Proceedings of the 1991 Winter Simulation Conference
Barry L. Nelson, W. David Kelton, Gordon M. Clark (eds.)

GENERATING CUSTOMIZED TEXT AND GRAPHICS IN THE COMET EXPLANATION TESTBED

Michael Elhadad
Steven Feiner
Kathleen McKeown
Dorée Seligmann

Department of Computer Science
Columbia University
New York, New York 10027

ABSTRACT

For many applications that must explain reasoning, processes, or plans to their users, multimedia explanations are more effective than text or pictures alone: the use of two or more modalities makes it possible to communicate the same or complementary information in different ways. While hypermedia is a currently available technology for providing such explanations, conventional implementations do not adequately address the needs of different users. In this paper, we show how COMET, a system that uses natural language and graphics generation components to produce the text and pictures of its explanations dynamically, can create a variety of different explanations to explain the same concepts, and thus better meets the needs of different users. We focus on four ways in which COMET produces different explanations for the same concepts, describing how it is influenced by the plan containing the concept, by user background knowledge, by previous dialogue, and by interactive exploration of an explanation.

1 INTRODUCTION

Many applications need to provide explanations to their users. Expert systems must be able to explain their reasoning to users, programming environments must be able to provide help about their software tools, simulation systems must be able to explain the causal processes that are modeled, and diagnosis systems must be able to instruct users how to detect and repair system or equipment failures. Particularly in cases where the underlying concepts include physical actions or processes, explanations are more effective if they integrate graphics and language: two or more modalities allow different ways of communicating the same or complementary information. Graphics allow the user to "step into" the world that is modeled to view and explore spatial representations of objects, while language can better convey the abstract relations between objects and actions.

A number of systems are beginning to use hypermedia to provide explanations (Nielsen, 1990). Although hypermedia is a currently available technology, conventional implementations have significant drawbacks. Typically, all pictures and text are authored by people in advance and stored for retrieval. Consequently, hypermedia systems that use "canned" text and graphics cannot do a good job of addressing the needs of users with different backgrounds asking questions in different situations. There are two problems here. First, because all material is authored by hand, customization for a diversity of users would result in a combinatorial explosion in the amount of material that would have to be created and stored. Therefore, when designing material, hypermedia authors typically address only coarse variations in users and situations by grouping them into a few broad equivalence classes. Second, the user's freedom to choose what to do next, upon which hypermedia is based, means that presentation continuity between pieces of text or images cannot be taken into account when they are created, because the author does not know the order in which they will be presented. Thus, material does not flow as well as that designed for a particular presentation order.

In this paper, we present an alternative approach that generates multimedia explanations interactively. We argue that a system that can dynamically determine both what information needs to be included in an explanation and how to express that information can produce a wider variety of explanations that better meet the needs of its users. In previous papers (Feiner and McKeown, 1990a, Feiner and McKeown, 1990b, Elhadad et al., 1989), we overviewed COMET (COordinated Multimedia Explanation Testbed), a system that uses natural language and graphics generation components to produce the text and pictures of its explanations dynamically. Rather than using pre-stored text and pictures, COMET produces its explanations at the time they are requested and, as a result, is able to generate a variety of different explanations to explain the same concepts. Here we focus on the different ways COMET can vary an explanation and the rationale on which such variation is based.

COMET produces explanations for equipment maintenance and repair. It diagnoses certain failures in a military communications radio and explains the complex tasks and actions that the user must perform in the course of diagnosis. In the sections that follow, we first present a brief overview of COMET's architecture and describe assumptions that we make about COMET's underlying knowledge base, highlighting the difference between entering information in the knowledge base and authoring multimedia explanations. We note that COMET's techniques are domain independent and can be applied to other knowledge bases containing information about actions, objects, and causal processes. We then describe four ways in which COMET produces different explanations for the same concepts, showing how the same action can be described in different ways depending on the overall plan it is part of, how the user's background knowledge influences explanation generation, how the dialogue of which the explanation is part has an influence, and finally, how dynamic generation provides opportunities for further interactive exploration of an explanation by its users.

2 ARCHITECTURE

COMET's architecture includes an expert-system diagnostic component, several knowledge sources, a single content planner, a media coordinator, media generators for text and graphics, a media layout manager, and interactive typesetting and rendering components (Feiner and McKeown, 1990b). Each of COMET's major components runs in parallel in its own process on up to five networked workstations. A simple menu interface allows users to enter requests for explanations. On receiving a request, the *content planner* uses text plans, or *schemas* (McKeown, 1985), to determine which information from a set of underlying *knowledge sources* should be included in the explanation. The content planner produces the full content for the explanation, which is represented as a hierarchy of *logical forms* (Allen, 1987). The logical forms specify the communicative goals the explanation must fulfill and content needed to achieve those goals, but not its form. They are passed to the *media coordinator*, which determines which pieces of information are to be communicated in text, and which in graphics. It annotates the logical forms with directives that specify these assignments.

Each annotated logical form is passed to a set of *media generators* that create material that expresses the information the media coordinator has assigned to them. COMET currently has a *text generator* and a *graphics generator*. The text generator (McKeown et al., 1990) includes both a *lexical chooser* that selects vocabulary and a *sentence generator* that forms the sentences. The graphics generator, IBIS (Intent-Based Illustration System) (Seligmann and Feiner, 1991), designs 3D color illustrations that fulfill a set of communicative goals derived from the logical forms. The resulting text and graphics are passed to the *media layout* component, which formats the final presentation, determining where on the display information is to be presented. Real-time 3D rendering and typesetting components then produce the finished display.

3 KNOWLEDGE BASE REQUIREMENTS

COMET's primary knowledge base, represented using LOOM (MacGregor & Bates, 1987), consists of approximately 700 concepts representing both objects and processes. Processes include plans and simple actions. Plans can contain substeps (which are, in turn, plans or actions), preconditions, and effects. For example, the plan for loading the radio's transmission frequency contains four substeps, each of which is itself a plan and is represented as a separate concept in the knowledge base. The plan's parameters are objects, which are also represented as separate concepts containing information about various attributes of the object. In addition, COMET contains information needed to generate graphics depicting each object and action. While abstract graphics, such as circuit schematics or box and arrow diagrams, are well-suited for many applications, illustrations for use in maintenance and repair are often intended to show the user the location or appearance of objects, or to instruct the user how to perform complex physical tasks. Generating these illustrations (and related text) requires that our knowledge base include information about the actual geometric and physical properties of objects: their size, shape, position, material, color, and surface treatment. With the possible exception of information about visual properties, however, this information is a subset of that needed for physical simulation. This is also the kind of information that can be obtained from a CAD/CAM database produced in the course of designing the objects being depicted.

To increase the system's coverage, new plans and actions can be added to the knowledge base. For our domain, we did this based on the diagnostic flowcharts in an existing repair manual (DOA, 1986). Each box in a flowchart translated directly into a step of a plan. Plans describing each substep were also provided. Since the same simple actions and objects are often used in more than one plan, adding new plans often made use of existing knowledge. After developing the system using one flowchart, we recorded the procedure for adding the rest of the manual's flowcharts to the knowledge base. The main difficulties that we encountered involved finding out how substeps were performed. This required locating the pages in the manual that define how the step in a flowchart box is accomplished. (Since there are no explicit page references in the manual and it includes several volumes, this was sometimes a time-consuming procedure.) Other than some debugging problems with our prototype version of LOOM, we found that a new flowchart and its substeps could be entered in less than a day's time.

Once the knowledge is added, COMET can generate different explanations from these same plans, actions, and objects. As explained in the following sections, knowledge about the user and the discourse situation in which an explanation is requested can determine both what information about the plan and its substeps is included as well as how it is described. In addition, the description of a simple action is influenced by the plan of which it is part. Thus, from the same steps, which correspond in a very straightforward way to flowchart boxes, COMET can generate a variety of different descriptions. The steps only need to be entered once. In contrast, in a hypermedia approach that relied on "canned" material, although the user could pick and choose steps, the system would not be able to knit them together smoothly.

While the techniques that we use for determining explanation content are specific to a task-based domain, they are not specific to equipment maintenance and repair. Thus, they can be used to provide descriptions of physical objects, of actions that operate upon these objects, and of complex plans. Our content planner was adapted from a system that generated descriptions of processes (Paris, 1987). Therefore, we expect it could be applied to any domain including physical objects, actions, plans, processes, and relations (e.g., causal, enablement, and effect) between these concepts. For example, we expect that the same architecture and techniques could be used to provide explanations for simulation systems. Such explanatory facilities would be an important complement to the use of natural language to design models for and to control simulation (Beck and Fishwick, 1989, Badler et al., 1991).

In the following sections, we demonstrate how COMET can adapt the explanations it generates to the user and situation.

4 INFLUENCE OF PLAN CONTEXT

COMET produces explanations for complex procedures by dynamically combining primitive actions. The content planner determines which substeps in its hierarchically structured plans should be decomposed and included in the explanation, in the process deciding which primitive actions are described. There are two main advantages in using such a compositional approach to represent actions:

- It concisely encodes knowledge that can be used to represent a large number of distinct plans by allowing the same primitive action to be reused in several different plans.
- The context of a primitive action can be used to produce more relevant explanations. The explanations for the same action may be different when it appears as a substep of different plans. Furthermore, the explanation does not need to consider all the other plans in which a single action could appear. Thus, the user is never presented with the choice of which alternative steps to follow.

For example, consider the difference between the explanation generated by COMET for loading the frequency into the radio's memory and the explanation available in the printed technical maintenance manual. Another operation, clearing the frequency, shares the first five substeps with the frequency-loading operation. The printed manual exploits the similarity between these two distinct operations and merges the instructions for performing both into a single description. Consequently, the manual is forced to use a conditional construct to describe the steps that differ and their corresponding (and different) consequences:

> If you are clearing the frequency, enter 00000, and push the Sto ENT button. The display will show FILL3 and blink. If you are loading a new frequency, enter the new frequency and push the Sto ENT button. The display will show the new frequency and blink. (DOA, 1986)

Such a strategy is misleading and can result in ambiguous explanations. In contrast, COMET can generate a complete description for each plan, customized to the specific needs of a user. COMET's explanation for the last three steps of loading the frequency is shown in Figure 1. Dynamically combining primitive actions also enhances the cost effectiveness of interactive explanation generation as compared to stored explanations, as only a relatively small number of simple primitive actions need to be encoded in the knowledge base.

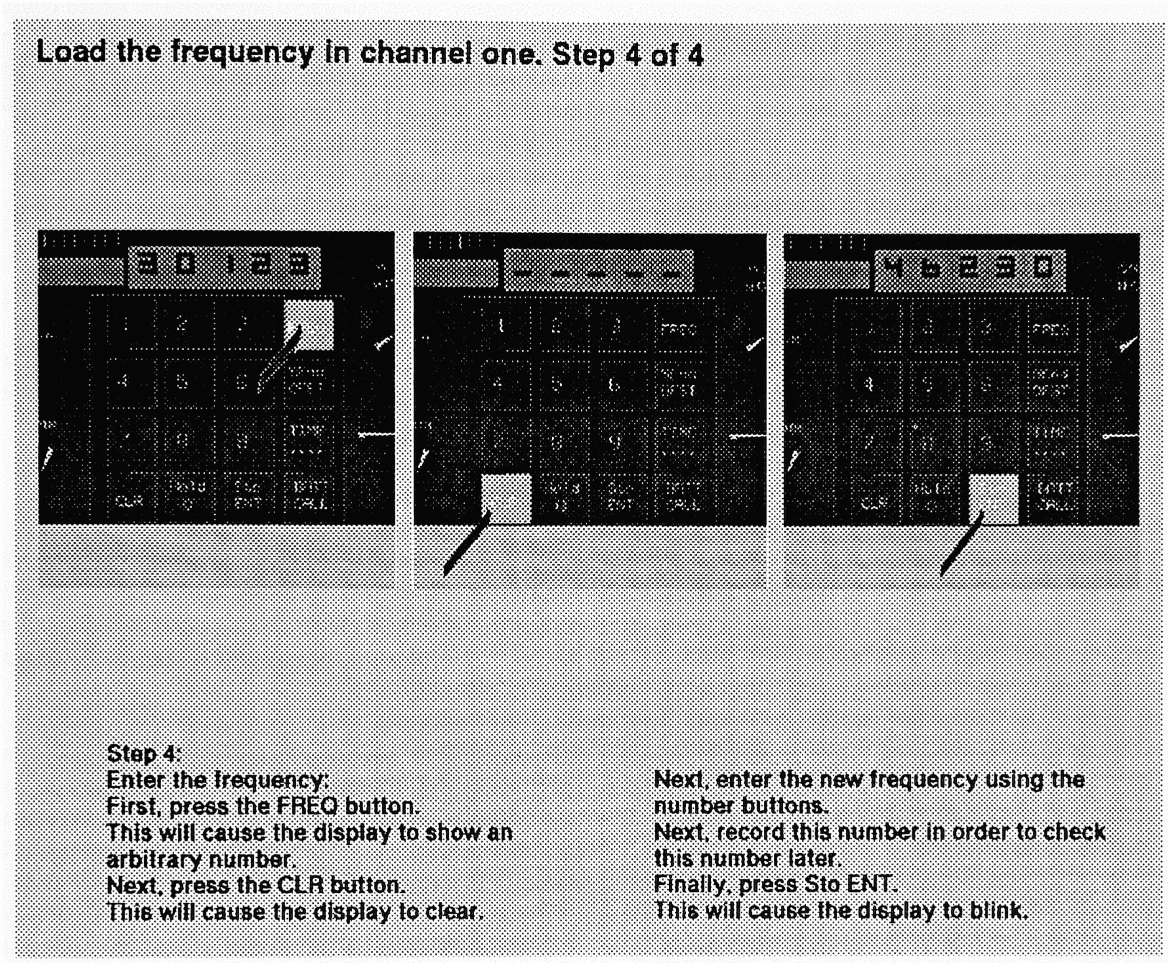

Figure 1: Explanation generated by COMET for loading the transmission frequency.

5 TAILORING EXPLANATIONS TO DIFFERENT USERS

While a printed manual is restricted to present the exact same information to all users, COMET has the potential to adapt its explanations to different users, based on their backgrounds or goals. Given a model of the current user, COMET can generate appropriate explanations.

One characteristic that we consider is the vocabulary with which the user is familiar. For example, the word "arbitrary" in Figure 1 is replaced by the word "some" when the user model indicates that the user is not familiar with the word "arbitrary".

Similarly, different illustrative styles are appropriate to certain audiences. For example, technical drawings use a rigid vocabulary of line weights and symbols (Giesecke, Mitchell, and Spencer, 1936) that is not necessarily meaningful to the layperson. The IBIS graphics generator uses a rule base of different design and style devices; these could be categorized for different types of users. Illustrations are designed using a generate-and-test approach. Stylistic and design devices are ordered so that those included in the user's vocabulary of visual cues are tried first. An arrow is used to show the concept *press* in Figure 1. An alternative design option is to use a sequence of two illustrations, the first showing a button in its unpressed state, the second showing it in its pressed state. By tailoring the textual and graphical vocabulary used in explanations to the level of the user, COMET is able to generate more appropriate, understandable, and concise explanations.

6 PAST DISCOURSE

Just as it is important to tailor a presentation for a particular user, it is also important that an explanation be sensitive to the discourse context in which it appears. An explanation that ignores what has previously been presented can be overly long, inconsistent, and misleading. To take past discourse into account, design decisions must be made dynamically since the user is not restricted to an a priori planned presentation, but can interactively change the order and level of detail with which information is presented. As an explanation is generated, each generator records not only *what* has already been presented, but also *how* it has been presented. The generators then use different strategies to ensure that the description of a step is responsive to the influence of the past discourse.

For the text generator, one of the effects of considering the previous discourse is to determine certain word choices. For example, one field maintenance procedure involves installing a new holding battery (the battery that maintains the radio's memory) when the radio experiences a memory failure. To test the new holding battery, the primary battery is removed for a brief period of time, during which the holding battery takes over. COMET generates the following text for this procedure:

> Install the new holding battery. Load the frequency in channel one. Set the FCTN knob to SQ OFF. Remove the primary battery. This will cause the radio to rely on the new holding battery for power. Wait 30 seconds. Reinstall the primary battery.

Consider the choice of the word "reinstall" in the last sentence. The knowledge base concept corresponding to this primitive action is the same one (*install*) realized as "install" in the paragraph's first sentence. The difference is that the battery has been removed in a preceding step. We store in the knowledge base the relationship between inverse actions. The lexical chooser can therefore check if the *install* concept in the last step is the inverse of one of the actions described previously. Note that using the word "install" in the last instruction could trigger the false inference that a new battery should be used. This is one of several techniques that are applied by the natural language generator to produce a coherent text over the course of the interaction and to avoid creating false inferences.

The IBIS graphics generator maintains the consistency of visual cues in several ways. For example, the arrows appearing in Figure 1 are generated to indicate concepts such as focus or direction. IBIS has available several different styles for such objects, which determine what icon to use, its orientation, and its color. These stylistic decisions are made according to the constraints associated with the individual illustration within which the object appears. For example, the color of an arrow is chosen to contrast with the background objects. In the context of an extended session, including several illustrations, such decisions are further restricted in order to maintain consistency. Therefore, if in a first illustration, the arrows were assigned a red color to indicate pressing a button, then all subsequent arrows indicating a pressing action will be red.

IBIS can also relax constraints based on the information previously displayed. For example, it realizes the concept *location* through one of several different design strategies. These include showing the entire context object (such as the object's parent in the object hierarchy) or showing parts of the context that include objects that could serve as landmarks (Feiner, 1985, Seligmann and Feiner, 1989, Seligmann and Feiner, 1991). When making design decisions, IBIS can also relax one of the input goals if it has already been satisfied in one of the previous illustrations. For example, if an object's location has been communicated in a previous illustration, IBIS is allowed to relax the constraint of including this information in subsequent illustrations. Thus, by taking past discourse into account, COMET is able to produce a more cohesive explanation, avoiding redundancies and ambiguities.

7 EXPLORING AN EXPLANATION

No matter how good the presentation, the user may still want to know additional information, perhaps in direct response to the presentation itself. Therefore, we have provided several ways in which COMET can change a presentation dynamically according to the user's focus of interest.

COMET organizes the explanation in a hierarchical manner, corresponding to the hierarchy of actions produced by the planner. At each step, the user can decide whether to ask for more detail about an action or to proceed to the next step.

COMET also gives the user fine-grain control over the illustrations by allowing user-controlled navigation in the 3D illustrated world. Although each illustration is first displayed using a viewing specification chosen by the system, the user may wish to see it from another viewpoint, for example to see objects that are hidden or to examine visible objects from a different perspective. In traditional 3D user-controlled navigation, the user changes the values that define the view and the scene changes only in the viewing transformation applied. In contrast, IBIS binds an illustration to the communicative goals with which it was specified and treats these as constraints (Seligmann and Feiner, 1991, Feiner and Seligmann, 1991). When the user changes the values that define the view, IBIS adjusts the illustration so that these goals remain satisfied. For example, if an illustration has been designed to show the holding battery, the illustration's communicative goals will in turn generate a required visibility constraint for the holding battery. As the user moves around in the illustration, however, other objects may potentially occlude the holding battery. IBIS detects when objects will occlude objects that must be visible and dynamically selects different rendering styles to depict these would-be obscuring objects.

Figure 2 was generated by IBIS to satisfy a visibility constraint for the holding battery. IBIS's rule base allows

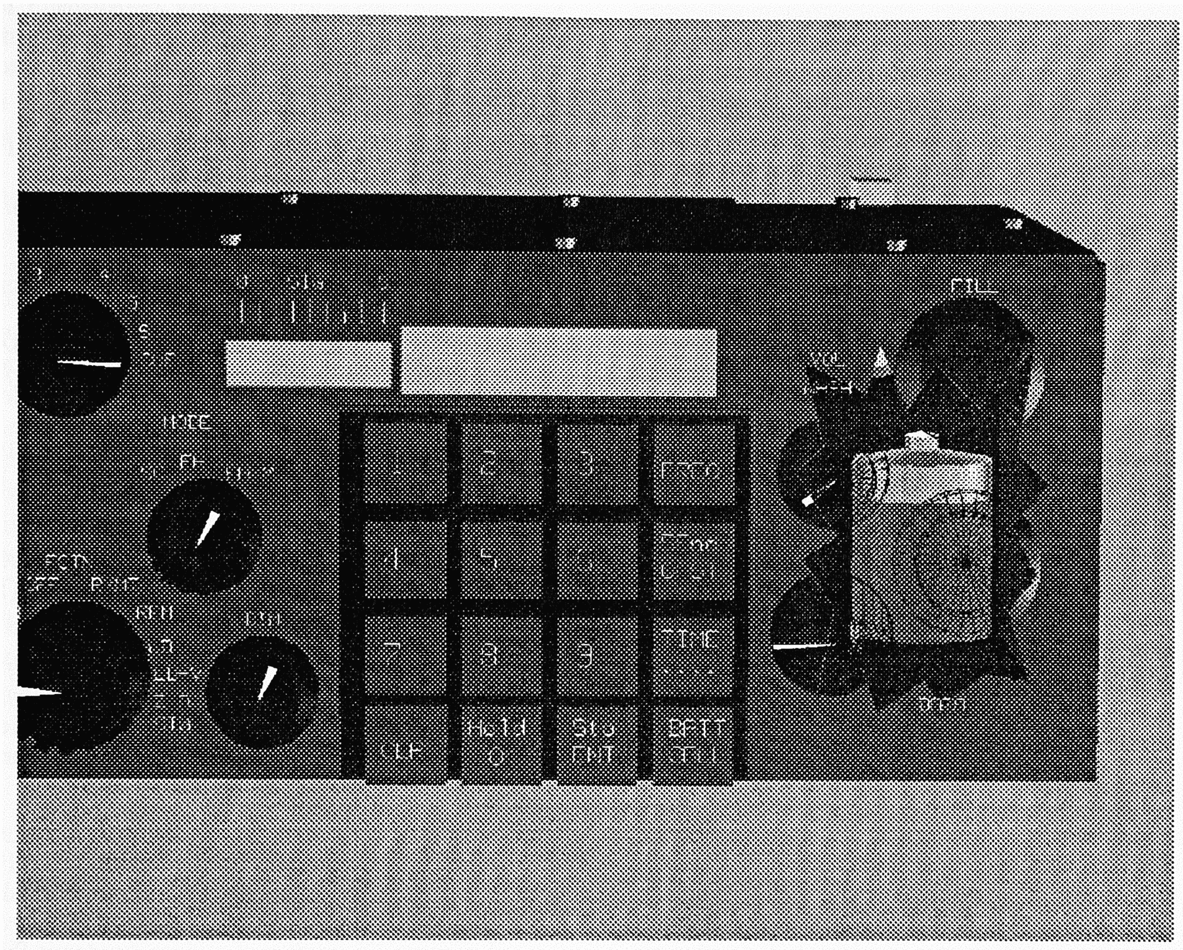

Figure 2: Cutaway illustration designed by IBIS to show the holding battery.

it to use a "cutaway view" in which the battery is seen through a cutout in the radio's faceplate, with any potentially occluding objects drawn in a wireframe style where they would otherwise obscure the cutout. Sometimes, no illustration rendered from a single viewpoint can satisfy all the input communicative goals. In these cases, IBIS tries to design a composite illustration that consists of a set of simpler illustrations that are adjacent, overlapping, or inset inside others.

8 CONCLUSIONS AND FUTURE WORK

We have described our work on COMET, a multimedia system that interactively generates written text and user-controlled dynamic 3D graphics to explain how to perform physical tasks. Our emphasis has been on the many ways in which COMET can vary the explanations that it generates: representing and explaining complex plans in terms of a composition of primitive actions; tailoring explanations to individual users by selecting appropriate lexical and graphical vocabulary; relying on information from past discourse to constrain how information is presented to make the explanation more consistent, concise, and cohesive; and making possible user navigation in the 3D illustrations.

Much of the future work on COMET is directed toward generating presentations that include complex temporal information (Feiner, Litman, McKeown, and Passonneau, 1991). While COMET currently represents and explains only sequential relationships among actions, many of the actions that take place in maintenance and repair and other domains overlap in time and have constraints on when they begin and end relative to other actions. Not only do we need to represent these relationships, but we also need to modify our generation facilities to communicate them. Therefore, COMET's current set of media generators will be augmented with additional generators for animated graphics and speech, media that are well-suited for expressing temporal relationships. Currently, all the information in one of COMET's displays is presented simultaneously to the user, with direct user manipulation of the viewing specification provided once an illustration has been generated. In contrast, speech and animation present information at specific points in time, which must be taken into account in presentation planning.

ACKNOWLEDGMENTS

This work is supported in part by the Defense Advanced Research Projects Agency under Contract N00039-84-C-0165, the Hewlett-Packard Company under its AI University Grants Program, the National Science Foundation under Grant IRT-84-51438, and the Office of Naval Research under Contract N00014-89-J-1782. COMET's development is an ongoing group effort that has also benefited from the contributions of Michelle Baker, Cliff Beshers, Andrea Danyluk, David Fox, Yumiko Fukumoto, Laura Gabbe, Jong Lim, Christine Lombardi, Jacques Robin, Frank Smadja, Michael Tanenblatt, and Tony Weida.

REFERENCES

Allen, J. (1987). *Natural Language Understanding.* Menlo Park, CA: Benjamin Cummings Publishing Company, Inc.

Badler, N., Webber, B., Kalita, J., and Esakov, J. (1991). Animation from Instructions. In Badler, N., Barsky, B., and Zeltzer, D. (Eds.), *Making Them Move: Mechanics, Control, and Animation of Articulated Figures.* San Mateo, CA: Morgan Kaufmann.

Beck, H. and Fishwick, P. (March 1989). Incorporating Natural Language Descriptions into Modeling and Simulation. *Simulation*, *52*(3), 102–109.

Department of the Army. (June, 1986). *TM 11-5820-890-20-1 Technical Manual: Unit Maintenance for Radio Sets AN/PRC-119, ...* Headquarters, Department of the Army.

Elhadad, M., Seligmann, D., Feiner, S., and McKeown, K. (August 22 1989). A Common Intention Description Language for Interactive Multi-media Systems. *A New Generation of Intelligent Interfaces: Proc. IJCAI-89 Workshop on Intelligent Interfaces.* Detroit, MI.

Feiner, S. (November 1985). APEX: An Experiment in the Automated Creation of Pictorial Explanations. *IEEE Computer Graphics and Applications*, *5*(11), 29–38.

Feiner, S. and McKeown, K. (March 5–9 1990). Generating Coordinated Multimedia Explanations. *Proc. CAIA90 (6th IEEE Conf. on Artificial Intelligence Applications).* Santa Barbara, CA.

Feiner, S. and McKeown, K. (July 29–August 3 1990). Coordinating text and graphics in explanation generation. *Proc. Eighth National Conference on Artificial Intelligence.* Boston, MA.

Feiner, S. and Seligmann, D. (June 26–28 1991). Dynamic 3D Illustrations with Visibility Constraints. *Proc. Computer Graphics International '91.* Cambridge, MA.

Feiner, S., Litman, D., McKeown, K., and Passonneau, R. (July 15 1991). Towards Coordinated Temporal Multimedia Presentations. *Proc. AAAI-91 Workshop on Intelligent Multimedia Interfaces.* Anaheim, CA.

Giesecke, F.E., Mitchell, A., and Spencer, H.C. (1936). *Technical Drawing.* New York, NY: The Macmillan Company.

MacGregor, R. and Bates, R. (May 1987). *The LOOM Knowledge Representation Language* (Tech. Rep.)). Marina del Rey, CA: USC ISI.

McKeown, K.R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Studies in Natural Language Processing. Cambridge, England: Cambridge University Press.

McKeown, K.R., Elhadad, M., Fukumoto, Y., Lim, J., Lombardi, C., Robin, J., and Smadja, F. (1990). Language Generation in COMET. In Mellish, C., Dale, R., Zock, M. (Eds.), *Current Research in Language Generation.* London: Academic Press.

Nielsen, J. (1990). *Hypertext & Hypermedia.* Boston, MA: Academic Press.

Paris, C.L. (1987). *The Use of Explicit User models in Text Generation: Tailoring to a User's Level of Expertise.* Doctoral dissertation, Columbia University.

Seligmann, D.D., and Feiner, S. (November 13–15 1989). Specifying Composite Illustrations with Communicative Goals. *Proc. UIST89 (ACM SIGGRAPH Symposium on User Interface Software and Technology).* Williamsburg, VA.

Seligmann, D. and Feiner, S. (July 1991). Automated Generation of Intent-Based 3D Illustrations. *Computer Graphics (Proc. SIGGRAPH 91)*, *25*(3), 123–132.

AUTHOR BIOGRAPHIES

MICHAEL ELHADAD is a Ph.D. student in Computer Science at Columbia University. His research interests include natural language processing, language generation, unification grammars, and models of argumentation. Elhadad has developed FUF, an efficient functional unification system for sentence generation that has been used in a variety of applications at Columbia. His thesis research involves using FUF for the task of word choice for language generation.

STEVEN K. FEINER is an Associate Professor of Computer Science at Columbia University. His research interests include image synthesis, applications of artificial intelligence to computer graphics, user interfaces, animation, hypermedia, virtual worlds, and visualization. Feiner received a Ph.D. in Computer Science from Brown University in 1987. He is on the editorial boards of *Electronic Publishing* and *ACM Transactions on Information Systems*, and is a coauthor of *Computer Graphics: Principles and Practice, 2nd Edition* (Addison-Wesley, 1990). In 1991 he received an Office of Naval Research Young Investigator Award.

KATHLEEN R. MCKEOWN is an Associate Professor of Computer Science at Columbia University. Her research interests include natural language generation, user modeling, expert system explanation, and natural language interfaces. McKeown received a Ph.D. in Computer Science from the University of Pennsylvania in 1982. She was elected President of the *Association of Computational Linguistics* for a term beginning in January 1992 and currently serves as its Vice President. She has served on the Executive Council of the *Association for Artificial Intelligence* and was co-program chair of their annual conference in 1991. In 1985 she received a National Science Foundation Presidential Young Investigator Award.

DOREE SELIGMANN is a Ph.D. student in Computer Science at Columbia University. Her research interests involve computer graphics illustration, visual languages, modeling artistic techniques, user interfaces, and multimedia communication. Seligmann's thesis work concentrates on issues of visual communication, and on the development of the IBIS intent-based illustration system. At AT&T Bell Laboratories, Holmdel, she is part of the Rapport project, a multimedia conferencing system.