COMBINED GRAPHIC- AND NATURAL LANGUAGE- INTERACTION
(Design and Implementation)

K. H. Hanne, J. Ph. Hoepelman

Fraunhofer Institut für Arbeitswirtschaft und Organisation (FhG/IAO)
Holzgartenstraße 17, 7000 Stuttgart 1, West Germany
Tel: + 49 711 121 3825, Tx: 721 978 iao-d
(UUCP: !mcvax!unido!iaoobel!hanne)
(Eurokom: K H F)

Abstract

The human factor oriented task to improve Man-Computer-Interaction by the possibilities of multimodal communication, i.e by the possibility of a combination of modes like Natural Language, Direct (graphical) Manipulation, and Formal Language was the starting point for our investigations and implementations. Systems have been developed and implemented on SUN workstations in C and in PROLOG on UNIX, providing a structure of modules and a communication layer for combined (multimodal) interfaces. DIS-QUE, Deictic Interaction System - Query Environment, show two of these applications allowing natural language queries combined with graphical selection (deictic actions) on forms and technical drafts. Seen from the human factors-, the linguistics-, and (to a certain extend) from the graphics point of view this kind of combined interaction is an interesting interaction improvement.

Résumé

Le point de départ pour nos investigations et implémentations a été la tâche ergonomique d'améliorer l'interaction homme-machine à l'aide des possibilités des communications multi-modales, i.e. à travers les possibilités de combiner les modes de la manipulation graphique directe, le langage naturel et les langages formales. Des systèmes ont été devélopés et implémentés dans les langages 'C' et PROLOG dans le système d'exploitation d'Unix sur des stations de travail SUN fournissant une structure de modules et une cauche de communication pour des interfaces combinées (multi-modales). Les systèmes DIS-QUE (Deictic Interaction System - Query Environment) représent deux de ces applications permettant des interrogations en langage naturel combiné avec des séléctions graphique (actions deïctiques) sur des formulaires et des esquisses techniques. Du point de vue ergonomique, linguistique et (dans un certain mesure) graphique, cette sorte d'interaction combinée représent une amélioration interessante.

Keywords: Combined User Interface, Graphical Interaction, Natural Language, System

1 Introduction

Graphics at the user interface to computer systems offer considerable advantages and many improvements both in the task of depicting complex structures (or knowledge) and in the task of interacting with trained as well as unskilled users. The great success of systems like Macintosh and Xerox Star which are normally used by non-computer-professionals illustrates how easily direct (graphical) manipulation (DM) can be learned.

The effectiveness for professionals, e.g. in the field of artificial intelligence, can be seen in the acceptance of direct manipulative interfaces on programmers' workstations and Lisp machines.

From the human factors point of view the specific application domain is of minor importance (though, obviously, it determines the implementation).

Natural language (NL), on the other hand, presents undeniable advantages (compare Chapter 2.2) in both spoken and written applications. [4] Important features are e.g. negation which is nearly impossible in direct manipulative systems, the use of quanitfiers (all, exists) or the use of vague expressions which can be handled in (good) NL systems.

The basic reason for the work presented in this paper is the expectation that man-machine-interfaces with separated modes like 'pure' Direct Manipulation or Natural Language have advantages and disadvantages, but can be optimized by a combination of the respective interaction modes. The task of developing multimodal interfaces consequently leads to the question of retrieving knowledge e.g. in expert systems, to the problem of presenting the "user's world" (compare e.g.[14,18]) and to the task of referent resolution of natural language systems combined with graphic interfaces.

In our approach the work was focused on the effects of screen-oriented deictic phenomena. System-architecture and -structure are based on layered

---

The work described in this paper has been carried out in the context of the ESPRIT project (107) and is partly supported by the Commission of the European Communities.

models. The systems are developed and implemented on SUN workstations in 'C' and are PROLOG based on UNIX providing a set of modules and a communication layer for combined (multimodal) interfaces, thus allowing the inclusion of deictic/natural language references to objects represented on the screen. Several applications, a (pure) direct manipulative interface to an expert system in the domain of Aircraft Design and systems allowing natural language/direct manipulation queries have been developed.

The approach and the systems introduced in this paper are investigated and developed in the context of the ESPRIT (European Strategic Programme of Research and Development in Information Technology) project 107 'A Logic Oriented Approach to Knowledge- and Databases Supporting Natural User Interaction' (LOKI) and demonstrate a way of improving user interfaces concerning the design and realization of future man-computer-systems.

2 Generic Interaction Modes

Based on the interface model presented e. g. in [6], three generic (substantially different) communication modes can be isolated, and profiles can be fixed according to the various viewpoints.

2.1 Direct (Graphical) Manipulation

The basic idea of direct manipulation is the visual presentation of the working environment and the objects of immediate interest in a symbolical or mnemotechnical form on a suitable screen and a possibilty to interact directly with the screen objects [6, 16]. These systems are usually implemented in the new generation of screen working places, equipped with a high resolution bitmap-display, and a pointing (selecting) device (e.g. mouse, joystick, touch-screen, etc.).

These interfaces show a fixed and limited model of application, arranged according to the needs of specific user groups. Systems with direct manipulation offer metaphors which represent the system functionality. Typical operations are those which can be carried out quickly and stepwise, the effects of which on the objects are immediately conceivable, and usually reversible. Examples for the use and application of systems with direct manipulation in relation to AI methods are knowledge based information storage and retrieval, dialogue control, and dialogue representation on the screen. The connection of expert systems and systems with direct manipulation allows, among other things, the presentation of objects with unknown names, afeature which is especially useful in learning, for unskilled users, or for those who use the system only sporadically.

2.2 Natural Language

The most prominent medium in human communication is natural language (NL). The high degree to which natural language is mastered by most potential users of computer systems on the one hand, and a series of possible applications on the other, make natural language systems one of the main topics of MCI research, and in addition an important factor in Artificial Intelligence research.

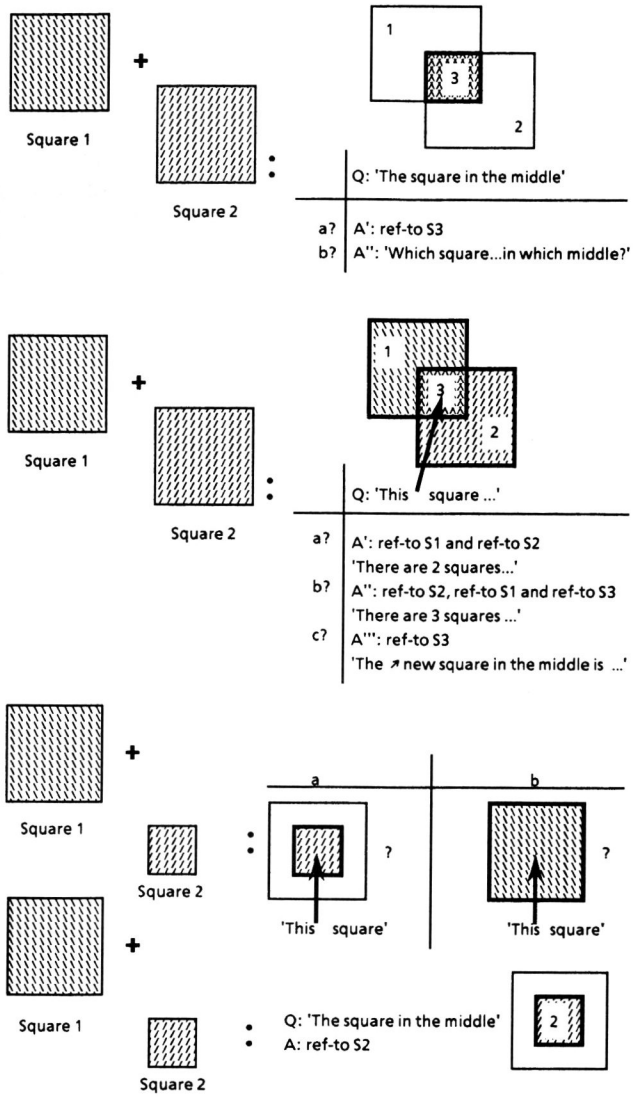
As an argument for the choice of natural language as a means of communication between man and computer, its 'naturalness' is frequently mentioned, the fact that (nearly) everybody is capable of using it, and thus does not have to learn a new, possibly difficult, formalism. Natural language has proved to be an adequate tool for the creation of references to our environment, to objects, actions, and abstract facts. Mechanisms have been developed that allow for effective dialogues, and for the creation of complex consistent texts. In addition, natural language contains means for the expression of, e.g., quantification, time reference, negation, deixis (pointing operations, like *'this'*, *'that'*, *'here'*, *'there'*).

Deixis is of relatively high importance in natural language interaction and is a crucial point in NL man computer interaction. NL is always used in a certain situation- at a certain time and at a certain place by people (or systems) who share a great deal of both, situational perception and general knowledge. These boundaries determine the comprehension of natural language with deictic expressions. The traditional categories of deixis are person, place and time [12,15]. The categories of discourse deixis and social deixis can be added, but in screen oriented MCI place deixis, in the first superficial approach to a pixel on a screen, is of prominent relevance. In human communication, deixis is oriented in an egocentrical way, i.e. both (or all), 'speakers' and 'listeners', have to share a common model of e.g. the viewpoints of the different participants. Of course, place deixis is possible without selection or pointing action on a screen (e.g. *'The square in the middle'* in picture 1).

Problems arising from the application of natural language in the user interface should not be ignored. The most natural means of communication between humans is not necessarily the most 'natural' one between man and machine, and many aspects of natural language are not yet understood theoretically, let alone described algorithmically. 'Good' Nl-systems are rare.

3 Combined (Multi-Modal) Communication

In conventional systems, the communication modes mentioned above are usually realized separately. Combined multimodal man-computer interfaces try to share the advantages of the different generic communication modes avoiding their disadvantages. A combination of these communication modes would be considerably more adequate for the respective tasks or user groups.



Picture 1: NL-Query vs. Graphic / NL-Query - 'Referential ambiguity'

Especially the combination of direct (graphical) manipulation and natural language interaction presents undeniable advantages (see e.g. [2,11,10,13]). In many computer applications already working with graphic objects, there is the simple possibility of combining natural language and graphic objects by pointing or selecting objects and operations. In MCI the expressive power of such combined natural language/direct manipulation interfaces lies in the possibility of allowing deictic interactions on screen-oriented objects.

Without going into detail we would like to mention as an obvious example for the power of deictic/DM interaction the possibility to 'talk' about objects the user has no name for, or to use simple pointing actions, like for example pointing on a map instead of giving complicated spatial descriptions. Applications on maps can help to talk about temporal aspects using spatial relations, if there is an implicit or explicit time axis. Objects which are already synthetic, e.g. business graphics, are well suited for combined nl-deictic interaction; since they are created by the system, an internal representation is implied. Picture 1 again show some of the effects of reference with NL and DM.

Unfortunately this combination of NL and DM cannot be resolved just on the display level, but needs a sort of common representation level, i.e. the system has to create a 'link' between the selection (e.g. 'there' associated with a selecting operation), and the user-intended object on the screen in order to understand for example a question.

4 Deictic Interaction Systems

In the following, deictic expressions will be defined as a reference to objects refered to in an utterance by means of a pointing gesture, e.g. John points at a square and says: *"This ↑ is a new square"* (compare picture 1) with the arrow marking approximately the chronological order of the pointing gesture in the course of the utterance.

Usually, conversational dialogues contain a rich variety of deixis (see e.g. [12],[15]) as humans try to refer deictically whenver possible, thus avoiding complicated, inexact, sometimes not even verbalizable object descriptions whereby the latter are of special importance when asking about unknown objects.

A form appears to be a very adequate instrument for the investigation of deictic phenomena in graphical interfaces. When questioning about the form, the user refers mostly to objects unknown to him, and of which he can at most describe the visual structure.

The FhG / IAO prototype DIS-QUE (Deictic Interaction System - Query Environment)[17] attempts the investigation of such phenomena in the light of the following main questions:

- How can deixis be included in a reasonable way in natural language dialogues between man and machine?
- Which strategy would be adequate to solve multiple references to concepts and overlaying objects?
- What should a question/answer system satisfying the mentioned dialogue requirements look like, how should it be structured?

4.1 Objective and Theory

Related approaches and theoretical work can be found in the investigations documented in [5,12]. Systems based on the philosphy of combined environments are e.g. [8]:

- the Intelligent Interface System FGS, Tokio,
- the ISOBAR System, Japan,
- the systems developed at BBN, USA,
- the VIEW System, USA,

- the " Put - That - There" System , MIT, USA,,
- the Krine System,
- the "Natural Language Graphics "Ohio, USA,
- some systems developed in the ESPRIT programme.

An overview of the field of deixis and its classification can be found e.g. in [12] where form deixis is classified into the category of place deixis subclassified for pointing gestures. A detailed investigation into pointing gestures can be found e.g. in [15].

Place deixis, i.e. pointing actions at e.g. directions, objects, and individuals, can be divided into the two following phases: Pointing, called reference specification, and recognition of the object pointed at, called reference identification, with the object pointed at called the demonstratum. Deixis would be quite easy to handle if there would always be just one possible demonstratum, as in cases in which the speaker points at one unambiguous object and the listener recognises exactly this object as demonstratum.

4.2 Graphical Semantics of Forms

In form deixis, consisting of pointing actions at objects on a form, the problem is simplified insofar as there is no space left between the pointer and the demonstratum. The pointing action is made in the context of a man machine dialogue with the help of a pointing device in the two-dimensional space of the terminal. This means that all those cases where several possible objects lie spatially behind or near each other are excluded from reference identification. However, this does not mean that deixis is no longer ambiguous. In the case of forms, there is the additional problem of interlocking similar objects because the apparent flat and unambiguous form can actually have a complex topology. If the user is not familiar with the terminology, even an additional description of the object is difficult, if not impossible. In short: In form deixis, reference is made to three different groups of objects :

I to directly visible objects on the form, such as input sections, texts, pages, etc.
II to data in the input sections,
III to concepts associated with the objects on the form or with the data in the input sections.

4.3 Deixis in DIS-QUE

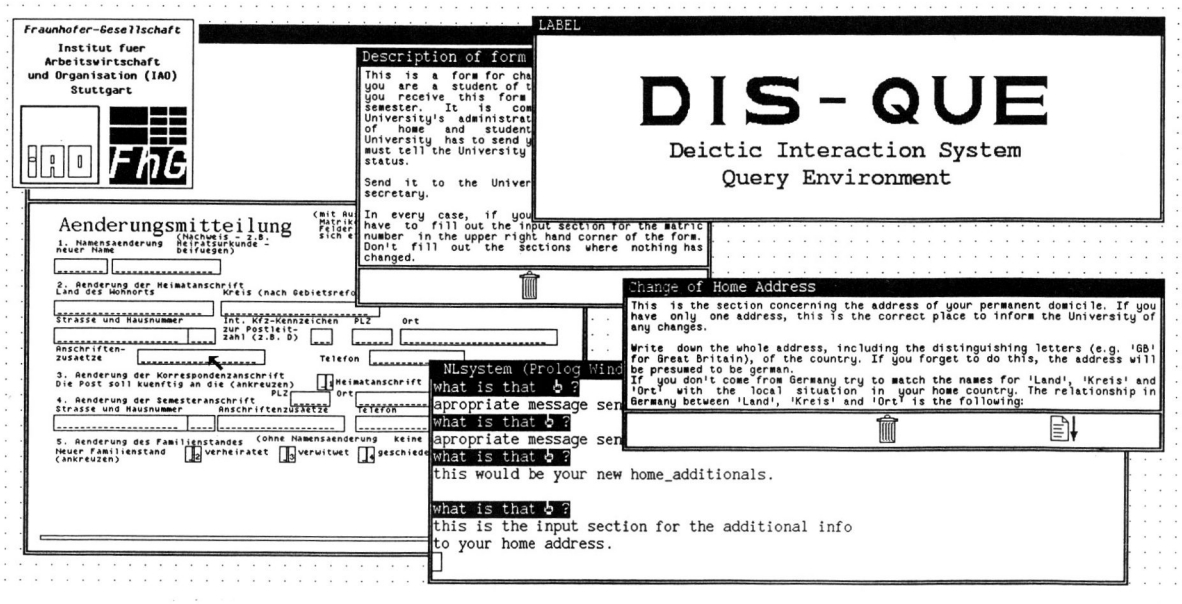
In DIS-QUE (see picture 2), deictic user requirements are reduced to a minimum:

1. the pointing action is included in the sentence, i.e. between two input words, and indicated by an arrow symbol ↑.
2. the user has to activate the actual pointing action by pressing a button on the mouse. Therefore pointing actions in DIS-QUE are always punctual.

The local information thus acquired is then used in combination with heuristics in order to figure out which object or concept was meant. Obviously, no heuristics is good enough to invariably find the object the user is referring to, this however being a problem that appears also in human conversation. If the inquirer receives a wrong answer, he will ask again or rephrase his question more exactly.

The heuristics of DIS-QUE is based on three principles:

- Start globally and go into detail.
- Try to find alternatives.
- Choose the hierarchically nearest object with respect to objects mentioned in the last question.



Picture 2: Example of the System's User Interface - Business Application (English)

5 Implementation of Combined Environments

The implementation was carried out in two phases: first in CProlog[4] on a VT240 graphics terminal under UNIX on Vaxes, and then transferred on a SUN workstation with a bitmap oriented screen under BIMProlog[1]/C and UNIX, offering a number of additional tools developed at the FhG/IAO [9] for the connection of window systems, graphics, and input.

Normally the bottleneck in the application of NL systems, which today are mostly knowledge based systems, is due to the communication between different processes that have to be separated in order to create independent but well defined modules that can be developed in parallel. Prolog in NL- or expert systems are in most cases well suited for reasoning tasks but cannot handle graphical input and output efficiently. In addition, it is worth including existing software developed in other languages. One approach is to separate (quasi) parallel processes and to establish a communication layer which has the advantages of independent processes, well defined interfaces, the inclusion of different languages, the possibility of the integration of an "intelligent" user interface manager which keeps track of the human factors and the history aspects of the interface and which allows for the inclusion of different interaction modes and different technical systems at the user as well as at the system frontend.

The Prolog/C communication system ProCi [7] facilitates the connection of different programming languages and processes (e.g. expert system task vs. help I/O),via process-process communication channels, and the inclusion of program parts written in C to the Prolog program (see picture 3). The same technique allows for the transfer of time consuming Prolog elements to C without changing the structure of the entire system. Additional tools allow graphic oriented in- and output. Other components presented in the diagram are based on Prolog.

6 Architecture of the FhG/IAO Prototypes

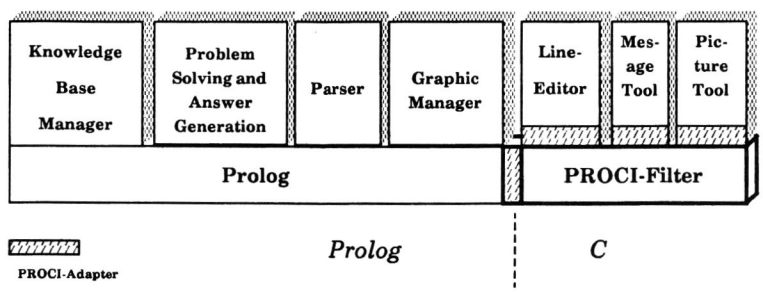
The structure of DIS-QUE is split into 7 different basic components shown concisely in picture 4. The parser works top-down and breadth-first. The grammar is a syntax and semantic controlled phrase structure grammar. The NL part of the system represents a fragment of the English language (and the German language in another application) restricted to simple query sentences, simple statements, and more complex nominal phrases in connection with pointing actions. The graphic manager handles the solution of positional attributes in natural language. Graphical data are stored in a file as a hierarchy of form objects together with the graphical and descriptional information. The data are structured in two hierarchies, a form object hierarchy, and a concept hierarchy representing the underlying concepts and their relations. The connection between the two hierarchies is established by values in the form's data section and by associative relations between objects of both hierarchies. The knowledge base manager pre-processes form data for problem solving. The problem solver (with the answer generating component) receives the semantic representation from the parser, analyzes it, and decides what was asked with the help of the graphic manager, and the knowledge base manager. The glossary is a term definition lexicon which is used by the system and contains about 200 words of the form's domain. For the text generation canned texts are used, which can be presented acoustically using a text-to-speech board.

7 User Interface

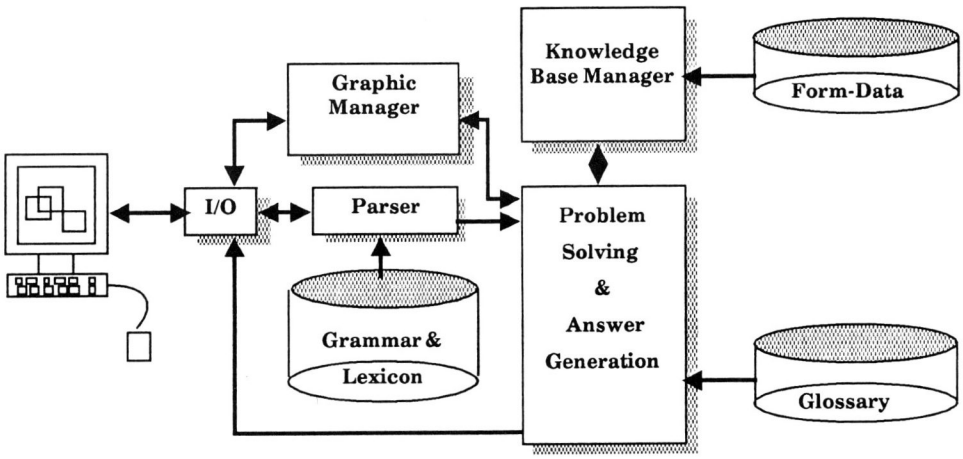
The user interface has been adapted to the possibilities of SUN workstations (see picture 2 and 5). The dialogue between user and program is presented in several different windows. Textual entries made by the user are simple English sentences which may contain pointing actions, e.g.

What is that ↑ ? What does this ↑ text mean ? What text is this ↑ ? What does the text above that section ↑ mean ? This ↑ is Stuttgart. What section is that ↑ ?

For simplification reasons, most questions are just built with the verb to be or to do being sufficient for a prototype since most facts can be expressed this way. The system permits some extent of complexity concerning positional information about the inquired object, e.g.



Picture 3: Implementation Structure



Picture 4: System Overview (DIS-QUE)

What does the text above that ↑ under that text ↑ beside that section ↑ mean ?

For system control there is also the possibility of command input. It is possible to add a text-to-speech board for an acoustic output of texts.

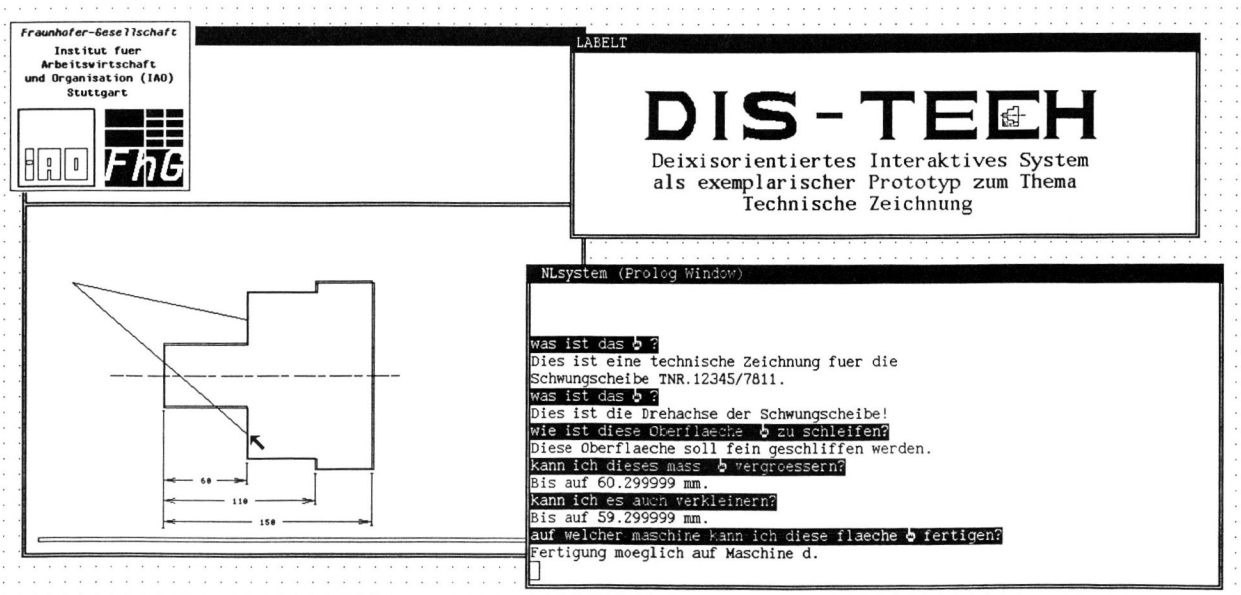
8 Technical Application

Picture 5 shows a sample dialogue step of another developed (technical) application and implemented system. The application area concerns technical drafts; the natural language used in this example is German. Developments in the technical areas can provide 'multimodal redundancy' in CAD tasks for e.g. unskilled users assisting them in learning and using specialized CAD 'programming' languages in a more natural way.

9 Discussion and Future Work

Despite its small performance capacity, DIS-QUE can demonstrate the importance of pointing actions in man-machine-dialogues being of particular interest in the domain of communication especially with unskilled users, as well as in explanatory systems. Here, the problem often occurs that the user has to talk about things he cannot or can only partially put into words, a problem that can be elegantly solved by deictic references. As one of the highest and most complex and at the same time most natural form of human dialogue, deixis will gain more and more importance, for example in the field of optical recognition of movement, and of recognition and production of speech.

The implemented tool-set facilitates the prototypical development of user interfaces according to known principle of MCI and provides a basis for future



Picture 5: Example of the System's User Interface - Technical Application (German)

applications dealing mainly with NL/DM dialogues on business graphics.

In the near future the natural language input will possibly be acoustical, using a (speaker dependent) speech recognognition system. This feature will enhance the acceptance of combined interaction by the intended users.

Acknowledgement

The work described in this paper has been carried out in the context of the ESPRIT project (107) and is partly supported by the Commission of the European Communities. The development of the DIS-QUE system was done by R.P. Wetzel, the development of the communication and graphics modules has been done by A. Graeble.

We would like to thank the partners and participants in the Loki Project (BIM, Belgium; SCICON, U.K.; SCS, Germany; Cretan Computer Institute, Greece; Technical University Munich, Germany; University of Hamburg and INCA, Germany; Cranfield Institute of Technology, U.K.) for their co-operation.

References

[1] BIM, (Belgium Institute of Management), BIM PROLOG Manual, (Everberg, Belgium, September 1986)

[2] Bijl, A; Szalapaj, P., Saying What You Want with Pictures and Words (in: Shackel, B. (Ed.), Proc. 1st Conference on Human-Computer-Interaction INTERACT 84, London, 1984)

[3] Bullinger H.J.; et al., An Integrated Approach to Language Processing (SpeechTechnology, Aug/Sept. 85, pp 62-68, 1985)

[4] Clocksin, W.F.; Mellish, C. S., Programming in Prolog, (Springer, 1985)

[5] Faehnrich, K.-P.; Hanne, K.-H.; Hoepelman, J.; Ziegler, J. Report E3: The Role of Graphics, (Loki Pilot phase FhG/IAO, 1984)

[6] Faehnrich, K.P.; Ziegler, J., Workstations Using Direct Manipulation as Interaction Mode - Aspects of the Design, Application, and Evaluation (in: Shackel, B. (Ed.), Proc. of the 1st Conf. on Human-Computer-Interaction INTERACT 84, London, 1984)

[7] Graeble, A.; Hanne, K.H., Integrated Prolog Graphics Direct Manipulation Interface System. Programmer Manual and Internal Manual (ESPRIT-LOKI Reports KRGR 6.2.2 and 6.2.3, FhG/IAO, Feb. 1987)

[8] Hanne, K.H., Investigation into the Literature of Combined Graphics Textual Systems(ESPRIT-ACORD Report, FhG/IAO, 1986)

[9] Hanne, K.H., Graeble, A.: Design and Implementation of Direct Manipulative and Deictic User Interfaces to Knowledge Based Systems, (in: Bullinger, H.J.; Shackel, B.: Human-Computer-Interaction Interact '87, p 1067- 1073, September 1987)

[10]Hanne, K.H.; Hoepelman, J.Ph.; Faehnrich, K.P., Combined Graphics/ Natural Language Interfaces to Knowledge-Based Systems (in: Proc. Conference on Artificial Intelligence and Advanced Computer Technology, TCM, Liphook, 1986)

[11]Hayes, Ph., Steps Towards Integrating Natural Language and Graphical Interaction for Knowledge-Based Systems (Proc. European Conference on Artificial Intelligence ECAI '86, Vol.1, pp. 456-465, 1986)

[12]Hoepelman, J.Ph.; Hanne, K.H., Oellinger, W., Classification of Deictic Phenomena in Natural Language (ESPRIT-LOKI Report KRGR 5.1A, FhG/IAO, January 1986)

[13]Kobsa, A.; Allgayer, J.; et al., Combining Deictic Gestures and Natural Language for Referent Identification (Proc. International Conference on Computational Linguistics, Bonn, pp. 356-561, 1986)

[14]Mackinlay, J.D., Automatic Design of Graphic Presentations (Stanford University Report STAN-CS-86-1138, 1986)

[15]Schmauks, D., Natural and Simulated Pointing. An Interdisciplinary Survey., (Working Paper Nr. 16, XTRA Project, University of Saarbrücken, March 1987)

[16]Shneiderman, B., The Future of Interactive Systems and the Emerge of Direct Manipulation (Behaviour and Information Technology, Vol.1, No.3, pp.237, 1982)

[17]Wetzel, R.P.; Hanne, K.H.; Hoepelman, J.Ph., DIS-QUE, Deictic Interaction System - Query Environment (ESPRIT-LOKI Report, KRGR 5.3 FhG/IAO, Jan. 1987)

[18]Zdybel, F., An Engine for Intelligent Graphics (in Bolc, L., Jarke, M., Cooperative Interfaces to Information Systems, Springer, New York, 1986)

Appendix: Some NL-Features of DIS-QUE

- English Grammar

The grammar presented here is restricted to the actually natural language part of DIS-QUE's grammar. In DIS-QUE 'Noun' does not only cover single nouns, but also compound nouns such as family status. The category 'Location' includes the pointing action as well as the possiblity to mark a location by using prepositional phrases, e.g

(5) Location → Loc

(6) Location → Locprep {Prep} NP

Rule (5) allows for direct deixis by means of the mouse, ↑ ; Rule (6) allows for relative statements. The facultative second preposition is used for the introduction of compound words orconstructions such as left of. The grammar can be extended so that deixis can occur at any place in the NP without changing the meaning.

- Knowledge Representation

The system contains knowledge about the form's topography, the hierarchy of the underlying form's concepts, the sections' contents, and their relation to the two hierarchies, namely the relation between the form's object and the concept hierarchy description.