

ERKLÄRUNGSKOMPONENTEN ALS DIALOGWERKZEUGE

Wolfgang Wahlster

Mai 1983

ABSTRACT

This paper discusses the relevance of explanation components for knowledge-based dialog systems and motivates the need for such modules. First, the various tasks of an explanation component are described and a general architecture for such a component is proposed. Then the broad range of applications of explanation components is illustrated by several implemented components of a backgammon program (QBKG), a computer-aided design system (CADHELP) and our natural language system HAM-ANS are examined. As a negative example we discuss the explanation component of a nuclear-power plant consultant and identify some serious flaws in its design. Finally, six guidelines for the design of explanation components summarize the main results of the paper.

ERKLÄRUNGSKOMPONENTEN ALS DIALOGWERKZEUGE

Wolfgang Wahlster

Forschungsstelle für Informationswissenschaft
und Künstliche Intelligenz
Universität Hamburg
Mittelweg 179
2000 Hamburg 13

1. EINLEITUNG

Die *Selbsterklärungsfähigkeit* wurde in den letzten Jahren von der Forschung zur Mensch/Maschine-Kommunikation als eines der zentralen Kriterien für die *Benutzerfreundlichkeit* von Dialogsystemen erkannt (vgl. z.B. Dzida et al. 1978). Im Normenentwurf für die Dialoggestaltung an Bildschirmarbeitsplätzen wird daher ausdrücklich gefordert, daß interaktive Systeme auf Verlangen bei jedem Dialogschritt Erklärungen anbieten können (vgl. DIN 66 234, Teil 8, Entwurf vom 21.6.1982).

Bei den derzeit auf dem Software-Markt angebotenen Dialogsystemen ist eine Selbsterklärungsfähigkeit jedoch bestenfalls ansatzweise vorhanden. Selten wird z.B. in Dialogsprachen für die Bedienung von Texteditoren oder Betriebssystemen mehr als eine einfache *HELP-Funktion* angeboten, mit der vorgefertigte Erläuterungen in der Form von Tabellen oder Texten abgerufen werden können. Der Erklärungswert solcher vorgefertigter Erläuterungen ist allerdings in den meisten Anwendungssituationen recht gering, da die Erklärungen weder auf das spezifische Informationsbedürfnis des jeweiligen Benutzers noch auf den bisherigen Dialogverlauf abgestimmt sind und daher häufig

- unverständlich (keine Berücksichtigung des beim Benutzer vermuteten Vorwissens),
- redundant und uninformativ (Wiederholung von bereits Bekanntem, irrelevante Details),
- kommunikativ inadäquat (falsche Erklärungsebene) und/oder
- inkohärent (fehlender Bezug zu den vorangegangenen Dialogschritten) sind.

Ziel des vorliegenden Aufsatzes ist es, den Leser auf neuartige Möglichkeiten zur Verbesserung der Selbsterklärungsfähigkeit von Dialogsystemen aufmerksam zu machen und auf einige Realisierungsmöglichkeiten mithilfe von Techniken aus dem Bereich der *Künstlichen Intelligenz* (Abk.: KI; vgl. Bibel/Siekman 1982) hinzuweisen.

2. DIE FUNKTION EINER ERKLÄRUNGSKOMPONENTE

Als *Erklärungskomponente* bezeichnet man diejenige Komponente eines Dialogsystems, deren Aufgabe es ist, eine für den jeweiligen Benutzer verständliche und im jeweiligen Dialogzustand angemessene Erklärung zu erzeugen. Während im Rahmen der Künstlichen Intelligenz zunächst hauptsächlich die automatische Erklärung

- (a) der Verarbeitungsergebnisse und
- (b) der Dialogschritte

des Systems im Vordergrund stand (vgl. Wahlster 1981), wird heute auch die Erklärung

- (c) der Funktion und Bedienung des Systems und
- (d) der vom Dialogsystem verwendeten Begriffe

angestrebt.

Die Integration einer Erklärungskomponente in ein Dialogsystem setzt voraus, daß im System das erklärungsrelevante Wissen vollständig in einem für die maschinelle Verarbeitung geeigneten Formalismus repräsentiert ist. Man spricht dann von *wissensbasierten Dialogsystemen*. Dabei wird in der KI unter Wissen eine Ansammlung von Kenntnissen, Erfahrungen und Problemlösungsmethoden verstanden, die den Hintergrund für komplexe Informationsverarbeitungsprozesse bildet. Die Gesamtheit in Wissen, die einem System zur Verfügung steht, nennt man *Wissensbasis*. Die Wissensbasis gliedert sich in *Wissensquellen*, deren Inhalt mithilfe von sog. *Wissensrepräsentationssprachen* codiert sind. Dialogsysteme, die in einer Programmiersprache mit Programm/Daten-Äquivalenz (z.B. LISP, vgl. Winston/Horn 1981) geschrieben sind, erweisen sich als besonders geeignet für den Aufbau von Selbsterklärungsfähigkeiten, da in ihnen die Erklärungskomponente Teile des ablaufenden Dialogprogramms lesen und mithilfe geeigneter zusätzlicher Wissensquellen erklären kann (vgl. Wahlster 1981).

Die meisten Erfahrungen mit dem Aufbau leistungsfähiger Erklärungskomponenten wurden bisher bei der Entwicklung von *Expertensystemen* (vgl. Barr/Feigenbaum 1982) gesammelt. Diesen Expertensystemen, die u.a. Chemikern bei der Strukturanalyse chemischer Verbindungen helfen, Juristen im Körperschaftsrecht beraten, Geologen bei der Suche nach Erzlagern unterstützen, Informatiker bei der Konfigurierung von Rechenanlagen entlasten und Ärzte bei ihrer diagnostischen Tätigkeit beraten, ist gemeinsam, daß sie für diffuse Gebiete Expertenwissen und darauf beruhende Fähigkeiten maschinell verfügbar machen. Typisch für *diffuse Wissensgebiete* ist, daß sie unvollständiges und vages Wissen umfassen, weitgehend durch heuristische Methoden bestimmt sind und nicht zu Theorien abstrahiert und systematisiert werden können (vgl. Raulefs 1982).

Da Systemauskünfte von Expertensystemen zwangsläufig mit Unsicherheit behaftet sind, ist für den Benutzer, der ja die Verantwortung bei der Weiterverwendung der Systemauskünfte trägt, ein Dialogwerkzeug notwendig, mit dem er das Systemverhalten kritisch auswerten und prüfen kann. Das Vorhandensein einer Erklärungskomponente ist daher für die Akzeptanz eines Expertensystems von großer Bedeutung. Eine neuere Studie, die im Zusammenhang mit der Anwendung eines Expertensystems für die Fehlerdiagnose großer Elektronik-Systeme durchgeführt wurde (vgl. McDermott/ Brooks 1982), ergab sogar, daß sich nur rd. 1/3 der Dialogschritte auf die Aktivierung der eigentlichen Problemlösungskomponente beziehen, während rd. 2/3 typischer Dialoge aus Aufrufen der Erklärungskomponente bestehen.

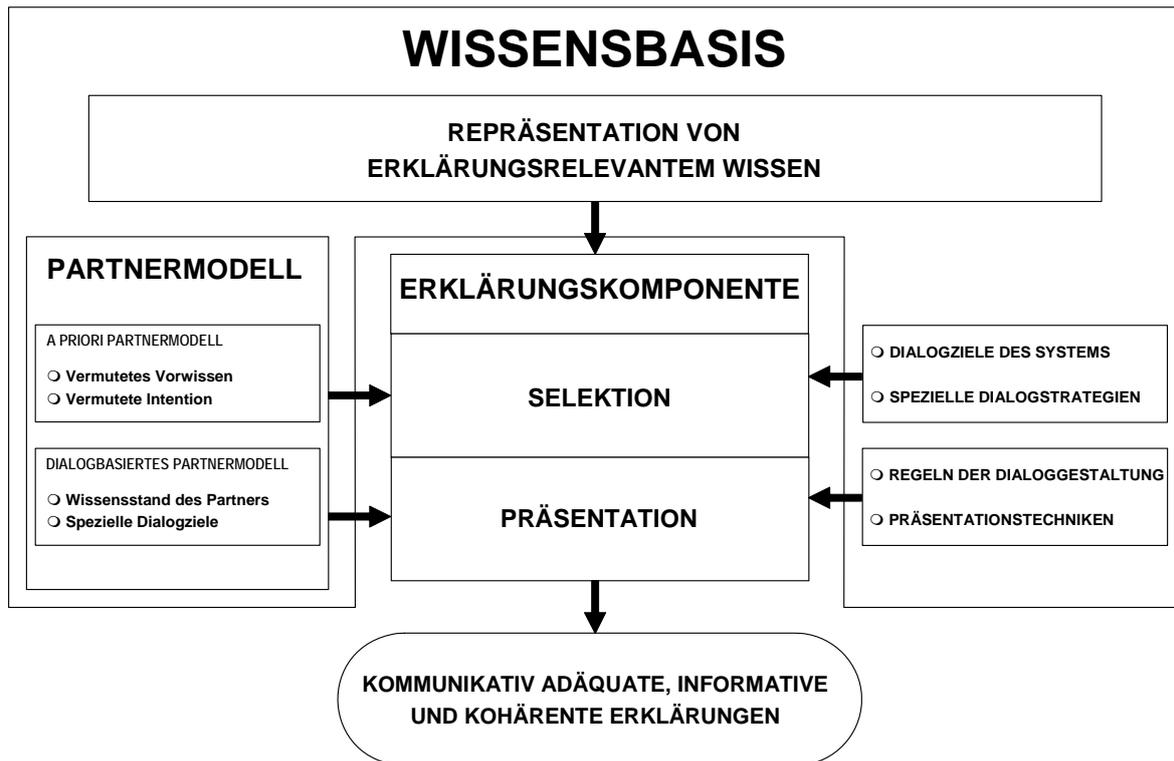


Fig. 1: Wissensquellen für eine Erklärungskomponente

Die Funktion einer Erklärungskomponente besteht u.a. darin, Teile der erklärungsrelevanten Wissensinhalte zu selektieren und diese anschließend situationsgerecht zu präsentieren (vgl. Fig. 1). Um kommunikativ adäquate Erklärungen generieren zu können, muß die *Selektion* und *Präsentation* durch ein Partnermodell gesteuert werden (vgl. Fig. 1). Das *Partnermodell* ist eine Wissensquelle, in der Information über Benutzerprofile (z.B. Laie, Auszubildender, Fachmann) gespeichert ist. Zusätzlich wird in einigen Systemen versucht, aufgrund einer Auswertung des jeweiligen Dialogverlaufs eine genauere Charakterisierung des Wissensstands und der Dialogziele des Gesprächspartners abzuleiten (vgl. Wahlster 1982).

Die Selektionsphase wird außerdem durch die jeweiligen Dialogziele des Systems (z.B. tutorieller Dialog, Beratungsgespräch, Auskunftsdiallog) und die damit verbundenen Dialogstrategien gesteuert. Schließlich sollte die Auswahl einer Präsentationsform von den in der Wissensbasis gespeicherten Regeln zur Dialoggestaltung und Präsentationstechniken (z.B. natürlichsprachlicher Text, Graphik, Tabellen, Bildsequenzen oder Kombinationen davon) abhängen.

Nach dieser einführenden Charakterisierung der Funktion von Erklärungskomponenten sollen im nächsten Abschnitt konkrete Gestaltungsmöglichkeiten für Erklärungskomponenten anhand implementierter Beispiele diskutiert werden.

3. BEISPIELE FÜR DIE GESTALTUNG VON ERKLÄRUNGSKOMPONENTEN

Um die Vielfalt der Einsatzmöglichkeiten des Dialogwerkzeugs Erklärungskomponente zu dokumentieren, betrachten wir in diesem Abschnitt die Selbsterklärungsfähigkeiten eines breiten Spektrums von wissensbasierten Systemen, das von einem Backgammon-Programm

über ein CAD-System bis hin zu einem Expertensystem für die Fehlersuche in Kernkraftwerken reicht (weitere Beispiele findet man in Wahlster 1981).

Das Backgammon-Programm QBKG (Berliner 1982) unterscheidet sich von den heute kommerziell erhältlichen Spielprogrammen nicht nur durch seine enorme Spielstärke, die zum Sieg über den Weltmeister ausreichte, sondern auch durch seine Fähigkeit, dem menschlichen Spielpartner die Spielzüge des Systems in verständlicher Form zu erklären. Fig. 2 zeigt eine aus dem Englischen übersetzte Interaktion, die vom Benutzer mit einem Erklärungswunsch für einen vorausgegangenen Spielzug des Systems eröffnet wird. QBKG selektiert dann aus der formalen Beschreibung des abgelaufenen Verarbeitungsprozesses, die in der Wissensbasis abgelegt ist, die relevanten Argumente für den zu erklärenden Zug und präsentiert diese in natürlicher Sprache. Allerdings verfügt QBKG nicht über eine voll ausgebaute Sprachgenerierungskomponente, sondern kommt durch die extreme Beschränkung des Gesprächsthemas mit kontextabhängigen Einsetzungen in die Leerstellen ausgewählter Argumentationsschemata aus.

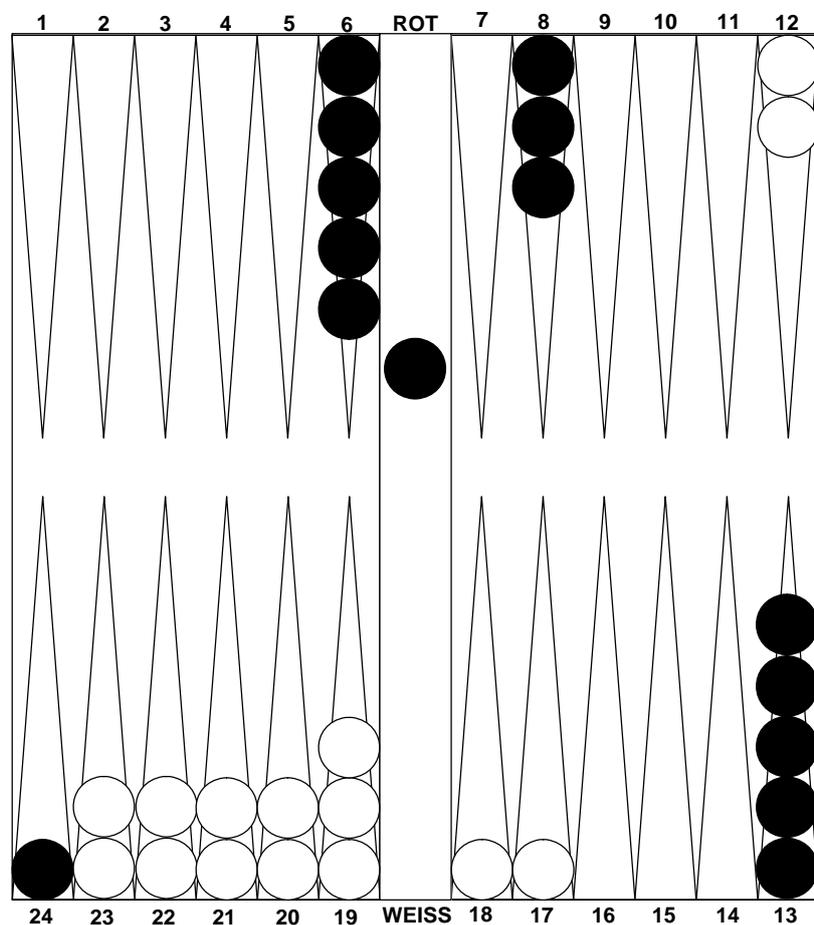


Fig. 2 Beispiel für die Selbsterklärungsfähigkeit von QBKG

BEN: WARUM HAST DU DEN ZUG 17-23, 23-24 DURCHGEFÜHRT STATT DEN ZUG 12-18, 17-18?

SYS: 1. IN DER GEGEBENEN STELLUNG HAT WEISS EINEN GROSSEN VORSPRUNG UND HAT GEWINNVORTEILE MIT DEUTLICHEN GAMMON-CHANCEN

2. DER DURCHGEFÜHRTE ZUG, 17-23, 23-24 (ZUG 1), IST VIEL BESSER ALS DER VORGESCHLAGENE ZUG 12-18, 17-18 (ZUG 2)
3. ES GIBT NICHTS, WAS FÜR ZUG 2 SPRICHT. DIE VORTEILE VON ZUG 1 SIND:
 - WEITAUS BESSERE CHANCEN, DEN GEGNER DAVON ABZUHALTEN, EIN BAND ZU ERZEUGEN (1)
 - VIEL BESSERER ANGRIFF VON WEISS (2)

Am Beispiel von QBKG kann man sich klar machen, daß aufgrund der Kombinatorik eines solchen Spiels die Ausgabe vollständig vorgefertigter Erklärungstexte für jeden möglichen Spielzug unmöglich ist.

Das System CADHELP (Cullingford et al. 1982) generiert nicht nur Text, sondern auch Graphik, um dem Benutzer die Funktionsweise eines komplexen CAD-Systems für den Hardware-Entwurf zu erklären. Das in FRANZ LISP implementierte System verwendet als Wissensrepräsentationssprache eine Variante der CD-Graphen (einen Überblick über die Repräsentation von Wissen für natürlichsprachliche KI-Systeme enthält Wahlster 1982), um Expertenwissen über die Bedienung des CAD-Werkzeugs zu codieren (vgl. Fig. 3 (a)).

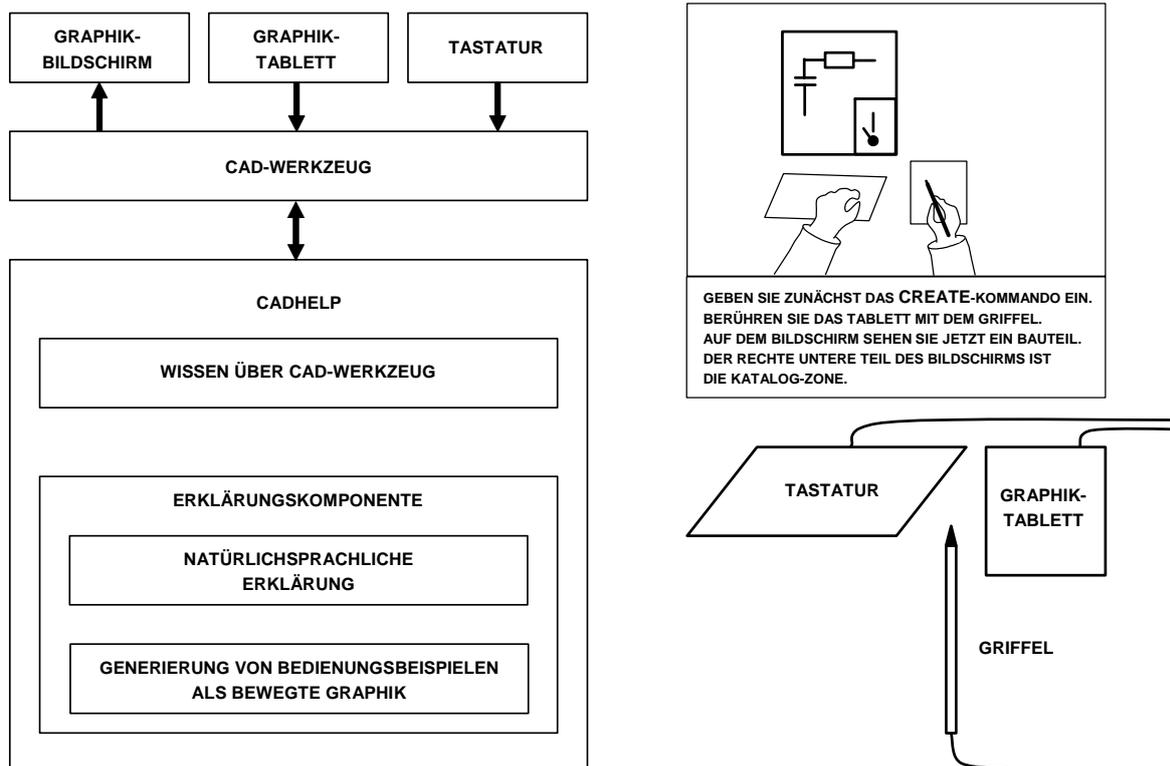


Fig. 3:

(a) Architektur von CADHELP

(b) Erklärungsbeispiel für CADHELP

Der Bildschirm ist so partitioniert, daß im oberen Teil Beispiele für Arbeitsabläufe am CAD-Arbeitsplatz mithilfe von Graphik-Sequenzen dargestellt werden, während im unteren Teil die entsprechenden natürlichsprachlichen Bedienungshinweise ausgegeben werden (vgl. Fig. 3 (b)). Wie Figur 3 (b) zeigt, bestehen diese Bedienungshinweise auch aus Begriffsdefinitionen (z.B. Katalog-Zone). Bei CADHELP wird die mit der Selbsterklärungsfähigkeit eines Systems stets verbundene *Reflexivität* dadurch besonders deutlich, daß der Graphik-Bildschirm

eine Darstellung von sich selbst zeigt (vgl. Fig. 3 (b)). CADHELP merkt sich, was dem Benutzer innerhalb eines laufenden Dialogs bereits erklärt wurde, und ist durch eine leistungsfähige Sprachgenerierungskomponente dazu in der Lage, spätere Wiederholungen von Details durch zusammenfassende Formulierungen zu vermeiden.

Es gehört zu den Besonderheiten der von uns entwickelten Erklärungskomponente, daß im Verlaufe eines Dialogs auch mehrschrittige und mehrsträngige Argumentationen zur Erklärung einer inferenzbasierten Systemantwort aufgebaut werden können (vgl. Wahlster 1981). Das Dialogbeispiel (1) - (6), in dem das System den

- (1) BEN: Ist der angebotene Mercedes billig?
- (2) SYS: Ich glaube ja.
- (3) BEN: Wieso?
- (4) SYS: Weil er reparaturbedürftig ist.
- (5) BEN: Nur deshalb?
- (6) SYS: Und weil er ziemlich alt ist.

Benutzer beim Gebrauchtwagenkauf berät, illustriert die Fähigkeit der Erklärungskomponente, aufgrund einer Auswertung der Informativität der einzelnen Ableitungsschritte zunächst nur den wichtigsten Argumentationsstrang für die mit einem gewissen Unsicherheitsfaktor behaftete positive Systemantwort anzuführen. Erst eine Nachfrage des Benutzers (vgl. (5)) führt dazu, daß eine weitere Erklärung für die Systemantwort angeführt wird.

Diese Erklärungskomponente ist Teil des natürlichsprachlichen KI-Systems HAM-ANS, das dem Benutzer über schriftliche Dialoge in deutscher Umgangssprache den Zugang zu anderen Software-Produkten (z.B. Datenbanksystem, Bildfolgenanalysesystem) ermöglicht. Die Erklärungskomponente von HAM-ANS kann nicht nur durch 'Warum'-Fragen des Benutzers angestoßen werden, sondern kann auch vom System selbst aktiviert werden, falls z.B. die Evidenz für eine Systemantwort aufgrund von un-

- (7) BEN: Ist das Bett einigermaßen hart?
- (8) SYS: Ich glaube ja, weil seine Matratze neu ist.

sicherem Wissen einen Schwellwert unterschreitet (vgl. (8)). Eine solche 'Überbeantwortung' wird von kooperativen Dialogsystemen auch erwartet, wenn die Antwort offensichtlich nicht den Erwartungen des Benutzers entspricht (vgl. Webber/Joshi 1982).

HAM-ANS verfügt auch über eine Komponente, die begriffliches Wissen des Systems erklären kann. Dabei wird in (11) eine Verbalisierung von Einträgen in das semantische Netzwerk vorgenommen.

- (9) SYS: Es gibt vier Sitzgelegenheiten.
- (10) BEN: Was verstehen Sie unter Sitzgelegenheit?
- (11) SYS: Eine Sitzgelegenheit ist ein Möbel. Folgende Ausprägungen von Sitzgelegenheiten sind bekannt: Stuhl und Sessel
Eine Sitzgelegenheit hat eine Sitzfläche, meistens vier Beine und oft eine Rückenlehne.

Als Beispiel für eine mißlungene Konzeption der Mensch/Maschine-Schnittstelle zu einer

Erklärungskomponente, betrachten wir zum Schluß dieses Abschnitts die Selbsterklärungsfähigkeiten eines Expertensystems, das für die Fehlersuche in Kernkraftwerken und für die Beratung der Bediener bei Zwischenfällen entwickelt wurde (vgl. Underwood 1982). Zu Beginn des zu behandelnden Beispieldialogs, der in (12) - (15) auszugsweise wiedergegeben wird, teilt der Operateur dem System einen Druckverlust im Kühlsystem des Reaktors mit. Nachdem das System beim Operateur einige Meßwerte erfragt hat, diagnostiziert es den Verlust von Kühlflüssigkeit und sagt das Anspringen des Notkühlsystems (ECCS) voraus. Nachdem das Notkühlsystem einige Zeit läuft, fragt der Operateur, ob er das Notkühlsystem ausschalten kann (vgl. (12)). Das System verneint dies und begründet seine ablehnende

- (12) BEN: (A (TURN OFF ECCS))?
(13) SYS: NO
(14) BEN: WHY?
(15) SYS: (S-COUPLE
 ((SC (AMT COOLANT 2000 1600 NEG))
 (S (ECCS OFF)))
 (S (REACTOR-CORE UNCOVERED)))

Haltung auf Nachfrage damit, daß bei einem Kühlflüssigkeitsverlust eine Abschaltung des Notkühlsystems dazu führt, daß schließlich der Reaktorkern frei liegt.

Bedenkt man, daß in der entsprechenden Anwendungssituation die Interpretation der Erklärung unter großem Stress erfolgt, so wird deutlich, daß die angebotene Erklärung untragbar ist. Der Bediener muß nicht nur Abkürzungen wie S für Zustand, SC für Zustandsübergang und S-COUPLE für Zustandskopplung korrekt dekodieren, sondern die logische Struktur der Erklärung aus der Listenstruktur eines LISP-Ausdrucks rekonstruieren. Auch für einen geübten Benutzer dürfte es unter den extremen psycho-physischen Bedingungen bei einem Reaktor-Zwischenfall kaum möglich sein, die Implikationsrelation und die darin verschachtelte Konjunktion der Prämissen aus der Klammerstruktur mit der notwendigen Geschwindigkeit zu erkennen. Bedenkt man die fatalen Folgen, die durch solche Kommunikationsprobleme an der Mensch/Maschine-Schnittstelle ausgelöst werden können, so wird die große Verantwortung des Informatikers bei der Entwicklung von Dialogwerkzeugen deutlich.

Obwohl alle in diesem Abschnitt vorgestellten Erklärungskomponenten Teile von Prototypen sind, die noch keine Produktreife erlangt haben, kann man erwarten, daß zukünftig ergonomisch gestaltete Erklärungskomponenten durch die damit erreichte größere Transparenz und Kontrollierbarkeit sowie die vereinfachte Handhabung eine Verbesserung der Akzeptanz, der Arbeitsproduktivität und der Arbeitszufriedenheit für komplexe Dialogsysteme bewirken.

4. ZUSAMMENFASSUNG: SECHS THESEN ZU ERKLÄRUNGSKOMPONENTEN

Im folgenden werden die wichtigsten Forderungen im Zusammenhang mit Erklärungskomponenten in Form von sechs Thesen zusammengefaßt.

These 1: Ohne Selbsterklärungsfähigkeit durch eine Erklärungskomponente sind komplexe Dialogsysteme (z.B. als Beratungssysteme konzipierte Expertensysteme) wertlos.

These 2: Erklärungskomponenten erfordern wissensbasierte Systeme, in denen die Information, die einer Verarbeitungsleistung zugrunde liegt, explizit dargestellt wird.

These 3: Erklärungen sollen nicht vordefiniert sein sondern, sie müssen *kontextabhängig* unter Berücksichtigung eines *Partnermodells* und ausgehend von einer expliziten Repräsentation erklärungsrelevanter Information erzeugt werden.

These 4: Die Leistungen einer Erklärungskomponente sollen nicht nur explizit durch den Benutzer abrufbar sein sondern ggf. auch *vom System selbst* ausgelöst werden können.

These 5: Die Präsentation der Erklärungen soll situationsabhängig aus Text, Graphik, Bildsequenzen oder Kombinationen dieser Darstellungsformen bestehen können.

These 6: Erklärungskomponenten sind komplexe Dialogwerkzeuge, deren Entwurf und Realisierung eine Verwendung von Methoden und Techniken der *Künstlichen Intelligenz* voraussetzen.

5. LITERATUR

BARR, A.; FEIGENBAUM, E.A. (eds.): The Handbook of Artificial Intelligence. Vol. II, Los Altos: Kaufmann 1982, S.76-294.

BERLINER, H.J.; ACKLEY, D.H.: The QBKG System: Generating Explanations from a Non-Discrete Knowledge Representation. In: Proc. of AAAI-82, Pittsburgh, S. 213-216.

BIBEL, W.; SIEKMANN, J.H. (eds.): Künstliche Intelligenz. (Informatik-Fachberichte 59), Berlin, Heidelberg, N.Y.: Springer 1982.

CULLINGFORD, R.E.; KRUEGER, M.W.; SELFRIDGE, M.; BIENKOWSKI, M.A.: Automated Explanations as a Component of a Computer-Aided Design System. In: IEEE Transact. on Systems, Man and Cybernetics, Vol. SMC-12, No. 2, März/April 1982. S. 168-181.

DZIDA, W.; HERDA, S.; ITZFELD, W.D.: Factors of User-Perceived Quality of Interactive Systems. GMD, IST Bericht Nr. 40 (2. Auflage), St. Augustin 1978.

McDERMOTT, D.; BROOKS, R.: ARBY: Diagnosis with Shallow Causal Models. In: Proc. of AAAI-82, Pittsburgh, S. 370-372.

RAULEFS, P.: Expertensysteme. In: Bibel, W.; Siekmann, J.H. (eds.): Künstliche Intelligenz, 1982. S. 59-98.

UNDERWOOD, W.E.: A CSA Model-Based Nuclear Power Plant Consultant. In: Proc. of AAAI-82, Pittsburgh, S. 302-305.

WAHLSTER, W.: Natürlichsprachliche Argumentation in Dialogsystemen. KI-Verfahren zur Rekonstruktion und Erklärung approximativer Inferenzprozesse. (Informatik-Fachberichte 48), Berlin, Heidelberg, N.Y.: Springer 1981.

WAHLSTER, W.: Natürlichsprachliche Systeme. Eine Einführung in die sprachorientierte KI-Forschung. In: Bibel, W.; Siekmann, J.H. (eds.); Künstliche Intelligenz, 1982. S. 203-283.

WEBBER, B.; JOSHI, A.: Taking the Initiative in Natural Language Data Base Interactions: Justifying Why. In: Proc. of 9th COLING, Prag, 1982. S. 413-418.

WINSTON, P.H.; HORN, B.K.P.: LISP. Reading: Addison-Wesley, 1981.