In: T. Catarci, M. F. Costabile, S. Levialdi (eds.): Advanced Visual Interfaces World Scientific, Singapore, pp. 243-257.

Automatic Design of Multimodal Presentations[1]

Wolfgang Wahlster

German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
D-W-6600 Saarbrücken 11, Germany
email: wahlster@dfki.uni-sb.de

Abstract:
We describe our attempt to integrate multiple AI components such as planning, knowledge representation, natural language generation, and graphics generation into a functioning prototype called WIP that plans and coordinates multimodal presentations in which all material is generated by the system. WIP allows the generation of alternate presentations of the same content taking into account various contextual factors such as the user's degree of expertise and preferences for a particular output medium or mode. The current prototype of WIP generates multimodal explanations and instructions for assembling, using, maintaining or repairing physical devices. This paper introduces the task, the functionality and the architecture of the WIP system. We show that in WIP the design of a multimodal document is viewed as a non-monotonic process that includes various revisions of preliminary results, massive replanning and plan repairs, and many negotiations between design and realization components in order to achieve an optimal division of work between text and graphics. We describe how the plan-based approach to presentation design can be exploited so that graphics generation influences the production of text and vice versa. Finally, we discuss the generation of cross-modal expressions that establish referential relationships between text and graphics elements.

1. Introduction

When explaining how to use a technical device humans will often utilize a combination of language and graphics. It is a rare instruction manual that does not contain illustrations. Multimodal presentation systems combining natural language and graphics take advantage of both the individual strength of each communication mode and the fact that both modes can be employed in parallel. It is an important goal of this research not simply to merge the verbalization results of a natural language generator and the visualization results of a knowledge-based graphics design component, but to carefully coordinate natural language and graphics in such a way that they generate a multiplicative improvement in communication capabilities. Allowing all of the modalities to refer to and depend upon each other is a key to the richness of multimodal communication.

We describe the basic methods used in our attempt to integrate multiple AI components such as planning, knowledge representation, natural language generation, and graphics generation into a functioning prototype called WIP (cf. [Wahlster et al. 91], [Wahlster et al. 92a]) that plans and coordinates multimodal presentations in which all material is generated by the system. We concentrate on the intercomponent interactions and synergies that arise from combining components.

The current prototype of WIP generates multimodal explanations and instructions for assembling, using, maintaining or repairing physical devices. WIP is currently able to generate simple German or English explanations for using an espresso machine, assembling a lawn-mower, or installing a modem, demonstrating our claim of language and application independence.

Since one of the design principles behind WIP is that the theoretical basis of all components should be sound enough to allow scaleup, we combined and extended only formalisms that have reached a certain level of maturity. The formal frameworks used in WIP are terminological logics, RST-based planning, constraint processing techniques, and tree adjoining grammars with feature unification.

One of the important insights we gained from building the WIP system is that it is actually possible to

[1] This is a revised and shortened version of material I have written for [Wahlster et al. 92b]

extend and adapt many of the fundamental concepts developed to date in AI and computational linguistics for the generation of natural language in such a way that they become useful for the generation of graphics and text-picture combinations as well. This means that an interesting methodological transfer from the area of natural language processing to a much broader computational model of multimodal communication seems possible. In particular, semantic and pragmatic concepts like coherence, focus, communicative act, discourse model, reference, implicature, anaphora, or scope ambiguity take an extended meaning in the context of text-picture combinations.

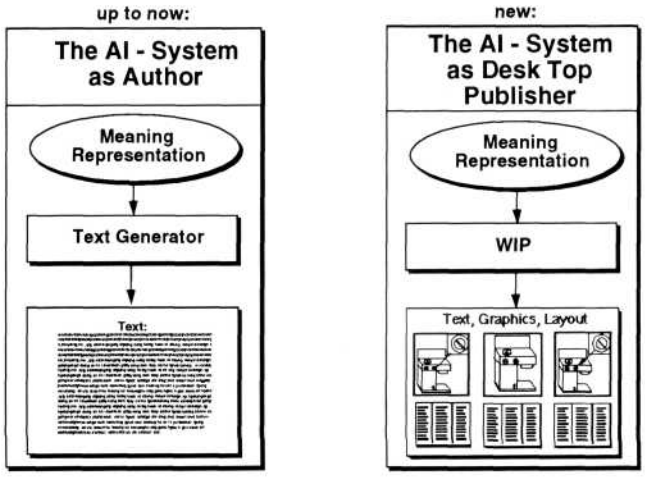

Fig. 1. From Text Generation to the Design of Multimodal Documents

A basic principle underlying the WIP model is that the various constituents of a multimodal presentation should be generated from a common representation of what is to be conveyed. This raises the question of how to decompose a given communicative goal into subgoals to be realized by the mode-specific generators, so that the modalities complement each other. Enforcing a consistent, harmonious and aesthetic integration of text and graphics is an essential subtask in automating the synthesis of multimodal presentations. To address this problem, we explored computational models of the cognitive decision process, coping with questions such as what should go into text, what should go into graphics, and which kinds of links between the verbal and non-verbal fragments are necessary. In addition, WIP deals with page layout as a rhetorical force, influencing the intentional and attentional state of the reader. In summary, systems like WIP shift the metaphor of "computer as author" (see Fig. 1) used in natural language generation to the broader view of "computer as desktop publisher" [see Dale 92].

The rest of the paper is organized as follows: Sections 2, 3 and 5 introduce the task, the functionality and the architecture of the WIP system, respectively. Section 4 provides a survey of related research and highlights the distinguishing features of the approach discussed in this paper. WIP's presentation planning process is described in section 6. In section 7 we discuss the generation of cross-modal expressions that establish referential relationships between text and graphics elements.

2. Generating Situated Presentations

WIP is a highly adaptive interface, since all of its output is generated on the fly and customized for the intended target audience and situation. The quest for adaptation is based on the fact that it is impossible to anticipate the needs and requirements of each potential user in an infinite number of presentation situations. Thus all presentation decisions are postponed until runtime. In contrast to hypermedia-based approaches to adaptive information presentation, WIP does not use any preplanned texts or graphics. That is, each presentation is designed from scratch by reasoning from first principles using commonsense presentation knowledge. Through its clear separation of content and form WIP goes well beyond

hypermedia systems.
We view the design of multimodal presentations including text and graphics design as a sub-area of general communication design. We approximate the fact that communication is always situated by introducing generation parameters (see Figs. 2 and 4) in our model. The current system includes a choice between user stereotypes (e.g. novice, expert), target languages (German vs. English), layout formats (e.g. hardcopy of instruction manual, screen display), and output modes (incremental output vs. complete output only). The set of generation parameters is used to specify design constraints that must be satisfied by the final presentation.

Presentation design can also be viewed as a relatively unexplored area of commonsense reasoning. Unlike most research on commonsense reasoning to date, the WIP project does not aim at metadomain research on general design principles, but focuses on logic-oriented methods capturing some of the reasoning in the design space of presentations for specific and realistic domains. A diverse set of evaluation knowledge for text, graphics and layout is necessary to select a particular design that satisfies the design specifications stated as generation parameters. WIP provides computationally tractable evaluations of candidate designs at various levels of the incremental generation process.The concept of tailoring presentations for the user can be seen as an extended version of the view concept known from database technology. One step on the way to intelligent interfaces for computer-supported collaborative work (CSCW) is to use multimodal systems like WIP as presentation experts that map fragments of a shared knowledge-base onto a variety of presentations satisfying the information needs of the individual group members (see Fig. 2).

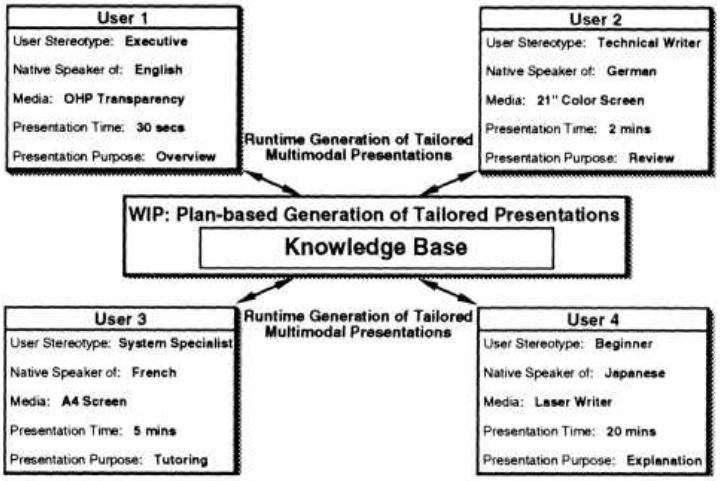

Fig. 2. Adaptive Multimodal Information Presentation in a Distributed Setting

In summary, WIP allows the generation of alternate presentations of the same content taking into account various contextual factors such as the user's degree of expertise and preferences for a particular output medium or mode.

3. Related Research

Over the past several years, a number of projects have entered the area between natural language processing and multimodal communication, often focusing on a single specific functionality, such as the use of pointing gestures parallel to verbal descriptions for referent identification ([Cohen et al. 89], [Kobsa et al. 86], [Neal & Shapiro 91]). The automatic design of complete multimodal presentations has only recently received significant attention in artificial intelligence research. The most extensive discussion of active research in this field can be found in the proceedings of a series of workshops on intelligent multimedia interfaces (e.g., [Arens et al. 89], [Sullivan & Tyler 91], [Maybury 91]).

We have been engaged in work in the area of multimodal communication for several years now, starting with the HAM-ANS ([Wahlster et al. 83]) and VITRA systems ([André et al 86], [Herzog et al. 89], [Wahlster 89])), which automatically create natural language descriptions of pictures and image sequences shown on the screen. These projects resulted in a better understanding of how perception interacts with language production. Subsequently, we have been investigating ways of integrating tactile pointing with natural language understanding and generation in the XTRA project ([Kobsa et al 86], [Wahlster 91]). WIP grew out of the experiences of our previous research into multimodal interaction, particularly in the VITRA and XTRA projects.

Various user interfaces to date combine natural language and graphics, but only a few of them ([Kerpedjiev 92], [Marks & Reiter 90], [McKeown & Feiner 90], [Roth et al. 91], [Wahlster et al. 91]) generate both forms of presentation from a common representation and thus can explicitly address the problem of media choice and coordination. For example, Kerpedjiev has designed a system that transforms a dataset about a particular weather situation into a multimodal weather report consisting of a text illustrated by tables and weather maps with various icons and annotations ([Kerpedjiev 92]). Whereas most systems combine text with informational graphics (e.g. maps, diagrams, charts), COMET ([McKeown & Feiner 90]) and WIP generate text illustrated by 3D graphics of physical objects.

The work closest to our own is done in the COMET project ([Feiner & McKeown 90]). Both projects share a strong research interest in the coordination of text and graphics. COMET generates directions for the maintenance and repair of a portable radio using text coordinated with 3D graphics. In spite of many similarities, there are major differences between COMET and WIP, e.g. in the systems' architectures, representation languages and processing strategies. While during one of the final processing steps of COMET the media layout component is supposed to combine text and graphics fragments produced by media-specific generators, in WIP's architecture a layout manager begins to interact with a presentation planner before text and graphics are generated, so that layout considerations can influence the early stages of the planning process and constrain the media-specific generators. In WIP we view layout as an important carrier of meaning. COMET uses a schema-based content planner while WIP uses an operator-based approach to planning. Another distinguishing feature of WIP's architecture is that it supports incremental output and a direct interaction of text and graphics design.

The importance of the layout dimension is also stressed in recent work by Hovy and Arens that involves the generation of formatted text exploiting the communicative function of headings, enumerations and footnotes ([Hovy & Arens 91]).

	Informational Graphics	**3D Graphics of Physical Objects**
Static Media	**Maps, Charts, Diagrams** **Example Systems: SAGE, FNN**	**Rendered Pictures** **Example Systems: WIP, COMET**
Dynamic Media	**Hypermedia Presentations** **Example Systems: AIFresco, IDAS**	**Animation** **Example Systems: VITRA-SOCCER, AnimNL**

Fig. 3. Combining Text Production with Four Types of Graphics Generation.

Whereas the majority of work has concentrated on combining static media, the VITRA-Soccer project ([Herzog et al. 89]; for details of VITRA's animation component see [Schirra 92]), the AnimNL project ([Badler et al. 91]) and recent extensions of COMET ([Feiner et al. 91]) and WIP additionally deal with dynamic media, such as animation. Systems like AlFresco ([Stock 91]) and IDAS ([Reiter et al. 92]) demonstrate that natural language generation can be enhanced by integration with hypermedia systems. In such systems the generated text may contain links to hypercards and canned text or images can be combined with generated text for a hypermedia presentation.

Figure 3 summarizes the various types of graphical presentations that have been combined with generated text in recent research prototypes. In all these projects the generation system is no longer only the author of a text, but plays the role of a desktop publisher, a hypertext designer, a multimodal interface designer or a commentator of animations.

4. A Functional View of WIP

The task of the knowledge-based presentation system WIP is the context-sensitive generation of a variety of multimodal documents from an input including a presentation goal. The presentation goal is a formal representation of the communicative intent specified by a back-end application system.

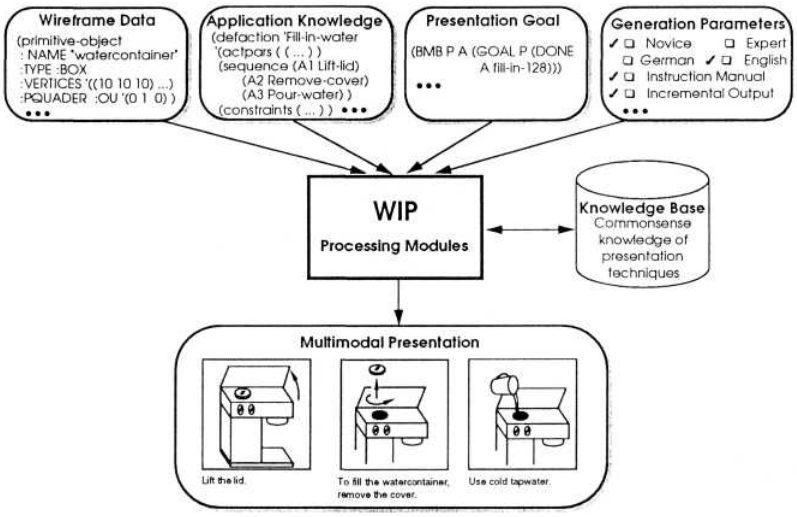

Fig. 4. WIP - A Functional View

The example of a presentation goal in Fig. 4 represents the system's assumption about the mutual belief (BMB) of the presenter P and the addressee A, that it is P's goal that A carries out a plan denoted by the constant fill-in-128. This is a concrete domain plan specified as part of WIP's application knowledge. In this case, the plan is a fully instantiated sequence of actions represented in the assertional part of the hybrid knowledge representation system RAT (Representation of Actions in Terminological Logics, cf. [Heinsohn et al. 92]). The terminological part of RAT is used to represent the ontology and abstract plans for a particular application domain (see Fig. 4).

In addition to this propositional representation, that includes the relevant information about the structure, function, behavior, and use of the technical device, WIP has access to an analogical representation of the geometry of the machine in the form of a wireframe model (see Fig.4).

WIP is a transportable interface based on processing schemes that are independent of any particular back-end system and thus requires only a limited effort to adapt to a new application. Obviously, for a new domain the application knowledge and the wireframe model must be transformed into WIP's representation schemes. While for each domain the application knowledge and the wireframe model

are fixed, the presentation goal and the generation parameters can be varied to tailor WIP's results to a particular communicative situation. WIP is designed for interfacing with heterogeneous back-end systems such as expert systems, tutoring systems, intelligent control panels, and help systems, which supply the presentation system with the necessary input.

Note that the incrementality mentioned in section 2 as one of the options for the generation of multimodal output, characterizes a likely application scenario for systems like WIP, since the intended use includes intelligent control panels and active help systems, where the timeliness and fluency of output is critical, e.g. when generating a warning. In such a situation, the presentation system must be able to start with an incremental output although it has not yet received all the information to be conveyed from the back-end system (cf. [Finkler & Schauder 92]).

WIP can also be used in a standalone fashion, where an author specifies the necessary domain information. This leads to the long-term vision of an intelligent authoring system, that forces one to specify information only once in a formal way and then allows the generation of a possibly infinite variety of presentations of this information tailored to various audiences and media. In contrast to the current situation in technical writing and document preparation, this approach - similar to the view concept in database design - could ensure consistency across all derived presentations, since the underlying content is stored only in one place.

.

5. The Architecture of the WIP System

The architecture of the WIP system guarantees a design process with a large degree of freedom that can be used to tailor the presentation to suit the specific context. During the design process, a presentation planner (cf. [André & Rist 90a,b]) and a layout manager (cf. [Graf 92]) orchestrate the mode-specific generators, and the design record (see Fig. 5) provides information about intermediate results of the presentation design that is exploited in order to prevent disconcerting or incoherent output. This means that decisions of the language generator may influence graphics generation and that graphical constraints may sometimes force decisions in the language production process.

Fig.5 shows a sketch of WIP's current architecture. Note that WIP includes two parallel processing cascades for the incremental generation of text and graphics. In WIP, the design of a multimodal document is viewed as a non-monotonic process that includes various revisions of preliminary results, massive replanning or plan repairs, and many negotiations between the corresponding design and realization components in order to achieve a fine-grained and optimal division of work between the selected presentation modes.

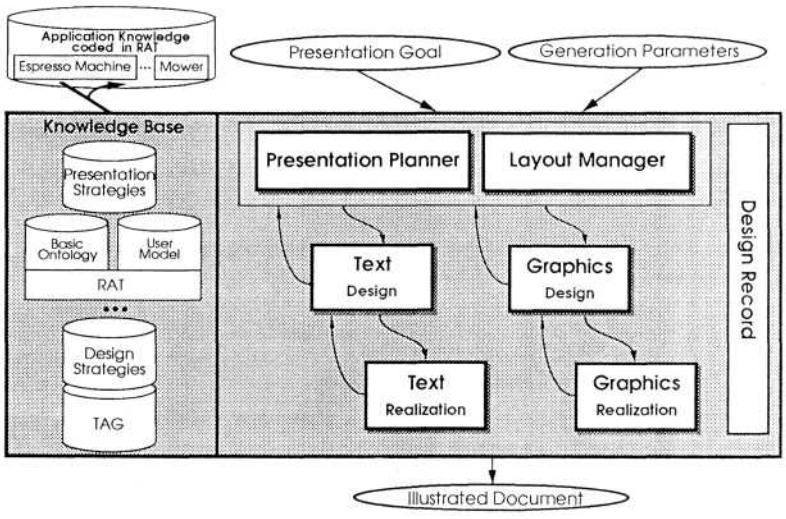

Fig. 5. The Architecture of the WIP system.

The incremental, cascaded architecture with feedback and negotiation among components supports self-monitoring and the anticipation of the addressee's interpretation (see [Wahlster 911). A diverse set of evaluation knowledge for text, graphics and layout is necessary to select a particular design that satisfies the design specifications stated as generation parameters. WIP's architecture enables computationally tractable evaluations of candidate design at various levels of the incremental generation process.

The presentation planner sets parameters that influence decision processes in the text and graphics design components. For example, the generation mode is set to 'noun phrase' for the synthesis of figure captions. The text and graphics design components can be seen as micro-planners of the what-to-say and what-to-show parts of the mode-specific generators. For example, lexical choice is not carried out by the presentation planner on the macro-plan level, but by the text design component.

The design record keeps the history of the design decisions on all levels of the incremental generation process (see Fig. 5). All components of WIP have access to the central design record. WIP's basic ontology and user model are represented in the terminological logic RAT (see Fig. 5). In addition, WIP's knowledge base includes declaratively coded presentation strategies, design strategies and a lexicalized Tree Adjoining Grammar (TAG, cf. [Harbusch et al. 91]). As the result of a 30 person-year effort the WIP prototype is fully implemented, comprising 5.5 MB of Common Lisp and CLOS source code.

6. The Presentation Planning Process

At the heart of the presentation system is a parallel top-down planner (cf. [André & Rist 92]) and a constraint-based layout manager. The presentation planner receives a request for communication in the form of a high-level presentation goal (see Fig. 4). It then accesses presentation knowledge to analyze this goal and to generate a refinement-style plan in the form of a directed acyclic graph (DAG). The leaves of the planning DAG are specifications for individual acts of presentation. These are sent to the appropriate task queue of the text or graphics design component. The text designer handles elementary acts, such as s-assert (generate a surface structure for an assertion) or s-request (generate a request). The graphics designer executes pictorial acts, such as s-depict (generate a picture) or s-annotate (label an object). The text and graphics design processes consist of a choice among several different possible realizations: the realizations that best achieve several goals are preferred. Here further refinements of individual presentation goals are possible.

Since the presentation planner has no direct access to knowledge concerning mode-specific realization, it cannot consider this information when building up a candidate document structure. Thus, it is not able to foresee in which way parts of a document are eventually combined by the generation components. This means that the initial plan often has to be revised to incorporate the results provided by the generators. When revising a first draft of a presentation it is not uncommon for the desire to change one word or graphics element to necessitate the replanning of an entire utterance or picture. This illustrates the problem of dependencies among choices. That is, in order to determine how to express one part of the input WIP must consider the way the surrounding part will turn out. This means that information is propagated not only top-down but also bottom-up in the DAG representing the current presentation plan. However, due to the distributed processing scheme of WIP it cannot be guaranteed that the results of the individual components are always available at a given time. In some situations, it might happen that the planner is not able to expand a node because it is still waiting for a generator to supply results. To avoid processing delays, WIP's presentation planner expands nodes not always in a depth-first fashion, but flexibly selects the nodes to be expanded using heuristics, such as the number of assumptions to be made. To allow for alternating revision and expansion processes, WIP's presentation planner is controlled by a plan monitor that also determines the next nodes to be expanded (cf. [André & Rist 92]).

7. Connecting Verbal and Pictorial Elements by Cross-Modal Expressions

In a multimodal presentation, cross-modal expressions establish referential relationships of representations in one modality to representations in another modality. The use of cross-modal deictic assertions such as (a) "The on/off switch is located in the upper left part of the picture" is essential for the efficient coordination of text and graphics in illustrated documents (see Fig. 6).

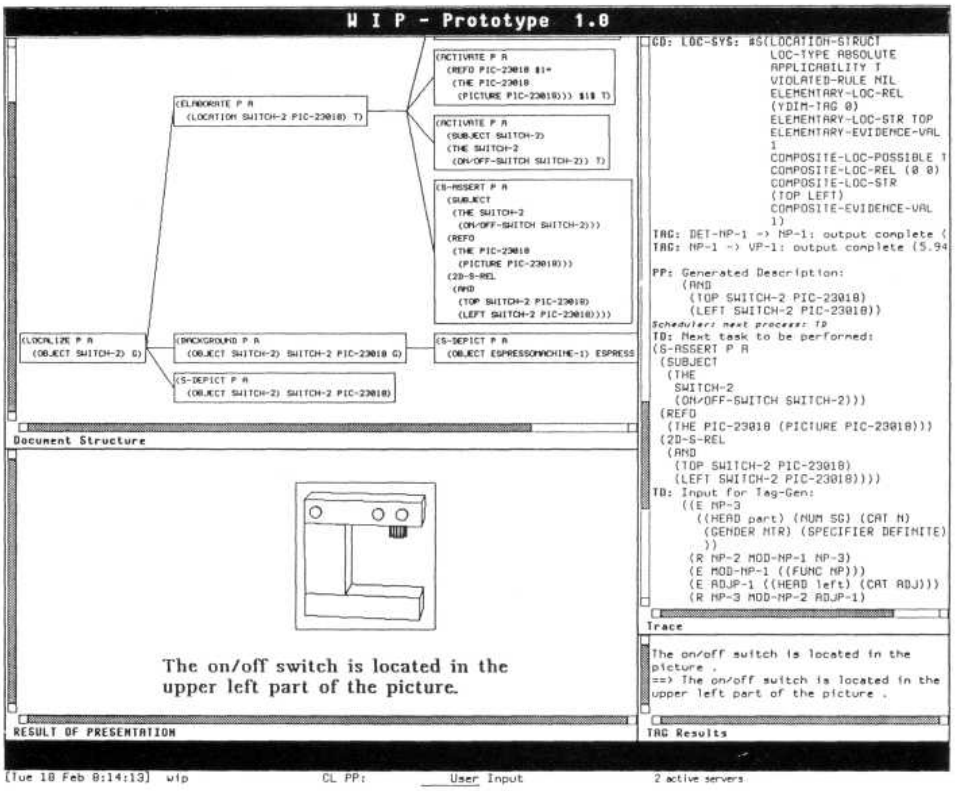

Fig. 6. Incremental Generation of a Cross-modal Reference.

Given the presentation goal (BMB P A (3D-LOC SWITCH-2)), the presentation planner designs the text-picture combination in the bottom left pane of Fig. 6 communicating the relevant information about the spatial position of the on/off switch.

In this example, WIP uses a spatial description to refer to an object shown in a synthetic picture of the espresso machine. Note that the multimodal referential act is only successful if the addressee is able to identify the intended knob of the real espresso machine. Obviously the depiction of the switch cannot itself be used as an on/off switch, but only refers to a physical object which is the result of a multi-level reference resolution. The cross-modal assertion in the text refers to a pictorial element that visualizes an instance of a concept represented by a RAT term as part of WIP's application knowledge. An additional coreferentiality relation exists between the individual constant SWITCH-2 in the ABox of RAT and an object in the wireframe model of the machine providing a description of the geometry of that knob. Finally, the depiction of the knob generated by WIP's graphics design component in turn refers to the corresponding switch of the real machine.

The generation of cross-modal expressions highlights the tight interaction between various components of WIP and the cross dependencies among decisions of the mode-specific generators. In our example after a first draft of the picture has been completed by the graphics designer, the presentation planner activates the text design component which calls the graphics component once again to ask for a localization of a pictorial element.

No serial architecture with a total ordering of the components for text and graphics generation would be adequate in this case. Obviously the spatial relations cannot be computed in advance because they depend on the viewing specification chosen by the graphics design component. It apparently makes no sense to generate spatial descriptions before a particular 2D projection for the 3D wireframe model is chosen. On the one hand it is impossible to generate a natural language expression with a spatial reference to a picture before that picture is designed. On the other hand the combinatorial explosion involved in the generation of all possible spatial relations between graphical elements of a designed picture excludes the synthesis of spatial descriptions without knowing whether they will be needed.

The top left pane in Fig. 6, labeled 'Document Structure', shows a fragment of the DAG produced by the presentation planner. Note that the LOCALIZE act is decomposed into three acts. The main act specifies the graphics designer's task, which is to depict SWITCH-2 in a picture. One subsidiary act tries to provide background information for the generated depiction by showing other salient parts of the machine as the visual context of the switch. The other subsidiary act is supposed to generate text that elaborates on the picture. Further refinements using presentation strategies for textual elaboration finally lead to the cross-modal expression discussed above. Although the mode flag is set to TEXT for this elaboration (coded as T in the corresponding node of the presentation plan, see Fig. 6), the graphics designer is used to compute a spatial relation describing the absolute localization of the switch in the picture.

The most important steps in the design process leading to the cross-modal assertion (a) are shown in the top right pane of Fig. 6 which displays a partial trace of the interaction between the major components of the presentation system. After the presentation planner (PP in the trace) has established a new node in the DAG that contains an unbound variable representing a description of the location of the switch in the picture, the graphics designer (GD in the trace) calls its localization component to determine the value of that variable. One of the basic ideas behind this component is that absolute localizations like 'in the upper left part of the picture' can be derived from relative spatial predicates like 'left-of(x,y)' and 'on-top-of(x,y)' through the use of virtual reference objects induced by the page layout. This means that objects depicted in a figure can be spatially related to the center, the corners, the borderline and even to the caption of that figure.

In the example shown, the image of the espresso machine is displayed by the graphics component in a rectangular picture region. This is used as a frame of reference for the spatial description encoding the position of SWITCH-2's depiction (see the bottom left pane of Fig. 8). The relative location of the on/off switch is described by the conjunction of the literals 'left-of(SWITCH-2, center(PIC-23018))' and 'on-top-of(SWITCH2,center(PIC-23018))', that use the center of the figure as a reference object. In WIP, the center of a picture is approximated by a virtual rectangle in the middle with one third of the horizontal and vertical extension of the whole figure (for more details see [Wazinski 92]). These relative localizations are then transformed into absolute ones by deleting the second argument. The presentation planner forwards the result of the localization process to the text design (TD) component for lexical choice (see top left pane of Fig. 6).

The generation of cross-modal expressions can involve various levels of recursion. One subtlety not illustrated by the example above is the use of different frames of reference for spatial relations in a single cross-modal expression. Suppose that in addition to the picture discussed in the previous example, another figure is placed on the same page. Then the generic localization methods of WIP will generate another relative description like 'right-of(PIC-23018, center(PAGE-l))1 leading to a recursive spatial reference such as 'in the upper left part of the figure on the right'. Since the layout constraints specified in WIP's input, together with revisions of the presentation planner force the layout manager to backtrack from time to time during the incremental design of a multimodal presentation, it may turn out that a figure must be repositioned and thus parts of the cross-modal expression must be revised. For example, 'the figure on the right' may become 'the figure at the top'.

Another level of recursion in the localization process is introduced by dealing with groups of objects. In this case, a group can serve at the same time as a frame of reference for one of its elements and as a perceptual unit that itself has to be localized using other reference objects in the figure (cf. [Wahlster et. al. 78]). For example, the generation of a localization for the group of two switches on the right part of the machine in Fig. 7 leads to a cross-modal expression like 'The left button on the right part of the picture is the selector switch' (see [Wazinski 92] for further details).

As illustrated by this example such verbal descriptions can get quite long-winded. Therefore WIP's presentation strategies include alternate methods to establish cross-modal referential relations. The graphics generator includes a module that supports various labeling techniques for placing text strings in a figure so that they annotate the parts of a composite object in an illustration. The generation of labels as a part of the graphics design is an example, where compared to the previous discussions about the localization component the dependency between graphics generation and text generation is reversed. In this case the text generator is activated during the graphics design process in order to produce a string that can be used for labeling a picture element. Note that the same terminology should be used to refer to objects in both text and graphics. It would lead to an incoherent text-picture combination, if a switch that is labeled 'on/off switch' in a picture is referred to as 'starting switch' in the corresponding text. This means that for the generation of multimodal presentations the design record plays the same role as the discourse model for verbal communication allowing the presentation planner to ensure the consistent use of referential expressions across modes.

Conclusion

We introduced the architecture of the knowledge-based presentation system WIP. It includes two parallel cascades for the incremental generation of text and graphics. We showed that in WIP the design of a multimodal document is viewed as a non-monotonic process that includes various revisions of preliminary results, massive replanning and plan repairs, and many negotiations between design and realization components in order to achieve an optimal division of work between text and graphics. We described how the plan-based approach to presentation design can be exploited so that graphics generation influences the production of text and vice versa. In particular, we showed how WIP can generate cross-modal references and revise text due to graphical constraints.

Acknowledgements

The WIP project is supported by the German Ministry of Research and Technology under grant ITW8901 8. The development of WIP is an ongoing group effort and has benefited from the contributions of my collaborators Elisabeth André, Thomas Rist, Winfried Graf, Wolfgang Finkler, Karin Harbusch, Jochen Heinsohn, Bernhard Nebel, Hans-Jürgen Profitlich, and Anne Schauder as well as our students Andreas Butz, Bernd Herrmann, Antonio Krüger, Daniel Kudenko, Thomas Schiffmann, Georg Schneider, Frank Schneiderlöchner, Christoph Schommer, Dudung Soetopo, and Detlev Zimmermann. Thanks go to Andrew Csinger for comments on an earlier draft of this paper.

References

[André et al. 86] E.André, G. Bosch, G. Herzog and T. Rist, Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions, in: Proceedings 7th European Conference on Artificial Intelligence (ECAI-86), Brighton, England (1986) Vol. 2, 1-8.

[André & Rist 90a] E. André and T. Rist, Towards a Plan-Based Synthesis of Illustrated Documents, in: Proceedings 9th European Conference on Artificial Intelligence (ECAI-90), Stockholm, Sweden (1990) 25-30.

[André & Rist 90b] E. André and T. Rist, Synthesizing Illustrated Documents: A Plan-Based Approach, in: Proceedings InfoJapan-90, Tokyo, Japan (1990) Vol. 2, 163-170. [André & Rist 92] E. André and T. Rist, The Design of Illustrated Documents as a Planning Task, in: M. Maybury, editor, Intelligent Multimedia Interfaces, AAAI Press, (1992).

[André et al. 92] E. André, W. Finkler, W. Graf, T. Rist, A. Schauder and W. Wahlster, WIP: The Automatic Synthesis of Multimodal Presentations, in: M. Maybury, editor, Intelligent Multimedia Interfaces, AAAI Press, (1992).

[Arens et al. 89] Y. Arens, S. Feiner, J. Hollan and B. Neches, eds., A New Generation of Intelligent Interfaces. Workshop, IJCAI-89, Detroit, MI (1989).

[Badler et al. 91] N. Badler, B. Webber, J. Kalita and J. Esakov, Animation from Instructions, in: N. Badler, B. Barsky and D.Zeltzer, eds., Making them Move: Mechanics, Control Animation of Articulated Figure, Morgan Kaufmann, San Mateo, CA, (1991) 51-93.

[Bandyopadhyay 90] S. Bandyopadhyay, Towards an Understanding of Coherence in Multimodal Discourse, Technical Memo TM-90-01, DFKI, Saarbrücken, Germany (1990).

[Cohen et al. 89] P.R. Cohen, J.W. Sullivan, M. Dalyrymple, R.A. Gargan, D.B. Moran, J.O. Schlossberg, F.C.N. Perreira and S.W. Tyler, Synergistic Use of Direct Manipulation and Natural Language, in: Proceedings CHI-89, Austin, (1989) 227-233.

[Dale 92] Visible Language: Multimodal Constraints in Information Presentation. In: Dale, R., Hovy, E., Rosner, D., Stock, O. (eds.): Aspects of Automated Natural Language Generation. Heidelberg: Springer (1992), 281-283.

[Feiner & McKeown 90] S.K. Feiner and K. McKeown, Coordinating text and graphics in explanation generation, in: Proceedings AAAI-90, Boston (1990) 442-449.

[Feiner et al. 91] S.K. Feiner, D.J. Litman, K.R. McKeown and R.J. Passonneau, Towards Coordinated Temporal Multimedia Presentations, in: M. Maybury, editor, Intelligent Mutimedia Interfaces, Workshop Notes from AAAI-91, Anaheim (1991).

[Finkler & Schauder 92] W. Finkler and A. Schauder, Effects of Incremental Output on Incremental Natural Language Generation, in: Proceedings 10th European Conference on Artificial Intelligence (ECAI-92), Vienna, Austria (1992)

[Graf 92] W. Graf, Constrained-Based Graphical Layout of Multimodal Presentations, in: Proceedings Advanced Visual Interfaces (AVI) Workshop, Rome, Italy (1992)

[Harbusch et al. 91] K. Harbusch, W. Finkler and A. Schauder, Incremental Syntax Generation with Tree Adjoining Grammars, in: Proceedings Fourth International GI Congress on Knowledge-based Systems, Springer, Berlin, Germany (1991) 363 - 374.

[Heinsohn et al. 92] J. Heinsohn, D. Kudenko, B. Nebel and H.-J. Profitlich, RAT -Representation of Actions in Terminological Logics, in: J. Heinsohn and B. Hollunder, eds., Proceedings DFKI Workshop on Taxonomic Reasoning, DFKI Document D-92-08, Saarbrücken, Germany (1992) 16-22.

[Herzog et al. 89] G. Herzog, C-K. Sung, E. André, W. Enkelmann, H.-H. Nagel, T. Rist, W. Wahlster and G. Zimmermann, Incremental Natural Language Description of Dynamic Imagery, in: W. Brauer and C. Freksa, eds., Proceedings Third International GI Congress on Knowledge-Based Systems, Springer, Berlin, Germany (1989) 153-162.

[Hovy & Arens 91] E.H. Hovy and Y. Arens, Automatic Generation of Formatted Text, in: Proceedings AAAI-91, Anaheim (1991).

[Kerpedjiev 92] S.M. Kerpedjiev, Automatic Generation of Multimodal Weather Reports from Datasets, in: Proceedings 3rd ACL Conference on Applied Natural Language Processing (ANLP-92), Trento, Italy (1992) 48-55.

[Kobsa et al. 86] A. Kobsa, J. Allgayer, C. Reddig, N. Reithinger, D. Schmauks, K. Harbusch and W. Wahlster, Combining Deictic Gestures and Natural Language for Reference Identification, in: Proceedings 1 lth International Conference on Computational Linguistics (COLING-86) Bonn, Germany (1986) 356-361.

[Marks & Reiter 90] J. Marks and E. Reiter, Avoiding Unwanted Conversational Implicatures in Text and Graphics, in: Proceedings AAAI-90, Boston (1990) 450-456.

[Maybury 91a] M. Maybury, editor, Intelligent Mutimedia Interfaces, Workshop Notes from AAAI-91, Anaheim (1991). [Maybury 91b] M. Maybury, Planning Multimedia Explanations Using Communicative Acts, in: Proceedings AAAI-91, Anaheim (1991) 61-66.

[McKeown & Feiner 90] K. McKeown and S. Feiner, Interactive Multimedia Explanation for Equipment Maintenance and Repair, in: DARPA Speech and Language Workshop, (1990)42-47.

[Neal & Shapiro 91] J.G. Neal and S.C. Shapiro, Intelligent Multi-Media Interface Technology, in: J.W. Sullivan and S.W. Tyler, eds., Intelligent User Interfaces, Reading: Addison-Wesley (1991) 11-43.

[Reiter et al. 92] E. Reiter, C. Mellish and J. Levine, Automatic Generation of On-Line Documentation in the IDAS Project. In: Proceedings 3rd Conference on Applied Natural Language Processing (ANLP-92), Trento, Italy (1992) 64-71.

[Rist & André 92] T. Rist and E. André, From Presentation Tasks to Pictures: Towards an Approach to Automatic Graphics Design, in: Proceedings 10th European Conference on Artificial Intelligence (ECAI-92), Vienna, Austria (1992).

[Roth et al. 91] S. Roth, J. Mattis and X. Mesnard, Graphics and Natural Language as Components of Automatic Explanation, in: J.W. Sullivan and S.W. Tyler, eds., Intelligent User Interfaces, Reading: Addison-Wesley (1991) 207-239.

[Schirra 92] J. Schirra, A Contribution to Reference Semantics of Spatial Prepositions: The Visualization Problem and its Solution in VITRA, in: C. Zelinsky-Wibbelt, editor, The Semantics of Prepositions - From Mental Processing to Natural Language Processing, Mouton de Gruyter, Berlin, Germany (1992).

[Stock 91] O. Stock, Natural Language and Exploration of an Information Space: The AlFresco Interactive System, in: Proceedings IJCAI-91, Sydney (1991).

[Sullivan & Tyler 91] J.W. Sullivan and S.W. Tyler, eds., Intelligent User Interfaces, Reading: Addison-Wesley (1991).

[Wahlster et al. 78] W. Wahlster, A. Jameson and W. Hoeppner, Glancing, Referring and Explaining in the Dialogue System HAM-RPM, American Journal of Computer Linguistics, Microfiche 77 (1978) 53-67.

[Wahlster et al. 83] W. Wahlster, H. Marburger, A. Jameson and S. Busemann, Over-answering Yes-No questions: Extended Responses in a NL Interface to a Vision System, in: Proceedings IJCAI-83, Karlsruhe, Germany (1983) 643-646.

[Wahlster 89] W. Wahlster, One Word Says More Than a Thousand Pictures. On the Automatic Verbalization of the Results of Image Sequence Analysis Systems. In: Computers and Artificial Intelligence 8 (1989) 479-492.

[Wahlster 91] W. Wahlster, User and Discourse Models for Multimodal Communication, in: J.W. Sullivan and S.W. Tyler, eds., Intelligent User Interfaces, Reading: Addison-Wesley (1991) 45-67.

[Wahlster et al. 91] W. Wahlster, E. André, W. Graf, T. Rist, Designing Illustrated Texts: How Language Production is Influenced by Graphics Generation, in: Proceedings Fifth Conference of the European Chapter of the Association for Computational Linguistics (EACL-91), Germany (1991) 8-14.

[Wahlster et al. 92a] W. Wahlster, E .André, S .Bandyopadhyay, W. Graf, T. Rist, WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation. In A. OrtonyJ . Slack and O.Stock ,eds., Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues. Springer: Heidelberg, pp. 121-144.

[Wahlster et al. 92b] W. Wahlster, E .André, W. Finkler, H.-J. Profitlich, T. Rist: Plan-based Integration of Natural Language and Graphics Generation. DFKI Report, Saarbrücken, Germany, 1992.

[Wazinski 92] P. Wazinski, Generating Spatial Descriptions for Cross-modal References, in: Proceedings 3rd Conference on Applied Natural Language Processing (ANLP-92), Trento, Italy (1992) 56-63.