Abstract: In this paper, a new natural language interface is presented that can be applied to make the use of intelligent robots more flexible. This interface was developed for the autonomous mobile two-arm robot KAMRO, which uses several camera systems to generate an environment model and to perform assembly tasks. A fundamental requirement in human-machine interaction for intelligent robots is the ability to refer to objects in the robot's environment. Hence, the interface and the intelligent system need similar environment models and it is necessary to provide current sensor information. Additional flexibility can be achieved by integrating the man-machine interface into the control architecture of the robot and to give it access to all internal information and to the models that the robot uses for an autonomous behaviour. In order to fully exploit the capabilities of a natural language access, we favour a dialogue-based approach, i.e., for the interface, KANTRA, presented here, the human-machine interaction is not restricted to unidirectional communication.

In this article, we want to report about the joint efforts of the University of Karlsruhe and the University of the Saarland at providing a natural language access to the autonomous mobile two-arm robot KAMRO which is being developed at IPR. In the second section, the state of the art in natural language access to intelligent technical systems is presented. In section three, our mobile robot KAMRO is briefly described. In section four and five, the main aspects of natural language access are summarized, and in section six, our conception for a dialogue-based natural language access system is presented. The article ends with an evaluation and our conclusion for future work.

2. STATE OF THE ART

Although natural language processing and robotics constitute two major areas of AI, they have been studied rather independently. Only a few works are concerned with natural language access for human-machine interaction and communication.

Sondheimer [6] focuses on the problem of spatial reference in natural language machine control. The well known SHAKEY system [7], a mobile robot without manipulators, is able to understand simple commands given in natural language. The work described in [8] concentrates on language-aided instruction for teleoperational control. Specific words can be utilized to simplify the specification of teleoperational functions for the instruction of a remote robot system. Specific words can be used to simplify the specification of teleoperational functions for the instruction of a remote robot system. Torrance [9] presents a natural language interface for a navigating indoor office-based mobile robot. In addition to giving commands and asking questions about the robot's plans the user can associate arbitrary names with specific locations in the environment. Some theoretical aspects of natural language communication with robot systems from the perspective of computer linguistics are discussed in [10]. Other approaches have been concerned with natural language control of autonomous agents within simulated 2D or 3D environments [11-13]. One salient aspect for natural language access to robot systems is the relationship between sensory information and verbal descriptions. Such issues have already been investigated in the field of integrated natural language and vision processing [14-17].
3. THE INTELLIGENT MOBILE ROBOT KAMRO

Higher intelligence and greater autonomy of more advanced robot systems increase the requirements for the design of a flexible interface to control the system on different levels of abstraction. In the KAMRO (Karlsruhe Autonomous Mobile Robot) project, for example, an autonomous mobile robot (Fig. 1) for assembly tasks is being developed which also has the capability of recovering from error situations [18]. The autonomous mobile robot KAMRO is a two-arm robot-system that consists of a mobile platform with an omnidirectional drive system, two Puma 260 manipulators, and different sensors for navigation, docking and manipulation.

KAMRO is capable of performing assembly tasks (Fig. 2) autonomously. The tasks or robot operations can be described on different levels: assembly precedence graphs, implicit elementary operations (pick, place) and explicit elementary operations (grasp, transfer, fine motion, join, exchange, etc.). Given a complex task, it is transformed by the control architecture (Fig. 3) from assembly precedence graph level to explicit elementary operation level. The generation of suitable sequences of elementary operations depends on position and orientation of the assembly parts on the worktable while execution is controlled by the real-time robot control system. Status and sensor data which is given back to the planning-system enable KAMRO to control the execution of the plan and correct it, if necessary.

4. HUMAN-ROBOT-INTERACTION

Tactile, acoustic, and vision sensors provide robots with perceptual capabilities, enabling them to explore and analyze their environment in order to behave more intelligently. The intelligent robot system KAMRO uses several camera systems to generate an environment model in order to perform complex assembly tasks. To interact with this robot system, it would be an insufficient proposal to use natural language just as a command-language for the operator of a robot system. The reason is that incomplete information must may completed by the human operator if the robot is not able to generate the missing information from its sensors or knowledge base.

Four main situations of human-machine interaction can be distinguished in the context of natural language access to the KAMRO system:
• **Task specification:**
  Operations and tasks which are to be performed by the robot can be specified on different levels. On high level the commands are like "assemble Cranfield Benchmark." The lowest level are explicit elementary operations as "finemotion" or "grasp." This is useful if a operator likes to control the robot or its components directly like teleoperation. On a higher level, the operator can also specify a suitable assembly sequence, in order to perform a certain task. Several different descriptions can lead towards a assembly precedence graph generated automatically from the sentences. The most abstract specification would just provide a qualitative description of the desired state, i.e., the final positions of the different assembly parts.

• **Execution monitoring:**
  The capability of autonomous behaviour gives the robot systems the ability to work up an assembly mission in different orders. Therefore, it is important to inform the operator what part of the mission the robot performs. Depending on the demands formulated by the user, it is possible to give the descriptions and explanations in more or less detail.

• **Explanation of error recovering:**
  The capability of recovering from error situations leads to dynamic adjustment of assembly plans during execution. This feature makes it more difficult for the operator to predict and understand the behaviour of the robot. Natural language explanations and descriptions of why and how a certain plan was changed would increase the cooperativeness of the intelligent robot. Similar descriptions are even more important in error situations which can not be autonomously handled by the robot.

• **Updating the environment representation:**
  Because of the dynamic nature of the environment, geometric and visual information can never be complete, even for an autonomous robot equipped with several sensors. In a natural-language dialogue, user and robot can aid each other by providing additional information about the environment and world model.

To completely implement the interaction possibilities on a high level, it would be necessary to take all four above mentioned situations into account.

### 5. DIALOG-ORIENTED USE OF NATURAL LANGUAGE

As mentioned above, simple instructions in a natural language syntax are not sufficient for interacting with an autonomous mobile robot. In any command-based approach the limitations of uni-directional communication will soon become obvious [19]. Flexible human-machine interaction is only possible if intelligent robots are made more responsive. A robot needs the capability to report task specific information, to explain its behaviour, and to provide information about the environment. We claim, that a dialog-based approach has to be taken for all the communicative situations mentioned above. That way, further queries can always be utilized to resolve ambiguities or misunderstandings.

Most existing natural language interfaces have been developed in order to provide access to databases or expert systems (e.g. [20-22]). In general, three main modules of a natural language dialog system can be distinguished:

1. **Analysis component:**
   A parser translates the natural language input from the user into a semantic representation encoded in a knowledge representation language. These propositions form the basis for further processing.

2. **Evaluation component:**
   Utterances have to be interpreted with respect to the internal world model of the intelligent system. Further interaction between natural language access system and application system is required to perform the tasks requested by the user. Information is fed back to the dialog system if queries have to be answered or if the user has to be informed about interesting state changes.

3. **Generation component:**
   The information to be expressed to the user needs to be translated from its propositional form into natural language utterances. Depending on the situational context, either a response, a query, or a textual description is generated.

Various knowledge sources constitute the knowledge base of a natural language dialog system. Lexicon and grammar provide the morpho-syntactic knowledge for the analysis and generation of natural language utterances. Domain-specific conceptual knowledge serves as a link between the linguistic level and the internal representation on the conceptual level of the application system. A record of the dialog contributions of both user and system is stored in the dialog memory which is necessary for the analysis and generation of pronouns, etc. In addition, a natural language interface maintains an explicit model of the user's goals, plans, and prior knowledge about the domain under consideration. Such a user model is a necessary prerequisite for a system to be capable of exhibiting cooperative dialog behaviour.

A peculiarity of a natural language interface to a robot, database, or expert system is the fact that the meaning of a natural language expression depends on the contents of the domain or world model of the application system. Thus, a referential semantics has to be defined, which relates objects on the linguistic level to entities in the knowledge base of the intelligent system. The access to sensory data in an autonomous mobile robot allows the definition of a referential semantics that is perceptually anchored.

Spatial reference, i.e., the way humans refer to the location, direction, orientation, and relation of objects and activities in space, is one of the most important phenomena here. The ability to refer to spatial aspects of the complex environment
is essential for human-robot interaction in any of the four communicative situations mentioned in the previous chapter.

6. THE NATURAL LANGUAGE INTERFACE KANTRA

The KAMRO system is an autonomous mobile two-arm robot situated in a complex dynamic environment. Flexible human-robot interaction is required in order to cope with diverse communicative situations like task specification, execution monitoring, explanation of error recovery, and updating the environment representation. Tasks can be specified on different levels of abstraction. Error recovery is explained and task execution is monitored by generating adequate natural language descriptions. Information about the environment and the world model can be exchanged in a more complex dialogue.

KAMRO's natural language interface KANTRA (KAmro Natural language TRAnslator) is founded on a dialog-based approach, as it has been proposed in the previous chapter. The architecture of our integrated system is shown in Fig. 4.

Natural language commands and queries from the user form the input for the natural language access system. The linguistic analysis translates the natural language expressions into propositions. The syntactic-semantic parser we use is a modified version of SB-PATR [23] which is based on a unification grammar with semantic information.

The propositions are further interpreted in the evaluation component, which is also responsible for the reference semantic interpretation. For this task, the natural language system must have access to all processed information and the environment representation inside the intelligent agent. The evaluation component realizes the interface to the robot and the shared knowledge sources. The encapsulated knowledge of the robot comprises that part of its knowledge base which pertains to low-level control of the mobile platform, manipulators, and sensors. Depending on the robot's reactions, like confirmations, error notifications, sensor input, information about changes in the world model etc., the evaluation component is also responsible for the appropriate reactions and responses of the access system.

Fig. 5: Different localisation expressions for robot and operator

The generation component translates selected propositions into natural language descriptions, explanations, and queries. An incremental generator, which is based on Tree Adjoining Grammars, generates the surface structures [24].

KANTRA is an extension of the VITRA (Visual Translator) system, which allows for natural language access to visual data [15]. A referential semantics has been defined which connects verbal descriptions to visual and geometric information. This approach provides powerful methods to treat the problem of spatial reference. In order to use and understand localization expressions, the interface has to take into account how the user perceives the assembly parts and
the robot (cf. Fig. 5). A more detailed description of the utilization of spatial relations in KANTRA can be found in [25].

7. EVALUATION

In the context of natural language access to the autonomous mobile robot KAMRO, many important applications are presented in the first part of the article: the interface can be used for the four main situations of human-machine interaction: task specification, execution monitoring, explanation of error recovering, and updating the environment. Information about the environment and the world model can be exchanged in a more complex dialogue. Natural language utterances have to be interpreted with respect to the robot's current environment, i.e., a reference semantics is required which connects verbal descriptions to visual and geometric information. Implementing interfaces for all of the mentioned four types of natural dialogue areas is a challenging task. The work described here is only a first step towards the first application, task description.

The KAMRO interface is based on the VITRA (Visual Translator) system, which allows natural language access to visual data. To use spatial relationships and to understand localization expressions, the interface must consider how the operator perceives the assembly parts and the robot. If the natural language interface is supposed to have access to all processed information as well as the environment representation inside the intelligent system, it has to be integrated into the robot's control structure. This way, both systems have access to tasks, plans and internal models.

To implement the interaction possibilities on a high level, it would be necessary to take all four above mentioned situations into account. The presented concept is one step towards future natural language interfaces for intelligent robots.

8. CONCLUSION

The aim of our joint efforts is the integration of the Karlsruhe autonomous mobile robot KAMRO and the natural language component VITRA developed in Saarbruecken. In this contribution, we have focused the problem of task specification or to be more specific, the spatial relations at task description. The specific need for generating and understanding localization expressions was shown, and it was also described how such natural language utterances can be processed taking into account the information provided by the vision sensors and the model of the environment. In this context, the presented concept is one step towards natural language access to intelligent robot systems.

9. ACKNOWLEDGEMENT

This research work was performed at the Institute for Real-Time Computer Systems and Robotics (IPR), Prof. Dr.-Ing. U. Rembold and Prof. Dr.-Ing. R. Dillmann, Faculty for Computer Science, University of Karlsruhe, and at the Department of Computer Science, University of the Saarland, Saarbruecken, Prof. Dr. W. Wahlster. The project is part of the nationally based research project on artificial intelligence (SFB 314) funded by the German Research Foundation (DFG).


