

Integrierte multimodale Mensch-Rechner-Kommunikation
- ein Beispiel eines kombiniert
natürlichsprachlichen und graphischen Systemes

K.H. Hanne, K.P. Fähnrich, J.Ph. Hoepelman

Fraunhofer-Institut für Arbeitswirtschaft und Organisation
(IAO), Stuttgart

Zusammenfassung: Der Beitrag befaßt sich mit der Integration verschiedener Dialogformen in der Mensch-Rechner-Interaktion. Ausgehend von einem geschichteten Modell, wird die Idee generischer Interaktionsformen eingeführt. Hypothetische Nutzungsprofile werden diskutiert. Die Motivation für kombinierte direkt manipulative und natürlichsprachliche Schnittstellen wird gegeben und eine erste Implementation vorgestellt.

1 Generische Interaktionsformen

Realisierte Mensch-Rechner-Schnittstellen werden heute häufig nicht nach allgemein gültigen Prinzipien des Software-Engineering und der Ergonomie entwickelt. Dies hat entscheidende Nachteile. Eine eingehende Diskussion findet sich in [1]. Dort wird u. a. auf den Vorteil von Schichtenarchitekturen für Benutzerschnittstellen eingegangen. Das dort vorgestellte Schichtenmodell versucht, Vorteile anderer Ansätze (siehe z.B.: [2],[3]) zu integrieren und als ein geschichtetes Modell - ähnlich dem ISO-OSI Modell (Open Systems Interconnection) - zu interpretieren. Weitere Details finden sich z.B. in [4, 5, 6].

Auf diesem Schnittstellenmodell basierend, lassen sich generische (grundlegend verschiedene) Interaktionsformen isolieren und Profile nach verschiedenen Gesichtspunkten bestimmen. Als Hypothese werden dabei formale Sprachen (im Sinne von Programmiersprachen), direkte Manipulation und natürliche Sprache als generische Interaktionsformen betrachtet. Es läßt sich die folgende kurze Charakterisierung geben:

- o Im Fall der **formalen Sprachen** muß der Benutzer ein anwendungsorientiertes, konzeptuelles Modell aufbauen. Er muß sich an eine komplexe Syntax anpassen, hat aber den Vorteil, mit demselben Mechanismus verschiedene konzeptuelle und semantische Modelle in effizienter Weise einführen zu können.
- o **Direkt manipulative** Systeme basieren auf einem festen konzeptuellen Modell im Anwendungssystem. Auf semantischem Niveau werden generische Objekte und Funktionen bereitgestellt. Die Syntax ist elementar und die Interaktion erfolgt auf eine relativ natürliche Art. Auf die Eigenschaften und einige empirische Ergebnisse zur direkten Manipulation wird

Die beschriebenen Arbeiten wurden z.T. im Rahmen der ESPRIT-Projekte LOKI (A Logic Oriented Approach to Knowledge and Databases Supporting Natural User Interaction) und HUFIT (Human Factors Laboratories in Information Technologies) durchgeführt und durch die Kommission der Europäischen Gemeinschaften gefördert. Neben den Autoren haben B. Arndt und A. Gräble bei der Implementierung mitgearbeitet.

in [6] detailliert eingegangen.

- o In **natürlichsprachlichen** Systemen können die Benutzer und das System ein gemeinsames konzeptuelles und semantisches Modell einführen und im Laufe der Kommunikation anpassen. Kommunikationsdefekte in niederen Ebenen können durch die reichhaltige gemeinsame Struktur auf den höheren Schichten kompensiert werden. Die verwendete Syntax ist komplex, jedoch dem Benutzer im Regelfall geläufig.

Ein Versuch, einen Kriteriensatz zur Bewertung der Qualität verschiedener Mensch-Rechner-Schnittstellen zu finden, wurde in der DIN-Normungsarbeit unternommen. Für die Bewertung werden im Entwurf zu DIN 66234, Teil 8 [7] Kriterien und Definitionen vorgeschlagen. In Anlehnung und teilweiser Erweiterung zeigt das folgende Bild 1 hypothetische Profile der diskutierten Interaktionsformen.

Kriterium	FS	DM	NL
Angemessenheit für geschlossene Aufgaben	-	+	-
Angemessenheit für offene Aufgaben	+	-	+
Selbsterklärungsfähigkeit, Erlernbarkeit, Fehlertoleranz	-	+	+
Leistungsfähigkeit bei geübten Benutzern	+	-	-
Steuerbarkeit, Verlässlichkeit	+	+	-
Adaptionsfähigkeit der Kommunikation	-	-	+

FS: Formale Sprache, NL: Natürliche Sprache
DM: Direkte Manipulation

Bild 1: Hypothetische Profile generischer Interaktionsformen

Diese Profile sind im weiteren empirisch zu evaluieren. Erste Arbeiten dazu sind wiederum in [6] dokumentiert.

2.1 Direkte Manipulation

Eine eingehende Charakterisierung der direkten Manipulation findet sich in z.B. in [4], [6].

2.2 Natürliche Sprache

Das wichtigste Medium der zwischenmenschlichen Kommunikation ist die natürliche Sprache. Der hohe Grad der Beherrschung der natürlichen Sprache durch die meisten potentiellen Benutzer von Computersystemen auf der einen Seite, und eine Reihe von möglichen Anwendungsgebieten auf der anderen Seite, lassen natürlichsprachliche Systeme eines der Hauptthemen der Künstlichen-Intelligenz-Forschung und zudem ein wichtiger Aspekt der

Mensch-Rechner-Kommunikation werden.

Als mögliche Anwendungsgebiete natürlichsprachlicher Systeme sind unter anderem zu erwähnen:

- o Dialogsysteme allg.
- o Natürlichsprachliches Erstellen und Abfragen von Datenbasen
- o Programmieren in natürlicher Sprache
- o Automatische Übersetzung
- o Linguistische Textverarbeitung

Einige bekannte amerikanische Systeme sind: **LIFER** [8], **PLANES** [9], **LUNAR** [10], **SHROLU** [11]. In Deutschland sind **PLIDIS** [12], **USL** [13] und **HAM-ANS** [14] zu erwähnen. Ein kommerziell zur Verfügung stehendes, amerikanisches System zur Datenbankabfrage ist **INTELLECT** [15]. Eine zusammenfassende Bewertung findet sich beispielsweise in [16].

Trotz hohem Forschungs- und Entwicklungsaufwand darf nicht übersehen werden, daß die natürlichsprachliche Benutzerschnittstelle große Probleme mit sich bringt. Das natürlichste Kommunikationsmittel zwischen Menschen muß nicht unbedingt auch das natürlichste Kommunikationsmittel zwischen Mensch und Maschine sein. Außerdem erfordern viele Anwendungen zumindestens teilweise stark formalisierte bzw. visualisierte Kommunikation.

Viele Eigenschaften der natürlichen Sprache werden bisher theoretisch nicht hinreichend verstanden. Die oben erwähnten Systeme umfassen denn auch bestenfalls ein beschränktes Fragment der natürlichen Sprache, das außerdem in den meisten Fällen auf ein bestimmtes Anwendungsgebiet eingeengt ist. Es besteht immer die Gefahr, daß der Benutzer, auf Grund der Illusion des Verstehens der natürlichen Sprache durch das System, übertriebene Erwartungen in die Fähigkeiten des Systems stellt. Dies kann zu Enttäuschungen und Fehlinterpretationen führen, zu Ineffizienz auf Grund von semantischen Übergeneralisierungen ('semantic overshoot' [17], [18]). Bei der Implementierung natürlichsprachlicher Schnittstellen ist es also ausgesprochen wichtig, daß die Fähigkeiten des Systems für die Benutzer jederzeit genau durchschaubar sind, und daß jede Systemleistung zusätzlich 'präzise' abgefordert werden kann [19]. Die in der natürlichen Sprache gegebenen Techniken einer effektiven Dialogführung sind für benutzerfreundliche Dialog- und Frage-Antwort-Systeme von großer Bedeutung. Gegenüber formalen Sprachen oder direkter Manipulation verfügt die natürliche Sprache außerdem über Mittel für den effizienten Ausdruck von z.B. Deixis (Zeigeoperationen, wie z.B. "dieser", "jener", "hier", "dort", usw.), Zeitreferenzen, Verneinung, Quantifizierung, etc..

Es bleibt festzuhalten, daß natürlichsprachliche Systeme einen wichtigen Platz in zukünftigen Mensch-Rechner-Schnittstellen einnehmen werden. Jedoch bestehen starke technologische Restriktionen und Einschränkungen aufgrund der Arbeitsaufgabe, die den isolierten Einsatz natürlicher Sprache oft nicht praktikabel erscheinen lassen.

2.3 Kommunikation mit formalen Sprachen

Der Kommunikationsmodus 'formale Sprache' beinhaltet im wesentlichen formale Sprachen im Sinne der Mathematik, insbesondere aber die Programmier- und Kommandosprachen und die anderen klassischen 'benutzerinitiierten' Interaktionsmodi mit viel stärker festgelegten syntaktischen Strukturen. Eine Übersicht findet sich beispielsweise in [20].

3 Adaptive, multimodale Kommunikation

In herkömmlichen Systemen sind die vorher aufgezeigten Kommunikationsmodi im wesentlichen separat realisiert. In vielen praktischen Situationen wäre jedoch ein Mix dieser Kommunikationsmodi der Aufgabenstellung und den Benutzergruppen wesentlich gerechter. Auch die menschliche Kommunikation im Arbeitsleben findet multimodal statt. Adaptive, multimodale Schnittstellen, die die jeweiligen Vorteile der generischen Kommunikationsmodi verbinden, die Nachteile jedoch vermeiden sollen, erfordern einen kontextabhängigen Umschaltmechanismus, der in das Benutzerinterface eingefügt werden muß. Der Startzustand des Systemes hängt von der Anwendung ab; so kann er entweder in der direkten Manipulation für ungeübte Benutzer, in der Eingabe und Ausgabe von natürlicher Sprache oder in einem kommandosprachlichen Interface realisiert sein. Die 'Steuerung' eines solchen Systemes kann von einem 'wissensbasierten Benutzerassistenten' übernommen werden (Details siehe z.B. [4]).

Wenn alle generischen Kommunikationsmodi in der Benutzerschnittstelle vorhanden sind, kann durch diese Kombination 'mehrdimensionale' Redundanz in die Schnittstelle gebracht werden - z.B. indem der Inhalt einer Businessgraphik als Tabelle oder als Graphik mit begleitendem Text präsentiert wird. Ebenfalls können Kommunikationsdefekte, die in einer Dialogform auftreten, durch Wechsel in eine andere behoben werden. Insbesondere die Verknüpfung von natürlicher Sprache und Graphik bringt unbestreitbare Vorteile (siehe z.B. [21], [22]). In vielen Rechneranwendungen, in denen mit graphischen Objekten gearbeitet wird, ist die Möglichkeit natürliche Sprache durch eine Zeige- oder Auswahloperation mit graphischen Objekten zu verknüpfen, gegeben. Die Bedeutung von Ausdrücken in natürlicher Sprache in der Realität ist, i. A., durch Mittel wie Deixis - Zeigen, Demonstrieren, Hinweisen etc. - in der 'Realität verankert' ([23, 24]). Die Semantik und die Funktion von deiktischen Ausdrücken wie 'dieses', 'das', 'dort', 'hier', 'dann', 'jetzt', aber auch die deiktische Benutzung von Beschreibungen (z.B. 'Der Mann an der Ecke') ist ein aktuelles Forschungsgebiet der Computerlinguistik ([25]). In Verbindung mit Mensch-Rechner-Schnittstellen, die die Möglichkeit bieten, Objekte (eines Diskursbereiches) graphisch zu repräsentieren, auf diese Objekte oder auf Teile dieser Objekte zu zeigen, diese interaktiv zu verändern oder zu benennen, könnte sowohl die Mächtigkeit natürlichsprachlicher als auch direkt manipulativer Interaktion vergrößert werden. Höhere Konzepte wie Quantifikation, Negation oder Fragen jedoch können mit graphischen Mitteln allein kaum ausgedrückt werden.

Es gibt erste Ansätze wissenbasierter-graphischer und graphisch-natürlichsprachlicher Systeme. **SWYSS** [26] von der Universität Hamburg kombiniert ein wissensbasiertes Szenenanalyse-System mit einem Dialogsystem. **ISOBAR** [27] ist ein in Japan entwickeltes System, das sowohl sprachlichen als auch 'pictoriellen' Input akzeptiert. Es existiert eine gemeinsame Repräsentation für eine Teil der graphischen Symbole, die in meteorologischen Karten verwendet werden, und ein Fragment der natürlichen Sprache, das sich auf meteorologische Daten bezieht. **MOLE** [28] aus Edinburgh ermöglicht den graphischen Entwurf architektonischer Objekte durch quasinatürlichsprachliche Spezifikationen. Eine Wissensbasis enthält Wissen über architektonische Objekte und kann gemäß den Wünschen des Benutzers abgeändert werden. Andere Systeme, die in Edinburgh entwickelt wurden, basieren auf Stiny's Shapegrammars [29]. Sie ermöglichen Graphiken auf Grund des Wissens über 'Formen' und deren Beziehungen zu konstruieren. Ein amerikanisches System zur Konstruktion und Manipulation graphischer Objekte ist in [22] beschrieben.

Zur Verdeutlichung der Möglichkeiten kombiniert natürlich-sprachliche und graphisch unterstützte Interaktion durchzuführen zeigt das folgende Bild eine Dialogumgebung mit einem ausgewählten Beispiel.

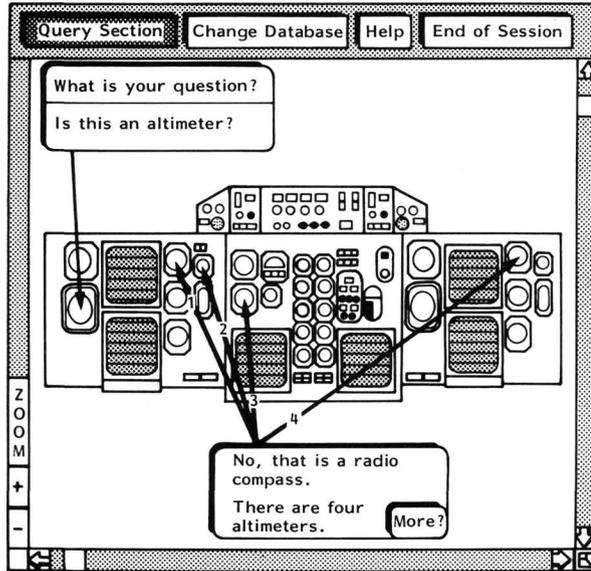


Bild 2: Beispieldialog mit lokalem deiktischen Ausdruck

Dem Benutzer wird mit Hilfe des schematisch-graphisch dargestellten Cockpitlayouts die Möglichkeit gegeben, in natürlicher Sprache eine Frage zu stellen und sie mit einer Zeigehandlung (Deixis) zu verknüpfen. Der deiktische Ausdruck ('this') wählt durch die Verbindung mit einem Zeigergerät eine Stelle des Bildschirms aus. Solche Systeme müssen, um die Anfragen richtig zu verstehen, eine Verbindung zwischen der Auswahlaktion 'this' und dem, vom Benutzer intendierten Objekt auf dem Bildschirm herstellen. Die natürlichsprachlichen Komponenten des Systems

müssen die Fragen aufbereiten und über die gemeinsame interne Wissensrepräsentation beantworten und wiederum an passender Stelle die natürlichsprachliche, graphische oder formale Antwort platzieren. Relevant ist in jedem Fall die Möglichkeit, auf eine gemeinsame semantische Repräsentation sowohl über die natürlichsprachliche Komponente wie auch über die graphische und formale Komponente zugreifen zu können.

3 Implementierung eines kombinierten natürlichsprachlichen graphische Systemes

Programmiersysteme und Workstations, die eine komfortable Programmierung und Implementierung mit relativ mächtigen Programmiersprachen erlauben, haben die Entwicklungen auf diesem Gebiet wesentlich beschleunigt. Die Programmiersprachen LISP und PROLOG haben sich als die Programmiersprachen der KI etabliert. Insbesondere im amerikanischen Raum sind LISP und auf LISP basierende Systeme auch heute noch die erste Wahl.

PROLOG versucht, logisches Programmieren zu unterstützen und durch den deklarativen Charakter und die impliziten Resolutionsmechanismen - im Gegensatz zu prozeduralen, traditionellen Programmiersprachen - der Denk- und Formulierungsweise vieler Experten näherzukommen.

Logische Programmiersprachen wie PROLOG eignen sich zur Entwicklung und zur Implementierung von kombinierten NL-Graphik Schnittstellen. Einerseits bietet die Logik und die logischen Programmiersprachen ein gutes und mächtiges Werkzeug zum Beschreiben und Implementieren von natürlichsprachlichen Phänomenen, andererseits legen auf Mechanismen der Wissensrepräsentation die Verwendung solcher Sprachen nahe.

Einen Lösungsweg (auf von-Neumann-Rechnern) zur Verknüpfung von PROLOG, graphischen Systemen und der Einbettung eines natürlichsprachlichen Systemes wurde am IA0 im Rahmen eines Esprit-Projektes beschrritten.

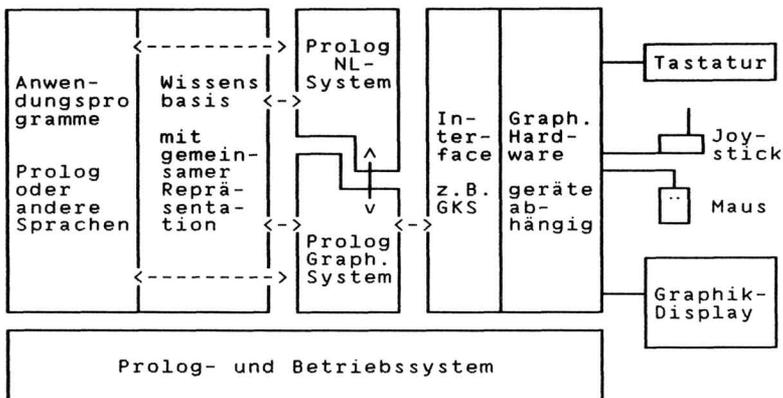


Bild 3: Eine Architektur für ein kombiniertes natürlichsprachliches und graphisches Dialogsystem.

Das Bild 3 zeigt eine Architektur eines Systemes, daß eine adaptive multimodale Benutzerschnittstelle zuläßt.

Die Abbildung 4 zeigt Dialogschritte eines am IA0 als Prototyp implementierte Systemes. Dieses System baut auf einem, in Prolog II implementierten, bestehenden Expertensystem ORBIS [30] mit natürlichsprachlicher Schnittstelle auf. Das System wurde um eine deutsche Version (zu Englisch und Französisch) erweitert und in CProlog reimplementiert. Graphische Interaktionsmöglichkeiten wurden in Prolog realisiert und zusammen mit der Möglichkeit interaktiv (über ein Graphic Data Base Management System) eine Repräsentation graphischer Daten und Objekte zu erstellen, abzufragen und zu manipulieren, als ein erster Prototyp installiert. Momentan greift die DM-Schnittstelle über einen Preprozessor auf die gemeinsame Wissensbasis des Systemes zu. Die entsprechenden deiktischen Ausdrücke werden aufgelöst und dem eigentlichen NL Expertensystem zur 'Beantwortung' überlassen. Die Bilder geben den im Original vierfarbig möglichen Bildschirmdialog wieder. Möglich sind verschiedene einfache, mit deiktischen Ausdrücken verknüpfte Anfragen, die verbunden mit einer durch die Frage aktivierten Auswahlaktion an das Expertensystem weitergereicht werden. (siehe z.B. [31]).

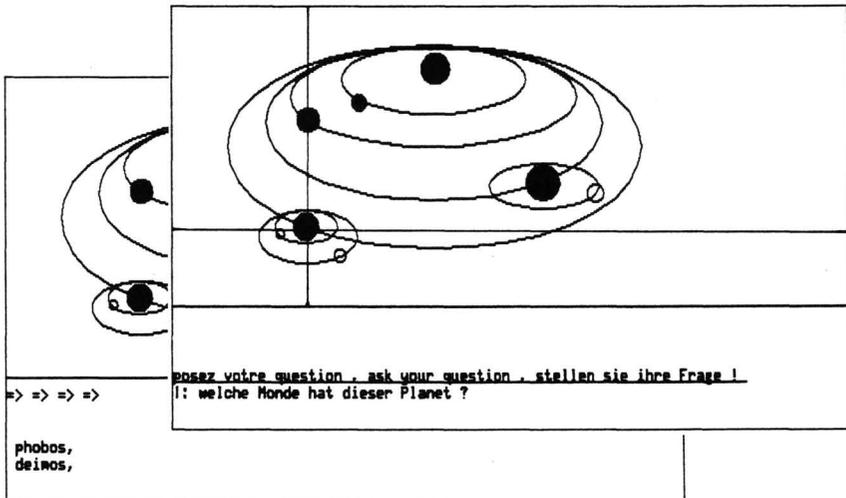


Bild 4: Beispieldialog mittels des IA0-Graphik-NL Prototypes

Die jetzt realisierte Version hängt in hohem Masse von den relativ bescheidenen Möglichkeiten des zugrunde liegenden natürlichsprachlichen Systemes ab. Im Verlauf des Projektes werden andere, erweiterte Schnittstelle auf hochauflösende Bildschirmen unter Einschluß von Fenstertechniken realisiert werden, sowie die von anderen projektbeteiligten entwickelten NL- und Prolog- Systeme eingesetzt werden.

4 Literatur

- [1] Fähnrich, K.P.; Kärcher, M.:
Software Architekturen für Mensch-Computer-Schnittstellen
- dargestellt am Beispiel eines multilingualen Text-
systems, Proc. Software-Ergonomie '85, Stuttgart, 1985.
- [2] Williamson, H.:
User Environment Model. Bericht des 1. Treffens der
European Environment Subgroup der IFIP WG 6.5, 1981.
- [3] Moran, Th.:
The Command language Grammar: A Representation of the
User interface of Interactive Computer Systems, Int. J.
Man-Machine-Studies, 15, 1981, S. 3-50.
- [4] Bullinger, H.J.; Fähnrich, K.P.:
Symbiotic Man-Computer Interfaces and the User Agent
Concept, In: Salvendy, G. (Hrsg.): Proc. 1st USA-Japan
Conf. on Human-Computer-Interaction, Honolulu, Elsevier,
Amsterdam, 1984.
- [5] Fähnrich, K.P.; Ziegler, J.:
Workstations using Direct Manipulation as Interaction
Mode- Aspects of Design, Application and Evaluation,
Proc. IFIP Conf. INTERACT '84, London, 1984.
- [6] Fähnrich, K.P.; Ziegler, J.:
Direkte Manipulation als Interaktionsform an
Arbeitsplatzrechnern, Proc. ACM-Tagung Software Ergonomie
'85, Stuttgart, Sept. 1985.
- [7] DIN: DIN 66234, Teil 8, Entwurf, 1984.
- [8] Hendrix, G.C.; Sacerdoti, E.D.; Sagalowicz, D.; Slocum, J.:
Developing a Natural Language Interface to Complex Data,
ACM TODS, 3, 2, June 1978, S. 105-147.
- [9] Waltz, D.; Finin, T.; Green, F.; Conrad, F.; Goodman, B.;
Hadden, G.:
The PLANES system: Natural Language Access to a Large
Database, Report T-34, Univ. Illinois, Urbana, 1976.
- [10] Woods, W.A.:
Semantics and Quantification in Natural Language Question
Answering, Advances in Computers, Vol. 17, Acad. Press,
1972.
- [11] Winograd, T.:
Understanding Natural Language, Acad. Press, 1972.
- [12] Berry-Rogghe, G.L.; Kolvenbach, M.; Lutz, H.D.:
Interacting with PLIDIS, a Deductive Question Answering
System for German, in Bolc, L. (Hrsg.): Natural Language
Question Answering Systems, Hanser, 1980.
- [13] Zoeppritz, M.:
Syntax for German in the User Specialty Language System,
Tübingen, 1984.
- [14] Nebel, B.; Marburger, H.:
Das Natürlichsprachliche System HAM-ANS: Intelligenter
Zugriff auf heterogene Wissens- und Datenbasen, In: J.
Nehmer (Hrsg.): Proc. 12. GI-Jahrestagung, Springer, 1982.
- [15] o.V.:
INTELLECT Query System User's Guide, Artificial
Intelligence Corp, Waltham, Mass., 1980.
- [16] Hoepelman, J., Ostler, N.:
Evaluation of NL-Sites, Teil 1 und 2, ESPRIT-LOKI-Report,
1984.
- [17] Schefe, P.:
Natürlichsprachlicher Zugang zu Datenbanken?, Angewandte
Informatik, 10, 1983.

- [18] Codd, F.F.:
How About Recently? (English Dialog with Relational
Databases using Rendezvous Version 1.), In Shneiderman,
B. (Hrsg.): Database: Improving Usability and Responsive-
ness, Acad. Press, N.Y., 1978.
- [19] Shneiderman, B.:
Natural vs. Precise Concise Languages for Human
Operations on Computers: Research Issues and Experimental
Approaches. In: Proc. 18th ACL, 1980.
- [20] Fähnrich, K.P.; Ziegler, J.:
Dialogue Techniques - An Overview, erscheint in: Shackel,
B. (Hrsg.): Proc. NATO Workshop on Research Needs in
Human-Computer-Interaction, Loughborough, 1985.
- [21] Bijl, A; Szalapaj, P:
Saying what You want with Pictures and Words, in:
Shackel, B.: Proc 1st Conf. Human-Computer-Interaction,
INTERACT 84, London, 1984.
- [22] Brown, D; Chandrasekaran, B.:
Design Considerations for Picture Production in Natural
Language Graphics System, Dept. of Comp. & Inform.
Science, Ohio State Univ., 1983.
- [23] Kaplan, R.:
Indexicals, Mimeo, Stanford, 1977.
- [24] Kamp, H.:
Deixis, Mimeo, Universität Stuttgart, 1982.
- [25] Kamp, H.:
Formal properties of 'Now', Theoria, Vol. XXVII, 1972.
- [26] Hussmann, M.; Schefe, P.:
The Design of SWYSS, a Dialogue System for Scene
Analysis, in Bolc, L. (Hrsg.): Natural Language
Communication with Pictorial Information Systems,
Springer, 1984.
- [27] Masdo, Y; et.al.:
Language-Picture Question-Answering though Common
Semantic Representation and its Application to the World
of Weather Report, in Bolc, L. (Hrsg.): Natural Language
Communication with Pictorial Information Systems,
Springer, 1984.
- [28] Krishnamurti, R; Giraud, C.:
Towards a Shape Editor, EdCAAD, Dept of Architecture,
Univ. of Edinburgh, 1984.
- [29] Stiny, C.:
Pictorial and Formal Aspects of Shape and Shape Grammars,
Basel, 1975.
- [30] Colmerauer, A; Knittredge, R.:
Dialogue pour le system ORBIS, Programmsource, 1983.
- [31] Hanne, K.H.:
ESPRIT-LOKI-Interim-Report Kr-Gr 2&4, Part 1-4, July
1985.
- Fraunhofer-Institut für Arbeitswirtschaft und
Organisation (IAO), Holzgartenstr. 17, D7000 Stuttgart 1