Some Layout Issues for Multimedia Argument Generation

Nancy Green
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
nancy.green@cs.cmu.edu

Abstract

We are interested in the problem of text layout as it relates to our work on the intelligent generation of integrated text and information graphics (e.g., tables, bar charts, plots, maps) in the AutoBrief system (Kerpedjiev *et al.* 1997; Green, Carenini, & Moore 1998; Green *et al.* 1998). Intelligent layout generation is related to the problems of media allocation and cross-media cue generation. We propose an approach to these problems that makes use of argument structure to define domain-independent policies for a multimedia generation system.

Introduction

We are interested in the problem of text layout as it relates to our work on the intelligent generation of integrated text and information graphics (e.g., tables, bar charts, plots, maps) in the AutoBrief system (Kerpedjiev *et al.* 1997; Green, Carenini, & Moore 1998; Green *et al.* 1998). We distinguish lower level problems of layout (for example, choice of margin width, type font, and other physical attributes of text) from higher level problems of layout such as

- choice of format, e.g., the decision to express some information in the form of a text outline versus a text paragraph versus a graphic),
- planning physical organization such as hypertext structure, and
- ordering of presentation elements, including sentence order and the relative position of figures and text.

The choice of format overlaps with the problem of *media allocation*, the problem of deciding which parts of a presentation to express in which media, i.e. in the case of AutoBrief, text in paragraph format and/or information graphics. (Information graphics are used to visualize abstract rather than physical data (Card, Mackinlay, & Shneiderman 1999).) Furthermore, the problem of where to place a figure containing an information graphic with respect to text overlaps with the problem of generating *cross-media cues*, textual cues (e.g., *as shown in Figure 1*) to help the user recognize the role of a figure in an argument and locate it in the presentation. We have found extensive use of such cues in a corpus of human-authored print documents that express arguments in coordinated text and information graphics. In this paper, we discuss work in progress on the use of argument structure in media allocation and in generating cross-media cues.

Figure 1 gives an example, from the domain of transportation scheduling, of a multimedia presentation that might be created by AutoBrief in response to a user's request for a recommendation on how to improve a schedule. The presentation begins with a description of the problem: the first sentence summarizes the problem and Figure A provides more detail. Media allocation is responsible for having chosen text and graphics to express these two parts of the presentation, respectively. The cross-media cue in the first sentence *(Figure A)* is intended to make the relationship of the claim expressed in text to the graphic explicit. Next, the presentation provides a description of a hypothesized contributing factor to the problem: *This may be due in part to insufficient aircraft for bulk cargo.* The next sentence, *Most of the late cargo is bulk (Figure B)*, provides a generalization supporting this hypothesis and a cross-media cue to Figure B. Figure B shows the data upon which the generalization was based. Again, media allocation is responsible for the decision to express the support as a graphic, and the cue is intended to make the relationship of text to graphic explicit. Lastly, the presentation concludes with a recommendation.

Note that although each figure was placed immediately following the text which the figure was intended to elaborate upon or support, our approach is to generate explicit cross-media cues. Without these linking cues, the user would have to recognize the connection between a section of text and a graphic on his own. Also, it is not always feasible to provide a graphic next to related text, e.g., when a graphic plays a role in more than one place in a presentation.

A total of 125 tons of cargo will be late (Figure A).

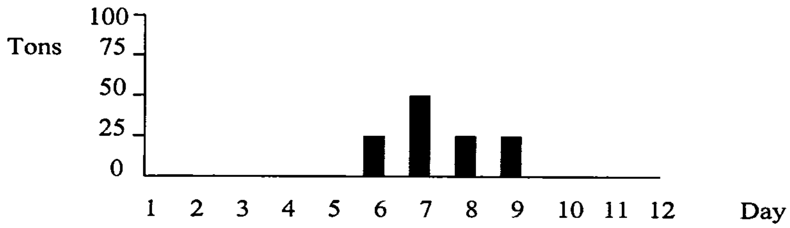

Figure A: Late Cargo Arriving on each Day

This may be due in part to insufficient aircraft for bulk cargo.

Most of the late cargo is bulk (Figure B).

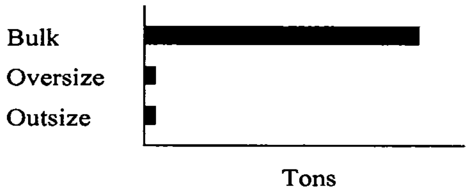

Figure B: Amount of late cargo by cargo type

You might consider increasing the number of aircraft that can carry bulk cargo.

Figure 1: Example

System Overview

We have implemented a prototype system, AutoBrief, in the domain of transportation scheduling that generates presentations in text and information graphics. As in several multimedia generation systems (McKeown *et al.* 1992; Towns, Callaway, & Lester 1998; Wahlster *et al.* 1993), the content and high-level organization of a presentation is first planned using a goal-driven planner (Young 1994). The presentation planner attempts to satisfy presentation goals using information about the domain, and presentation strategies encoded as media-independent plan operators. (We refer to the plan operators as being media-independent since no knowledge of media is encoded in them.) The resulting plan, which may include ordering constraints, describes the content and structure of the presentation.

For example, the plan underlying the sample presentation of Figure 1 is shown in Figure 2. For readability, only the goal hierarchy is shown. English paraphrases of the plan's content language expressions are given. Nodes are labelled numerically for reference later in the paper. Subgoals labelled *core* contain the information that is realized as the core (Moser & Moore 1996) of a segment in the presentation shown in Figure 1, and are identifiable as core goals in the argumentation operators defined in our system. The analysis of the argument encoded in the plan is based upon argument strategies for three types of arguments, factual, causal, and recommendations, described in a textbook on writing arguments (Mayberry & Golden 1996).

In the next phase of generation, media allocation is performed. During media allocation, terminal nodes of the plan are assigned to text and/or graphics (represented by the annotations *TEXT* and *GRAPHIC* in Figure 2). Then, the graphics generator is invoked to process the nodes that have been allocated to graphics. All graphic design decisions are made by the graphics generator, so properties of graphics to appear in the presentation are not known until the graphics generator has finished its designs. For example, the graphics generator may design a single graphic to achieve multiple goals of the plan. Thus, generation of cross-media cues cannot be performed until graphic design is finished.

After the graphic design phase, generation of cross-media acts is performed. The text generator is responsible for realizing cross-media cues and acts of the plan allocated to text.

Media Allocation

For background, first we present an overview of some general media allocation issues that are important for AutoBrief. Then we propose a novel media allocation policy based on argumentation goals.

To give an example of a case when graphics might be preferable to text (Roth & Hefley 1993), consider

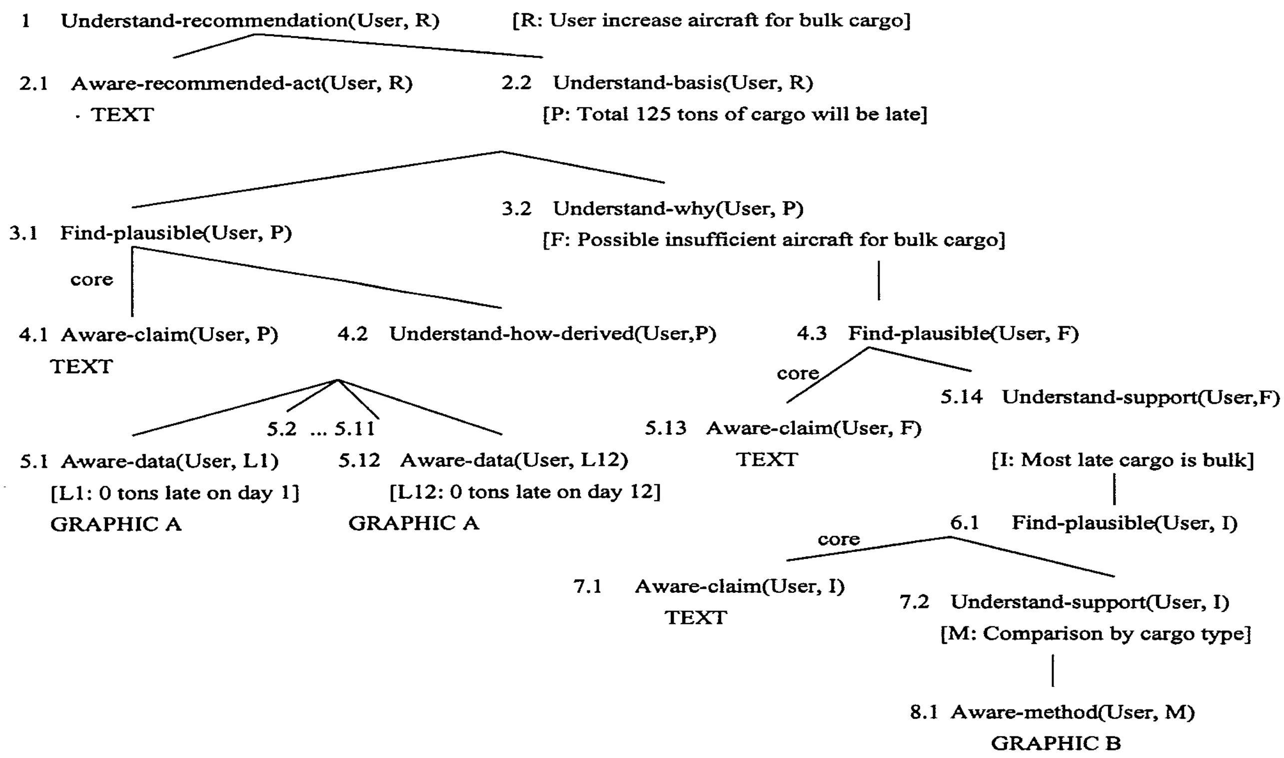

Figure 2: Plan for Argument in Figure 1

subgoals 5.1 to 5.12 of the plan shown in Figure 2. (To save space, only subgoals 5.1 and 5.12 are fully shown. To paraphrase the subgoals, subgoals 5.1 to 5.5 are to make the user aware that 0 tons of late cargo arrived on days 1 to 5, respectively; subgoals 5.6 to 5.9 are to make the user aware that 25, 50, 25, and 25 tons of late cargo arrived on days 6 to 9, respectively; and subgoals 5.10 to 5.12 are to make the user aware that 0 tons of late cargo arrived on days 10 to 12, respectively.) Each of these subgoals is to be achieved by performing an act of the same type (*Assert-data*[1]) and the content of each of these acts has the same propositional form (which can be paraphrased as *Tons(i) of late cargo arrive on day i*, for *i* from 1 to 12). These 12 acts are expressed in Graphic A of Figure 1. Although they also could be expressed in a paragraph of text, the text would be repetitious,[2] and its linear form would not facilitate operations the user might wish to perform such as looking up the late tons arriving on a certain day.

On the other hand, there are reasons for not expressing certain communicative acts in graphics, including: some types of act have no obvious graphic representation (e.g., *Recommend*), the graphics generator may have no way of expressing the propositional content of an act, or the graphical representation of an act may exceed acceptable limits on complexity or conciseness. In our system, we assume that text is always available as a medium of expression if graphics is not selected.

In addition to its above role as a "fallback" medium, text also can serve an important argumentative function. Textbooks on technical writing and information visualization, e.g., (Cleveland 1994), stress that the text of a document should express the main point of a graphic or the conclusions which the reader is expected to draw from it. For example, in Figure 1, the information that 125 tons of cargo will be late could be recognized by an ideal rational agent from Figure A alone. However, there may be other conclusions that could be drawn that do not play a role in the argument (no late cargo arrives before day 6, etc.). Also, a less-than-ideal reader may not draw all required conclusions.

Thus, to ensure that the goal of an argument expessed in graphics can be easily recognized by the au-

[1] The acts are not shown in the figure, due to space limitations. *Assert-data* is a primitive act whose goal is *Aware-data*.

[2] Although text generators can use text aggregation strategies such as we used above to produce more concise text, there are limits to the amount of aggregation that can be performed before information is lost. Also, although text generators can use stylistic variation to make a text less repetitious, there is the danger that the user will draw unintended inferences based on Grice's Maxim of Manner from the use of variation (Grice 1975).

dience, regardless of whether it is inferrable from a graphic, we adopt the following media allocation policy: *Allocate to text any leaf of the plan that is the core of a segment with details or support to be expressed in graphics.* This policy accounts for the allocation of nodes 4.1, 5.13, and 7.1 of Figure 2 to text. Other policies (such as described above) are applied independently of this policy. (For example, node 2.1 would be allocated to text because of the type of act, *Recommend.*)

Cross-Media Integration

As mentioned in the System Overview section, cross-media cue generation is performed after graphic design. Cross-media cue generation is the process of generating an act whose goal is to make the reader aware of the role of a graphic in the argument. The goal can be achieved by expressing the relationship of a graphic to the appropriate part of the text, e.g., *Figure 5 suggests that port capacity is insufficient at hour 8.* The novel approach described here is based upon the representation of arguments in our system, and presupposes the media allocation policy described above. Our approach to generating these acts is as follows: *If N is the core of a segment containing additional information, to be realized by a graphic G, and G provides details or support for N, then we want the user to recognize the relationship of N to G. Thus, generate a cross-media cue relating N to G.* For example, this approach results in the generation of cross-media cues relating nodes 4.1 and 7.1 of the plan in Figure 2 to Figures A and B, respectively, in the example in Figure 1.

Note that since the same graphic may be designed by the graphics generator to fulfill multiple goals, the above approach may result in references to a graphic from more than one part of the presentation. For instance, after the presentation in Figure 1 is shown, the system might continue with a presentation about days on which late cargo arrives. If the plan for this new presentation includes the claim that *the most late cargo arrives on day 7,* supported by the data previously presented in Figure A, and if the graphics generator decides to allocate the supporting data to the same graphic, then another cross-media integration goal would relate the new claim to Figure A.

Related Work

Previous work in media allocation by intelligent multimedia presentation systems has not addressed the role of argument structure in media allocation. Instead, media allocation is based upon which types of information can be conveyed by illustrations of objects and processes (Feiner & McKeown 1991; Towns, Callaway, & Lester 1998; Wahlster *et al.* 1993), or in terms of abstract features of media, datatypes, producer, and audience (Arens, Hovy, & Vossers 1993). However, the generation of cross-media referring expressions (expressions that refer to other parts of a presentation) (McKeown *et al.* 1992) does need to be addressed in our future work. For example, cross-media cues can be combined with referring expressions that link elements of the text (e.g. *needed capacity* below) to elements of the graphic (e.g. referred to as *the blue line* below): *As shown in Figure 6, needed capacity (the blue line) exceeds available capacity (the red line).*

There has been work on generating text coordinated with information graphics, but none addressing the role of argument structure. One caption generation system (Fasciano & Lapalme 1996) produces text describing the emphasis of a graphic. However, the mechanism producing these captions is based upon templates provided by the user rather than reasoning on the role of a graphic in the presentation. Another caption generation system (Mittal *et al.* 1998) is capable of automatically generating extended captions to help the user understand potentially complex information graphics designed by an automated design system. However, the captions are limited to descriptions of graphic design and data-to-grapheme mappings, since the system does not represent the communicative goals to be achieved by a graphic. Nevertheless, in future work, we would like to integrate that type of information with the information provided in our approach.

Conclusions

In conclusion, media allocation and cross-media cue generation are problems related to intelligent layout generation. We have proposed an approach to these problems that makes use of argument structure to define domain-independent policies for a multimedia generation system.

Acknowledgments

This project was supported by DARPA, contract DAA-1593K0005. Other members of the AutoBrief project are Stephan Kerpedjiev, Joe Mattis and Steve Roth of Carnegie Mellon University, Giuseppe Carenini of the University of Pittsburgh, and Johanna Moore of the University of Edinburgh.

References

Arens, Y.; Hovy, E.; and Vossers, M. 1993. On the knowledge underlying multimedia presentations. In Maybury, M., ed., *Intelligent Multimedia Interfaces.* AAAI Press. 280–306.

Card, S. K.; Mackinlay, J.; and Shneiderman, B. 1999. Information visualization. In Card, S. K.; Mackinlay, J.; and Shneiderman, B., eds., *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann Publishers, Inc.: Cambridge University Press. chapter 1, 1–33.

Cleveland, W. S. 1994. *The Elements of Graphing Data.* Summit, New Jersey: Hobart Press.

Fasciano, M., and Lapalme, G. 1996. PostGraphe: a System for the Generation of Statistical Graphics and

Text. In *Proceedings of the 8th International Natural Language Generation Workshop*, 51–60.

Feiner, S., and McKeown, K. 1991. Automating the generation of coordinated multimedia explanations. *IEEE Computer* 24(10):33–40.

Green, N.; Carenini, G.; Kerpedjiev, S.; Roth, S.; and Moore, J. 1998. A media-independent content language for integrated text and graphics generation. In *Proceedings of the Workshop on Content Visualization and Intermedia Representations (CVIR'98) of the 17th International Conference on Computational Linguistics (COLING '98) and the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98).* Coling-ACL'98 and ACL SIGMEDIA.

Green, N.; Carenini, G.; and Moore, J. 1998. A principled representation of attributive descriptions for integrated text and information graphics presentations. In *Proceedings of the Ninth International Workshop on Natural Language Generation.*

Grice, H. P. 1975. Logic and conversation. In Cole, P., and Morgan, J. L., eds., *Syntax and Semantics III: Speech Acts.* New York: Academic Press. 41–58.

Kerpedjiev, S.; Carenini, G.; Roth, S.; and Moore, J. 1997. AutoBrief: a multimedia presentation system for assisting data analysis. *Computer Standards and Interfaces* 18:583–593.

Mayberry, K. J., and Golden, R. E. 1996. *For Argument's Sake: A Guide to Writing Effective Arguments.* New York: HarperCollins College Publishers. Second Edition.

McKeown, K.; Feiner, S.; Robin, J.; Seligmann, D.; and Tanenblatt, M. 1992. Generating cross-references for multimedia explanation. In *Proceedings of AAAI*, 9–16.

Mittal, V.; Moore, J.; Carenini, G.; and Roth, S. 1998. Describing complex charts in natural language: A caption generation system. *Computational Linguistics* 24(3):431–467.

Moser, M., and Moore, J. D. 1996. Towards a synthesis of two accounts of discourse structure. *Computational Linguistics* 409–419.

Roth, S., and Hefley, W. 1993. Intelligent multimedia presentation systems: Research and principles. In Maybury, M., ed., *Intelligent Multimedia Interfaces*, 13–58. AAAI Press.

Towns, S. G.; Callaway, C. B.; and Lester, J. C. 1998. Generating coordinated natural language and 3d animations for complex spatial explanations. In *Proceedings of AAAI98.*

Wahlster, W.; Andre, E.; Finkler, W.; Profitlich, H.-J.; and Rist, T. 1993. Plan-based integration of natural language and graphics generation. *Artificial Intelligence* 63:387–427.

Young, M. R. 1994. A developer's guide to the Longbow discourse planning system. Technical Report ISP TR Number: 94-4, University of Pittsburgh, Intelligent Systems Program.