G. Ferguson und L. Galescu: Towards Conversational Human-Computer Interaction. *AI Magazine*, 22(4):27–37 2001.
- Allen, James F.: The TRAINS Project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI (JETAI)*, 7:7–48 1995.
- Allen, James F. und C. Raymond Perrault: Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178 1980.
- Andersen, C.; D. Traum; K. Purang; D. Purushothaman und D. Perlis: Mixed initiative dialogue and intelligence via active logic. In: *In Proceedings of the AAAI'99 Workshop on Mixed-Initiative Intelligence*, 1999. 1999.
- Anttila, A. T.: Variation and phonological theory. In: *Handbook of Language Variation and Change*, Hg. J. K. Chambers; Peter Trudgill und Natalie Schilling-Estes, Blackwell, Oxford 2001.
- Austin, John: *How to Do Things with Words*. Cambridge, MA: Harvard University Press 1962.
- Baum, Micha; Gregor Erbach; Markus Kommenda; Georg Niklfeld und Estela Puig-Waldmüller: Speech and Multimodal Dialogue Systems for Telephony Applications based on a Speech Database of Austrian German. *OGAI*, 20(1):29–34 Jan. 2001.
- Bernsen, Niels Ole; Laila Dybkjær und Hans Dybkjær: A dedicated task-oriented dialogue theory in support of spoken language dialogue system design 1994.
- Beskow, Jonas: *Animation of Talking Agents* 1997.
- Bilange, E.: A Task Independent Oral Dialogue Model. In: *Proceedings of the 5th. Meeting of the European Chapter of the Association for Computational Linguistics*. S. 83–87 1991.

Blaylock, Nate und James Allen: Statistical goal parameter recognition. In: *14th International Conference on Automated Planning and Scheduling (ICAPS'04)* jun 3-7 2004.

Bohus, Dan und Alex Rudnicky: LARRI: A language-based maintenance and repair assistant. *IDS-2002*, paper 44:13 June 17-19 2002.

Bohus, Dan und Alexander I. Rudnicky: Constructing accurate beliefs in spoken dialog systems. *Workshop on Automatic Speech Recognition and Understanding, 2005 IEEE*, S. 272–277 27.Nov.-1.Dec. 2005.

Bos, J.; E. Klein; Oliver Lemon und T. Oka: DIPPER: Description and Formalisation of an Information-State Update Dialogue System Architecture 2003.

Bremes, Christian: Philosophie der Bedeutung: Bedeutung als Bestimmung und Bestimmbarkeit, Bd. 1 von *Trier Studien zur Kulturphilosophie*. Königshausen & Neumann 1997.

Bretier, Philippe und David M. Sadek: A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. In: *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, Hg. Jörg P. Müller; Michael Wooldridge und Nicholas R. Jennings. Springer, Bd. 1193/1997 von *Lecture Notes in Computer Science*, S. 189–203 1997.

Brinker, Klaus: Linguistics of Text and Conversation, In: *Text- und Gesprächslinguistik*, Hg. Klaus Brinker; Gerd Antos; Wolfgang Heinmann und Sven F. Sager, Walter de Gruyter, Berlin - New York, Bd. 1, Kap. III, S. 164–175 2000.

Burger, Brice; Isabelle Ferrané und Frédéric Lerasle: Multimodal Interaction Abilities for a Robot Companion. In: *Int. Conf. on Computer Vision Systems (ICVS'08)* 2008.

Carletta, J.; S. Isard; Doherty G. Sneddon; A. Isard; J. C. Kowtko und A. H. Anderson: The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31 1997.

Carlson, Lauri: Dialogue games: An Approach To Discourse Analysis, Bd. 17. D.Reidel Publishing Company, Dordrecht: Holland 1983.

Chu-Carroll, J. und Sandra Carberry: Conflict Resolution in Collaborative Planning Dialogues. *International Journal of Human-Computer Studies*, 53(6):969–1015 2000.

- Chu-Carroll, J. und J. Nickerson: Evaluating automatic dialogue strategy adaptation for a spoken dialogue system 2000.
- Cohen, Philip R. und Hector J. Levesque: Rational interaction as the basis for communication. In: *Intentions in Communication*, Hg. Philip R. Cohen; Jerry Morgan und Martha E. Pollack, MIT Press, S. 221–256 1990.
- Cohen, Philip R. und C. R. Perrault: Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212 1979.
- Cohen, William W.: Abductive Explanation-Based Learning: A Solution to the Multiple Inconsistent Explanation Problem. *Machine Learning*, 8:167–219 1992.
- Constantinides, Paul C. und Alexander I. Rudnicky: Dialog Analysis In The Carnegie Mellon Communicator 1999.
- Cuayáhuítl, Heriberto: *Spoken Dialogue Management Using Hierarchical Reinforcement Learning and Dialogue Simulation*. Dissertation, University of Edinburgh, School of Informatics, Institute for Communicating and Collaborative Systems Okt. 2005.
- Damasio, Antonio: Descarte's Irrtum: Fühlen, Denken und das menschliche Gehirn. Berlin: List 1995.
- Damasio, Antonio: Der Spinoza-Effekt: Wie Gefühle unser Leben bestimmen. Berlin: List 2005.
- Danieli, Morena und Elisabetta Gerbino: Metrics for Evaluating Dialogue Strategies in a Spoken Language System. In: *Proceedings of the 1995 AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*. S. 34–39 1995.
- Eckert, W.; T. Kuhn; H. Niemann; S. Rieck; A. Scheuer und E.G. Schukat-Talamazzini: A Spoken Dialogue System for German Intercity Train Timetable Inquiries. In: *EUROSPEECH*. Berlin, S. 1871–1874 1993.
- Elhadad, Michael: *Using argumentation to control lexical choice: a functional unification-based approach*. Dissertation, Graduate School of Arts and Sciences, Columbia University 1992.
- Farfán, Fernando; Alberto Portilla und Heriberto Cuayáhuítl: Evaluating Dialogue Strategies In A Spoken Dialogue System For Email. *Artificial Intelligence and Applications*, S. 403–104 Sept.Okt. 2003.

- Ferguson, George und James F. Allen: TRIPS: An Integrated Intelligent Problem-Solving Assistant. In: *In Proc. 15th Nat. Conf. AI*. AAAI Press, S. 567–572 1998.
- Ferguson, George M.: *Knowledge Representation and Reasoning for Mixed-Initiative Planning*. Dissertation, University of Rochester 1995.
- Ferguson, George M.; James F. Allen; Brad W. Miller und Eric K. Ringger: *The Design and Implementation of the TRAINS-96 System: A Prototype Mixed-Initiative Planning Assistant*. Techn. Ber. TN96-5, The University of Rochester 1996.
- Flammia, G.: *Discourse segmentation of spoken dialogue: An empirical approach*. Dissertation, MIT 1998.
- Fritz, Gerd: Kohärenz. Grundfragen der Linguistischen Kommunikationsanalyse. Narr. Tübingen 1982.
- Fritz, Gerd: Formale Dialogspieltheorien, In: *Handbuch der Dialoganalyse*, Hg. Gerd Fritz und Franz Hundsnurscher, Tübingen: Niemeyer, S. 131–152. XV, 576 S. ; 23 cm : Ill., graph. Darst. Literaturangaben 1994a.
- Fritz, Gerd: Grundlagen der Dialogorganisation, In: *Handbuch der Dialoganalyse*, Hg. Gerd Fritz und Franz Hundsnurscher, Tübingen: Niemeyer, S. 177–201. XV, 576 S. ; 23 cm : Ill., graph. Darst. Literaturangaben 1994b.
- Garson, James W.: *Modal logic for philosophers*. Cambridge Univ. Press 2006.
- Georgila, Kallirroi und Oliver Lemon: Adaptive Multimodal Dialogue Management Based on the Information State Update approach. In: *Online Proceedings of W3C Workshop on Multimodal Interaction* Juli 2004.
- Glass, James R.: *Challenges for Spoken Dialogue Systems* 1999.
- Gloning, Thomas: Praktische Semantik und Linguistische Kommunikationsanalyse, In: *Handbuch der Dialoganalyse*, Hg. Gerd Fritz und Franz Hundsnurscher, Tübingen: Niemeyer, S. 113–129. XV, 576 S. ; 23 cm : Ill., graph. Darst. Literaturangaben 1994.
- Goddeau, D.; Helen Meng; Joseph Polifroni; Stephanie Seneff und S. Busayapongchai: A Form-Based Dialogue Manager for Spoken Language Applications. In: *Proc. ICSLP '96*. Philadelphia, PA, Bd. 2, S. 701–704 1996.
- Grice, Paul H.: Logik und Konversation, In: *Handlung, Kommunikation, Bedeutung*, Hg. G. Meggle, Frankfurt am Main: Suhrkamp, S. 522 1993.

- Grosz, Barbara J. und Candace L. Sidner: Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204 1986.
- Harel, David: First-Order Dynamic Logic. Springer-Verlag New York, Inc., Secaucus, NJ, USA 1979.
- Henderson, J.; Oliver Lemon und K. Georgila: Hybrid Reinforcement/Supervised Learning for Dialogue Policies from COMMUNICATOR data 2005.
- Hintikka, Jaakko: Knowledge and Belief: An Introduction into the logic of the two notions. Cornell University Press, Ithaca 1962.
- Hintikka, Jaakko: Logic of conversation as a logic of dialogue, In: *Philosophical grounds of rationality*, Hg. R.E. Grandy und R. Warner, Oxford, S. 259–276 1986.
- Hughes, George E. und M. J. Cresswell: A new introduction to modal logic. Routledge 1996.
- Jefferson, Gail: Side sequences., In: *Studies in social interaction*, Hg. D.N. Sudnow, New York, NY: Free Press., S. 294–337 1972.
- Jefferson, Gail; Emanuel A. Schegloff und Harvey Sacks: The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, 53(2):361–382 Juni 1977.
- Jokinen, Kristiina; Antti Kerminen; Mauri Kaipainen; Tommi Jauhiainen; Graham Wilcock; Markku Turunen; Jaakko Hakulinen; Jukka Kuusisto und Krista Lagus: Adaptive Dialogue Systems — Interaction with Interact 2004.
- Jordan, Pamela W. und Marilyn A. Walker: Deciding to Remind During Collaborative Problem Solving: Empirical Evidence for Agent Strategies. In: *Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference, Vol. 2*, Hg. Howard Shrobe und Ted Senator. AAAI Press, Menlo Park, California, S. 16–23 1996.
- Jurafsky, Daniel und James H. Martin: Automatic Speech Recognition, In: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Prentice Hall, NJ, Kap. 9. 2. Aufl. Okt. 2007.
- Kamm, Candace A.; Marilyn A. Walker und Diane J. Litman: Evaluating Spoken Language Systems 1999.

- Kobsa, Alfred und Wolfgang Wahlster, Hg.: User models in dialog systems. Springer-Verlag New York, Inc., New York, NY, USA 1989.
- Larchey-Wendling, Dominique: Combining Proof-Search and Counter-Model Construction for Deciding Gödel-Dummett Logic. In: *CADE-18: Proceedings of the 18th International Conference on Automated Deduction*. Springer-Verlag, London, UK, S. 94–110 2002.
- Larsson, Staffan und David Traum: Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. In: *Natural Language Engineering Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*, Cambridge University Press, S. 323–340 2000.
- Lemon, Oliver; Anne Bracy und Alexander Gruenstein: A Multi-Modal Dialogue System for Human-Robot Conversation 2001.
- Lemon, Oliver; K. Georgila; J. Henderson und M. Stuttle: An ISU dialogue system exhibiting reinforcement learning of dialogue policies: generic slot-filling in the TALK in-car system 2006.
- Lemon, Oliver; Alexander Gruenstein; Alexis Battle und Stanley Peters: Multi-tasking and Collaborative Activities in Dialogue Systems. In: *Proceedings of 3rd SIGdial Workshop on Discourse and Dialogue*. S. 113–124 2002.
- Levin, Erland; S. Narayanan; R. Pieraccini; K. Biatov; E. Bocchieri; Giuseppe Di Fabrizio; W. Eckert; S. Lee; A. Pokrovsky; M. Rahim; P. Ruscitti und Marilyn A. Walker: The AT&T DARPA Communicator mixed-initiative Spoken Dialog System. *INTER-SPEECH 2005 — ICSLP 2005*.
- Levin, Esther und Roberto Pieraccini: A Stochastic Model of Computer-Human Interaction for Learning Dialogue Strategies. In: *Proc. Eurospeech '97*. Rhodes, Greece, S. 1883–1886 1997.
- Lippmann, R.: Speech recognition by machines and humans. *Speech Communication*, 22(1):1–15 Juli 1997.
- Litman, Diane J. und Shimei Pan: Empirically Evaluating an Adaptable Spoken Dialogue System 1999.
- Litman, Diane J.; Satinder Singh; Michael Kearns und Marilyn A. Walker: NJFun: a reinforcement learning spoken dialogue system. In: *Proceedings of the ANLP/NAACL 2000 Workshop on Conversational Systems*. Association for Computational Linguistics, Seattle, S. 17–20 2000.

- Loos, Bernike und Hans Peter Zorn: Combining Information Extraction and Knowledge Acquisition for Spoken Dialogue Systems 2007.
- McRoy, Susan W.: Achieving Robust Human-Computer Communication 1998.
- McRoy, Susan W. und Syed S. Ali: A Practical, Declarative Theory of Dialogue. In: *Proceedings of the IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Hg. Jan Alexandersson. International Joint Conference on Artificial Intelligence, Murray Hill, New Jersey, S. 97–103 1999.
- McRoy, Susan W.; Syed S. Ali; Angelo C. Restificar und Songsak Channarukul: Building intelligent dialog systems. *Intelligence*, 10(1):14–23 1999.
- McRoy, Susan W.; S. Haller und Syed S. Ali: Mixed Depth Representations for Dialog Processing 1998.
- McRoy, Susan W. und Graeme Hirst: The Repair of Speech Act Misunderstandings by Abductive Inference. *Computational Linguistics*, 21(4):435–478 1995.
- McTear, M.: Modelling spoken dialogues with state transition diagrams: experiences of the CSLU toolkit 1998.
- Melin, Hkan: ATLAS: A generic software platform for speech technology based applications. In: *TMH-QPRS, Quarterly Progress and Status Report*. S. 200–1 2001.
- Milward, David und Martin Beveridge: Ontology-Based Dialogue Systems. In: *Proceedings of the 3rd Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. IJCAI, Bd. IJCAI-03, S. 9–18 2003.
- Moore, Robert; Douglas Appelt; John Dowding; J. Mark Gawron und Douglas Moran: Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS. In: *Proceedings of the January 1995 ARPA Spoken Language Systems Technology Workshop*. Morgan Kaufmann, Austin, TX, S. 261–264 1995.
- Otero, Nuno; Joe Saunders; Kerstin Dautenhahn und Chrystopher L. Nehaniv: Teaching Robot Companions: The Role of Scaffolding and Event Structuring. *Connection Science*, 0(0):1–26 Jan. 2008.
- Paek, Tim: Empirical methods for evaluating dialog systems. In: *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics, Morristown, NJ, USA, S. 1–9 2001.

- Perrault, C. Raymond: An application of default logic to speech act theory. In: *Intentions in Communication*, Hg. Philip R. Cohen; Jerry Morgan und Martha E. Pollack, MIT Press 1990.
- Pierrehumbert, J.: Prosody, Intonation, and Speech Technology. In: *Challenges in Natural Language Processing*, Hg. M. Bates und R. Weischedel, Cambridge University Press, Cambridge, UK, S. 257–282 1993.
- Polifroni, J. und S. Seneff: Galaxy-II as an architecture for spoken dialogue evaluation 2000.
- Polifroni, Joseph; Stephanie Seneff; James Glass und Timothy J. Hazen: Evaluation methodology for a telephone-based conversational system. In: *In Proc. First International Conference on Language Resources and Evaluation*. S. 43–49 1998.
- Reiter, R.: A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132 Apr. 1980.
- Restificar, Angelo C.; Syed S. Ali und Susan W. McRoy: Argument Detection and Rebuttal in Dialog 1999.
- Rudnicky, A. und W Xu: An agenda-based dialog management architecture for spoken language systems. In: *IEEE ASRU Workshop*. S. 337–340 1999.
- Sacks, Harvey; Emanuel A. Schegloff und Gail Jefferson: A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735 1974.
- Sadek, David M.: A study in the logic of intention. In: *Proceedings of Knowledge Representation and Reasoning*, Hg. C. Rich; W. Swartout und B. Nebel. Bd. KR & R-92, S. 462–473 1992.
- Sadek, David M.: Towards an artificial agent as the kernel of spoken dialog system: A progress report. In: *Proceedings of the AAAI Workshop on Integration of Natural Language and Speech Processing* 1994.
- Schank, Roger C. und Christopher K. Riesbeck: Inside Computer Understanding. Lawrence Erlbaum Associates Inc, US, 2. Aufl. 1982.
- Schlegoff, E. A. und H. Sacks: Opening up closings. *Semiotica*, 7(4):289–327 1973.
- Searle, John R.: Speech Acts. Cambridge: Cambridge University Press 1969.
- Searle, John R.: Meaning, Communication and Representation, In: *Philosophical grounds of rationality*, Hg. R.E. Grandy und R. Warner, Oxford, S. 209–226 1986.

- Seneff, S.; E. Hurley; R. Lau; C. Pao; P. Schmid und V. Zue: Galaxy-II: A reference architecture for conversational system development. In: *Proc. ICSLP-98*. Bd. 3, S. 931–934 1998.
- Seneff, Stephanie; R. Lau und Joseph Polifroni: Organization, Communication, and Control in the GALAXY-II Conversational System 1999.
- Singh, Satinder; Diane J. Litman und Michael Kearns: Optimizing dialogue management with reinforcement learning: Experiments with the njfun system 2002.
- Sperber, Dan und Deirdre Wilson: On defining 'Relevance', In: *Philosophical grounds of rationality*, Hg. R.E. Grandy und R. Warner, Oxford, S. 243–258 1986.
- Sutton, Stephen und Ron Cole: Universal Speech Tools: the CSLU Toolkit 1998.
- Toulmin, S.: The uses of argument. Cambridge, England: Cambridge University Press. 1958.
- Traum, David und Staffan Larsson: The Information State Approach to Dialogue Management: Theory and Implementation, In: *The Trindi Book*, Hg. Jan van Kuppevelt und Ronnie W. Smith, Springer, Kap. 15, S. 325–350. 1. Aufl. 2003.
- Traum, David R.: Speech acts for dialogue agents 1999.
- Traum, David R. und James F. Allen: Discourse Obligations in Dialogue Processing. In: *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, Hg. James Pustejovsky. Morgan Kaufmann, San Francisco, S. 1–8 1994.
- Traum, David R. und Carl Andersen: Representations of Dialogue State for Domain and Task Independent Meta-Dialogue. In: *Proceedings of the IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Hg. Jan Alexandersson. International Joint Conference on Artificial Intelligence, Murray Hill, New Jersey, S. 113–120 1999.
- Traum, David R.; Lenhart K. Schubert; Massimo Poesio; Nathaniel G. Martin; Marc N. Light; Chung Hee Hwang; Peter A. Heeman; George M. Ferguson und James F. Allen: *Knowledge Representation in the TRAINS-93 Conversation System*. Techn. Ber. 633/TN96-4, The University of Rochester, Rochester, New York 14627 1996.
- Turing, A.M.: On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, Series 2(42):230–265 1936-1937.

- Turing, A.M.: Computing Machinery and Intelligence. *Mind*, LIX(236):433–460 Okt. 1950.
- Turunen, M. und J. Hakulinen: Agent-Based Adaptive Interaction and Dialogue Management Architecture for Speech Applications. *Lecture Notes in Computer Science*, 2166:357–365 2001.
- Wahlster, Wolfgang, Hg.: SmartKom: Foundations of Multimodal Dialogue Systems. Cognitive Technologies. Springer 2006.
- Walker, M. und L. Hirschman: Evaluation for DARPA communicator spoken dialogue systems 2000.
- Walker, Marilyn A.: Centering, anaphora resolution, and discourse structure. In: *Centering in Discourse*. Oxford University Press, S. 401–435 1998.
- Walker, Marilyn A.; Donald Hindle; Jeanne Fromer; Giuseppe Di Fabbrizio und Craig Mestel: Evaluating Competing Agent Strategies for a Voice Email Agent. In: *Proc. Eurospeech '97*. Rhodes, Greece, S. 2219–2222 1997a.
- Walker, Marilyn A.; Diane Litman J.; Candace A. Kamm und Alicia Abella: PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In: *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, Hg. Philip R. Cohen und Wolfgang Wahlster. Association for Computational Linguistics, Somerset, New Jersey, S. 271–280 1997b.
- Walker, Marilyn A.; Aravind K. Joshi und Ellen F. Prince: Centering Theory in Discourse. Clarendon Press, Oxford 1998a.
- Walker, Marilyn A.; Diane J. Litman; Candace A. Kamm und Alicia Abella: Evaluating Interactive Dialogue Systems: Extending Component Evaluation to Integrated System Evaluation. In: *Interactive Spoken Dialog Systems: Bridging Speech and NLP Together in Real Applications*, Hg. Julia Hirschberg; Candace Kamm und Marilyn Walker, Association for Computational Linguistics, New Brunswick, New Jersey, S. 1–8 1997c.
- Walker, Marilyn A.; Diane J. Litman; Candace A. Kamm und Alicia Abella: Evaluating spoken dialogue agents with PARADISE: Two case studies 1998b.
- Walker, Marilyn A. und R. Passonneau: DATE: A Dialogue Act Tagging Scheme for Evaluation of Spoken Dialogue Systems 2001.

- Warner, Richard und R.E. Grandy: Paul Grice: A View of his Work, In: *Philosophical grounds of rationality*, Hg. R.E. Grandy und R. Warner, Oxford, S. 1–44 1986.
- Wilks, Yorick; Roberta Catizone und Markku Turunen: Dialogue Management. *COMPANIONS Consortium: State Of The Art Papers 2*. January 2006.
- Wittgenstein, Ludwig: *Philosophische Untersuchungen*. Suhrkamp 2003.
- Xu, Wei und Alexander I. Rudnicky: Task-Based Dialog Management Using an Agenda. In: *ANLP/NAACL Workshop on Conversational Systems*, Hg. Candace Sidner et al. Association for Computational Linguistics, Somerset, New Jersey, S. 42–47 2000.
- Yang, Fan; Peter A. Heeman und Kristy Hollingshead: Towards understanding mixed-initiative in task-oriented dialogues. In: *INTERSPEECH-2004*. S. 217–220 2004.
- Zue, Victor; Stephanie Seneff; James R. Glass; Joseph Polifroni; Christine Pao; T. Haen und Lee Hetherington: JUPITER: A telephone-based conversational interface for weather information 2000.

Anhang A

Vorhandene Systeme

Diese Beispielsammlung von Systemen ist vor allem anhand eigener Recherchen und der Informationen entstanden, die auf folgenden Webseiten zu finden sind:

D. Bohus (Dialogsysteme 1: http://www.cs.cmu.edu/~dbohus/SDS/links_page_systems.html)

S. Larsson (Dialogsysteme 2: http://www.ling.gu.se/~sl/dialogue_links.html) am *GU Dialogue Systems Lab*.

Dabei wurden die Systeme auf Grund der vorhandenen Information überprüft und die Ergebnisse in tabellarischer Form zusammengefasst. Die Auflistung erfolgt nach Typ wie in Kap. „Beispiele und Anwendungsmöglichkeiten“ auf Seite 99 beschrieben. Die Beispielsammlung stellt keinen Anspruch auf Vollständigkeit.

A.1 Assitenten

A.1.1 Auskunft

Beschreibung	Multimodales Dialogsystem für gesprochene Sprache für die Suche nach Wohnungen auf dem Immobilienmarkt in Stockholm. Die Benutzerin kann sich mit einer virtuellen Person unterhalten und Wohnungen in Stockholm suchen (siehe auch „Bayesian Receptionist“ auf Seite A-4). Der Nachfolger des Systems ist „HIGGINS“ auf Seite A-8.
Domain	Der Immobilienmarkt in Stockholm
Typ	multimodal
Sprache	Schwedisch
Entwickelt bei	KTH - Royal Institute of Technology, Stockholm
Entwickelt von	TMH/CTT: Linda Bell, Jonas Beskow, Rolf Carlson, Jens Edlund, Joakim Gustafsson (currently TeliaSonera), Anna Hjalmarsson, Magnus Nordstrand Telia Research: Johan Boye, Mats Wirén
URL	http://www.speech.kth.se/adapt/
Publikationen	http://www.speech.kth.se/adapt/#bibliography
Plattform	-

Tabelle A.1: AdApt

Beschreibung	<i>Automated Multilingual Interaction with Information and Services (AMITIÉS)</i> AMITIÉS wurde mit dem Ziel entwickelt, den EU und USA Bürgerinnen zu erlauben, Information und Hilfe von kommerziellen Anbietern über Telefon und Internet zu beziehen, indem sie ihre Muttersprache verwenden in der Form einer interaktiven Konversation. AMITIÉS würde die heutigen Call-Zentren enorm unterstützen.
Domain	Auskunft
Typ	gemischte Initiative
Sprache	multilingual
Entwickelt bei	EU/USA
Entwickelt von	http://www.dcs.shef.ac.uk/nlp/amities/amities_partners.htm
URL	http://www.dcs.shef.ac.uk/nlp/amities/
Publikationen	http://www.dcs.shef.ac.uk/nlp/amities/amities_pubs.htm
Plattform	-

Tabelle A.2: Amities

Beschreibung	BusLine ist eine VoiceXML basierende Anwendung, die auf der Plattform von Alex Rudnicky beruht. Die Benutzerin kann sich nach dem Busverkehr rund um CMU, Oakland und Squirrel Hill zu erkundigen.
Domain	Information, öffentliche Verkehrsmittel
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University (CMU) und Tellme Networks, Inc.
Entwickelt von	-
URL	http://www.cs.cmu.edu/~aria/BusLine.html
Publikationen	http://www.cs.cmu.edu/~aria/publications.html
Plattform	Tellme

Tabelle A.3: BusLine

Beschreibung	Busman ist ein System für gesprochene Sprache für den Busverkehr in Tampere, Finland. Busman spricht daher Finnisch (verwendet ASR und TTS für Finnisch).
Domain	Telefonauskunft, öffentliche Verkehrsmittel
Typ	gemischte Initiative, multilingual
Sprache	Finisch/Englisch
Entwickelt bei	Speech-based and Pervasive Interaction Group, Tampere, USA
Entwickelt von	Markku Turunen and Jaakko Hakulinen
URL	http://www.cs.uta.fi/hci/spi/Postimies/
Publikationen	http://www.cs.uta.fi/hci/spi/publications.html
Plattform	Jaspis

Tabelle A.4: Bussimies / Busman

Beschreibung	Jupiter ist ein Dialogsystem für Wetterauskunft.
Domain	Wetterauskunft
Typ	Systeminitiative
Sprache	Englisch
Entwickelt bei	MIT, USA
Entwickelt von	-
URL	http://web.mit.edu/is/isnews/v14/n03/140301.html
Publikationen	http://publications.csail.mit.edu/ai/browse/completebrowse.shtml
Plattform	-

Tabelle A.5: Jupiter

Beschreibung	<i>The Bayesian Receptionist</i> wurde mit zwei Zielen entwickelt: erstens einen Rezeptionisten den Benutzerinnen anzubieten und zweitens eine Testumgebung zu schaffen.
Domain	Rezeption am Campus
Typ	multi-modal
Sprache	English
Entwickelt bei	Microsoft
Entwickelt von	Eric Horvitz, Nuria Oliver, Tim Paek, Matt Rhoten, Eric Ringger, Kentaro Toyama, Scott Meredith
URL	http://research.microsoft.com/adapt/conversation/
Publikationen	http://www.speech.kth.se/ctt/publications/
Plattform	Quartet (http://research.microsoft.com/adapt/conversation/systems.html#Quartet)

Tabelle A.6: Bayesian Receptionist

Beschreibung	NBA Game Update Line ist eine VoiceXML Anwendung, die auf der Tellme Plattform aufgebaut wurde, um Informationen über die NBA-Spiele zu geben.
Domain	NBA Auskunft
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	CMU, USA
Entwickelt von	Stefanie Shriver
URL	-
Publikationen	http://www.cs.cmu.edu/~awb/papers/ICSLP2002/voicexml.pdf
Plattform	-

Tabelle A.7: NBA Game Update Line

Beschreibung	Ein Agent, welcher über Telefon Auskünfte gibt, was man in New Jersey unternehmen kann.
Domain	Stadtauskunft
Typ	multimodal, mixed-initiative
Sprache	Englisch
Entwickelt bei	AT&T, USA
Entwickelt von	Satinder Singh, Diane Litman, Michael Kearns, Marilyn Walker
URL	http://www.cs.pitt.edu/~litman/
Publikationen	http://www.cs.pitt.edu/~litman/
Plattform	-

Tabelle A.8: NJFUN

Beschreibung	Bildata ist ein schwedisches Dialogsystem über gebrauchte Wagen. Fragen Sie das System nach Preis, Baujahr, etc.
Domain	Gebrauchtwagenhandel
Typ	Auskunft
Sprache	Schwedisch
Entwickelt bei	NLPLAB
Entwickelt von	Lars Ahrenberg, Nils Dahlbäck, Annika Silvervarg (Flycht-Eriksson), Genevieve Gorrell, Maria Holmqvist, Arne Jönsson, Magnus Merkel, Pernilla Qvarfordt, Håkan Sundblad, Pontus Wärnestål (Johansson)
URL	http://www.ida.liu.se/~nlplab/ (derzeit offline)
Publikationen	http://www.ida.liu.se/~nlplab/publications/index.shtml
Plattform	-

Tabelle A.9: Bildata

Beschreibung	BirdQuest ist ein schwedisches Dialogsystem über skandinavischen Vögel. Fragen Sie das System nach Migrationsrouten, Farben etc.
Domain	Vogelkunde (Ornithologie)
Typ	gemischte Initiative, Auskunft
Sprache	Schwedisch
Entwickelt bei	NLPLAB
Entwickelt von	Lars Ahrenberg, Nils Dahlbäck, Annika Silvervarg (Flycht-Eriksson), Genevieve Gorrell, Maria Holmqvist, Arne Jönsson, Magnus Merkel, Pernilla Qvarfordt, Håkan Sundblad, Pontus Wärnestål (Johansson)
URL	http://www.ida.liu.se/~nlplab/
Publikationen	http://www.ida.liu.se/~nlplab/publications/index.shtml
Plattform	-

Tabelle A.10: BirdQuest

Beschreibung	Animierter sprechender Agent namens Olga. Sie ist eine 3-dimensionale animierte Figur, die sich mit den Benutzerinnen gerne unterhält. Sie kann aber auch Information aus Datenbanken abfragen. Sie kann auch die Initiative übernehmen und Ratschläge und Tips geben.
Domain	Information-/Datenbanksuche
Typ	mixed-initiative, multimodal
Sprache	Schwedisch
Entwickelt bei	CID NADA KTH, TMH KTH, Lingvistik SU, SICS, & Nordvis AB
Entwickelt von	-
URL	http://www.nada.kth.se/~osu/olga/e_index.html
Publikationen	http://www.nada.kth.se/~osu/olga/e_dokumentation.html
Plattform	-

Tabelle A.11: OLGA

Beschreibung	Humanoid, der Informationen über Wohnungen geben kann.
Domain	Rea's task domain is real estate sales (REA is an acronym for „Real Estate Agent“)
Typ	mixed-initiative
Sprache	Englisch
Entwickelt bei	Gesture and Narrative Language, MIT Media Lab
Entwickelt von	Justine Cassell
URL	http://www.media.mit.edu
Publikationen	http://www.soc.northwestern.edu/justine/jc_papers.htm http://www.lcc.gatech.edu/~mateas/nidocs/BickmoreCassell.pdf http://web.media.mit.edu/~justine/research.html
Plattform	-

Tabelle A.12: REA

Beschreibung	OVIS : Openbaar Vervoer Informatie Systeem (Auskunft für öffentliche Verkehrsmittel).
Domain	Verkeersinformatiesysteem
Typ	multimodal, gemischte Initiative
Sprache	Niederländisch
Entwickelt bei	University of Nijmegen
Entwickelt von	http://lands.let.ru.nl/staff/intro.php
URL	http://lands.let.kun.nl/TSPublic/strik/ovis.html
Publikationen	http://lands.let.kun.nl/TSPublic/strik/ovis.html
Plattform	-

Tabelle A.13: OVIS

Beschreibung	Sprachsystem für Fluginformationen. Die Benutzerin kann über das System Flugtickets kaufen (über American Airlines EAASY SABRE).
Domain	Air Travel Information Service (ATIS)
Typ	systeminitiative (Dialog-status)
Sprache	Englisch
Entwickelt bei	MIT, USA
Entwickelt von	-
URL	http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA460588
Publikationen	http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA460588&Location=U2&doc=GetTRDoc.pdf
Plattform	-

Tabelle A.14: PEGASUS

Beschreibung	ConQuest-IJCAI'07 (Conference Questions IJCAI 2007) eingesetzt während der IJCAI-2007 Konferenz, um Auskünfte zu geben und Testdaten zu sammeln.
Domain	Auskunft über laufende Konferenz
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Rohit Kumar, Dan Bohus, Gopala Krishna, Venkatesh Keri, Sergio Grau Puerto, Stefanie Tomko, Antoine Raux
URL	http://www.conquest-dialog.org/
Publikationen	(siehe RavenClaw/Olympus, CMU, Interspeech-2006 and IJCAI-2007)
Plattform	RavenClaw/Olympus

Tabelle A.15: ConQuest

Beschreibung	Verkeraskunft über den Busverkehr in Port Authority of Allegheny County in der Nähe von Pittsburgh.
Domain	öffentliche Verkehrsmittel
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Antoine Raux, Dan Bohus, Brian Langner, Alan Black, Maxine Eskenazi, Lori Levin, Rita Singh
URL	http://www.speech.cs.cmu.edu/letsgo/
Publikationen	http://www.speech.cs.cmu.edu/letsgo/#pubs
Plattform	RavenClaw/Olympus

Tabelle A.16: Let's Go

Beschreibung	WAXHOLM ist ein anderes Auskunftssystem über Schiffe in der Umgebung von Stockholm.
Domain	Seefahrt für die Waxholm Gesellschaft
Typ	Systeminitiative
Sprache	Schwedisch
Entwickelt bei	TMH, Stockholm, Sweden
Entwickelt von	Jonas Beskow, Mats Blomqvist, Rolf Carlson, Kjell Elenius, Björn Granström, Joakim Gustafson, Sheri Hunicut, Niko Ström
URL	http://www.speech.kth.se/waxholm/waxholm2.html
Publikationen	http://www.speech.kth.se/waxholm/waxholm2.html
Plattform	-

Tabelle A.17: Waxholm

Beschreibung	Voyager ist ein Dialogsystem, welches Auskünfte über Tourismus in Boston geben kann. Voyager-Systeme wurden für Englisch, Japanische und Italienisch entwickelt.
Domain	Tourismus, Reiseauskunft
Typ	multilingual
Sprache	Englisch
Entwickelt bei	MIT, USA
Entwickelt von	-
URL	offline
Publikationen	http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA460630&Location=U2&doc=GetTRDoc.pdf
Plattform	-

Tabelle A.18: Voyager

Beschreibung	Navigationssystem für 3-D Welten.
Domain	Navigation für Fußgänger
Typ	gemischte Initiative
Sprache	Schwedisch
Entwickelt bei	CTT, KTH, Stockholm, Schweden
Entwickelt von	Jens Edlund, Gabriel Skantze
URL	http://www.speech.kth.se/higgins/
Publikationen	http://www.speech.kth.se/higgins/publications.shtml
Plattform	Galatea discourse modeller

Tabelle A.19: HIGGINS

Beschreibung	Das System erfüllt die Rolle einer Telefonistin, die erkennt welche Dienststelle die Anruferin sprechen möchte.
Domain	Information/Telefonauskunft
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	AT&T, USA
Entwickelt von	Alicia Abella, Tirso Alonso, Allen Gorin, Giuseppe Riccardi and Jerry Wright
URL	http://www.corp.att.com/attlabs/reputation/timeline/01hmiyh.html
Publikationen	-
Plattform	-

Tabelle A.20: How May I Help You?

Beschreibung	DiSCoH (<i>Dialogue System for Conference Help</i>) ist ein Aufgaben-orientiertes System für Konferenzankunft.
Domain	Auskunft über Tagungen
Typ	aufgaben-orientiert, gemischte Initiative
Sprache	Englisch
Entwickelt bei	AT&T Labs, ICSI, and Edinburgh University
Entwickelt von	-
URL	http://www.icsi.berkeley.edu/~dilek/discoh/index.html
Publikationen	-
Plattform	-

Tabelle A.21: DiSCoH

A.1.1.1 Das DARPA Communicator Projekt

Beschreibung	Luftverkehrsauskunft, Reservierungen von Hotelzimmern und Autos.
Domain	Information, Reservierungen, Reisen
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	AT&T, USA
Entwickelt von	-
URL	http://www.research.att.com/viewProject.cfm?projID=34
Publikationen	http://www.research.att.com/viewProject.cfm?projID=34
Plattform	DARPA Communicator

Tabelle A.22: AT&T Communicator

Beschreibung	Luftverkehrsauskunft, Reservierungen von Hotelzimmern und Autos.
Domain	Information, Reservierungen, Reisen
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Alex Rudnick, Wei Xu, Tina Bennett, Ananlada Chotimongkol, Dan Bohus, Kevin Lenzo, Alan Black
URL	http://www.speech.cs.cmu.edu/Communicator/
Publikationen	http://www.speech.cs.cmu.edu/air/papers/ http://www.cs.cmu.edu/%7Ecbennett/main.html http://www.cs.cmu.edu/%7Eawb/
Plattform	DARPA Communicator/RavenClaw
Verwandte Projekte	Roomline, Movieline
Bemerkung	Der CMU Communicator ist durch andere Projekte ersetzt worden wie das Roomline oder das Movieline System.

Tabelle A.23: CMU Communicator

Beschreibung	Ein Dialogsystem, um nach Information über Filme in den Kinos um Pittsburg zu suchen.
Domain	Flim/Kinoauskunft
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Alex Rudnick, Paul Constantinides
URL	http://www.speech.cs.cmu.edu/Movieline/
Publikationen	http://www.speech.cs.cmu.edu/Communicator/links.html
Plattform	CMU Communicator

Tabelle A.24: MovieLine

Beschreibung	Auskunftssystem im Luftverkehrsbereich, Reservierungen von Autos, Hotelzimmer u.a.
Domain	Informationssuche im Bereich Luftverkehr
Typ	Plattform
Sprache	Englisch
Entwickelt bei	Colorado University
Entwickelt von	Wayne Ward, Bryan Pellom, Kadri Hacioglu, Sameer Pradhan
URL	http://cslr.colorado.edu/beginweb/cumove/cucommunicator.html
Publikationen	http://cslr.colorado.edu/beginweb/cumove/cucommunicator.html
Plattform	DARPA Communicator

Tabelle A.25: Colorado University Communicator

Beschreibung	Auskunftssystem im Luftverkehrsbereich.
Domain	Informationssuche im Bereich Luftverkehr
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Bell Labs, USA
Entwickelt von	-
URL	http://www.bell-labs.com/project/communicator/
Publikationen	-
Plattform	DARPA Communicator

Tabelle A.26: Bell Labs DARPA Communicator

Beschreibung	Reiseberater, Planungsassistent. Das System handelt wie eine Reiseberaterin und kann komplexe Flugrouten planen in Kombination mit Auto- und Hotelreservierungen.
Domain	Informationssuche im Bereich Luftverkehr und Reisen
Typ	gemischte Initiative, Agenten-orientiert
Sprache	Englisch
Entwickelt bei	Stanford Research Institute (now called SRI International), USA
Entwickelt von	-
URL	http://www.ai.sri.com/~communic/
Publikationen	http://www.ai.sri.com/~communic/euro-eval7.doc
Plattform	DARPA Communicator

Tabelle A.27: SRI Communicator

A.1.2 E-Mail

Beschreibung	E-Mail Access through the Telephone Using Speech Technology Resources
Domain	E-Mail
Typ	multimodal, gemischte Initiative
Sprache	-
Entwickelt bei	E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid
Entwickelt von	Projektleiter: Luis Hernández
URL	http://www.upm.es/investigacion/actividades/V_ProgramaMarco/E-MATTER_Ing.html
Publikationen	-
Plattform	-

Tabelle A.28: E-MATTER

Beschreibung	DUMAS (Dynamic Universal Mobility for Adaptive Speech Interfaces) wurde gestartet mit dem Ziel elektronischen Geräten Sprachfähigkeiten zu verleihen.
Domain	E-Mails, SMS, Kommunikation
Typ	gemischte Initiative
Sprache	multilingual
Entwickelt bei	Swedish Institute for Computer Science
Entwickelt von	Markku Turunen and Jaakko Hakulinen
URL	http://www.cs.uta.fi/hci/spi/Postimies/
Publikationen	http://www.cs.uta.fi/hci/spi/publications.html
Plattform	Jaspis

Tabelle A.29: Athosmail/DUMAS

Beschreibung	Mailman (Postimies in Finnisch) ist ein multilinguales System und wird als E-Mail-Client eingesetzt.
Domain	E-Mails
Typ	gemischte Initiative, multilingual
Sprache	Finisch/Englisch
Entwickelt bei	Speech-based and Pervasive Interaction Group, Tampere, USA
Entwickelt von	Markku Turunen and Jaakko Hakulinen
URL	http://www.cs.uta.fi/hci/spi/Postimies/
Publikationen	http://www.cs.uta.fi/hci/spi/publications.html
Plattform	Jaspis

Tabelle A.30: Postimies / Mailman

A.1.3 Organizer

Beschreibung	VERA (Voice Enabled Reminder Assistant) ist ein Agent, der sich um die Verwaltung von Terminen kümmert und dazugehörige Tätigkeiten ausführt. Zwei Systeme wurden entwickelt - <i>VeraIn</i> , die Aufgaben in einer Datenbank speichern konnte und <i>VeraOut</i> , die auch proaktiv handeln kann
Domain	Organizer
Typ	Systeminitiative
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Ulas Bardak, Scott Judy, Vasco Pedro, Tal Blum, Jeongwoo Ko, Ryosuke Miyata
URL	http://www.ravenclaw-olympus.org/vera.html
Publikationen	http://www.cs.cmu.edu/~dbohus/ravenclaw-olympus/links.html
Plattform	RavenClaw-Olympus

Tabelle A.31: Vera

A.1.4 Reparatur

Beschreibung	LARRI (<i>A Language-based Agent for Retrieval of Repair Information</i>) ist ein multi-modales Dialogsystem zur Unterstützung der Mechaniker der F18 Flugzeuge.
Domain	Reparatur (F18 Flugzeuge)
Typ	multi-modal, gemischte Initiative
Sprache	Englisch
Entwickelt bei	CMU
Entwickelt von	Dan Bohus, Alexander Rudnicky, Yitao Sun, Rong Zhang, Kevin Lenzo, Alan Black
URL	http://www.cs.cmu.edu/~dbohus/ravenclaw-olympus/larri.html
Publikationen	http://www.cs.cmu.edu/~dbohus/ravenclaw-olympus/papers/larri.ps
Plattform	RavenClaw/Olympus

Tabelle A.32: LARRI

Beschreibung	IPA ist ein multi-modales Dialogsystem, das dafür bestimmt ist, die Astronauten auf der ISS während Reparaturarbeiten zu unterstützen.
Domain	Weltraumstation
Typ	multi-modal
Sprache	Englisch
Entwickelt bei	RIALIST Group, RIACS/NASA, Ames Research Center
Entwickelt von	Greg Aist, Brad Boven, Steven Phan, Dan Bohus, Ellen Campana, Beth-Ann Hockey, John Dowding, Jim Hieronymous
URL	http://www.riacs.edu/navroot/Research/detail/RIALIST_ver4/index.htm
Publikationen	http://www.riacs.edu/navroot/Research/detail/RIALIST_ver4/publications.htm
Plattform	RavenClaw/Olympus
Nachfolger/Verwandt:	Clarissa (Tab.: A.41)

Tabelle A.33: Intelligent Procedure Assistant (IPA)

A.1.5 Problemlösungen, Routenplanung

Beschreibung	TRAINS ist eines der ältesten Systeme und ist für die Analyse von Aufgaben-orientierten kollaborativen Systemen konzipiert worden.
Domain	Routenplanung
Typ	kollaborative Planung, gemischte Initiative, Agenten-orientiert
Sprache	Englisch
Entwickelt bei	Rochester University, USA
Entwickelt von	James Allen, Peter Heeman, David Traum
URL	http://www.cs.rochester.edu/research/cisd/projects/trains/
Publikationen	http://www.cs.rochester.edu/research/cisd/pubs/
Plattform	-

Tabelle A.34: TRAINS-91,93,95, 96

Beschreibung	TRIPS ist der Nachfolger von TRAINS.
Domain	Routenplanung
Typ	kollaborative Planung, gemischte Initiative, Agenten-orientiert
Sprache	Englisch
Entwickelt bei	Rochester University, USA
Entwickelt von	James Allen, Peter Heeman, David Traum
URL	http://www.cs.rochester.edu/research/cisd/projects/trips/
Publikationen	http://www.cs.rochester.edu/research/cisd/pubs/
Plattform	-

Tabelle A.35: TRIPS

A.1.6 Steuerung

Beschreibung	Dico wird entwickelt, als ein modernes System für die Evaluierung von Multimodalität.
Domain	ermöglicht Interaktion mit peripheren Geräten und Diensten über gesprochene Sprache und erweiterte Dialoge.
Typ	multi-modal, gemischte Initiative
Sprache	mehrere Möglichkeiten
Entwickelt bei	Göteborg University, KTH, Volvo AB, Volvo Cars, TeliaSonera
Entwickelt von	Projektleiter: Johan Jarlengrip
URL	http://www.dicoproject.org
Publikationen	http://www.dicoproject.org/publications.htm
Plattform	GoDiS

Tabelle A.36: DICO

Beschreibung	Das USI (<i>Universal Speech Interface</i>) Projekt (a.k.a. „ <i>speech graffiti</i> “) versucht ein universelles Sprachinterface zu entwickeln, um Menschen dadurch die Möglichkeit zu geben, effizient und gemütlich mit Maschinen und Informationsdiensten zu kommunizieren.
Domain	Steuerung von Anwendungen, Auskunft
Typ	Interface (multi-modal)
Sprache	Englisch
Entwickelt bei	Carnegie Mellon University
Entwickelt von	Roni Rosenfeld, Stefanie Shriver, Jerry Zhu, Arthur Toth, Thomas K. Harris, James Sanders
URL	http://www.cs.cmu.edu/~usi/
Publikationen	http://www.cs.cmu.edu/~usi/pubs.htm
Plattform	-
Verwandte Projekte	ApartmentLine (Immobilien), MovieLine (Kino/Filmauskunft)

Tabelle A.37: USI

Beschreibung	Multi-modale Oberfläche (general purpose). Das System kontrolliert einen autonomen mobilen Robot-Hubschrauber, der unterschiedliche Tätigkeiten ausführen kann.
Domain	Gerätesteuerung
Typ	multi-modal
Sprache	Englisch
Entwickelt bei	CSLI (Computational Semantics Laboratory), Stanford University
Entwickelt von	Oliver Lemon
URL	http://www-csli.stanford.edu/semlab-hold/witas/
Publikationen	http://www-csli.stanford.edu/semlab-hold/witas/
Plattform	-

Tabelle A.38: WITAS

Beschreibung	D'Homme ist ein Projekt mit dem Ziel, die Steuerung über gesprochene Sprache von vernetzten Geräten zuhause oder ausserhalb zu untersuchen.
Domain	Gerätesteuerung
Typ	multimodal
Sprache	Schwedisch, Spanisch, Englisch
Entwickelt bei	Department of Linguistics University of Gothenburg
Entwickelt von	Projektleiter Dr. David Milward. http://www.ling.gu.se/projekt/dhomme/Partners/partners.html
URL	http://www.ling.gu.se/projekt/dhomme/index.html
Publikationen	http://www.ling.gu.se/projekt/dhomme/Publications/publications.html
Plattform	

Tabelle A.39: D'Homme

A.1.7 Kontoverwaltung

Beschreibung	CTT-Bankk ist ein durch Sprache gesteuertes Dialogsystem und verwendet unter anderem Sprachauthentifizierung.
Typ	multi-modal
Sprache	Schwedisch
Entwickelt bei	CTT - Centre for Speech Technology, Stockholm
Entwickelt von	Magnus Ihse, Håkan Melin, Anna Sandell
URL	http://www.speech.kth.se/ctt/
Publikationen	http://www.speech.kth.se/ctt/publications/
Plattform	ATLAS

Tabelle A.40: CTT-Bank

A.1.8 Unterstützung in der Mensch-Mensch-Interaktion

Beschreibung	Ein Assistent für die Unterstützung von Astronauten in ihrer Tätigkeit auf der ISS.
Domain	Unterstützung, mehrere Aufgaben, Freisprechoperation
Typ	multimodal
Sprache	Englisch
Entwickelt bei	NASA, Philips, Xerox u.a.
Entwickelt von	-
URL	http://ti.arc.nasa.gov/projects/clarissa/ http://ti.arc.nasa.gov/projects/clarissa/iss_report.html
Publikationen	http://ti.arc.nasa.gov/projects/clarissa/publications.php?ta=&gid=&pid=
Plattform	-

Tabelle A.41: Clarissa

Beschreibung	Computer in the Human Interaction Loop
Domain	Unterstützung in der Mensch-Mensch-Interaktion
Typ	multi-modal
Sprache	multilingual
Entwickelt bei	Universität Karlsruhe (TH) und Fraunhofer Institut IITB
Entwickelt von	Scientific Coordinator: Prof. Alexander Waibel und Dr. Rainer Stiefelhagen Universität Karlsruhe (TH) Interactive Systems Labs Financial Coordinator: Prof. Hartwig Steusloff und Dr. Kym Watson Fraunhofer Institute for Information and Data Processing
URL	http://chil.server.de/servlet/is/101/
Publikationen	http://chil.server.de/servlet/is/350/ http://chil.server.de/servlet/is/3158/
Plattform	-

Tabelle A.42: CHIL

A.2 Tutoring

Beschreibung	Das Projekt LISTEN (L iteracy I nnovation that S peech T echnology E Nables) der CMU (Carnegie Mellon University) ermöglicht die Alphabetisierung durch einen interaktiven Tutor, der gesprochene Sprache verwendet. Zielgruppe sind vor allem Kinder, aber auch Analphabeten.
Domain	Reading Tutor
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	CMU, USA
Entwickelt von	Jack Mostow, Greg Aist, Andrew Cuneo, et al.
URL	http://www.cs.cmu.edu/~listen/
Publikationen	http://www.cs.cmu.edu/~listen/pubs.html
Plattform	-

Tabelle A.43: LISTEN

Beschreibung	NICE (<i>Natural Interactive Communication for Edutainment</i>) demonstriert einen natürlichen Zugang für Kinder und Erwachsene zu Konversationen über unterschiedliche Autoren, Literatur und historische Werke.
Domain	Tutoring
Typ	gemischte initiative
Sprache	multilingual
Entwickelt bei	NATURAL INTERACTIVE SYSTEMS LABORATORY (NISLAB), PHILIPS GMBH, CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, Telia Research Ab, Liquid Media A.B, SYDDANSK UNIVERSITET
Entwickelt von	Niels Ole Bernsen, Ulrik Lindahl, Anders Lindstr, Jean-Claude Martin, Rainer Siemung
URL	http://www.ist-world.org/ProjectDetails.aspx?ProjectId=470a9893b3ca4ef989751237d01a0c39
Publikationen	http://www.ist-world.org/Browse.aspx?A=1&DataSelectionDataType=ResultPublication&DataSelectionSelectionType=advanced&GeneralSearchTerm=&DataSelectionSourceDatabaseId=
Plattform	-

Tabelle A.44: NICE

Beschreibung	Untersucht die Möglichkeiten der Interaktion mit einem intelligenten Tutoring System über gesprochene Sprache.
Domain	Tutoring
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	Stanford University
Entwickelt von	Elizabeth Owen Bratt, Tina Chen, Herb Clark, John Fry, David Haley, Stanley Peters, Gabe Recchia, Karl Schultz
URL	http://www-csli.stanford.edu/semlab-hold/muri/
Publikationen	http://www-csli.stanford.edu/semlab-hold/muri/
Plattform	-

Tabelle A.45: MURI

Beschreibung	Das Ziel von dem Projekt ist das Lernen der Studenten am PC durch eine sprachliche Interaktion zu ermöglichen.
Domain	Tutoring
Typ	gemischte Initiative, multi-modal
Sprache	Englisch
Entwickelt bei	University of Pittsburgh, USA in Kooperation mit AT&T, USA
Entwickelt von	Michael Kearns, Charles Isbell, Satinder Singh, Diane Litman, Jessica Howe
URL	http://www.cs.pitt.edu/~litman/itspoke.html
Publikationen	http://www.cs.pitt.edu/~litman/itspoke.html#papers http://www.cs.pitt.edu/%7Elitman/
Plattform	Why2, a text-based tutoring system in the domain of qualitative physics

Tabelle A.46: ITSPOKE

Beschreibung	Natürlichsprachliche Oberfläche, um alles über das UNIX-Betriebssystem lernen zu können
Domain	Tutoring
Typ	gemischte Initiative
Sprache	Englisch
Entwickelt bei	University of California, Berkeley
Entwickelt von	Robert Wilensky, David N Chin, Marc Luria, James Martin, James Mayfield, Dekai Wu
URL	http://techreports.lib.berkeley.edu/accessPages/CSD-89-520?abstract=
Publikationen	http://techreports.lib.berkeley.edu/accessPages/CSD-89-520?abstract=
Plattform	-

Tabelle A.47: The Berkeley UNIX Consultant

Beschreibung	Virtual Language Tutor - Konversationsagent für Computer Assisted Language Learning (CALL)
Domain	Tutoring
Typ	Systeminitiative, gemischte Initiative
Sprache	skandinavische Sprachen und Englisch
Entwickelt bei	KTH, Stockholm, Sweden University of Southern Denmark. University of Oslo, Norway.
Entwickelt von	-
URL	http://www.speech.kth.se/ville/
Publikationen	http://www.speech.kth.se/ville/publications.html
Plattform	-
Verwandte Projekt:	DEAL (http://www.speech.kth.se/deal/)

Tabelle A.48: Ville

A.3 Übersetzen und Dolmetschen

Beschreibung	Das Verbmobil-System erkennt gesprochene Spontansprache, analysiert die Eingabe, übersetzt sie in eine Fremdsprache, erzeugt einen Satz und spricht ihn aus. Mit Verbmobil wurde erfolgreich angestrebt, durch Fokussierung und Zusammenschluss möglichst vieler Know-How-Träger, Deutschland in diesem Jahrtausend eine internationale Spitzenposition in der Sprachtechnologie und ihrer wirtschaftlichen Umsetzung zu sichern.
Domain	simultane Übersetzung gesprochener Spontansprache
Typ	passive Dialogteilnahme
Sprache	Deutsch, Englisch, Japanisch
Entwickelt bei	DFKI
Entwickelt von	Gesamtleiter: Prof. Dr. Wolfgang Wahlster
URL	http://verbmobil.dfki.de/
Publikationen	(offline) http://verbmobil.dfki.de/
Plattform	-

Tabelle A.49: Verbmobil

A.4 Robotik

Beschreibung	<i>The cognitive robot companion</i> - Das Ziel des Projektes ist die Lernfähigkeiten und die dazu gehörenden Prozesse zu untersuchen.
Domain	Robotik
Typ	open-ended learning processes
Sprache	-
Entwickelt bei	http://www.cogniron.org/About.php
Entwickelt von	http://www.cogniron.org/About.php
URL	http://www.cogniron.org/Home.php
Publikationen	http://www.cogniron.org/wiki/PublicationsPage
Plattform	-

Tabelle A.50: Cogniron

A.5 Plattformen

Beschreibung	Ariadne ist ein Dialogsystem, welches Interaktionen mit der Benutzerin anhand von Anwendungsbeschreibungen generiert. Es ist nicht von Aufgaben oder Sprache abhängig. Das bedeutet, dass die Entwicklerin einer Anwendung nicht mehr Dialoge designen muss, sondern die Anwendung wird sprachfähig gestaltet, so dass das Dialogsystem, weiß was zu tun ist.
Domain	offen
Typ	flexibles Design (multimodal)
Sprache	Englisch
Entwickelt bei	NTT Computer Science Laboratories
Entwickelt von	Matthias Denecke
URL	http://www.opendialog.org/index.html
Publikationen	http://www.opendialog.org/info_papers.html
Plattform	Ariadne

Tabelle A.51: Ariadne

Beschreibung	Jaspis ist eine offene Umgebung für adaptive Sprachanwendungen.
Domain	offen
Typ	flexibles Design (multi-modal)
Sprache	Englisch
Entwickelt bei	NTT Computer Science Laboratories
Entwickelt von	Matthias Denecke
URL	URL: http://www.cs.uta.fi/hci/spi/Jaspis/
Publikationen	PDF: http://acta.uta.fi/pdf/951-44-5909-1.pdf
Plattform	Plattform
Verwandte Projekte	Busman, Mailman, Doorman

Tabelle A.52: Jaspis

Beschreibung	Ein langfristiges Projekt, um eine generische multimodale Mensch-Maschine-Oberfläche zu entwickeln.
Domain	Dialogue Management und Knowledge Acquisition
Typ	multi-modal
Sprache	-
Entwickelt bei	-
Entwickelt von	Paul Piwek
URL	offline
Publikationen	http://mcs.open.ac.uk/pp2464/denk.html http://cat.inist.fr/?aModele=afficheN&cpsidt=2042324
Plattform	-

Tabelle A.53: DENK

Beschreibung	DIPPER ist eine kollaborative Initiative der „Language Technology Group“ der Universität Edinburgh und CSLI Stanford mit dem Ziel, eine gemeinsame Datenbank zu gestalten, um die Entwicklung von Dialogsystemen zu unterstützen.
Domain	-
Typ	Plattform, Information-status
Sprache	-
Entwickelt bei	Edinburgh Language Technology Group, University of Edinburgh
Entwickelt von	Johan Bos, Alex Gruenstein, Oliver Lemon, Tetsushi Oka
URL	http://www.ltg.ed.ac.uk/dipper/
Publikationen	http://www.ltg.ed.ac.uk/dipper/
Plattform	-

Tabelle A.54: Dipper

Beschreibung	Der Galaxy Communicator ist der erweiterte Nachfolger des MIT Galaxy-Systems und ist in der Version 4 auf der Sourceforge-Seite derzeit frei erhältlich.
Domain	-
Typ	Plattform
Sprache	-
Entwickelt bei	AT & T, MIT, MITRE Corporation
Entwickelt von	-
URL	http://communicator.sourceforge.net
Publikationen	http://communicator.sourceforge.net/sites/MITRE/distributions/GalaxyCommunicator/docs/manual/index.html#Reference
Plattform	DARPA Communicator

Tabelle A.55: GALAXY Communicator

Beschreibung	GoDiS ist eine Implementierung der Theorie „issue-based dialogue management“ (Larsson und Traum, 2000; Traum und Larsson, 2003)
Domain	issue-based dialogue management
Typ	gemischte Initiative, Informationstatus-orientierte
Sprache	mehrere möglich
Entwickelt bei	Göteborg University
Entwickelt von	Projektleiter: Larsson
URL	http://www.ling.gu.se/grupper/dialoglab/godis/
Publikationen	http://www.ling.gu.se/grupper/dialoglab/godis/publications.html
Plattform	TrindiKit
Verwandte Projekt	AgendaTalk, DJGoDiS, Dico, VideoGoDiS, GotTram, GoDiSDelux etc.

Tabelle A.56: GoDiS

Beschreibung	August ist ein schwedisches System mit einem animierten Kopf. Damit kann man sich über alles mögliche unterhalten.
Domain	offen
Typ	multi-modal
Sprache	Schwedisch
Entwickelt bei	Basiert auf Waxholm, entwickelt von CTT http://www.speech.kth.se/ctt/
Entwickelt von	Joakim Gustafson, Nikolaj Lindenberg und Magnus Lundenberg
URL	http://www.speech.kth.se/august/
Publikationen	http://www.speech.kth.se/ctt/publications/
Plattform	Waxholm

Tabelle A.57: AUGUST

Beschreibung	Multimodales, multilinguales Auskunftssystem für kleine Terminale wie z.B. das mobile Telefon
Domain	Mobiltelefonie und Internet
Typ	multi-modal
Sprache	multilingual
Entwickelt bei	EURESCOM
Entwickelt von	Projektleiter: Peter Stollemaier
URL	http://www.eurescom.de/public/projects/P1100-series/p1104/default.asp
Publikationen	http://www.eurescom.de/public/projects/P1100-series/p1104/default.asp
Plattform	-

Tabelle A.58: MUST

Beschreibung	Ein System zum Erforschen der Integration und Adaption in Dialogsystemen
Domain	In " <i>Implementing a Natural Command Language Dialogue System</i> ", wurde ein System auf Spanisch implementiert, das als <i>Natural Command Language</i> über Telefon betrieben werden kann.
Typ	multimodal, information-state
Sprache	Spanisch
Entwickelt bei	Department of Linguistics, University of Gothenburg
Entwickelt von	Technischer Projektleiter: Prof. Robin Cooper Administrative Projektleiterin: Mareike Schmitt
URL	http://www.ling.gu.se/projekt/siridus/
Publikationen	http://www.ling.gu.se/projekt/siridus/Publications/publications.html
Plattform	TRINDI

Tabelle A.59: Siridus

Beschreibung	RavenClaw ist eine Dialogmanagementumgebung für Aufgaben-orientierte Dialogsysteme.
Domain	Plattform
Typ	Aufgaben-orientiert
Sprache	-
Entwickelt bei	Carnegie Mellon University, USA
Entwickelt von	Dan Bohus, Antoine Raux Gopala Krishna, Venkatesh Keri, Sergio Grau Puerto, Stefanie Tomko, Thomas Harris, Alex Rudnicky, Jahanzeb Sherwani, Andrew Hoskins, Scott Judy, Yitao Sun, David Huggins-Daines
URL	http://www.ravenclaw-olympus.org/
Publikationen	(siehe RavenClaw/Olympus, CMU, Interspeech-2006 and IJCAI-2007)
Plattform	-
Verwandte Projekte:	RoomLine, the Let's Go! Public (Bus Information System)

Tabelle A.60: RavenClaw/Olympus

Beschreibung	Dialogische Mensch-Technik-Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten.
Domain	Auskunft, Gerätesteuerung
Typ	multi-modal, gemischte Initiative
Sprache	Deutsch/Englisch
Entwickelt bei	DFKI GmbH, European Media Laboratory GmbH, ICSI Friedrich-Alexander-Universität Erlangen-Nürnberg, Ludwig-Maximilians-Universität München, Philips GmbH, SONY Deutschland GmbH
Entwickelt von	Projektleiter: Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster (DFKI)
URL	http://www.smartkom.org/
Publikationen	http://www.smartkom.org/papers_de.html
Plattform	-

Tabelle A.61: SmartKom

Beschreibung	Das EU Projekt TALK beschäftigt sich mit der Entwicklung von Technologien für adaptive multi-modale und multilinguale Dialogsysteme
Domain	Gerätesteuerung, Aufgabenbewältigung
Typ	multi-modal, Informationstatus-orientiert
Sprache	multilingual
Entwickelt bei	http://www.talk-project.org/index.php?id=241
Entwickelt von	Projektleiter: Prof. Manfred Pinkal
URL	http://www.talk-project.org/
Publikationen	http://www.talk-project.org/index.php?id=2668
Plattform	SIRIDUS und TRINDI
Verwandte Projekte:	Sammie, DJGoDiS, AgendaTalk, Mimus etc.

Tabelle A.62: Talk

Beschreibung	COLLAGEN: Java Middleware für die Entwicklung von kollaborativen Agenten
Domain	kollaborative Agenten (Unterstützung beim Lösen von Problemen)
Typ	aufgaben-orientiert, agenten-orientiert
Sprache	-
Entwickelt bei	MERL - Mitsubishi Electric Research Laboratories
Entwickelt von	http://www.merl.com/people/
URL	http://www.merl.com/projects/collagen/
Publikationen	http://www.merl.com/projects/collagen/
Plattform	JAVA

Tabelle A.63: Collagen

Beschreibung	Dieses Projekt beschäftigt sich mit Argumentation in multimodaler computerunterstützten Kollaboration.
Domain	grounding in multi-modal collaboration
Typ	multi-modal
Sprache	multilingual
Entwickelt bei	Faculté de Psychologie et des Sciences de l'Education ("School of Psychology and Education") of the University of Geneva
Entwickelt von	Pierre Dillenbourg, David Traum, Patrick Mendelsohn und Daniel Schneider
URL	http://tecfa.unige.ch/tecfa/research/cscps/bootnap.html
Publikationen	http://tecfa.unige.ch/tecfa/research/cscps/bootnap.html
Plattform	-

Tabelle A.64: BOOTNAP

Beschreibung	TRINDI (Task Oriented Instructional Dialogue) ist Projekt LE4-8314 in dem Bereich „Language engineering“ des TELEMATICS APPLICATIONS Programms. Das Programm deckt folgende Arbeitsgruppen * Telematics for Services of Public Interest * Telematics for Knowledge * Telematics for Improving Employment and Quality of Life * Anwendungunterstützende Tätigkeiten http://cordis.europa.eu/telematics/
Domain	Telematik
Typ	Toolkit
Sprache	-
Entwickelt bei	Universität Göteborg
Entwickelt von	Koordinator: Robin Cooper http://www.ling.gu.se/projekt/trindi/people.html
URL	http://www.ling.gu.se/projekt/trindi/
Publikationen	http://www.ling.gu.se/projekt/trindi/publications.html
Plattform	-

Tabelle A.65: TRINDI

Beschreibung	<i>Cooperative, coordinated natural language utterances</i> Das langfristige Ziel des Projekts war eine unifizierte Architektur für einen interaktiven Diskurs aufzubauen, mit den zwei Aspekten Interpretation und Generierung aus der Perspektive der Information und der Intentionen. Es werden vier theoretischen Probleme behandelt: 1) Erkennung von Sprecher/Hörerplänen, 2) Formalismen im Bereich der Konversation, 3) Computerverarbeitung von Diskurs, 4) Untersuchungen im Bereich Interaktion zw. Generierung und Interpretation von Diskurs.
Domain	Diskurs, Verstehen von Intentionen
Typ	Evaluierung
Sprache	Englisch
Entwickelt bei	The University of Pittsburgh Intelligent Systems Program The Natural Language Group at SRI International
Entwickelt von	Barbara Di Eugenio, Jerry Hobbs, Pamela Jordan, Johanna Moore, Megan Moser, Richmond Thomason
URL	http://www.pitt.edu/~coconut/coconut.html
Publikationen	http://www.pitt.edu/~coconut/research-papers.html
Plattform	-

Tabelle A.66: COCONUT

Beschreibung	ATLAS ist eine JAVA Entwicklerumgebung (<i>framework</i>) für die Gestaltung von Sprachsoftwarelösungen, entwickelt und verwendet am TMH und KTH in Schweden.
Domain	-
Typ	Plattform
Sprache	Schwedisch
Entwickelt bei	Centre for Speech Technology (CTT), TMH, KTH
Entwickelt von	-
URL	http://www.speech.kth.se/speech/
Publikationen	http://www.speech.kth.se/~melin/papers/qpsr1_2001-atlas.ps
Plattform	-

Tabelle A.67: ATLAS

Lebenslauf

Persönliche Daten:

Name: Emanoil Kotsev

Adresse: Seidlgasse 28/13
1030 Wien, Österreich

Geburtsort: Sofia, Bulgarien

Geburtsdatum: 24.02.1973

Ausbildung:

Ab 1997 ordentliches Studium an der Universität Wien. Studienrichtung Germanistik, Fächerkombination statt zweite Studienrichtung (Computational Linguistics, Übersetzungswissenschaften und Niederlandistik)

1994 - 1997 Germanistik, Computerlinguistik und Informatik an der Sofioter Universität „St. St. Kliment Ochridski“, zweite Sprache Niederländisch.

1992 - 1994 Bundeswehrdienst

1987 - 1992 Deutschsprachiges Gymnasium Sofia mit Maturabschluss

Sprachkenntnisse:

Deutsch sehr gut in Wort und Schrift

Englisch gut in Wort und Schrift
Niederländisch gut in Wort und Schrift
Russisch schwach, vor allem mündlich
Bulgarisch Muttersprache

Beruflicher Werdegang und projektbezogene Uniarbeit:

- 2007/8 Bulgarisch-Deutsch-Bulgarisches online Wörterbuch (Evaluierung der TEI4 Definition anhand eines elektronischen bilingualen Wörterbuchs) am Institut für Translationswissenschaft (Dokumentation: <http://www.unet.univie.ac.at/~a9706152/freedict-bul-deu-bul.pdf>).
- 2007 Konsultant bei Sphinx Managed Services GmbH <http://www.sphinx-ms.at>, Bereiche: Telekommunikation, Application Support
- 2004 Summary (IMKAI) - Programm für automatische Zusammenfassung von deutschen und englischen Texten (Dokumentation: <http://www.unet.univie.ac.at/~a9706152/summary.pdf>).
- Ab 2002 Konsultant bei TTTech Computertechnik AG <http://www.tttech.com>, Bereiche: User-Support, OS-Deployment und Migration, Hardware- und Softwareerfassung, Server und Netzwerksupport, Datenbankentwicklung, Skripten.

Hobby

- ▷ Literatur und Geschichte
- ▷ Musik
- ▷ Wandern, Tennis

Zusammenfassung

Das Thema der vorliegenden Arbeit kann mit den folgenden drei Hauptfragen zusammengefasst werden: (1) was ist Dialog, (2) welche charakteristische Merkmale hat der Dialog und (3) was ist notwendig, um Dialog erfolgreich mit Hilfe von Maschinen (Computersystemen) zu simulieren. Die Arbeit wurde in der Hoffnung geschrieben, dass sie als Basis und Ausgangspunkt für weitere Forschungen in den bezeichneten wissenschaftlichen Gebieten benutzt werden kann, wie auch in der Hoffnung, dadurch die Aufmerksamkeit von Sprachwissenschaftlerinnen und Germanistinnen auf Forschungsperspektiven und Herausforderungen in diesem Gebiet zu lenken. Dabei wird auch gezeigt, wie weit die theoretischen Grundlagen, die sich auf zwischenmenschliche Dialogformen beziehen, in interaktiven Dialogsystemen implementiert werden können; welche zusätzliche Mittel notwendig sind, um solche Systeme menschenähnlicher in der Dialogkommunikation zu machen.

Die Frage nach interaktiven Dialogsystemen wird vor allem aus pragmatischen Gründen gestellt, denn mit der Entwicklung der Computertechnologie werden auch die Einsatzmöglichkeiten für Computer immer mehr. Das führt dazu, dass viel mehr Menschen, vor allem Nicht-Experten, am Computer arbeiten. Die Kommunikation mit dem Computer läuft allerdings, mit Ausnahme der Erfindung und Einführung der Maus in den 1960er/1970er, bis heute noch vor allem über die Konsole (Tastatur/Monitor).

Die menschliche Dialogfähigkeit zeichnet sich durch die Fähigkeit aus, typische Organisationsprinzipien des Dialogs zu erkennen und anzuwenden. Die Prinzipien betreffen einerseits die kognitiven Fähigkeiten des Menschen, andererseits unterschiedliche formale Aspekte. Ein Dialogsystem sollte also diese menschliche Dialogfähigkeit in unterschiedlichen kommunikativen Situationen nachahmen können. Eben aus diesem Grund steht einerseits der Dialog als Kommunikationsmittel im Zentrum der vorliegenden Diplomarbeit, andererseits aber die Frage, wie ein gesprochener Dialog mit einem Computersystem und in welchem Ausmaß möglich wäre.