

Task Communication Through Natural Language and Graphics*

Norman Badler and Bonnie Webber

With increases in the complexity of information that must be communicated either by or to computers comes a corresponding need to find ways to communicate that information simply and effectively. It makes little sense to force the burden of communication on a *single* medium, restricted to just one of spoken or written text, gestures, diagrams, or graphical animation, when in many situations information is only communicated effectively through *combinations* of media. For example,

Directions

In response to requests for directions, respondents often choose to provide both a sketch map (for visual indications of relative distance, spatial relationships, etc.) as well as verbal guidance as to landmarks to attend to, obstacles to watch out for, opportunities to take, etc.

Instructions

Instructors training a subject in a new task often choose to present the task in at least two ways: they demonstrate what motions the trainee is supposed to carry out, using direct training, film or graphic media, and they convey what intentional actions those motions are meant to represent, through natural-language text or speech.

Situation Assessment

Graphic media (diagrams and animation) can provide a way of visualizing significant patterns in situations (cf. the current interest in *Scientific Visualization*), while natural-language text (either spoken or written) can provide needed information on what the patterns may mean, why they may have developed, or what may be done to deal with them. (For example, it is well-known that narration-less scientific visualizations can be pretty, but nevertheless unilluminat-

ing to anyone but experts. Natural-language narration is necessary to convey the meaning and significance of such visualizations.)

Task Design

As in situation assessment, graphic media can enable a task designer to see how agents of different size, agility and strength would perform the same task in range of different workplaces, while natural-language text can explain the reasons for such behavior or for an agent's inability to perform the task as specified.

In each of these cases, there is more than one sort of information to convey—for example, spatial location, motion dynamics, temporal relations, intentions, causal relations, conditional actions, disjunctive alternatives, etc. Theoretically, one *can* make one medium responsible for conveying all relevant information, establishing new conventions for communicating particular types of information where they don't already exist (cf. setting up temporary conventions for maps and diagrams via accompanying legends). But this can overload a recipient's cognitive faculties, making it even more difficult for him or her to integrate all the information. Using multiple media, one can take advantage of both the individual strengths and efficiencies of each medium and the fact that several can be employed in parallel.

Interest has been growing in both the natural language and the computer graphics communities in getting computers to produce and even understand such multi-media communication. To tap this interest and initiate sharing of both problems and results, a small workshop entitled Task Communication through NL and Graphics was held at the University of Pennsylvania 30-31 May 1990, sponsored by the Army Research Organization and attended by researchers from Harvard, MIT, DEC Cambridge Research Labs, Columbia,

Systems Exploration Inc. (SEI), University of Chicago and University of Pennsylvania.

The workshop lasted one and a half days with the first day taken up with presentations by the various groups and the following half-day with discussion of some key issues and possible future actions (one of which was the decision to produce and circulate this summary).

The presentations began with Barbara Grosz and Joe Marks of Harvard University talking about their work [3, 4, 5] on using text and diagrams in the context of multi-agent planning. This work investigates how two agents (here, a human network manager and a computerized support system) cooperate to assess a situation (here, the load on a computer network) and plan their best course of action. The goal of the work is to enable the system to describe a network load situation in natural language, with the key features made visible through network diagrams. The manager and system should then be able to negotiate possible courses of action through natural language, with the system illustrating consequences of those actions through diagrams as well as words. Grosz described their framework and approach to planning that enables such negotiation, while Marks described his system for automatically creating appropriate and graphically well-formed network diagrams from symbolic representations.

After this, Steve Feiner and Kathy McKeown of Columbia University described their work [2] on automatically generating task instructions which coordinate 3D color graphics with natural-language captions. The particular task they demonstrated was that of changing the battery in a particular Army radio. The instructions were generated from a common representation, with a "media coordinator" deciding which information would be communicated through graphics alone, which through natural-language alone, and which through both.'

Following lunch, David Zeltzer of the MIT Media Lab described his work [6] on developing a taxonomy of tools for use in defining and interacting with objects and agents (here, broadly construed as independently acting objects) in virtual environments (here, graphical simulations). In this taxonomy, tools are classified along two

*Requests for further information can be made to badler@central.cs.upenn.edu or bonnie@central.cs.upenn.edu

dimensions: (1) explicitness of behavioral control—at one end, completely guided, at the other end, specified algorithmically—and (2) level of control—from machine-level specifications to task-level specifications (possible action abstractions and intentions within the virtual world). Zeltzer's *pièce de résistance* was an animation (in progress) entitled "Grinning Evil Death", which illustrated to what use some of these tools could be put.

The fourth talk of the day was given by Norm Badler and Bonnie Webber of the University of Pennsylvania. They described their work [1] with Mark Steedman on *Animation from Instructions*, which has as its goal the automatic production of narrated animated simulations of agents carrying out tasks specified through natural-language instructions. The system is being built primarily bottom-up, and Badler showed a videotape of its current abilities to vary agents' visible behavior in carrying out a fixed (simple) task, depending on their ascribed size, strength and speed. Webber then discussed features of natural-language instructions, in order to make clear what distinguishes them from other discourse types and what is involved in using them to drive animation. (Eventually, the representation derived from such instructions will also contribute to the animation's narration.)

Three short talks were then given by Phil Agre (University of Chicago), Candy Sidner (DEC Cambridge Research Labs), and Medhat Korna (SEI). Agre talked about "The Body in Representational Practice", describing two realms—deixis and gesture—in which talking about activity and carrying out activity partially merge. He related this to the more general problem of shared representations for communication. Sidner presented data in support of the idea that tasks are more often *negotiated* rather than simply *described*, and discussed the consequences of this view for task communication, independent of what media are used. Korna ended the day by illustrating real instructions that F-16 aircraft maintenance crews are confronted with, and how problems with such instructions could be avoided through the use of anthropomorphically valid animations in the course of task design.

Discussion the next day focussed primarily on two topics—*representation* and *communicative choice*.

Representation.

Participants were generally in favor of the "shared representation" approach to multi-media communication taken by Feiner & McKeown in generating illustrated instructions and by Marks in generating network diagrams with accompanying legends. In this approach, the computer is assumed to have a single representation of the complex information it is to convey (as communicating agent) or to acquire (as recipient). Discussion centered around several issues:

Acquisition

If the computer is not the original source of the complex information to be conveyed (e.g., the results of an event simulation that it has run), how is it to acquire the desired common representation? Badler and Webber advocated using incrementally richer subsets of natural-language to drive independently acquired and specified motor skills (basic actions like sitting, lifting, pushing, etc.). The question was raised whether our current difficulty in directing an agent in realistic behavior through a combination of programming and manual control meant that one should (or could) use natural-language to either augment or modify the description.

A related issue was whether information obtained through one type of media could be understood meaningfully by another—for example, whether changes made graphically to Marks' network diagrams could automatically be made available for natural-language expression. (Currently his system does not have this ability.)

Media Dependence

Can we design media-independent shared representations of situations or tasks that can be used appropriately by whatever resources were currently available for their communication? For example, suppose that a third medium (say, animation) were made available to a system already using natural-language and diagrams. Similarly, with respect to a given communicative medium, can a representation be designed to take advantage of improvements in its communicative powers (e.g., if a natural-language generator was made more sensitive to the discourse context or a display facility was enabled to use multiple screens rather than a single one)?

Expressive Power

Can today's representational formalisms actually provide the expressive power we need? Such expressivity is needed both for describing situations and actions and for specifying the criteria on which the choice of communicative medium and communicative features provided by that medium depends. A complaint was raised that mainstream AI currently seems more interested in worst case complexity analyses than in expressivity. Participants hoped that AI researchers could be convinced that not all significant aspects of action could be trivially embedded in logical formalisms designed for state-space representations.

Communicative Choice

Participants seemed in agreement that the *type* of information to be communicated should be a major factor in deciding what medium to use in communicating it (e.g., Feiner and McKeown currently assign the communication of location and physical attributes to graphics alone, and Badler and Webber plan to use text to communicate why particular actions may have failed). Open for discussion were other issues though, including:

Conventions

There appear to be conventions that apply to the presentation of information. Although these conventions may vary from field to field, within a field they reliably engender a predictable interpretation. For example, in graphic presentations, that which is focussed or in the center of the presentation is taken as most significant. Sometimes these conventions may involve more than one medium, as in the use of both language and gesture in deictic communication, e.g., "Put that there". Participants agreed that whatever systems we build for communicating through multiple media, they will have to respect the conventions standardly used in a field, lest recipients be misled, e.g., placing significance on the wrong things. That means we will have to identify and encode such conventions as presentational constraints.

Target Audience

Projects varied as to the intended target audience of their multi-media

communication. Both Feiner & McKeown and Grosz & Marks take people as the target audience of their multimedia presentations; Zeltzer, animating synthetic agents, takes a computer system as the target audience of his action directives, while Badler and Webber take the computer as target audience for natural-language instructions, and people as target audience of the narrated animated simulations that their system is meant to produce. Clearly, there are both correspondences and differences in communicating complex information to computers and to people, and participants were interested in what could be learned by considering the differences as well as the similarities.

Related to this is another issue that was discussed—that of communicating with different *human* audiences. It is clear from work on natural-language generation and on user modeling that one needs to shape the content, the presentation, the level of detail, etc. of a text to its intended audience, depending on (at least) their previous knowledge, their goals, their communicative preferences, and perhaps even their cultural background. It was suggested that in multi-media communication, we may also have to vary our use of communicative media and the features they make available, depending on the target audience.

Discourse

We know that for effective natural-language discourse, one's communicative choices should reflect what should be taken as known or as salient from the previous discourse. It was noted that the same appears to apply to graphic discourse. For example, decisions about how to present one situation as a network diagram should take account of previous diagrams used in the same discourse: that is, the diagrams should be such that true parallels are brought out and false parallels avoided. Discourse using multiple media will have to be even more sensitive to what discourse participants know or take as salient, given what has been conveyed by each medium. Workshop participants found this a fascinating topic, about which it appeared worth educating each other, as well as doing joint research.

Conclusion

Despite initial trepidation, by the end of the morning, participants seemed to agree that getting together had been a worthwhile experience and worth doing again. The authors are currently preparing for another workshop in 1991.

References

- [1] Norman Badler, Bonnie Webber, Jeff Esakov and Jugal Kalita. Animation from Instructions. In N. Badler, B. Barsky and D. Zeltzer (eds.), *Making Them Move: Mechanics, Control and Animation of Articulated Figures*. Los Altos CA: Morgan-Kaufmann Publishers, 1990. (Also appears as Technical Report CIS-90-17, Dept. of Computer and Information Science, Univ. of Pennsylvania, Philadelphia PA, 1990.)
- [2] Steven Feiner and Kathleen McKeown. Coordinating Text and Graphics in Explanation Generation. Proc. AAAI-90, Boston MA, July 1990. Menlo Park CA: AAAI Press.
- [3] Karen Lochbaum, Barbara Grosz and Candace Sidner. Models of Plans to Support Communication: An Initial Report. Proc. AAAI-90, Boston MA, July 1990. Menlo Park CA: AAAI Press.
- [4] Joseph Marks. A Syntax and Semantics for Network Diagrams. In Proc. 1990 IEEE Workshop on Visual Languages, Skokie IL, Oct. 1990.
- [5] Joseph Marks and Ehud Reiter. Avoiding Unwanted Conversational Implications in Text and Graphics. Proc. AAAI-90, Boston MA, July 1990. Menlo Park CA: AAAI Press.
- [6] David Zeltzer. Task Level Graphical Simulation: Abstraction, Representation and Control. In N. Badler, B. Barsky and D. Zeltzer (eds.), *Making Them Move: Mechanics, Control and Animation of Articulated Figures*. Los Altos CA: Morgan-Kaufmann Publishers, 1990.

About the Authors

Dr. Norman I. Badler is the Cecilia Fidler Moore Professor and Chair of Computer and Information Science at the University of Pennsylvania and has been on that faculty since 1974. He also directs the Computer Graphics Research Facility with two full time staff members and about 40 students. Badler received his Ph.D. in Computer Science in 1975 from the University of Toronto.

Bonnie Lynn Webber is an associate professor at the University of Pennsylvania. She has conducted extensive research in Natural Language Processing and has co-edited several books on the field, including *Readings in Natural Language Processing* (Morgan Kaufmann, 1986).