

**Forschungsberichte der  
Technischen Fakultät  
Abteilung Informationstechnik**

**Kognitive Aspekte bei der Realisierung  
eines robusten Robotersystems  
für Konstruktionsaufgaben**

Bernd Hildebrandt, Alois Knoll,  
Christian Scheering, Jianwei Zhang

Report 98-01

**Impressum:**

Herausgeber:

Robert Giegerich, Alois Knoll, Helge Ritter, Gerhard Sagerer,  
Ipke Wachsmuth

Anschrift:

Technische Fakultät der Universität Bielefeld,  
Abteilung Informationstechnik, Postfach 10 01 31, 33501 Bielefeld

ISSN 0946-7831

## Zusammenfassung

*Im Rahmen des SFB 360 entwickeln wir einen Situierten Künstlichen Kommunikator, im vorliegenden Fall ein Robotersystem, für Konstruktionsaufgaben. Das Robotersystem kann durch spontansprachliche Anweisungen gesteuert werden. Die Hauptkomponenten des Systems sind eine Sprachverstehenskomponente und eine Zwei-Arm-Roboterkomponente. Die Sprachverstehenskomponente basiert auf der Combinatory Categorical Grammar und ermöglicht eine inkrementelle und interaktive Sprachverarbeitung. Die Roboterkomponente verfügt über eine Vielzahl von Sensoren und beherrscht komplexe Montageoperationen wie Stecken und Schrauben. Durch die gewählte Architektur und die zugrundegelegten kognitiven Verarbeitungsprinzipien können Teilkomponenten des Systems der aktuellen Situation entsprechend interagieren und Informationsredundanz nutzen. Das System erhält dadurch eine hohe Robustheit.*

## Abstract

*In the framework of the SFB 360 we are developing a Situated Artificial Communicator for assembly tasks. The robot system can be instructed using spontaneous speech. The main components of the system are a speech understanding module and a two-arm-robot module. The speech understanding module is based on Combinatory Categorical Grammar, which makes incremental and interactive speech understanding possible. The robot module is provided with multiple sensors and it masters complex assembly operations like peg-in-hole or screwing a nut into a bolt. The architecture and the cognitive principles enable interactive processing that depends on the actual situation and allows the system to take advantage of redundant items of information. It is due to these principles that our Situated Artificial Communicator is highly robust.*

## 1 Einleitung

Der Sonderforschungsbereich *Situierte Künstliche Kommunikatoren* (SFB 360) hat sich neben der Schaffung theoretischer Grundlagen auch die prototypische Realisierung eines Situierten Künstlichen Kommunikators (SKK) zum Ziel gesetzt [19, 18]. Um die Vielfalt und Komplexität natürlicher Kommunikationssituationen überschaubar und modellierbar zu halten, wurden als Forschungsgegenstand aufgabenorientierte Dialoge gewählt und als gemeinsame empirische Basis ein Konstruktionsszenario. Das Szenario deckt alle relevanten Aspekte wie sprachliches Handeln, visuelle Perzeption und Objektmanipulation ab. Der menschliche Instrukteur erteilt anhand eines Bauplans einem künstlichen Konstrukteur (einem Robotersystem) Anweisungen, um aus Holzbauteilen ein Modellflugzeug zu konstruieren.

Innerhalb des SFB gibt es sich ergänzende Ansätze, um sich der Realisierung eines SKK zu nähern. Mit *CODY*, dem virtuellen Konstrukteur, lassen sich aus Bauteilen, die 3D-computergraphisch visualisiert werden, beliebige Aggregate auf einer virtuellen Montagefläche konstruieren [28]. Hierzu wird ein hybrider Repräsentationsformalismus für imaginale und attributive Information entwickelt [12, 11]. *CoRA* ist ein virtueller Roboter, der in einer hybriden Architektur Sprache, Perzeption und Handlung integriert, und der auf Anweisungen unterschiedlicher Komplexität situiert und flexibel reagiert [16, 27]. Zusätzlich werden bis zu 15 heterogene Software-Module, die von den jeweiligen Teilprojekten erstellt wurden, zu einem Gesamtsystem integriert [3, 4].

Von zentraler Bedeutung ist die konkrete Realisierung eines SKK. Im SFB wird deshalb ein Robotersystem für Konstruktionshandlungen entwickelt, das durch spontan gesprochene Anweisungen gesteuert werden kann (z.B. [31]). Die zugrundegelegten Verarbeitungsprinzipien und die Architektur des Systems sind kognitiv motiviert (z.B. [8, 13]), wobei die technischen Rahmenbedingungen teilweise enge Grenzen setzen. So verfügt das System beispielsweise über zwei Roboterarme und erscheint einem menschlichen Konstrukteur als anthropomorph bezüglich des Aktionsradius, der etwa dem eines erwachsenen Menschen entspricht (Abbildung 1). Dies gilt aber beispielsweise nicht für die Greifer, die nur über zwei Finger verfügen. Im folgenden wird dieses Robotersystem für Konstruktionsaufgaben und seine kognitiven Grundlagen vorgestellt und gezeigt, wie Robustheit erreicht wird.

## 2 Prinzipien kognitiver Informationsverarbeitung

Um die Prinzipien menschlicher Text- und Sprachverarbeitung zu verstehen, wurde eine Vielzahl psycholinguistischer Experimente durchgeführt. Häufig lagen den Experimenten sprachlich ambige Strukturen zugrunde, die Rückschlüsse auf unterschiedliche Verarbeitungsprozesse zulassen. Verfeinerte Meßmethoden ermöglichen inzwischen multimodale Experimente, bei denen beispielsweise die Wechselwirkungen zwischen visueller und sprachlicher Information untersucht wird [30]. Die Ergebnisse legen eine sprachunabhängige Verallgemeinerung der Verarbeitungsprinzipien nahe. Demnach sollte ein kognitiv motiviertes Robotersystem zumindest folgende Prinzipien berücksichtigen:

- die Verarbeitung ist inkrementell,
- die Verarbeitung ist interaktiv und
- die Verarbeitung ist modular.

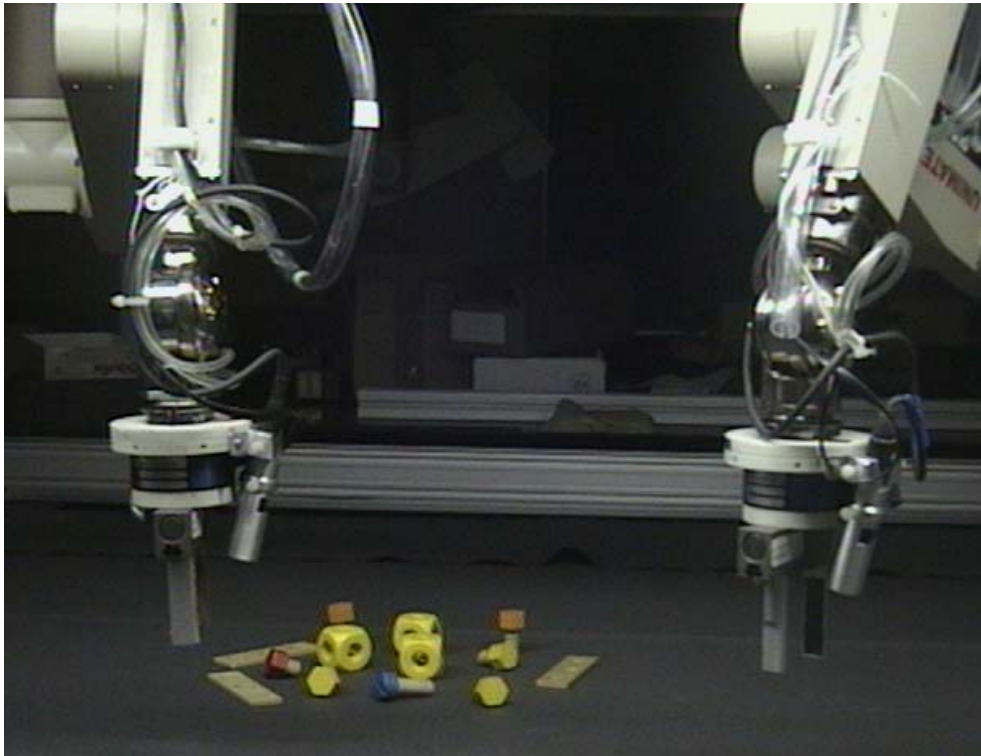


Abbildung 1: Roboterzelle

Bei einer *inkrementellen* Verarbeitung wird Information möglichst umgehend verarbeitet, ohne diese zuvor zu sammeln oder auf bestimmte Schlüsselinformationen zu warten. Bezüglich der Sprachverarbeitung heißt dies, daß Wörter in der Regel unmittelbar und interaktiv verarbeitet werden, ohne zuvor auf das Satzende zu warten. Verarbeitungsrelevante Informationseinheiten, die Inkremente, können neben Wörtern auch Konstituenten im syntaktischen Bereich oder Konzepte bzw. Referenten im mentalen Modell im semantischen Bereich sein. Inkremente im visuellen Bereich können objektbezogene Regionen, die durch Farbe oder Form eine Einheit bilden, oder Aufmerksamkeitsbereiche sein. Obwohl die Granularität gegenwärtig unterspezifiziert ist [30], wird von einer vom Verarbeitungsstand und -ziel abhängigen Einteilung ausgegangen.

Eng verknüpft mit der inkrementellen Verarbeitung ist das Prinzip der *interaktiven* Verarbeitung, bei der unterschiedliche Informationsbereiche frühzeitig berücksichtigt werden. Dies können neben Interaktionen zwischen syntaktischer und semantischer Information auch Interaktionen zwischen sprachlicher und visueller Information sein. Beispielsweise konnte [23] durch Augenbewegungsexperimente zeigen, daß die Reihenfolge von Lokalangaben durch Adjektive (*das rechte obere Objekt* bzw. *das obere rechte Objekt*) die Blickbewegungen beeinflussen können, wenn Versuchspersonen in einer symmetrischen Konstellation das benannte Objekt ansehen sollten. In welchem Maße und unter welchen Bedingungen sprachliche Verstehensprozesse mit den verschiedenen Informationsbereichen interagieren, ist größtenteils noch unerforscht. Dennoch muß ein kognitiv motiviertes Robotersystem Möglichkeiten zur Interaktion und Integration vorsehen; technisch bedeutet dies die Möglichkeit zur Datenfusion. Wenn man menschliche Informationsverarbeitung als interaktive Verarbeitung auffaßt, dann gibt es Verarbeitungsebenen, Komponenten oder Module, die miteinander interagieren. *Modularität* wird seit [5] meist als ein kognitives Architekturmodell verstanden, bei dem einzelne kognitive Module isoliert voneinander sind und sequentiell durchlaufen werden. Wir unterscheiden hingegen funktional unterschiedliche Verarbeitungsberei-

che, wie beispielsweise Syntax und Semantik, Diskurs- und Weltwissen [10] oder visuelle und taktile Wahrnehmung [15].

### 3 Systemkomponenten und Architektur

Die Hauptkomponenten der Realisierung sind ein Spracherkenner und ein Sprachverstehensmodul sowie das Zwei-Arm-Robotersystem, das über eine Vielzahl von Sensoren verfügt. Neben einem externen Multi-Kamera-System zur Objekterkennung und groben Lagepositionierung [20] ist auf jedem Arm eine „Eye-on-Hand“ Kamera montiert, durch die im Nahbereich mittels eines visuellen Regelungsprozesses Bauteile exakt gegriffen werden können [14, 21]. Zusätzlich sind an den Handgelenken der Arme Kraftmomentensensoren montiert. Sie erlauben dem System komplexe Montageoperationen wie Stecken und sogar Schrauben. Durch die schritthaltende Auswertung sowohl der Kräfte wie der Bilder der Handkameras werden dabei Fehler in Position und Orientierung der Greifer weitgehend ausgeglichen [2].

Der Spracherkenner basiert auf dem Diktiersystem IBM *Via Voice Gold*. Um die Erkennungsleistung zu erhöhen, wurde das Vollformen-Lexikon dem Vokabular des Szenarios angepaßt und das System um grammatische Bedingungen und einfache Wortkompositionsregeln erweitert [17]. Der Spracherkenner ist als „Frontend“ zur Sprachverstehenskomponente angelegt. Falls eine Äußerung nur partiell korrekt erkannt wurde, ist in der Regel dennoch eine zumindest teilweise Interpretation möglich.

Die Sprachverstehenskomponente basiert auf einer von Mark Steedman entwickelten Variante der Kategorialgrammatik, der *Combinatory Categorical Grammar* (z.B. [24, 25, 26]). Um die kognitiven Verarbeitungsprinzipien der Sprachrezeption, wie z.B. der inkrementellen Verarbeitung, und der flexiblen Konstituentenstellung im Deutschen zu entsprechen, wurde die Grammatik modifiziert und an das Deutsche angepaßt [9, 7]). Dadurch können selbst komplexe und (lokal) ambige Anweisungen (z.B. *steck die lange Schraube in das zweite Loch von links der siebenlöchrigen Leiste*) effizient interpretiert werden. Aus den sprachlichen Anweisungen leitet die Sprachverstehenskomponente semantische Strukturen ab, aus denen wiederum elementare Operationen für den Roboter generiert werden.

Entscheidend ist die Architektur, die die Berücksichtigung des situativen Umfelds ermöglicht (Abbildung (2)). Die passiven Interaktionen (Perzeption) und aktiven Interaktionen (Aktion) sind angedeutet. Der menschliche Instrukteur kann sprachliche Anweisungen erteilen, Objekte durch Zeigegesten spezifizieren und neben der gewollten Aktion sprachliche Reaktionen des SKK erhalten. Die Interpretation einer Anweisung basiert zusätzlich zum linguistischen Wissen auch auf konzeptuellem Wissen über Objekte und Handlungen. Gegenwärtig werden nur vollständig interpretierte Anweisungen an ein Modul *Identifikation* weitergereicht, das unter Berücksichtigung situativen Wissens Objekte für die maßgebliche Handlung identifiziert und diese dadurch spezifiziert. Geplant ist hier eine inkrementelle Schnittstelle. An die Identifikation schließt sich eine situativ abhängige Handlungsplanung an, die auf die einzelnen Handlungsmöglichkeiten des Roboters, wie Greifen oder Schrauben, abgestimmt ist. Die Bereiche *Kontrolle*, *Koordination*, *Operationensequenz* und *Interpreter* dienen der Informationsverwaltung und Ablaufsteuerung und können als eigener Bereich des *Scheduling* aufgefaßt werden [31]. Entsprechend des aktuellen Roboterzustands werden aus kognitiven Operationen elementare Roboterdirektive generiert. So wird aus der Handlungsprimitive schrauben(Nut,Bolt) die folgende Operationensequenz:

1. `find_contact(nut, bolt)` (Stelle Kontakt her),

2. `find_thread()` (Finde die Bohrung),
3. `find_notch_point()` (Finde den Gewindegang),
4. `screw()` (Schraub ein).

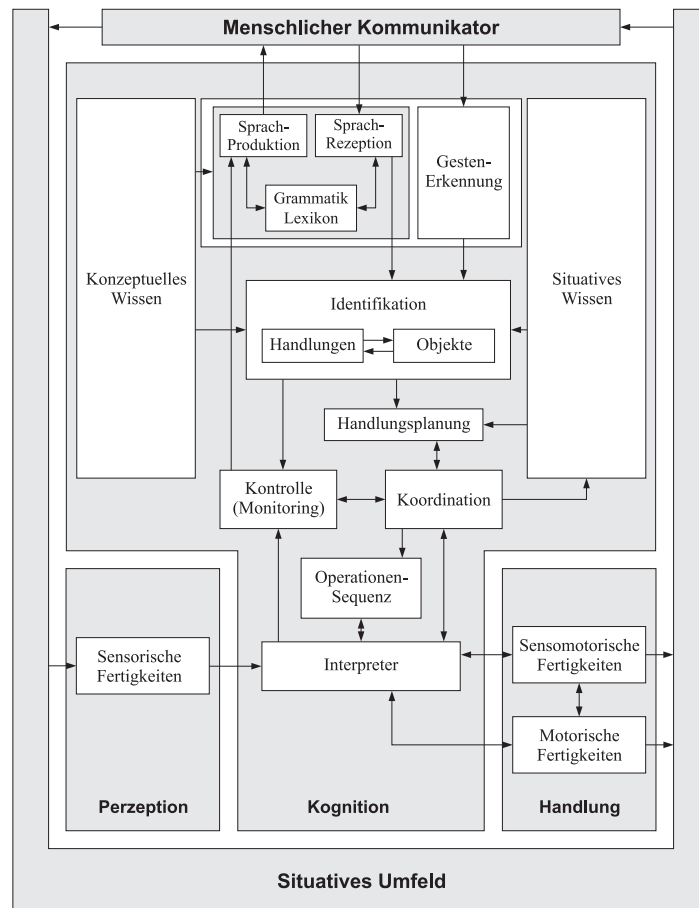


Abbildung 2: Architektur des Konstruktionsroboters

Zur Abarbeitung einer solchen Operationensequenz verwendet der Interpreter sogenannte Fertigkeiten (*skills*). Aus dem perceptiven Bereich stellen sensorische Fertigkeiten Information über die Situation zur Verfügung. Einige davon müssen echtzeitfähig sein, um die technischen Randbedingungen für geschlossene Regelkreise der Robotersteuerung zu erfüllen. Als positiver Nebeneffekt wird eine unmittelbare Reaktion des Systems auf menschliches Eingreifen möglich und somit eine wirklichkeitsnahe Interaktion mit dem Menschen. Die Handlungen werden über motorische Fertigkeiten umgesetzt, die einzelne Roboterbewegungen wie das Öffnen, Schließen oder Drehen des Greifers steuern. Als direktes Bindeglied zwischen Perzeption und Aktion dienen sensomotorische Rückkopplungen, die beispielsweise für kraftüberwachte Bewegungsregelungen benötigt werden. Unerwartete Ereignisse, wie beispielsweise eine Kraftüberschreitung oder der Ausfall eines Moduls, werden an den Interpreter gemeldet, der im Zusammenspiel mit Kontrolle und Kooperation geeignete Maßnahmen wie z.B. Rückfragen an den Instrukteur veranlaßt. Darüber hinaus wird über den Interpreter eine ständige Aktualisierung des Wissens über die Situation vorgenommen. So zählt zum Wissen über die Szene unter anderem, ob sich ein Objekt auf dem Tisch

oder in der Hand des Konstruktionsroboters befindet. Die Veränderung der Objektposition erfolgt in solchen Fällen meist durch intendierte Handlungen. Eine Veränderung der Szene, die eine Aktualisierung der Szenenrepräsentation erfordert, kann allerdings auch unbeabsichtigt auftreten, beispielsweise wenn eine bisher stehende Schraube umfällt, während ein anderes Objekt in ihrer Nähe gegriffen wird.

## 4 Robustheit durch Interaktion und Redundanz

Während der Handlungsinterpretation und -ausführung kommt es zu einem komplexen Interaktionsgeflecht zwischen den Modulen. Die vorgestellte Architektur und die zugrundegelegten Verarbeitungsprinzipien bilden die Basis für ein robustes System, das unerwartete, unvollständige und auch widersprüchliche Information verarbeiten kann, ohne handlungsunfähig zu werden. Voraussetzung hierfür ist, daß das System

- instabile Zustände erkennt,
- die Situation diagnostizieren kann und
- eigenständig Fehlerbehandlungsroutinen einleitet oder über den Zustand berichtet und den Instrukteur interaktiv um Unterstützung bittet.

Eine Grund für die Notwendigkeit einer robusten Informationsverarbeitung sind die sprachlichen Anweisungen. So kann gegenwärtig kaum mehr als 85% Wortakkuratheit von einem Spracherkenner erwartet werden. Daraus folgt, daß das Sprachverstehenssystem lexikalische Einfügungen, Ersetzungen und Auslassungen bewältigen muß. Außerdem zeigen die im SFB erhobenen Sprachdaten, daß eine Äußerung nicht immer einen Satz bildet, zumindest nicht gemäß einer üblichen Grammatik [22]. Ein Sprachverstehenssystem muß also selbst bei einem perfekten Spracherkenner robust bezüglich der sprachlichen Eingabe sein. Ähnliches gilt für die Perzeptions- und die Handlungskomponente des SKK. Insgesamt muß das System eine Reihe qualitativ unterschiedlicher Schwierigkeiten überwinden, die im folgenden ausgeführt werden.

**Wortersetzungen:** Der Spracherkenner ist oft unsicher bezüglich der Flexionsendungen *-em* und *-en*, so daß beide leicht vertauscht werden. Da durch die Flexion u.a. Kasus und Genus markiert sind, kann bei einer Vertauschung die vorliegende Nominalphrase nicht als syntaktisch kongruent analysiert werden. Als Fehlerbehandlungsroutine liegt es daher nahe, auf Kasus- und Genuskongruenz innerhalb einer Nominalkonstituente ganz zu verzichten, wenn eine konsistente alternative Konstituente fehlt. Um Übergeneralisierungen zu vermeiden, kann eine solche Strategie allerdings nur vereinzelt eingesetzt werden.

**Unbekannte Wörter:** Falls das erkannte Wort nicht im linguistischen Lexikon vorkommt, wird es von der Sprachverstehenskomponente übergangen. Solange das unbekannte Wort für die semantische Interpretation unerheblich ist, wie bei Höflichkeitsfloskeln, ist dies Verfahren erfolgreich. Langfristig sollte die Sprachverstehenskomponente jedoch unbekannte Wörter lernen können, um inhaltlich relevante Information nicht zu verlieren. Auch wenn das System manchmal auf die syntaktische Funktion eines Wortes aufgrund seines Umfelds schließen kann (z.B. *und die Leiste steckst du auf die rote Unbekannt*), wird ein Klärungsdialog mit dem Instrukteur im Regelfall notwendig sein.

**Wortauslassungen:** Ein Klärungsdialog mit dem Instrukteur sollte außerdem geführt werden, wenn bei einer elliptischen Äußerung oder einer Wortauslassung semantische Information vom



System nicht inferiert werden kann. Selbst wenn eine Äußerung nicht vollständig interpretiert werden kann, generiert das Sprachverstehenssystem zumindest partielle Interpretationen, so daß Nachfragen möglich sind, wie z.B. *Was soll ich mit 'die Leiste' und 'auf die rote Schraube' tun? bzw. die rote Was?*.

**Semantische Inkonsistenz:** Eine Anweisung kann semantisch inkonsistent sein oder zumindest dem System so erscheinen. Im Zweifelsfall muß das System beim Instrukteur um eine Korrektur oder eine detaillierte Spezifikation nachfragen. Manchmal kann es eine vollständige Handlungsanweisung ableiten, indem das System weitere Informationen aggregiert und fehlende Information inferiert. Wörtlich genommen kann beispielsweise die Anweisung *schraub die Leiste auf den Würfel* nicht ausgeführt werden, da weder die Leiste noch der Würfel die Funktion einer Schraube ausübt. Wird allerdings das benötigte Instrument vom System inferiert (*schraub die Leiste mit der roten Schraube auf den Würfel*), ist die Handlung möglich. Bei einer unterspezifizierten Anweisung wird also aufgrund semantischen Wissens das benötigte Instrument inferiert.

**Sprachliche Ambiguität:** Das Sprachverstehenssystem muß darüber hinaus verschiedene Formen der Ambiguität bewältigen. Auf der lexikalischen Verarbeitungsebene kann ein Wort unterschiedliche syntaktische Kategorien haben, z.B. kann *schrauben* sowohl Nomen wie Verb sein (die entsprechende Orthographie liefert der Spracherkenner nicht, da dazu eine syntaktische Voranalyse nötig wäre). In den meisten Fällen wird eine solche lexikalische Ambiguität durch den syntaktischen Kontext aufgelöst; d.h. es wird jene Kategorie gewählt, mit der eine syntaktische Analyse möglich ist. Außerdem kann auf der syntaktischen Verarbeitungsebene die Anbindung der Konstituenten ambig sein. Ein klassisches Beispiel hierfür ist die Anbindung der Präpositionalphrase [9]. Hierbei nutzt das Sprachverstehenssystem eine Vielzahl von Informationsquellen. Unter anderem wird geprüft, ob für das polyseme Verb eine präferierte Lesart vorliegt, ob das Objekt semantisch mit der Funktionalität des Verbaruments übereinstimmt und für welche Interpretation es ein Denotat im situativen Kontext gibt. Kann das Sprachverstehenssystem die Ambiguität nicht selbst auflösen, sollte es beim Instrukteur nachfragen.

**Situative Ambiguität:** Eine Anweisung, die von der Sprachverstehenskomponente semantisch vollständig interpretiert wurde, führt nicht unbedingt zu einer gleichfalls erfolgreichen Handlung. Beispielsweise können bei einer Anweisung wie *nimm eine Schraube* mehrere Schrauben in der Szene als Denotat zur Verfügung stehen. Im Regelfall folgt die Handlungskomponente dann einem Ökonomieprinzip: Es wird jenes Objekt gegriffen, das am leichtesten zu erreichen ist und dessen Positionierung am wenigsten Greifprobleme bereitet. Je nach Verteilung der Objekte in der Szene kann das System nachfragen, ob beispielsweise eine lange oder eine kurze Schraube benötigt wird.

**Sprachliche Intervention:** Aus einer Vielzahl von Ursachen kann es dazu kommen, daß die Handlungskomponente eine Handlung ausführen will, die bezüglich der Intention des Instrukteurs falsch ist. Dies ist beispielsweise immer dann der Fall, wenn das System eigenständig entscheidet und ein Objekt greifen will, das der Instrukteur nicht gemeint hat. Während der Greifarm sich auf das Objekt zubewegt, hat der Instrukteur allerdings die Möglichkeit, verbal zu intervenieren. Gegenwärtig kann die Handlung nur durch den Befehl *stop* unterbrochen werden. Ergänzend sollen künftig auch korrigierende Information wie *eine andere Schraube* bzw. *besser die vordere Schraube* einfließen können. Die Handlung wird hierdurch lediglich modifiziert, womit vermieden werden soll, daß die Handlung völlig neu aufsetzen muß.

**Objektposition:** Die Detektion von Objekten und die Bestimmung ihrer jeweiligen Position sowie Ausrichtung in einer komplexen Szene mit Objektüberdeckungen und Hintergrundobjekten ist überaus schwierig (z.B. [6]). Die visuelle Komponente benutzt daher mehrere Kameras, die die Szene aus unterschiedlichen Perspektiven aufnehmen. Hierdurch wird einerseits eine höhere Toleranz gegenüber schlechter Signalqualität, wie sie z.B. bei ungünstigen Lichtverhältnissen gegeben

ist, oder gar gegenüber Bildausfall erreicht. Andererseits besteht die Hoffnung, daß durch ein verteilt operierendes Multi-Sensoren-Agenten-Netzwerk [20] verdeckte Objekte sicherer erkannt und ihre Lage präziser bestimmt werden kann.

**Objektverschiebung:** Verändert sich die Position oder Lage eines Objektes während einer Feinpositionierung, z.B. wenn ein Objekt gegriffen werden soll oder wenn zwei Objekte zueinander ausgerichtet werden, dann kann die visuelle Komponente sofort nachregeln. Wie ein Mensch muß auch der SKK auf Veränderungen in der Szene sofort reagieren können, damit er handlungsfähig bleibt (*real-time capability*). Damit die Handlungen trotz Veränderungen in der Szene erfolgreich ausgeführt werden können, bedarf es also einer schnellen Kommunikation zwischen den einzelnen Komponenten [1].

Die Anforderung an das Gesamtsystem, echtzeitfähig zu sein, hat für die Sprachkomponente zur Folge, daß sie nicht beliebig lange Zeit für die Interpretation benutzen darf. Sie muß vielmehr jederzeit in der Lage sein, zumindest Teilinterpretation an andere Komponenten liefern zu können (*any-time capability*). Hierzu benötigt sie Kriterien, um eine laufende Analyse und Interpretation abbrechen zu können. Ein quantitatives Abbruchkriterium ist die bisher verstrichene Zeit und die Anzahl konkurrierender Lösungswege. Ein qualitatives Abbruchkriterium ist erreicht, wenn die zur Handlung benötigte Information vorliegt, also beispielsweise der Typ der Handlung und alle obligatorischen Argumente. Zugrundegelegt wird hierbei eine inkrementelle und interaktive Verarbeitung sowie das Prinzip, nicht allen möglichen Interpretationen zu folgen, sondern nur der stabilsten und informativsten (*intelligent pruning*). Die Operationalisierung dieses Prinzips ist allerdings nicht unproblematisch und hängt oft von spezifischen Konstellationen ab.

Die Robustheit des gesamten Systems beruht auf Redundanz, da sich verschiedene Informationsquellen für auf denselben Bereich beziehen, wie beispielsweise bei der visuellen Perzeption. Darüber hinaus bedient sich das System reaktiver und interaktiver Robustheitsstrategien:

- systemextern, indem mit dem Instrukteur kommuniziert wird, wobei die Initiative vom Situiereten Künstlichen Kommunikator oder vom Instrukteur ausgehen kann,
- systemintern bzw. intermodular, indem verschiedene Systemkomponenten interagieren und
- modulintern, indem verschiedene Arten linguistischer Information oder visuelle Information aus unterschiedlichen Perspektiven integriert werden.

## 5 Zusammenfassung und Ausblick

Der im Rahmen des SFB 360 entwickelte Konstruktionsroboter wird durch spontansprachliche Anweisungen gesteuert. Er kann beliebige Objekte des Szenarios greifen und durch Fügehandlungen wie Stecken und Schrauben miteinander verbinden (s. Abbildung (3)). Aufgrund der gewählten Architektur, der zugrundegelegten kognitiven Verarbeitungsprinzipien und der Berücksichtigung von Situativität erhält das System ein hohes Maß an Robustheit. Weiterführende Arbeiten beziehen sich sowohl auf die Sprachverstehenskomponente wie auf den Konstruktionsroboter. So kann die Sprachverstehenskomponente unterspezifizierte Handlungsanweisungen zwar in Roboterdirektiven übertragen, die der Konstruktion komplexer Aggregate dienen. Doch kann die Struktur der Aggregate oder allein ihr Gewicht unter Umständen dazu führen, daß eine Realisierung mit zwei Greifbacken und ohne weitere Hilfsmittel nicht möglich ist. Dasselbe gilt für das rechtwinklige Ausrichten zweier Objekte: Der Winkel zwischen den Objekten kann so ungünstig sein, daß mit



Abbildung 3: Umsetzung einer Schraub-Anweisung

den Greifbacken eine Ausrichtung nicht möglich ist. Für solche Fälle ist vorgesehen, daß der SKK den Instrukteur um aktive Hilfe bittet, d.h. der SKK erkennt seine Grenzen und fordert den Instrukteur auf, die gewünschte Aktion zumindest partiell selbst auszuführen. Die dafür benötigte Dialogkomponente muß noch entwickelt werden, wobei die Ergebnisse und die Erkenntnisse anderer Teilprojekte des SFB einfließen werden. Die Grammatik und das Lexikon wird fortlaufend ausgebaut und den erweiterten Möglichkeiten des Konstruktionsroboters angepaßt. Hierbei werden die empirischen Befunde der psycholinguistischen Experimente zum Sprachverstehen berücksichtigt (z.B. [29]), um ein Gesamtsystem zu erhalten, das im Sinne der experimentell-simulativen Methode neue Fragestellungen aufwirft und gegebenenfalls seinerseits als Experimentalumgebung dienen kann.

## Literatur

- [1] COLLANI, Y. VON und A. KNOLL: *OCCL - Object Communication C++ Library (Manual)*. Univ. Bielefeld, AG Technische Informatik.
- [2] COLLANI, Y. VON, J. ZHANG und A. KNOLL: *A neuro-fuzzy solution for fine-motion control based on vision and force sensors*. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, 1998.
- [3] FINK, G. A., N. JUNGCLAUS, F. KUMMERT, H. RITTER und G. SAGERER: *A distributed system for integrated speech and image understanding*. In: SOTO, R. (Herausgeber): *Proceedings of the International Symposium on Artificial Intelligence*, Seiten 117–126, 1996.
- [4] FINK, G. A., N. JUNGCLAUS, H. RITTER und G. SAGERER: *Integration verteilter Systeme mit DACS*. In: *20. DAGM-Symposium*. Springer-Verlag, 1998.
- [5] FODOR, J.: *The modularity of mind*. Bedford, Cambridge, MA, 1998.
- [6] GRAF, T., A. KNOLL und A. WOLFRAM: *Fuzzy invariant indexing: A general indexing scheme for occluded object recognition*. In: *Proceedings of the 4th International Conference on Signal Processing*, Beijing, 1998.
- [7] HILDEBRANDT, B. und H.J. EIKMEYER: *Effizientes Sprachverstehen mit Combinatory Categorical Grammar*. Technischer Bericht, SFB360, Univ. Bielefeld, in Vorb.
- [8] HILDEBRANDT, B., R. MORATZ, G. RICKHEIT und G. SAGERER: *Integration von Bild- und Sprachverstehen in einer kognitiven Architektur*. *Kognitionswissenschaft*, (4):118–128, 1995.
- [9] HILDEBRANDT, B. und G. RICKHEIT: *Verarbeitung von Präpositionalphrasen in der Combinatory Categorical Grammar*. Technischer Bericht 97/6, SFB360, Univ. Bielefeld, 1997.
- [10] HILDEBRANDT, B., G. RICKHEIT und H. STROHNER: *WORT - Eine Computersimulation kognitiver Textverstehensprozesse*. *Kognitionswissenschaft*, (3):95–106, 1993.
- [11] HOFFHENKE, M., K. KESSLER, B. JUNG, G. RICKHEIT und I. WACHSMUTH: *Imaginale Prototypen von mechanischen Objekten und Aggregaten und ihre Rolle bei der Konzeptverarbeitung*. *Kognitionswissenschaft*, im Druck, 1999.
- [12] HOFFHENKE, M. und I. WACHSMUTH: *Dynamische Konzeptualisierung mit imaginalen Prototypen*. Technischer Bericht 97/9, SFB360, Univ. Bielefeld, 1997.
- [13] KNOLL, A., B. HILDEBRANDT und J. ZHANG: *Instructing cooperating assembly robots through situated dialogues in natural language*. In: *Proceedings IEEE Conference on Robotics and Automation*. IEEE Press, 1997.
- [14] MEINICKE, P. und J. ZHANG: *Calibration of a "self-viewing" eye-on-hand configuration*. In: *Proceedings of the IMACS Multiconference on Computational Engineering in Systems Applications, Lille, France*, 1996.
- [15] MORATZ, R., H.-J. EIKMEYER, B. HILDEBRANDT, F. KUMMERT, G. RICKHEIT und G. SAGERER: *Integrating speech and selective visual perception using a semantic network*. In: *AAAI-95, Fall Symposium on Computational Models for Integrating Language and Vision*, 1995.
- [16] PETERS, K., S. STRIPPGEN und J.T. MILDE: *CoRA – An instructable robot*. In: LÜTH, T., R. DILLMANN und H. WÖRN (Herausgeber): *Distributed Robotic Systems*, Band 3. Springer-Verlag, 247–256, 1998.

- [17] REINSCH, M., B. HILDEBRANDT und J. ZHANG: *RCRC – Ein flexibles System zur Spracherkennung*. Technischer Bericht 98/11, SFB360, Univ. Bielefeld, 1998.
- [18] RICKHEIT, G. und B. HILDEBRANDT: *Der Sonderforschungsbereich „Situierete Künstliche Kommunikatoren“*. Linguistik – Die Bielefelder Sicht, Seiten 145–148, 1997. Bielefeld: Aisthesis Verlag.
- [19] RICKHEIT, G. und I. WACHSMUTH: *Collaborative Research Center “Situating Artificial Communicators” at the University of Bielefeld, Germany*. Artificial Intelligence Review, 10:165–170, 1996.
- [20] SCHEERING, C. und B. KERSTING: *Uncalibrated hand-eye coordination with a redundant camera system*. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, 1998.
- [21] SCHMIDT, R.: *Ein robustes Verfahren zur Roboter-Feinpositionierung durch visuelles Lernen*. Diplomarbeit, Univ. Bielefeld, Arbeitsgruppe Technische Informatik, 1998.
- [22] *Arbeitsmaterialien SFB 360 – Konstruieren im Dialog: „Wir bauen jetzt also ein Flugzeug ...“*. SFB 360, Universität Bielefeld, in Vorb.
- [23] SICHELSCHMIDT, L.: *Der Augenblick des Verstehens: Motorische Indikatoren unmittelbarer semantischer Verarbeitung*. Technischer Bericht 4/95, Zentrum für interdisz. Forschung (ZiF), Univ. Bielefeld,, 1995. ZiF-Mitteilungen.
- [24] STEEDMAN, M.: *Combinatory grammars and human language processing*. In: GARFIELD, J. L. (Herausgeber): *Modularity in knowledge representation and natural-language understanding*, Seiten 187–205. MIT-Press, 1987.
- [25] STEEDMAN, M.: *Categorial grammar*. *Lingua*, 90:221–258, 1993.
- [26] STEEDMAN, M.: *Surface structure and interpretation*. MIT-Press, 1996.
- [27] STRIPPGEN, S., K. PETERS und J.T. MILDE: *Situating communication with a simulated robot*. In: ZOBEL, R. und D. MOELLER (Herausgeber): *ESM98, Simulation-Past, Present And Future, Proceedings of the European Simulation Multiconference*, Seiten 448–452. SCS Publications, 1998.
- [28] WACHSMUTH, I. und B. JUNG: *Dynamic conceptualization in a mechanical-object assembly environment*. Artificial Intelligent Review, 10:345–368, 1996.
- [29] WEISS, P., B. HILDEBRANDT und G. RICKHEIT: *Der Einfluß von Verbspezifität und situativem Kontext auf die Verarbeitung von Handlungsanweisungen*. In: KRAUSE, W., U. KOTKAMP und R. GÖRTZ (Herausgeber): *KogWis97: Proceedings der 3. Fachtagung der Gesellschaft für Kognitionswissenschaft*, Seiten 229–231, 1997.
- [30] WEISS, P., K. KESSLER, B. HILDEBRANDT und H.-J. EIKMEYER: *Inkrementelle und integrative Sprachverarbeitung*. Kognitionswissenschaft, im Druck, 1999.
- [31] ZHANG, J., Y. VON COLLANI und A. KNOLL: *Development of a robot agent for interactive assembly*. In: *Proceedings of 4th International Symposium on Distributed Autonomous Robotic Systems, Karlsruhe, Germany*, 1998.