Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Int. J. Human-Computer Studies 61 (2004) 32–70

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

AutoBrief: an experimental system for the automatic generation of briefings in integrated text and information graphics

Nancy L. Green[a,*], Giuseppe Carenini[b], Stephan Kerpedjiev[c], Joe Mattis[c], Johanna D. Moore[d], Steven F. Roth[c,e]

[a] *Department of Mathematical Sciences, University of North Carolina at Greensboro, 383 Bryan Building, Greensboro, NC 27402-6170, USA*

[b] *Computer Science Department, University of British Columbia, Vancouver, BC, Canada CA V6T 1Z4*

[c] *MAYA Viz, 2100 Wharton St., Suite 702, Pittsburgh, PA 15203, USA*

[d] *Human Communication Research Centre, The University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, UK*

[e] *School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*

Received 22 October 2002; received in revised form 3 October 2003; accepted 13 October 2003

Abstract

This paper describes AutoBrief, an experimental intelligent multimedia presentation system that generates presentations in text and information graphics in the domain of transportation scheduling. Acting as an intelligent assistant, AutoBrief creates a presentation to communicate its analysis of alternative schedules. In addition, the multimedia presentation facilitates data exploration through its complex information visualizations and support for direct manipulation of presentation elements. AutoBrief's research contributions include (1) a design enabling a new human–computer interaction style in which intelligent multimedia presentation objects (textual or graphic) can be used by the audience in direct manipulation operations for data exploration, (2) an application-independent approach to multimedia generation based on the representation of communicative goals suitable for both generation of text and of complex information graphics, and (3) an application-independent approach to intelligent graphic design based upon communicative goals. This retrospective overview paper, aimed at a multidisciplinary audience from the fields of human–computer

*Corresponding author. Tel.: +1-336-2561133; fax: +1-336-3345949.

E-mail addresses: nlgreen@uncg.edu (N.L. Green), carenini@cs.ubc.ca (G. Carenini), kerpedjiev@maya.com (S. Kerpedjiev), mattis@mayaviz.com (J. Mattis), j.moore@ed.ac.uk (J.D. Moore), roth+@cs.cmu.edu (S.F. Roth).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2003.10.007

interaction and natural language generation, presents AutoBrief's design and design rationale.

1. Introduction

AutoBrief is an experimental intelligent multimedia presentation system (Roth and Hefley, 1993) that automatically generates presentations in integrated text and information graphics. *Information graphics* are non-pictorial graphics such as bar charts and line graphs, and are distinguished from illustrations of physical objects or diagrams (Card et al., 1999). It is noted commonly that graphs can serve two purposes (e.g. Cleveland, 1994). First, presenting data in a graph may enable the viewer to discover new information, i.e. patterns, trends, or associations. Second, graphs are used as a communication medium:

> A graph is essentially a rhetorical device; it is a form of persuasion…A graph is constructed to illustrate particular patterns found in data. Many other graphs could be drawn from a particular data file, but they rarely are. Only those graphs are produced that seem important to the analyst in order to understand and to communicate what the data mean (Iverson and Gurgen, 1997, p. 73).

Likewise, AutoBrief is designed both to facilitate data exploration *by* the user and to communicate its analysis of data *to* the user. Acting as an intelligent assistant, AutoBrief analyses the user's data and creates a presentation to communicate the analysis. AutoBrief's presentations are expressed in integrated text and information graphics in order to exploit the strengths of the two media for communication, both individually and in combination. For example, a well-designed information graphic can display large amounts of supporting data compactly. Furthermore, it may enable the skilled viewer to grasp the designer's communicative goal in a glance. On the other hand, text enables complex arguments and fine distinctions to be communicated precisely. Used in conjunction with graphics, text can reinforce the designer's message (e.g. *Stock prices are rebounding*) and help the user to interpret the graphic's visual vocabulary (e.g. *The red line shows expenses and the black line shows income*).

In addition to its communicative function, the multimedia presentation generated by AutoBrief facilitates data exploration by the user in two ways. First, the information graphics in an AutoBrief presentation may enable the user to perceive trends or relationships that were not detected by AutoBrief's analyser. Second, AutoBrief provides a direct manipulation style of interaction. For example, the user may select certain phrases of text or elements of graphics in the presentation and drag them into an integrated visual data exploration environment. The data exploration environment interprets the items as having the same meaning as they did in the context of the AutoBrief presentation from which they came. Thus, the user need not know about the system's representation of the underlying data, nor need know a formal query language, to perform data exploration on entities referred to in an AutoBrief presentation.

AutoBrief's research contributions include (1) a design enabling a new human–computer interaction style in which intelligent multimedia presentation objects (textual or graphic) can be used by the audience in direct manipulation operations for data exploration, (2) an application-independent approach to multimedia generation based on the representation of communicative goals suitable for both generation of text and of complex information graphics, and (3) an application-independent approach to intelligent graphic design based upon communicative goals. This retrospective overview paper presents AutoBrief's design and design rationale. Previous papers have reported on specific aspects of its evolving design (Green et al., 1998a-c; Kerpedjiev et al., 1998a, b, 2000; Kerpedjiev and Roth, 2000). This paper is aimed at a multidisciplinary audience from the fields of human–computer interaction and natural language generation (NLG). (The members of the AutoBrief project were drawn from these two fields as well.) Thus, in Section 2 we begin by providing background information and rationale from each field that influenced the design (and designers) of AutoBrief. Section 2 ends with a description of related work in intelligent multimedia presentation systems. Section 3 describes AutoBrief from the user's point of view in terms of the application that was used as a testbed for our research, transportation logistics analysis. Section 4 presents the internal design of AutoBrief. Section 5 presents discussion of issues arising from the research and suggestions for future directions.

2. Background

2.1. Communicating via graphics

As mentioned in the introduction, it is often the case that many different visual representations of the same data are possible. In the field of information graphic design, statistics researchers and visual design experts have provided guidelines on selection of appropriate graphical techniques as a function of the type of data to be displayed (e.g. Bertin, 1983; Tufte, 1983; Cleveland and McGill, 1985). Many guidelines have been motivated by cognitive theories of graph comprehension (e.g. Kosslyn, 1989; Pinker, 1990). One sort of guideline concerns choosing an appropriate type of graphic format, i.e. bar chart, line graph, etc., for the type of data to be displayed, e.g. categorial, ordinal, or quantitative. Other types of guidelines concern the elements of graphic design such as choice of encoding technique for an attribute, e.g. position on the horizontal axis, color, shape, etc. Lastly, convention may constrain design of a graphic.

To illustrate data constraints on choice of graphic format, a statistics textbook (Moore, 1997) gives the following guidelines: a pie chart is appropriate for showing part–whole relationships for a single categorial variable (a variable whose values have no intrinsic ordering and that partition a data set into distinct groups) but is not appropriate for showing ordinal and quantitative data. On the other hand, a bar graph can be used to show categorial, ordinal, or quantitative data. An example of a guideline related to the design of graphic elements is that increasing the distance between scale marks on the vertical axis of a line graph may increase the slope of the line and thereby suggest a greater rate of change than would otherwise have been conveyed. To give an example of guidelines based upon convention, when designing a Cartesian graph to display a possible causal relation between two variables, it is customary to display the independent variable on the horizontal axis and the dependent variable on the vertical axis.

The first system to automatically design information graphics based upon such guidelines was developed by Mackinlay (1986). Subsequent research extended that approach by taking into consideration an additional sort of design constraint, namely, that the design should enable the user to perform information seeking tasks efficiently (Roth and Mattis, 1990; Casner, 1991; Beshers and Feiner, 1993). The rationale for task-influenced design is as follows. After analysing the needs of users, a designer can evaluate the efficiency of alternate graphic designs based upon which perceptual and computational tasks that a user must perform to accomplish his or her goals. For example, suppose that a horizontal bar chart displays employee salaries, where employee names are given on the vertical axis and salaries on the horizontal axis. If the user's goal is to determine the salary of an employee given the employee's name, then displaying the bars in alphabetical order by employee name is more helpful than some other ordering such as random or sorted in order of increasing salary. On the other hand, if the user's goal is to discover which employee earns the maximum, median, or minimum salary, then arranging the bars in order by salary would be more helpful. To give another example, although both pie charts and bar graphs can be used to show categorial data, if the user's goal is to compare the size of each group, then a bar chart would be more helpful than a pie chart. This is because it is a more difficult perceptual task to discriminate differences in the angles of wedges in a pie chart than it is to compare lengths of bars in a bar chart.

2.2. Data exploration

The direct manipulation paradigm of interaction rose to prominence in the 1980s. The paradigm is characterized as follows (Shneiderman, 1998): the interface creates a visual metaphor of the user's task objects and actions; instead of typing commands, the user acts directly upon the visual surrogates; the results of the user's actions on the underlying system are immediately visible through the visual metaphor; the user is able to interact rapidly and incrementally; and it is easy for a user to reverse the effects of his or her actions.

Applying the paradigm to user interaction with databases resulted in the development of graphical techniques to support a user's information seeking tasks. For example, the dynamic query technique presents users with a graphical representation of an overview of the data and allows users to "explore and conveniently filter out unwanted information. Users fly through information spaces by incrementally adjusting a query (with sliders, buttons, and other filters) while continuously viewing the changing results" (Shneiderman, 1994). Brushing and painting techniques are another example of a direct manipulation technique: by selecting elements of a graphic and changing their color, a user can specify ad hoc subsets of a data set for subsequent data operations (Roth et al., 1997). Direct manipulation enables users to search for information without knowing *how* to express a query in a formal query language. Moreover, it enables users to discover information without knowing *what* to look for. In contrast to *data mining*, which "emphasizes the use of automatic mechanisms to search for patterns", an interactive *data exploration* system supports an interactive, iterative process of data exploration by the user (Goldstein et al., 1994).

Visage (Kolojejchick et al., 1997; Roth et al., 1997) is a data exploration environment providing direct manipulation operations on data integrated with information visualizations created by an automatic graphic design tool, SAGE (Roth et al., 1994). The goal of information visualization is to enable a user to gain insight into data by presenting multiple data attributes in a single graphic. (Information visualization differs from scientific visualization in that the latter primarily addresses the display of physical objects and spatial data, e.g. medical imaging.) SAGE can design complex graphics by encoding multiple attributes using different visual dimensions (e.g. position, color, and shape), superimposing data from multiple graphics, and aligning multiple graphs along a common axis. Elements of the graphics can be acted upon by the user through direct manipulation operations. For example, when two graphics presenting different views of the same data are created, the user may select a subset of the elements in one visualization through a paint operation; Visage automatically coordinates painting of elements related to the same underlying data objects; thus, the corresponding elements in the other visualization will change color accordingly; or if the user manipulates a slider, corresponding elements of both visualizations will be affected.

2.3. Discourse-theoretical approaches to automated graphics design

The related fields of discourse interpretation and generation are concerned with the understanding and production, respectively, of discourse by an intelligent agent (human or computational). *Discourse* may refer to multi-sentence text or dialogue. (In this paper, we interpret *discourse* more broadly to include presentations in integrated text and graphics. Thus, we shall use the more medium-neutral term *audience* to refer to the intended recepient in discussion of multimedia issues later in the paper, rather than the terms most often used in language studies, *hearer* and *reader*.) Grice (1957) argued that communication involves the hearer's recognition of what the speaker intended to communicate. Thus a speaker's meaning may go beyond the strict semantic interpretation of what he or she actually said. Interpreting the speaker's intentions may take into consideration knowledge that is presumed to be shared by speaker and hearer, e.g. the preceding discourse or shared background knowledge. In addition, Grice (1975) proposed that a set of conversational Maxims (see Table 1) are usually observed in cooperative interaction. He pointed out that sometimes recognizing the speaker's intention requires the assumption that the speaker has been adhering to a particular maxim. For example, if someone answers

Table 1
Grice's conversational maxims

- *Quality*:

(a) Do not say what you believe to be false.
(b) Do not say that for which you lack adequate evidence.
- *Relevance*: Make your contribution relevant to the conversation.
- *Quantity*:

(a) Make your contribution as informative as is required (for the purpose of the exchange).
(b) Do not make your contribution more informative than is required.
- *Manner*: Avoid obscurity of expression and ambiguity, be brief, and be orderly.

Two to the question, *How many children do you have*? then by the Maxims of Quantity-a and Relevance one normally would interpret the response as meaning that the speaker has exactly (i.e. no more than) two children.

Today, much work in intelligent discourse generation (e.g. Moore, 1995) presupposes that communication requires the audience's recognition of the generating agent's intention. One of AutoBrief's design assumptions is that the Gricean model provides insight into an idealized viewer's interpretation of the communicative goals of a well-designed information graphic. For example, applying the maxims of Relevance and Quantity-b to Fig. 1a (which is based on a graphic appearing in a report on undergraduate education[1]) a viewer could analyse the designer's intentions as follows.

The height of each stacked vertical bar represents the total percentage of ethnic/minority freshmen in each of the categories listed along the x-axis (*Biological Science*, *Engineering*, etc.). Since the y-axis stops at 25% and the percentage of all other freshmen in each category is not shown (although it could be derived from the graphic), from Quantity-b one can reason that the designer does not intend for the viewer to focus on the proportion of ethnic/minority freshmen to all freshmen in the same x-axis category. On the other hand, the scale of the y-axis and the proximity of the vertical bars on the x-axis make it easy to discriminate the relative heights of the bars. Thus, from Grice's maxim of Relevance, one could reason that the designer intends for the viewer to make comparisons among the total percentages of ethnic/minority freshmen in various x-axis categories. However, since the graphic does not enable a viewer to determine these percentages with precision, from Grice's maxim of

[1] We have constructed Fig. 1a following the design of Fig. 10 in the report, "An Exploration of the Nature and Quality of Undergraduate Education in Science, Mathematics and Engineering: A Report of the National Advisory Group of SIGMA XI, the Scientific Research Society", published by Sigma Xi, The Scientific Research Society, New Haven, CT, 1989. Our version omits the data for the category Social Sciences for reasons of space and uses approximations of the other data since the raw data was not available to us. We designed Figs. 1b and c ourselves for the purpose of illustrating how alternate designs of information graphics conveying the same data may support different interpretations of the designer's communicative goals.

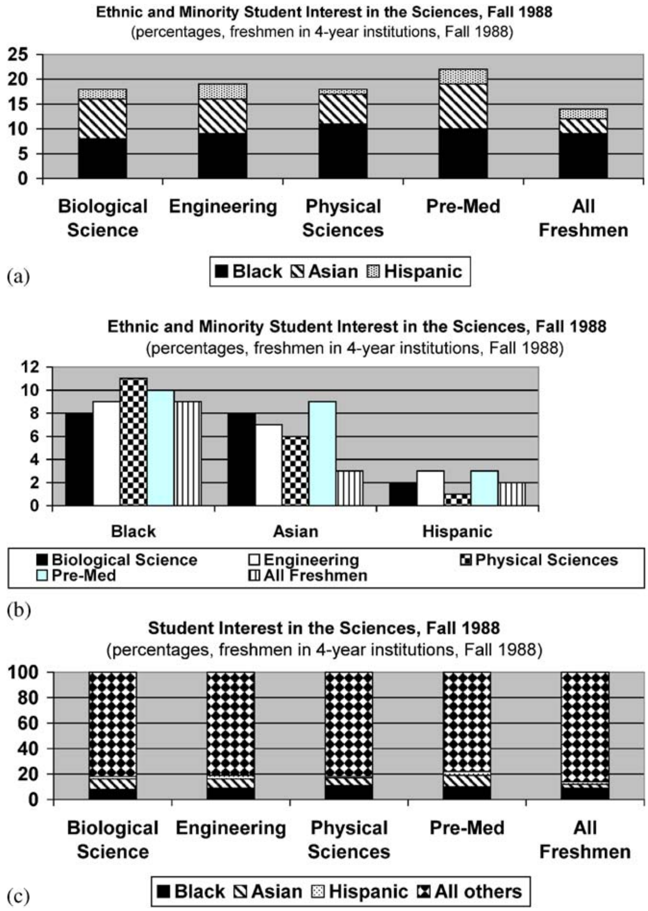

Fig. 1. (a, b) Ethnic and minority student interest in the sciences, Fall 1988 (percentages, freshmen in 4-year institutions, Fall 1988). (c) Student interest in the sciences, Fall 1988 (percentages, freshmen in 4-year institutions, Fall 1988).

Quantity-b one could reason that the designer's goals do not require the reader to know the exact value of each total.

Furthermore, since all of the bars except the rightmost bar (the one labelled *All Freshmen*) are ordered alphabetically by x-axis category name rather than bar height, from Quantity-b one could reason that the designer's goal does not involve a comparative ranking of the data in the leftmost four x-axis categories; this design choice facilitates, instead, looking up the vertical bar for any one of those categories given its name. (This advantage would be more apparent if there were more categories.) From Relevance one could reason also that the bar labelled *All Freshmen* plays a special role since it is not presented in alphabetical position but is presented at one end. Therefore, it is plausible that the designer intends for the viewer to be aware of the relation between the total for the category *All Freshmen* and the total for each of the other four categories, e.g. the total percentage of ethnic/minority freshmen interested in *Biological Science* compared to the total percentage of all ethnic/minority freshmen (shown in *All Freshmen*).

So far this interpretation does not account for why each bar has been divided into *Black*, *Asian*, and *Hispanic* segments. The design does not enable the viewer to easily grasp the proportion of *Black*, *Asian*, and *Hispanic* segments in each vertical bar, except in cases where there is a highly visible difference, such as that between *Black* and *Hispanic* segments in *Physical Sciences*. Nor, except for the group (*Black*) whose segments line up with the x-axis, does it enable the viewer to easily compare the same group across x-axis categories, e.g. to compare *Hispanic Engineering* to *Hispanic All Freshmen*. Thus, although by Relevance one could reason that the data for each of the specific minority groups must play a role in the designer's communicative goals, by Quantity-b, one could reason that it must be secondary to the designer's goal of enabling comparisons between the total for *All Freshmen* and the total for each of the other x-axis categories.

Now let us compare our Gricean-style interpretation of the goals of Fig. 1a with the text that actually accompanied the graphic: "The CIRP data on minority interest in the sciences suggest that there has been some improvement at the front-end of the pipeline over the past decade. Interest in engineering, physical science, and life science majors among Black and Hispanic freshmen all posted gains between 1978 and 1988. Moreover, in some cases these gains push minority students past many of the commonly used measures of parity often employed to assess representation and progress (Fig. 10). For example, Blacks represent 9.8% of the first-time, full-time freshmen enrolled in the nation's 4-year colleges and universities in Fall 1988; however, Blacks also account for 11.5% of the freshmen planning to pursue physical science majors (e.g. chemistry, physics, mathematics) in Fall 1988. Similarly, Hispanic students represent 1.8% of the first-time, full-time freshman population this past fall and 2.1% of the aspring freshman engineering students" (p. 38).

The author's claim that in Fall 1988 the percentage of minority students in engineering, physical science, and life science majors each exceeds one benchmark, the percentage of all freshmen who are minority students, is consistent with our Gricean interpretation that the designer intended to enable the viewer to compare the total represented by each of the left-hand four bars to the total represented by the bar labelled *All Freshmen*. Also, the author's claim that the total gain can be attributed to gains in the percentages of Black and Hispanic freshmen in those majors is consistent with our Gricean interpretation that the designer intended for data about particular minority groups to play some kind of secondary role. In this case, the segmentation of the vertical bars provides a visual explanation for exceeding the measure of parity shown in the graphic. However, without the text explicitly drawing the audience's attention to comparisons between Black and Hispanic freshmen in those majors to all Black and Hispanic freshmen, respectively, the visual explanation might not be as effective.

This secondary goal could be conveyed more effectively by use of an information graphic such as Fig. 1b. For example in Fig. 1b, it is easier to see that the percentage of *Hispanic* freshmen in *Engineering* exceeds the percentage of all *Hispanic* freshmen than it would be to see this in Fig. 1a. On the other hand, Fig. 1b could not be used to make the author's point about all minority students as effectively as Fig. 1a. Not only would that require the viewer to compute the totals shown in Fig. 1a, without the text the viewer might not even realize that the author intended to make a point about the totals. We suggest that the design of Fig. 1a represents a practical compromise, enabling the first point to be communicated effectively in graphics, while at the same time providing a view of the data about Black and Hispanic freshmen that is discussed in more detail in the text.

To give a final example of how the design of a graphic may influence what communicative goal is ascribed to its designer, consider Fig. 1c, which presents the same data as shown in Fig. 1a. The design of this graphic makes visible the percentages for non-ethnic/non-minority freshmen; i.e. the data labelled *All Others* in Fig. 1c is not displayed in Fig. 1a although it can be derived from the totals displayed in Fig. 1a. By relevance, a viewer could reason that the designer's goal must involve the proportion of each of the four segments in a bar to total bar height. Furthermore, in this design the extent of the difference in the percentages of the three ethnic/minority groups compared to *All Others* is much more prominent than differences among different ethnic/minority groups, whether compared to each other or in same group comparisons across different x-axis categories. Thus, this graphic would be less effective to convey the goals stated in the text. Instead, use of this design could be used for making the point that the representation of the three ethnic/minority groups, in each of the sciences as well as in the freshman class as a whole, is low compared to the group *All Others*.

In general, there may be more than one plausible interpretation of a graphic's designer's intentions. However, some interpretations, while accurate descriptions of information that can be "read" from the graphic, are less plausible as summaries of the designer's communicative goals. For example, as noted above, the information represented by *All Others* in Fig. 1c could be derived by a viewer from the information presented in Fig. 1a. Although it can be derived, it is not a plausible interpretation of the designer's communicative goals in Fig. 1a. This is analogous to the distinction in language between Gricean speaker meaning on the one hand, and the semantic interpretation, entailments, and defeasible pragmatic implications of a sentence on the other hand (Levinson, 1983).

Others have argued that a Gricean model is applicable to graphics (Marks and Reiter, 1990; Oberlander, 1996). Some guidelines for graphical design are consistent with this model as well. Table 2 compares the Principles of Graphical Excellence proposed in a popular book on the design of data graphics, The Visual Display of Quantitative Information (Tufte, 1983), with Grice's Maxims.

The communicative goal-driven view of discourse generation has been extended to automatic graphic design. For example, the IBIS system designs enhancements to illustrations, such as highlighting, based upon the communicative goals of a graphic (Seligman and Feiner, 1991). Zhou and Feiner (1997) developed a communicative goal-driven approach to generate coherently connected visual displays of computer network diagrams. Mittal (1997) surveyed visual techniques used to signal the designer's intention in a corpus of human-designed information graphics.

Table 2
Comparison of Tufte's Principles to Grice's Maxims

Tufte (1983, p. 51)	Related Gricean Maxims
"Graphical excellence is the well-designed presentation of interesting data—a matter of *substance*, of *statistics*, and of *design*."	Manner, Relevance
"Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency."	Manner
"Graphical excellence is that which gives to the viewer the greatest number of ideas …"	Quantity-a
"in the shortest time with the least ink in the smallest space."	Manner, Quantity-b
"Graphical excellence is nearly always multivariate."	Quantity-a
"And graphical excellence requires telling the truth about the data."	Quality

In conclusion, we found the Gricean model invaluable for informal analysis of the relationship between communicative goals and graphic design techniques. However, Grice's maxims have not been operationalized in computer systems for natural language or graphics generation. In the next section we discuss how communicative goal-driven generation has been implemented computationally.

2.4. Intelligent multimedia generation

As Reiter and Dale (2000) point out, NLG systems frequently are based upon a view of language generation as a goal-driven process. In this view, primitive communicative actions, such as informing, requesting, and persuading, are used by an agent in order to achieve communicative goals to affect the knowledge/belief state or intentions of the hearer. In many NLG systems that generate multi-sentence text, e.g. (Moore, 1995), higher-level communicative goals, such as to explain something, correspond to structural/rhetorical units of a text; a planning process decomposes the higher-level goals into lower-level goals corresponding to structural sub-units and so on down to the level of primitive communicative goals and actions, resulting in the construction of a hierarchically structured communicative plan.

This approach has been extended to the generation of text with pictorial graphics and maps, e.g. for generating instructions with illustrations for tasks such as radio repair (Feiner and McKeown, 1991) and coffee machine operation (Wahlster et al., 1993); maps with narrated directions (Maybury, 1991); and, more recently, language and animation (e.g. Dalal et al., 1996; Stone and Lester, 1996; André et al., 1999; Cassell, 2000; Rickel and Johnson, 2000). In intelligent multimedia generation systems, the higher-level goals of the plan are usually *media-independent*, i.e. they do not presuppose or dictate the medium to be used. In some systems, the lower-level goals in the hierarchical plan are *media-dependent*, and the plan may specify rhetorical relations between parts of the plan to be realized in text and those to be realized in graphics (e.g. Maybury, 1991; André and Rist, 1994). In other systems, the entire plan is media-independent and a post-planning process (called *media allocation* or *media selection*) is used to select the medium for realizing different parts of the plan (e.g. Feiner and McKeown, 1991; Wahlster et al., 1993).

While recognizing the importance of intentions, early work in the generation of integrated text and information graphics did not adopt the approach of constructing a plan for the entire presentation starting from a media-independent communicative goal. One early system, Postgraphe, creates a simple graph and caption based upon information from the user about his or her intentions, e.g. to emphasize a rising trend between 1980 and 1990 in the values of the price attribute in the user's data set (Fasciano and Lapalme, 1999). In Postgraphe design heuristics are used to design the graphic based upon the type of intention specified by the user (e.g. trend, comparison, etc.) and the type of data to be displayed. However, no communicative plan is constructed, and the range of possible designs is much more limited than those that could be generated by an automatic graphic design system that supports data exploration, such as SAGE (described in Section 2.2). Also the text generated is limited to describing the specifications provided by the user and the type of graphic. Another early effort in this area, the Caption Generation System (Mittal et al., 1998) generates captions for information graphics created by SAGE. After SAGE has designed a graphic, the Caption Generation System identifies complex features of the graphic. Then the Caption Generation System uses a variety of communicative strategies to construct a communicative plan with the goal of helping the user to understand the graphic. However, the plan is used primarily for text generation. (The exception is that the plan may contain media-specific actions for modifying the original graphic to aid the explanation.)

In contrast to Postgraphe and the Caption Generation System, as we will describe in more detail shortly, in AutoBrief the generation process begins with a media-independent main goal that is decomposed to create a media-independent plan for the entire presentation. Subsequently, different parts of the plan are designated for realization in text and/or graphics by a media selection component. AutoBrief's graphics generator transforms the goals of the plan selected for graphic realization into a specification of tasks that the graphic should enable the user to perform; then, taking properties of the data and human visual capabilities into consideration, a graphic is designed to enable the user to perform those tasks efficiently. The rationale is that by being enabled to perform those tasks efficiently, the user will recognize the communicative goals designated for realization in graphics. In parallel with and independently of graphics generation, the text generator produces text to achieve the goals of the communicative plan designated for realization in text. Also, in contrast to the other multimedia generation systems surveyed in this section, AutoBrief introduces the human–computer interaction style of enabling its presentation elements, both textual and graphical, to be used as objects of direct manipulation actions.

3. AutoBrief prototype: user interface and application domain

A prototype of the AutoBrief architecture has been implemented in the domain of transportation scheduling. The prototype is designed to assist transportation logistics analysts while they use an incremental scheduling system, DITOPS (Smith et al., 1996).

Typically, an analyst uses DITOPS to create many trial versions of schedules. A schedule may contain an overwhelming amount of quantitative data in complex relationships. One role of AutoBrief is to act as an intelligent assistant providing briefings that summarize important features of and possible problems in the trial schedules. For example, Fig. 2a shows part of a briefing created by AutoBrief that presents information about how much of the cargo would be late at a certain port in one schedule selected by the user. In order to give the user control over what parts of a presentation are shown, AutoBrief's presentations contain controls that resemble hypertext navigation links. When the user clicks on a control, a new screen is designed automatically to fulfill a communicative goal associated with the control. For example, the user may click on the link labelled *Details* on the screen shown in Fig. 2a to request AutoBrief to provide related information; the resulting presentation is shown in Fig. 2b. Also, the user may request to see AutoBrief's analysis of the cause of the problem described in Fig. 2b, which is presented in Fig. 2c. (The causal explanation is that there are not enough lift assets, i.e. air transportation, allocated for the two types of cargo that are late.)

Another role of AutoBrief is to enable the user to perform data exploration, since the user may question AutoBrief's analysis or may be able discover information that AutoBrief's analyser cannot detect. Thus, AutoBrief's automated graphic designer uses advanced techniques such as those used in SAGE (described in Section 2.2), e.g. presenting multiple data attributes in a single graphic, superimposing data, and aligning multiple graphs along a common axis. For example, the graphic in Fig. 2b uses the technique of superimposing data from multiple data sets. In particular, the data points for needed lift capacity and the data points for available lift capacity at the port are plotted (using different colors) in the same graphic to enable the audience to perceive the relationship of needed to available capacity. Notice that part of the briefing expressed in text in Fig. 2b is intended to bring the audience's attention to the day with the worst problem: *The day with the largest difference between needed and available lift capacity is Day* 4. However, the graphic has been designed to show the relationship between the two variables for all of the days, thereby supporting the user's initiative in data exploration. For example, the user might be interested in exploring the entire time period during which needed capacity exceeded available capacity.

To further facilitate data exploration, AutoBrief's briefings are presented within Visage, the visual data exploration environment described in Section 2.2. Through Visage, the analyst can control DITOPS, AutoBrief, and data exploration/visualization tools in an integrated manner. In one scenario, an analyst (1) supplies DITOPS with schedule requirements, resources, and constraints, (2) runs DITOPS, (3) asks AutoBrief for a briefing on the resulting schedule, and (4) acting on his or

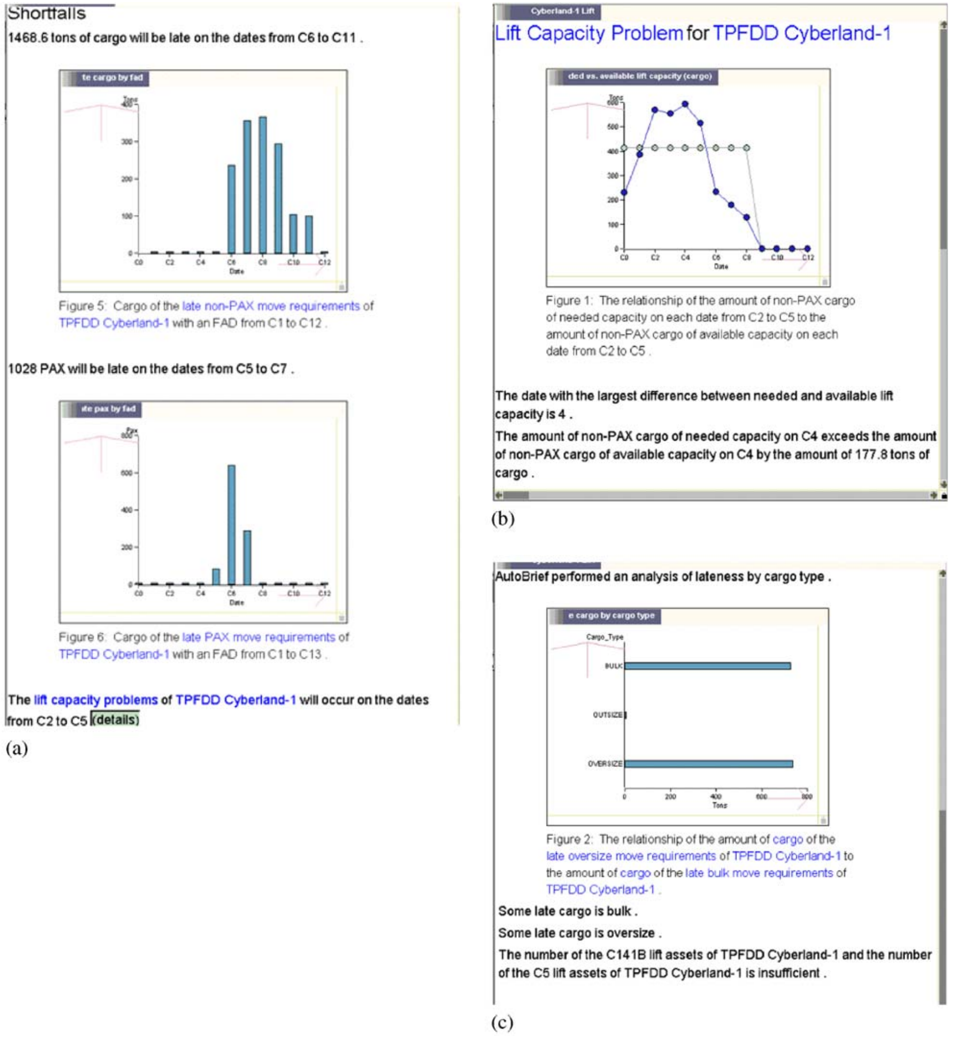

Fig. 2. (a) AutoBrief's summary of the shortfalls (i.e. late cargo) of the schedule named *TPFDD Cyberland*-1. First AutoBrief describes in text the total tons of cargo that will be late. The graph at the top of the screen shows how much of this late cargo arrives on each day (*FAD*). Then AutoBrief describes in text how many personnel (*PAX*) will be late, followed by a similar graph. The last sentence of the summary describes lift (air transportation) capacity problems that may be responsible for the lateness. If the user selects the *details* link, then the presentation shown in Fig. 2b is generated. (b) AutoBrief's more detailed description of the lift capacity problems that may be the cause of the lateness discussed on the preceding screen. The graph shows needed versus available lift capacity. In text, AutoBrief points out the date with the largest gap between needed and available capacity and the extent of the difference. (c) AutoBrief presents its analysis of the cause of the lift capacity problems: that there are not enough lift assets to carry the late bulk and oversize cargo.

her insight gained from the briefing, modifies the requirements and/or resources and/or constraints. The cycle (2) through (4) is repeated until a satisfactory schedule is produced.

Data exploration is facilitated also through AutoBrief's support for direct manipulation of the presentation: both highlighted text expressions and elements in graphics. As we mentioned in the Introduction, the system's interpretation of an item dragged by the user out of an AutoBrief presentation is determined by its meaning in the context of that presentation. Thus, the user need not know about the system's representation of the underlying data, nor need know a formal query language, to perform data exploration on entities in an AutoBrief presentation. To illustrate this type of scenario, the analyst (1) drags highlighted text from AutoBrief to a data visualization tool in Visage to explore the underlying data, (2) as a result of data exploration, decides to reschedule a certain cargo shipment, (3) to do this drags a bar representing that cargo shipment from a bar chart currently displayed by AutoBrief (such as the chart shown in Fig. 2a) to the DITOPS graphical user interface, and (4) directs DITOPS to reschedule just that shipment. (For a more detailed description of a scenario with screen shots, see http://www.cs.cmu.edu/~sage/ab-tour/start.html.)

AutoBrief can generate many different presentations. The number is a function of the properties of the data (i.e. the schedules created by DITOPS and the analyses produced by AutoBrief's Schedule Analyser) as well as the knowledge encoded in AutoBrief's Presentation Generator as plan operators, media selection heuristics, text and graphics aggregation policies, lexicon, and graphical primitives (see Section 4). To give a rough idea of the scope of the presentations that may be generated by the current implementation, presentations are composed using combinations of 58 genre-specific plan operators; and the system's graphic designer is capable of using grapheme properties (x-position, y-position, color, size, etc.) of five grapheme types (mark, bar, line, gauge, and text) and four visualization disciplines (chart, table, map, network); these graphic design primitives may be composed by clustering, single axis composition (alignment), and double axis composition. Although no formal evaluations have been performed, the prototype has been demonstrated to potential users, transportation logistics analysts, who reacted very favorably.

Furthermore, an indirect evaluation of AutoBrief's associated data exploration facilities (through Visage) and lower level graphic design component is suggested by the experience of participants in eight exercises with six to 20 participants each that were performed to evaluate the Command Post of the Future (CPOF) research prototype system (Chuah and Roth, 2003). The CPOF provides collaborative visualization tools. In these exercises teams used CoMotion, a descendant of Visage, to create and choose from alternate plans for handling simulated crisis scenarios. CoMotion provides a highly interactive visualization environment in which data objects can be manipulated by the same operations as provided by Visage, such as brushing and painting and dragging data objects from one visualization to another. (CoMotion's additional functionality enables distributed users to manipulate shared visualizations in real time.) Although some aspects of the design of the visualizations created in the CPOF exercises were predetermined and others were determined by user control, lower level graphic design in CoMotion is automated and follows an approach similar to that implemented in AutoBrief. The CPOF participants became adept in use of the data exploration tools in several hours and evaluated the system very favorably. In summary, it is plausible that AutoBrief users would evaluate the corresponding features of AutoBrief in a similar way.

The experience gained from the CPOF exercises also supports the potential usefulness of a system like AutoBrief that can analyse and filter data and then communicate its analysis to the user as a starting point for the user's decision making. Instead, the participants in the CPOF exercises were required to spend hours on this analysis and summarization process themselves. A related question is whether a system such as AutoBrief that presents information selectively to the user could hinder the user from finding a satisfactory solution at times. For example, although AutoBrief is capable of producing a large variety of presentations, there are limits to what questions it can answer. Thus, if the solution to the user's problem were outside of the scope of AutoBrief's analysis and summarization capabilities, its presentations would be of questionable value to the user. We would argue that this is a good justification for designing AutoBrief to support a high degree of integration between the system-generated briefings and user-controlled data exploration. The user is free to question and go beyond AutoBrief's analysis at any time by moving from AutoBrief's presentation to the integrated data exploration environment.

A comprehensive user evaluation of AutoBrief's intelligent presentation generation components has not been performed. However, in Section 5 we discuss theoretical limitations of AutoBrief's Presentation Generator.

4. Internal design

4.1. System architecture

AutoBrief has two main components, a Schedule Analyser and a Presentation Generator. The Schedule Analyser, which uses domain-specific knowledge and heuristics, analyses the user's transportation schedules, identifies potential problems, and suggests possible solutions. The Presentation Generator, which plans presentations to achieve communicative goals, designs text and information graphics to communicate the output of the Schedule Analyser to the user and to enable the user to perform data exploration. The Schedule Analyser is application specific, while the Presentation Generator is the focus of our research (and this paper) and is intended to be portable to other domains. Fig. 3a shows the flow of information between components of the prototype system. The scheduling system's (DITOPS) inputs and outputs, as well as data created by AutoBrief's Schedule Analyser, are stored in the System Database. AutoBrief's Presentation Generator gets information needed to fulfill its communicative goals from the System Database.

The output of the Presentation Generator is rendered in a graphical user interface by Visage, the data exploration/visualization environment manager described in Section 2. Direct manipulation user actions on text and graphic objects in the

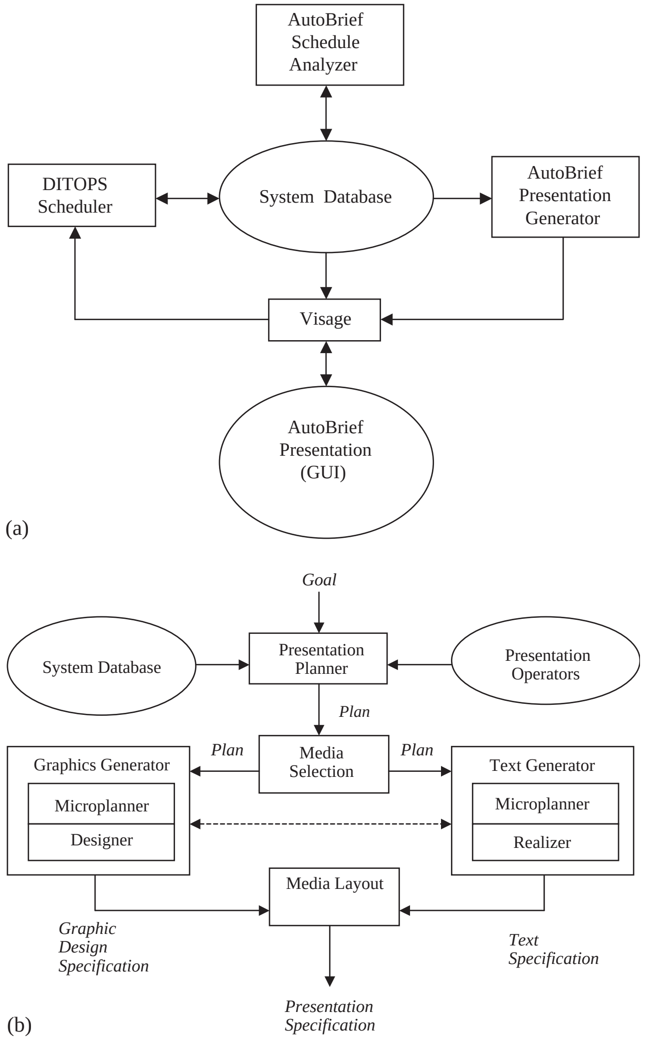

Fig. 3.

presentation (depicted by the arrow in Fig. 3a from Visage to DITOPS) is enabled through AutoBrief's integration with Visage. In particular, elements of AutoBrief's internal representation of the presentation are annotated with System Database identifiers for use by Visage. Visage manages and coordinates presentation objects in the user interface generated by any application in the environment (including AutoBrief) with their related data objects in the System Database.

The flow of information within AutoBrief's Presentation Generator is shown in Fig. 3b. In the first stage of generation, the content and organization of a presentation is planned by the Presentation Planner. The Planner constructs a media-independent Presentation Plan for achieving the presentation's main communicative goal. By *media-independent*, we mean that the plan does not specify whether a goal is to be achieved through text or information graphics. The Plan specifies a hierarchy of partially ordered subgoals, as well as media-independent communicative acts for achieving each of the subgoals. Next, the Media Selection module decides which subgoals are to be achieved in which media, and annotates the Plan accordingly. From an engineering point of view, this approach has several benefits over an approach where media properties are directly encoded in communicative goals and actions. First, it offers greater flexibility. Instead of using a fixed set of graphic designs encoded in plan operators, AutoBrief can harness the power of a general-purpose automated graphic designer embodied in its Graphics Generator. Second, it offers greater modularity. In the AutoBrief architecture, the same representation of communicative goals is used to drive two independent modules, the Text Generator and the Graphics Generator. Third, it simplifies the problem of media coordination, i.e. the problem of coordinating parts of the presentation expressed in different media, by providing a single representation of the presentation's goals.

After Media Selection, each media generator, the Text Generator and the Graphics Generator, decides how to achieve the goals assigned to it. The output of the Text Generator consists of sentences or phrases of English. The output of the Graphics Generator is a design for the presentation's information graphics. The dotted arrow between the Text and Graphics Generators represents information flow for the generation of text (such as captions) that appears in figures with a graphic rather than in the body of the presentation. Next, the outputs of the Text and Graphics Generators are interleaved by the Media Layout Module according to the ordering constraints of the Presentation Plan. The output of the Media Layout Module is then rendered by Visage. We now describe each of the components of the Presentation Generator in more detail.

4.2. Presentation planner

The Presentation Planner is responsible for planning the content and organization of a presentation. The input to the planning process is a top-level communicative goal triggered by some user action. For example, requesting a briefing about the most recent version of the schedule triggers a top-level goal for the user to be aware of a summary of the available resources, schedule requirements (e.g. amounts, types, dates, origins and destinations of different cargo shipments), and possible shortfalls (shipments arriving late) of that schedule. After seeing the briefing created by AutoBrief, the user could request further details about shortfalls mentioned in the briefing; this would trigger a new top-level goal for the user to be aware of the facts underlying AutoBrief's analysis of that problem. The prototype system makes use of a small number of genre-specific top-level goals, which can be classified as follows:

- Summarize information related to a schedule, including its shortfalls.

- Analyse lift (i.e. air transportation) capacity problems that may have contributed to the shortfalls.
- Analyse port capacity problems that may have contributed to the shortfalls.
- Analyse causes of the lift capacity problems.
- Analyse causes of the port capacity problems.

The Presentation Planner uses a general-purpose planner (Young et al., 1994), information stored in the System Database, and a set of presentation plan operators to construct a Presentation Plan to satisfy the top-level goal. AutoBrief provides two sorts of plan operators (both of which are media-independent): genre-specific and genre-independent plan operators. The genre-specific operators encode genre-specific knowledge of the organization and type of information required for transportation logistics briefings. These operators were written based on information gathered from potential users of the prototype system, and are used to decompose the top-level goal and higher-level subgoals of the plan. (The prototype described in Section 3 employs 58 of these operators.) During planning, variables in a plan operator are instantiated with information from the System Database that satisfies the plan operator's constraints. Information in the System Database may include quantitative information from the schedule (e.g. cargo amounts and arrival dates), as well as summary statistics and other information added by AutoBrief's application-specific Schedule Analyser. For example, a presentation goal to analyse a port capacity problem would be refined into a plan presenting information found in the System Database that was the result of the Schedule Analyser's analysis of any port capacity problem associated with the schedule.

The other sort of plan operator, which comprises the leaves of the plan, is the genre-independent plan operator. In the current implementation of the prototype, the only one of this kind is the *Assert* speech act operator;[2] the goal (intended effect) of the act, *Assert p*, is for the audience to be aware of *p*. For example, *p* could be information paraphrasable as *The largest single cargo shipment between Day* 1 *and Day* 3 *arrives on Day* 1. The information content of communicative goals and acts such as *Assert*, denoted by *p* in the preceding example, is expressed in a representation scheme called the *plan content language*. Plan content language expressions in the leaves of a plan are created during refinement of the genre-specific presentation plan operators by instantiating content language expression templates in the plan operators' constraints with data from the System Database. In the next section we describe how the plan content language enables communicative goals to be expressed.

[2] *Assert* is the only speech act type defined by a planning operator in the current implementation. However, some of the leaves of the plans that are generated in this implementation could be viewed as performing other types of speech act, such as *Recommend* and *Warn*. Also, as discussed in Section 4.3.2, an *Assert* act may specify a hierarchical structure of referential and attributive subgoals. However, those goals were not defined using separate plan operators in the current implementation in order to reduce time and space requirements for planning.

4.3. Representation of communicative goals

AutoBrief uses an approach to representing communicative goals, adapted from research on discourse, that is unique in its extension to multimedia generation. AutoBrief's approach involves two different though related aspects of the representation of communicative goals. The plan content language (Section 4.3.1) enables AutoBrief to represent, among other things, the main point of an information graphic. The referential–attributive goal distinction (Section 4.3.2) enables AutoBrief to express information about a discourse entity (in text or graphics) that is needed for achieving non-referential communicative goals. We introduce this section with constructed examples from the domain of logistics in order to illustrate systematic distinctions among communicative goals and their possible realization in graphics. In particular, we will show how different visualizations of the same data can convey different messages. (These figures will be used for illustration in Section 4.3.1 as well.)

First we apply a Gricean analysis to Fig. 4a similar in many ways to the analysis given for Fig. 1a (Section 2.3). The viewer could reason that since the bars for each day have been combined into two stacked bars, one showing the total amount of apples and the other the total amount of bananas, then the designer may be trying to say something about the two totals. Furthermore, since the lengths of the two stacked bars can be compared visually with little effort, the designer may be trying to say something about the relationship between these totals. Although information on the exact amount of fruit shipped each day is available (since the bars are segmented and since the vertical axis has labelled ticks), it is less accessible than the totals for each type of fruit. Thus, it may be playing a supporting role, e.g. to identify the origin of the difference in total amounts of apples and bananas by showing that the daily amounts of apples and bananas are the same for each day but Day 1. Finally, the use of an arrow in the graphic draws attention to the Bananas bar. Thus, the text above the graphic, *More bananas* (*a total of* 3 *tons between Days* 1 *and* 3) *than apples* (*a total of* 2.5 *tons in the same period*) *are shipped, due to the smaller amount of apples shipped on Day* 1, is a plausible interpretation of the designer's communicative goals.

We argue that the graphic in Fig. 4b is less effective than the graphic in Fig. 4a for conveying this message. That is, by an argument similar to the one given for Fig. 1b, given Fig. 4b instead of Fig. 4a the viewer must expend considerably more effort to determine the total number of bananas and the total number of apples and to compare the totals. In fact, a viewer might not even recognize that the designer intended the viewer to be aware of the two totals and the relation between them. A more plausible interpretation of the designer's intention underlying the Fig. 4b graphic (further reinforced by the arrow) is the message in the caption to Fig. 4b.

The graphics in Figs. 4c and d differ from the preceding bar graphs in encoding arrival day on the horizontal axis. Whereas the graphic in Fig. 4a displays the total amount of each fruit, the graphic in Fig. 4c displays the total cargo per day (i.e. combining shipments of apples and bananas for each day). Thus, we argue that the design of Fig. 4c graphic better conveys the message stated in the caption to Fig. 4c than does the design of Fig. 4a graphic, and, conversely for the message stated in

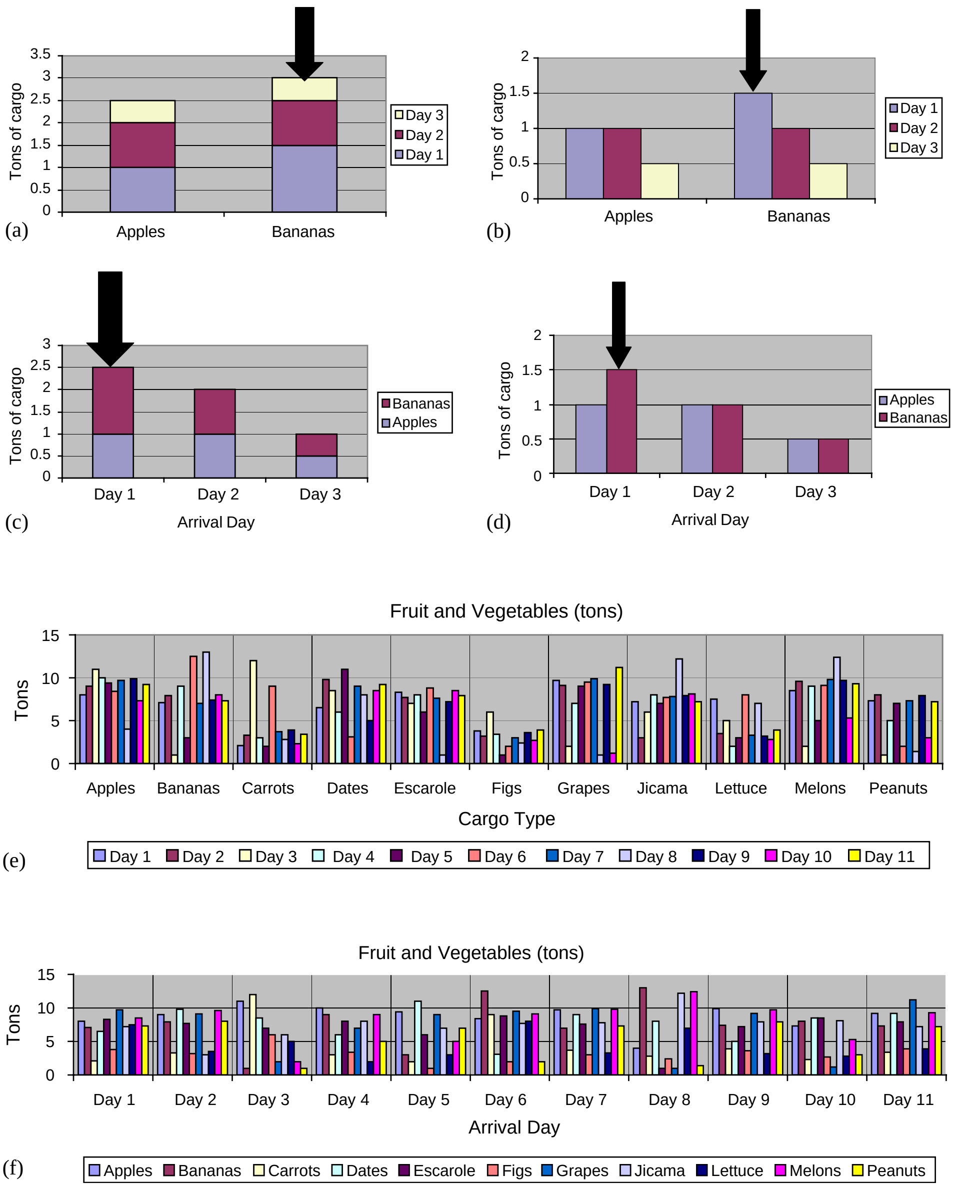

Fig. 4. (a) More bananas (a total of 3 tons between Days 1 and 3) than apples (a total of 2.5 tons in the same period) are shipped, due to the smaller amount of apples shipped on Day 1. (b) The largest single cargo shipment between Days 1 and 3, a 1.5 ton shipment arriving on Day 1, is bananas. (c) The largest total cargo shipment between Days 1 and 3, 2.5 tons of apples and bananas, arrives on Day 1. (d) The largest single cargo shipment between Days 1 and 3, a 1.5 ton shipment of bananas arrives on Day 1. (e) The largest single shipment (13 tons on Day 8) is bananas. (f) The largest single shipment (13 tons of bananas) arrives on Day 8.

Fig. 4a. The argument for why the Fig. 4d graphic would be less effective than the Fig. 4c graphic for conveying the stated message of the caption to Fig. 4c is similar to the argument we gave for Fig. 4b with respect to Fig. 4a.

Finally, although one might argue that the stated messages of the captions to Figs. 4b and 4d could be realized by either graphic, notice that as information for more types of cargo and more days is added to the graphics, as shown in Figs. 4e and f, the need for the different designs becomes more apparent. Because arrival day is encoded less prominently than cargo type in the design for Fig. 4e and b than it is in Fig. 4f and d, the latter design would be preferable, e.g. if the designer's intention were to convey that the largest single shipment of any fruit is on *Day* 8. In other words, the horizontal axis of the graphic in Fig. 4f has been extended to show *Day* 4 through *Day* 11, and achieving this goal would only require the audience to notice that the tallest bar is in a cluster over the label *Day* 8. Conversely, if the designer's intention were to convey that the largest single shipment between *Day* 1 and *Day* 11 was of cargo type bananas in a domain having eleven types of cargo, then the graphic design in Figs. 4e and b, with a cluster of bars (one for each Arrival Day) for each of the cargo types, would be more effective. Since cargo type is more prominent than day in this design, achieving this goal would only require the audience to notice that the tallest bar is in the cluster over the label *Bananas*. In the next section, we will refer back to these figures to illustrate how different graphic designs correspond to distinctions among communicative goals that can be represented in AutoBrief's internal representation scheme, the *plan content language.*

4.3.1. Plan content language

The information content of communicative goals and acts, e.g. the object of an *Assert*, is expressed in a representation scheme called the *plan content language*. The plan content language must be capable of representing anything to be expressed by AutoBrief in text or graphics. More specifically, one requirement for the plan content language is the ability to represent abstract and complex discourse entities (e.g. 25*% of available capacity between Day* 1 *and Day* 3). A *discourse entity* is an entity in the audience's conceptual model evoked by the presentation (Webber, 1983). In addition to domain-specific terms, this requires a suitable representation for natural language quantifiers (e.g. *most*), relations (e.g. *is greater than*), sets (e.g. *shipments arriving between Day* 1 *and Day* 3), and aggregate properties of sets (e.g. *the total amount of cargo arriving between Day* 1 *and Day* 3) for describing discourse entities. A second requirement is to represent complex discourse entities compositionally in order to facilitate generation. By *compositionally* we mean that the representation of a complex expression may be decomposed into meaningful subexpressions. For example, the internal representation of the concept, *the largest single cargo shipment between Day* 1 *and Day* 3, should enable the graphics generator to design a graphic such as the graphic in Fig. 4d showing the set of individual cargo shipments while making the largest of the set salient to the audience.

A third requirement for the plan content language is to represent an important *pragmatic* distinction between *semantically equivalent* expressions, which is discussed shortly. These three requirements led us to model the content language on first-order logic with restricted quantification (RQFOL). RQFOL, which has been used for representing natural language queries involving complex referring expressions (Webber, 1983; Woods, 1983) satisfies these three requirements. The final requirement is that the content language should be media-independent so that it can be used by either of AutoBrief's media generators. This requirement is met by not encoding media-specific information in the plan content language.

To illustrate the third requirement, consider two semantically equivalent representations of situations, shown in (1.1) and (2.1). By *semantically equivalent* we mean that they describe the same situation as represented by data in the System Database. (To make these and other examples more understandable, we shall use English-like paraphrases of content language expressions in most cases.) Although (1.1) and (2.1) describe the same situation, they describe the content of two different communicative goals. Thus, the two representations are not equivalent in the plan content language. The difference is that they evoke two different discourse entities, which are underlined in paraphrases (1.1) and (2.1), and each makes a different predication of that discourse entity. The distinction between (1.1) and (2.1) can be appreciated, for example, by considering Figs. 4b and d. Conceptually speaking, the discourse entities underlined in (1.1) and (2.1) correspond to the vertical bars under the arrows in the graphics in Figs. 4b and d, respectively. A simplified version of the representation of (1.1) and (2.1) in the plan content language is given in italics under each; we describe the syntax in more detail at the end of this section.

(1.1) <u>A 1.5 ton shipment arriving on Day</u> 1 is of cargo type banana.
*Main Predication: d*1 *is banana cargo type*
*d*1*: a* 1.5 *ton shipment arriving on Day* 1
(2.1) <u>A 1.5 ton shipment of banana cargo type</u> arrives on Day 1.
*Main Predication: d*1 *arrives on Day* 1
d1: a 1.5 *ton shipment of banana cargo type*

The ability to express this sort of subtle distinction in the content language is important for successful realization of communicative goals. As we argued previously, the designs of the graphics in Figs. 4b and d are not interchangeable; each design is more effective than the other for conveying the message shown in the caption. The graphic in Fig. 4b could be generated from an underlying plan whose leaves consist of six supporting assertions, whose content is shown in (1.1)–(1.6), and another assertion conveying the main point, whose content is shown in (1.7).[3]

(1.1) <u>A 1.5 ton shipment arriving on Day 1</u> is of cargo type banana.

[3] Currently, none of AutoBrief's plan operators explicitly encode rhetorical relations such as *Evidence*, although the operators were designed to embody rhetorical and argumentation structures. For example, the supporting relation between the set of assertions (1.1)–(1.6) and assertion (1.7) is not represented in the current set of plan operators. See Section 5 for a discussion of why making these relations explicit could be beneficial. Also, note that although (1.7) is in fact a logical implication of the set of assertions (1.1)–(1.6), a broader notion of what constitutes an acceptable argument, e.g. as in (Toulmin, 1969), seems to be needed to characterize this genre.

(1.2) A 1 ton shipment arriving on Day 2 is of cargo type banana.
(1.3) A 0.5 ton shipment arriving on Day 3 is of cargo type banana.
(1.4) A 1 ton shipment arriving on Day 1 is of cargo type apple.
(1.5) A 1 ton shipment arriving on Day 2 is of cargo type apple.
(1.6) A 0.5 ton shipment arriving on Day 1 is of cargo type apple.
(1.7) The largest single cargo shipment between Days 1 and 3 is of cargo type banana.

Similarly, the graphic in Fig. 4d could be generated from an underlying plan whose leaves consist of six supporting assertions, whose content is shown in (2.1)–(2.6), and another assertion conveying the main point, whose content is shown in (2.7).

(2.1) A 1.5 ton shipment of banana cargo type arrives on Day 1.
(2.2) A 1 ton shipment of apple cargo type arrives on Day 1.
(2.3) A 1 ton shipment of banana cargo type arrives on Day 2.
(2.4) A 1 ton shipment of apple cargo type arrives on Day 2.
(2.5) A 0.5 ton shipment of banana cargo type arrives on Day 3.
(2.6) A 0.5 ton shipment of apple cargo type arrives on Day 3.
(2.7) The largest single cargo shipment between Days 1 and 3 arrives on Day 1.

The distinction between the two plans can be appreciated by considering when text or graphics generated from one plan and not from the other is more context-appropriate. For example, (3b) or the graphic in Fig. 4b, whose main point is expressed in text as (3c), are more appropriate than (4b) or the graphic in Fig. 4d, whose main point is expressed in text as (4c), in response to the user request given in (3a). Conversely, (4b), (4c), or the graphic in Fig. 4d are the more appropriate responses to (4a).

(3a) *Of what type of cargo is the largest single shipment between Days* 1 *and* 3?
(3b) *Banana.*
(3c) *The largest single cargo shipment between Days* 1 *and* 3 *is of cargo type banana.*
(3d) *Of what type cargo is a* 1.5 *ton shipment arriving on Day* 1?
(3e) *A* 1.5 *ton shipment arriving on Day* 1 *is of cargo type banana.*
(4a) *On what day between Days* 1 *and* 3 *does the largest single shipment of cargo arrive?*
(4b) *Day* 1.
(4c) *The largest single cargo shipment between Days* 1 *and* 3 *arrives on Day* 1.
(4d) *On what day does a* 1.5 *ton shipment of banana cargo arrive*?
(4e) *A* 1.5 *ton shipment of banana cargo type arrives on Day* 1.

In text, the distinction between the semantically equivalent (1.1) and (2.1) can be expressed by subtle differences in English syntax. For example, (1.1) could be expressed in text as (3e), and (2.1) as (4e). Although (3e) and (4e) are very similar, they are not freely interchangeable in the context of a presentation; (3e) would be more appropriate than (4e) in response to (3d) and (4e) would be more appropriate than (3e) in response to (4d). The difference in context-appropriateness of use of (3e) and (4e) can be described in terms of the theme/rheme distinction of linguistic pragmatics (Vallduvi and Engdahl, 1996). According to this view, (3d) establishes the theme of (3e); the rheme of (3e) is the new information added to the discourse context by (3e), which is paraphrased in (3b). Similarly, (4d) establishes the theme of (4e), and (4b) paraphrases the rheme of (4e). However, while linguistics helps characterize a pragmatic distinction that is represented in our content language, in some cases its implications for graphics generation may be greater than for text generation, as illustrated by comparing the effectiveness of Figs. 4e and f for achieving different communicative goals.

To give another example of the significance of the approach for information graphics generation, the graphic in Fig. 4a could be generated from an underlying plan whose leaves consist of supporting assertions, whose content is shown in (5.1)–(5.8), and another assertion conveying the main point, whose content is shown in (5.9). Similarly, the graphic in Fig. 4c could be generated from a plan whose assertions have the content shown in (6.1)–(6.10).

(5.1) The amount of banana type cargo arriving on Day 1 is 1.5 tons.
(5.2) The amount of banana type cargo arriving on Day 2 is 1 ton.
(5.3) The amount of banana type cargo arriving on Day 3 is 0.5 tons.
(5.4) The amount of apple type cargo arriving on Day 1 is 1 ton.
(5.5) The amount of apple type cargo arriving on Day 2 is 1 ton.
(5.6) The amount of apple type cargo arriving on Day 3 is 0.5 tons.
(5.7) The total amount of banana type cargo arriving on Days 1−3 is 3 tons.
(5.8) The total amount of apple type cargo arriving on Days 1−3 is 2.5 tons.
(5.9) The total amount of banana type cargo arriving on Days 1−3 is greater than the total amount of apple type cargo arriving on Days 1−3.
(6.1) 1.5 tons of banana arrives on Day 1.
(6.2) 1 ton of banana arrives on Day 2.
(6.3) 0.5 tons of banana arrives on Day 3.
(6.4) 1 ton of apple arrives on Day 1.
(6.5) 1 ton of apple arrives on Day 2.
(6.6) 0.5 tons of apple arrives on Day 3.
(6.7) A total of 2.5 tons of apples and bananas arrives on Day 1.
(6.8) A total of 2 tons of apples and bananas arrives on Day 2.
(6.9) A total of 1 ton of apples and bananas arrives on Day 3.
(6.10) The largest total amount of apples and bananas for Days 1−3 arrives on Day 1.

We shall describe the syntax of the plan content language briefly now by considering the representation of (5.9) in more detail. Internally, (5.9) would be represented as a *main predication* and its arguments, as shown in (7). In (7), the main predication is the relation *is greater than*, and its arguments are labelled as $d1$ and $d4$, corresponding to the two discourse entities whose paraphrases are underlined in (5.9), respectively. In addition to the main predication, a compositional description of the two discourse entities is given in (7). That is, the discourse entities $d1$ and $d4$

are represented as composed of other discourse entities, *d*2 and *d*5, respectively; *d*2 and *d*5 are composed of *d*3 and *d*6, respectively.

(7) Main Predication: d1 is greater than d4
d1: the total of d2
d2: the set of weights of d3
d3: all cargo shipments that arrive on Day 1 or Day 2 or Day 3 and whose cargo type is banana
d4: the total of d5
d5: the set of weights of d6
d6: all cargo shipments that arrive on Day 1 or Day 2 or Day 3 and whose cargo type is apple

As this example shows, the first two requirements for the plan content language are met through the use of RQFOL to construct compositional representations of discourse entities in terms of quantifiers, sets, and set aggregation operators. The third requirement is met through RQFOL's distinction between the main predication and the description of discourse entities. (The fourth requirement is met by not encoding media-specfic information in the plan content language.) In summary, AutoBrief's adoption of an RQFOL-style approach enables the representation of the content of communicative goals in a way that is very useful for information graphics generation involving abstract and complex discourse entities. AutoBrief's adoption of RQFOL for information graphic generation is a novel application of the formalism. At the same time, historically originating from linguistics, the representation serves the needs of text generation as well.

4.3.2. Referential versus attributive goals

This section describes a distinction in communicative goals that is captured in AutoBrief's representation scheme and that has important consequences for both text and graphics generation. First, we describe the roots of this distinction in the philosophy of language. As noted by Donnellan (1966), descriptions are used in two ways in language. The content of an *attributive* description directly contributes to the speaker's communicative goals, whereas the only function of a *referential* description is to enable the audience to identify a particular referent. Donnellan noted that the same description can be used on different occasions for referential or attributive goals.

For example, in an AutoBrief presentation the phrase, *the* 34 *ton shipment arriving on Day* 2, could be used either for referential or attributive goals. An example of referential use is as follows. Suppose that the presentation includes the following text: *Storage facilities at Norfolk are adequate for all but one of the shipments arriving between Day* 1 *and Day* 11. *As shown in the graph, the* 34 *ton shipment arriving on Day* 2 *exceeds the maximum daily intake capacity shown by the blue line.* (*You can drag the bar representing that shipment to the DITOPS window and reschedule it.*) Also, suppose that the presentation includes a graphic with a design similar to the design of Fig. 4f with a blue line added to show maximum daily intake capacity. In the text, the description *the* 34 *ton shipment arriving on Day* 2 is used to enable the audience to identify the bar representing that shipment in the graphic so that the audience can perform an interface action using that bar. Note that depending on the design of the graphic, other descriptions such as *the largest shipment* or *shipment id* #127839291020 might be more effective means of enabling the audience to identify the shipment in question.

On the other hand, the same description could be used for an attributive goal, as in the following argument: *Storage facilities at Norfolk are adequate for all but one of the shipments arriving between Day* 1 *and Day* 11. *The* 34 *ton shipment arriving on Day* 2 *exceeds the maximum daily intake capacity. Therefore, we suggest that you allocate two additional refrigeration units to Norfolk on Day* 2. To understand this argument, the audience must recognize the connection between the information given in the description, *the* 34 *ton shipment arriving on Day* 2, and the rest of the argument. Assuming that the reader knows that the maximum daily intake capacity is 30 tons and that each refrigeration unit has a capacity of two tons, the reader may interpret the information presented in the description as a justification for the system's recommendation to allocate two more refrigeration units. However, the user's ability to identify the referent of *the* 34 *ton shipment arriving on Day* 2 is not required for the above argument to be comprehended.

It is important to represent the referential–attributive distinction in the communicative goals of the Presentation Plan. If all describing goals are treated as purely referential (as is the case in other generation systems), then there is nothing to prevent AutoBrief's generators from substituting alternative descriptions since any description that would enable the audience to identify the referent should be good enough. (Here we are using *description* to include descriptions presented in text as well as descriptions presented in graphics.) However, that could lead to a less effective presentation if the description is actually needed to satisfy an attributive goal.

For example, this could arise in text generation when using the heuristic of preferring the shortest unique referential description known to be familiar to the audience (Dale and Reiter, 1995). In such a scenario, the Text Generator might substitute *the largest shipment* or *shipment id* #127839291020 for *the* 34 *ton shipment arriving on Day* 2 in the above argument. Assuming that the audience has no prior information about the shipment of interest then these alternate versions of the argument should be less effective than the previous one. (Even if this argument were accompanied by a graphic from which the audience could deduce the information that the shipment identified as *the largest shipment* or *shipment id* #127839291020 is 34 tons and arrives on Day 2, it would require additional inferences for the audience to comprehend the system's justification for its recommendation.)

The distinction between referential and attributive goals is just as important for graphics generation. If all describing goals are treated as purely referential, then there is no reason why the Graphics Generator would not be free to use alternate encodings denoting the same entity. For example, suppose that the Graphics Generator treats the attributive goal to describe *the* 34 *ton shipment arriving on Day* 2 as a referential goal; then there is nothing to prevent the Graphics Generator from encoding the shipment in question as a bar labelled *shipment id* #127839291020. However, if the goal is to convey the above argument in a graphic, the design of the graphic must include the information given in the description, *the* 34 *ton shipment arriving on Day* 2.

In summary, AutoBrief's use of the referential–attributive distinction ensures that information selected during presentation planning for its genre-specific or rhetorical function, i.e. information to be used in an attributive description, will be preserved through subsequent phases of generation and thus will be expressed in the final presentation.

The attributive/referential distinction is represented in the plan content language simply by the presence or absence, respectively, of a description of a particular discourse entity. (Technically speaking, this indicates whether the subgoal is attributive or referential.) In the latter case, i.e. when no description is included, the discourse entity is specified by its system identifier rather than by a discourse entity variable. The generators use system identifiers to select information about an entity from the system database in order to generate text or graphics that will enable the audience to identify the given entity, i.e. to achieve the referential goal.

For example in (8), the information given about *d*1 is needed to satisfy an attributive goal. However, the information about *d*1 refers to *NFK*, where *NFK* is a discourse entity for which no attributive goal has been specified. Since *NFK* is a system identifier, information in the system database about *NFK* will be used to generate referring expressions such as *Norfolk* or *the port in Virginia* to enable the audience to identify *NFK*. This example also illustrates that it is possible to specify an attributive goal whose subcomponents may be both attributive and referential.

(8) Paraphrase: A cargo shipment to Norfolk arrives on Day 6.
Main predication: d1 arrives on d2
d1: a cargo shipment with destination NFK
NFK (referential goal)
d2: Day 6

To contrast (8) with an example of when the system has an attributive goal to describe Norfolk as the port with the largest capacity, consider the description of the discourse entity *d*3 in (9).

(9) Paraphrase: A cargo shipment to the port with the largest capacity arrives On Day 6.
Main predication: d1 arrives on d2
d1: a cargo shipment with destination d3
d3: the port with capacity d4
d4: the maximum of d5
d5: the set of capacities of all ports
d2: Day 6

4.4. Media selection module

As shown in Fig. 3b, after Presentation Planning the Media Selection Module decides which subgoals of the Presentation Plan are to be assigned to which of the media generators, Text and/or Graphics, and annotates the Plan accordingly. The decision to use graphics is based upon heuristics encoding graphic design knowledge that are used also by the Graphics Generator as described in Kerpedjiev et al. (1998a). These heuristics determine whether a graphic could be designed to realize a group of subgoals. Text is selected for those subgoals that cannot be realized in graphics. In addition, genre-specific heuristics are used to assign some subgoals to text even though they are to be realized in graphics as well. To give an example, the six assertions in (2.1)–(2.6) would be tedious to express in text, but are conveyed effectively in graphics as in Fig. 4d. (They would be assigned to graphics by the heuristic of preferring graphics for multiple homogeneous goals to convey quantitative information.) However, the main point, encoded in (2.7), can be expressed effectively in both media.

4.5. Text generator

The Text Generator transforms subgoals of a Presentation Plan assigned to it by the Media Selection Module into sentences of English. As shown in Fig. 3b, the transformation is performed in two phases, as is common in NLG systems (Reiter and Dale, 2000). In the first phase, the Text Microplanner transforms the goals into sentence specifications describing the content words, argument structure, and major syntactic features of each sentence. While constructing sentence specifications, the Text Microplanner prefers options that allow multiple goals to be achieved in the same sentence. For example, multiple subgoals of the plan may be transformed into a single sentence of English by selection of a verb whose predicate-argument structure and adjuncts enable several subgoals to be lexically *aggregated* (Reiter and Dale, 2000). In addition, subgoals may be aggregated by use of certain syntactic structures such as conjoined noun phrases. The motivation for attempting to aggregate goals is to produce more concise text. For example, without use of text aggregation techniques, (10.1) and (10.2) would be produced instead of (10.3).

(10.1) Three tons of apples arrive on Day 1.
(10.2) Two tons of bananas arrive on Day 1.
(10.3) Three tons of apples and two tons of bananas arrive on Day 1.

In the second phase, Text Realization, a general-purpose sentence generator for English, FUF/SURGE (Elhadad et al., 1997), is used to transform each sentence specification into a sentence. SURGE uses general knowledge of English grammar to supply word endings and closed class words (such as articles) and to order the words of a sentence. (It was necessary for us to add the lexical entries for a set of domain-specific open class words to SURGE's lexicon.) The output of Text Generation, the

Text Specification, is a data structure consisting of sentences of English, the goals of the Presentation Plan to be achieved by each sentence, and information needed to support direct manipulation of parts of the text in the user interface (described in Section 3). The latter information consists of links from syntactic constituents of the sentences to identifiers of corresponding objects in the System Database.

Text to be used within figures such as captions and titles is also generated by the Text Generator. The goals for this text are forwarded from the Graphics Generator (as shown by the dashed arrow pointing right in Fig. 3b) as a result of a process described in the next section. After transforming these goals into phrases of English, the resulting text is sent back to the Graphics Generator (as shown by the dashed arrow pointing left in Fig. 3b) for incorporation into the Graphic Design Specification for the figure. Note that these goals are represented in the same media-independent formalism used in the Presentation Plan, which enables the Text Generator to transform these goals to text using the same process as described above.

4.6. Graphics Generator

The Graphics Generator transforms the subgoals of the Presentation Plan assigned to it by the Media Selection component into designs for graphics in a two-phase process, as shown in Fig. 3b. In the first phase, the Graphics Microplanner subcomponent of the Graphics Generator transforms the subgoals into specifications of conceptual tasks that a graphic must enable the user to perform. A complete description of the complex process of creating the task specification is presented in Kerpedjiev and Roth (2000). In the second phase of graphics generation, the Graphic Designer, a task-based automated graphic design system that is an extended version of SAGE (Kerpedjiev et al., 2000), transforms the conceptual task specifications into graphic design specifications. The goal of this phase is to design graphics that enable the user's conceptual tasks to be performed efficiently, e.g. by enabling the user to perform a comparison visually instead of by mentally computing the relation between two numbers. The rationale is that by being enabled to perform those tasks efficiently, the user will recognize the communicative goals to be achieved by the graphic (as argued in Section 2.3).

We shall describe the Graphics Generator in a little more detail now. The inputs to the Graphics Microplanner are subgoals of the Presentation Plan, which contain declarative expressions in the plan content language. The outputs of the Graphics Microplanner are procedural task specifications. A task specification is composed of conceptual tasks such as *search* for an object or *compare* attributes of some objects, as well as control structures indicating the order in which the user should perform the tasks (e.g. *sequence*, *disjoint*). Heuristics are used to map the main predication of a plan content language expression to a task. For example, the main predication of (5.9), *is greater than*, would be mapped to a *compare* task.

(5.9) The total amount of banana type cargo arriving on Days 1–3 is greater than the total amount of apple type cargo arriving on Days 1–3.

Other heuristics are used to map discourse entities in the plan content language to user tasks to be performed on database entities. For example, in (5.9) the underlined discourse entities each would be mapped to a user task to *compute* the total amount of the database *cargo weight* attribute for two sets of database entities.

An important capability of the Graphics Microplanner is the ability to aggregate subgoals of the Presentation Plan into a single task specification so that the aggregated subgoals can be realized (during the Graphics Realization phase) in a single graphic supporting multiple user tasks. This is analogous to the process of text aggregation performed by the Text Microplanner in order to make generated text more concise. For example, the graphic in Fig. 4a aggregates the assertions whose content is shown in (5.1)–(5.9). (Note that in this case a single graphic can be used to aggregate more communicative goals effectively than a sentence could.)

The input to the second phase of graphic generation (i.e. the Graphic Designer) is a task specification. An important goal of this phase is to design a graphic that enables the user to perform the conceptual tasks specified in the preceding phase as efficiently as possible. For example, the *compare* task formulated during the first phase of graphics generation for conveying (5.9) would be transformed into the specification of a graphic design that enables the audience to compare the entities in question visually, e.g. as in Fig. 4a. The output of the second phase, the Graphic Design Specification, is a data structure containing a set of graphic designs and the goals of the Presentation Plan to be achieved by each graphic. Each Graphic Design Specification includes identifiers of objects in the System Database in order to support direct manipulation of graphic elements in the user interface (described in Section 3). The graphic designs are not actually rendered until after the next phase of presentation generation, Media Layout, at which time they are rendered by Visage.

In the process of designing a graphic, if its communicative goals cannot be achieved through graphic resources alone, then the Graphics Microplanner may send a caption specification to the Text Generator to be transformed into text to be used as a caption for the graphic. The caption specification is represented in the same language as that used in the Presentation Plan. For example, consider the graphic in Fig. 2a. The graphic alone cannot convey all of the information that needs to be expressed. Therefore, the Graphics Generator sends the Text Generator a caption specification with a description of the set of discourse entities shown in the graphic, which can be paraphrased as *Late cargo, type non-PAX*, *with a feasible arrival date* (*FAD*) *from Day* 1 *to Day* 12.[4]

5. Design limitations and future research

In this section, we discuss some limitations of the current design of AutoBrief's Presentation Generator and suggestions for future research.

[4] To explain the domain terminology in this example, cargo is classified as *PAX or non-PAX*; *PAX* refers to personnel to be transported and *non-PAX* refers to all other cargo. A *feasible arrival date* is more or less the arrival date according to the schedule.

5.1. Genre-specific operators

One limitation is a consequence of the design decision to perform content selection via *genre-specific* discourse operators. This approach offers more flexibility than a template-based approach to NLG (Reiter and Dale, 2000). The 58 genre-specific operators used in AutoBrief can be combined in different ways and instantiated differently depending on the top-level presentation goal and the current database state (including data provided by AutoBrief's Schedule Analyser) to create a large set of possible plans. Furthermore, since the plans are media-independent, they may be realized in a variety of different ways in text and graphics. Nevertheless, this approach still constrains what presentations can be generated by any particular version of the system. To compensate for this, AutoBrief is designed to support the user in data exploration activities to allow the user to search for answers to questions that the current version of AutoBrief does not address. Furthermore, it should be straightforward for the system developers to add new operators for this domain as new user requirements are identified. Also, if the new operators contained expressions that could not be lexicalized in the current version of the system, it should be straightforward for the developers to add the necessary lexical entries to the system.[5]

Another problem related to using primarily genre-specific operators is that porting AutoBrief to a new domain would require definition of a new set of genre-specific operators for the domain. It might be possible to mitigate this ease-of-portability problem in the future by also defining a set of discourse operators that are applicable to more than one genre (from here on let us use the term *multi-genre* to mean *applicable to more than one genre*). For example, multi-genre rhetorical operators based on a general theory of discourse coherence, such as rhetorical structure theory (RST) (Mann and Thompson, 1988), could be used together with genre-specific operators to generate coherent text as in (Moore, 1995). However, many of AutoBrief's current genre-specific operators could be viewed as instances of argumentation structures, representing a different level of multi-genre discourse organization than the relations characterized by RST (Reed and Long, 1998). At the time that AutoBrief was designed, the state of computational argumentation theory was not sufficiently developed to serve as a basis for defining multi-genre operators for AutoBrief.

Nevertheless, in theory an explicit internal representation of the structure of a system's arguments could be exploited throughout the multimedia generation process. For example, given an argument structure, a media selection component could use argument-level heuristics such as ''realize the claim in text and graphics'' and ''realize the supporting data in graphics only'' (Green, 1999), instead of or in addition to the assertion-level heuristics currently used by AutoBrief for media selection (Section 4.4). Also, a text generator should consider argument structure in

[5] Although it was beyond the scope of our research, another area for future research would be to develop knowledge acquisition tools to allow users to participate in extending system coverage by defining new plan operators and lexical entries.

selection of discourse cue words such as ‘therefore’. Furthermore, it is plausible that graphic design should reflect argument structure. For example, an argument’s claim may be justified by data and a warrant, and the warrant itself may be justified by other data (Toulmin, 1969). Although the extent to which argumentation structure could be reflected in graphic design is an open question,[6] it is plausible that these two different uses of data should be treated differently in the visual design of a presentation. However, since in AutoBrief’s presentation plans, data provided to support a claim is not distinguished from data provided to support a warrant, the graphics generated by AutoBrief cannot be designed to reflect this distinction. In summary, AutoBrief’s current reliance on genre-specific operators is a constraint on portability as well as on the potential effectiveness of its presentations. Currently one of the authors is investigating multi-genre argumentation-level strategies for use within a system for generating multimedia presentations in the domain of genetic counseling (Green, 2003).

5.2. *Media coordination*

Another limitation of AutoBrief’s current design is a consequence of adopting an architecture in which media-independent presentation planning (Section 4.2) is followed by media selection (Section 4.4), which is followed by media generation, performed in parallel by text and graphics generators (Sections 4.5 and 4.6), ending with media layout. Since the power of planning is harnessed only at the beginning of this process, before graphic design, content selection for captions must be performed heuristically by the graphics generator. However, a more comprehensive, planning-based approach to caption generation such as used in Mittal et al. (1998) requires a set of discourse plan operators for captions and uses the design of a graphic as input to the caption planning process. Thus, to enable a similar approach to caption generation in AutoBrief would require a second stage of planning, for captions, after graphics generation; the caption plans could then be input to the text generator. However, it would be straightforward to add such a caption planning module to AutoBrief’s dataflow, i.e. between graphics generation and media layout.

On the other hand, it would be less straightforward to extend the current system to include references to the generated graphics in the main body of text, i.e. the text currently generated from the main presentation plan operators as opposed to text that could be generated from stand-alone caption plan operators. For example, suppose that the media-independent presentation planning phase creates a plan that includes an assertion that could be realized as *Fifty tons of cargo will be late on Day* 5. Suppose that next the media selection module decides for this assertion to be realized in both text and graphics. Suppose that next a graphic is designed for the presentation with a design similar to the design of Fig. 4d and in it a bar representing

[6] As far as we know, there has been little empirical, i.e. corpus-based, work relating graphic design techniques to rhetoric and argumentation. Mittal (1997) identified a taxonomy of graphical techniques used for rhetorical effect. An analysis of a corpus of technical arguments expressed in text and graphics showed that argument structure may be reflected in information graphic design (Green, 2001).

the late cargo on Day 5 is highlighted in red. Finally, suppose that next the media layout module decides that this graphic will be presented on the screen below the main body of text. The question is how to modify the current architecture of AutoBrief so that the following sentence could be generated for inclusion in the main body of text: *Fifty tons of cargo* (*see the red bar in the figure below*) *will be late on Day* 5.

Previous multimedia systems that have addressed similar problems of media coordination have followed either the approach of allowing plan operators to include media-specific information about realization in graphics or of modifying the presentation plan after graphics generation. In order not to lose the flexibility of media-independent planning, AutoBrief could adopt the second approach. However, two related issues remain to be addressed: (1) On what factors should a media coordination component base its decisions to add content referring to the graphics to the original plan? (2) On what factors should the text generator base its decisions on where in the text to place this new graphics-related content? To give an example of the first issue, a simple content selection heuristic such as "include content referring to a graphic whenever a concept in the text is depicted in the information graphic" could result in so many cross-references that they would be more distracting than helpful. As for the second issue, the text generator must decide where in the discourse structure to direct the audience's attention to an accompanying graphic, e.g. at the beginning or end of a paragraph discussing data shown in the graphic, or at the first reference to a particular feature of the graphic (which might be somewhere in the middle of a paragraph).

Ideally, the process of media coordination should be informed by a theory of multimedia cognitive processing, which is currently an open area of research (e.g. Mayer, 2001), or by empirical studies addressing specific questions related to this problem, e.g. Green (2002). Future advances in research on multimedia cognition could be applied to improve AutoBrief's media selection and media layout modules as well. Currently, the media selection component uses both domain-specific heuristics and domain-independent graphics knowledge to perform media selection. However, it has no corresponding set of domain-independent principles for determining when to use text or for choosing between the two media. Also, AutoBrief's simple approach to media layout does not address issues of document structure and layout described by Power et al. (2003).

5.3. Reasoning about output

AutoBrief's Text Microplanner's (Section 4.5) current compositional approach to handling content language expressions involving the complex quantification typical in this domain is designed to convey the semantics of an underlying content language expression precisely. Unfortunately, this may result in more verbose noun phrases. A text generator can produce more concise output by reasoning about what inferences the audience can derive from its output (Stone and Doran, 1996). Although the input to each of AutoBrief's text and graphics generators is encoded in a logic-based plan content language, neither generator adopts a logical inference based approach.

However, because of its plan content language AutoBrief should be able to adopt a logical inference based approach in the future to improve conciseness of the text without sacrificing precision.

While no logic-based approach ever has been developed for generation of information graphics as complex as those that can be created by AutoBrief, an important benefit in theory would be that the graphics generator's design process could use a theorem prover to infer if a proposed design would have unintended effects. However, an alternate approach for inferring unintended effects of graphics is suggested by recent research on intelligent recognition of the communicative goals of information graphics to assist computer users who have visual impairments (Elzer et al., 2003a, b). Adopting AutoBrief's generation model as an idealized model of the human designer of an information graphic, the proposed recognition system for the visually impaired works in the reverse of AutoBrief, mapping visual components of the graphic such as vertical bars (provided by an intelligent vision system) to tasks to plausible communicative goals. Such a recognition system could be employed by AutoBrief to perform checks for unintended effects of a graphic and would not require a change in AutoBrief's current approach to graphics generation.

5.4. Evaluation

Lastly, to transform AutoBrief from a research prototype to a deployed system, considerable testing would be required. Performance testing is required to determine if any features of the current system design or implementation prevent the system from attaining acceptable levels. For example, in some cases we have observed that the FUF/SURGE generator used for syntactic realization takes several seconds to realize a paragraph of text. Also, ablation testing is needed to systematically identify what features of generated text and graphics impact usability (Carenini and Moore, 2001). For example, ablation testing could be used as a way of testing alternative media selection strategies.

6. Conclusions

This article has presented the design and design rationale of AutoBrief, an intelligent multimedia presentation system that generates presentations in text and information graphics to achieve communicative goals. In addition, the multimedia presentation facilitates data exploration through its design of complex information graphics and by supporting direct manipulation of textual and graphical elements in the generated presentation. A prototype system demonstrating the feasibility of the design has been implemented in the domain of transportation logistics scheduling. Although no formal evaluations have been performed, the prototype has been demonstrated to its potential users, transportation logistics analysts, who reacted very favorably. In related evaluations of another system, direct manipulation of graphical elements similar to that supported in AutoBrief and information visualizations produced using an intelligent graphic designer similar to AutoBrief's were viewed favorably as well.

The focus of our research and of this paper is AutoBrief's Presentation Generator. AutoBrief's generation process begins with the creation of a genre-specific media-independent plan to achieve communicative goals. AutoBrief uses a two-fold approach to representing communicative goals, originally motivated by research on discourse, that is unique in its extension to multimedia generation. First, AutoBrief's plan content language supports information graphics generation for complex discourse entities, i.e. entities described compositionally in terms of quantification, sets, and aggregate properties of sets. Second, the distinction between attributive and referential goals in the plan ensures that information selected for its genre-specific or rhetorical function will be preserved through subsequent phases of generation.

After the media-independent presentation plan has been created, different parts of the plan are designated for realization in text and/or graphics by a media selection component. AutoBrief's graphics generator transforms the goals of the plan selected for graphic realization into a specification of tasks that the graphic should enable the user to perform; then, taking properties of the data and human visual capabilities into consideration, a graphic is designed to enable the user to perform those tasks efficiently. The rationale is that by being enabled to perform those tasks efficiently, the user will recognize the communicative goals designated for realization in graphics. In parallel with and independently of graphics generation, the text generator produces text to achieve the communicative goals designated for realization in text. The text and graphics specifications created by the media generators are combined by the media layout component using the partial ordering of acts specified in the presentation plan.

AutoBrief's Presentation Generator employs techniques that are for the most part application-independent. Porting the Presentation Generator to a new domain would require definition of new plan operators, some new plan content language terms for representation of communicative goals, and some new lexical entries. However, AutoBrief's approach to representing communicative goals in plans is application-independent (and vocabulary-independent). While some heuristics used in the media selection module are application-specific, most are based on graphic design knowledge. AutoBrief's text and graphics generators use general linguistic and graphics generation techniques, respectively, that should be applicable in other domains. In particular, the Graphics Designer's input is an application-independent, task-based specification. Finally, the techniques used to support direct manipulation of textual and graphical elements of the presentation are application-independent. As an aid to future researchers, we have provided a detailed discussion of the limitations of the current design and suggestions for addressing them in the future.

In summary, AutoBrief's primary research contributions include (1) a design enabling a new human–computer interaction style in which intelligent multimedia presentation objects (textual or graphic) can be used by the audience in direct manipulation operations for data exploration, (2) an application-independent approach to multimedia generation based on the representation of communicative goals suitable for both generation of text and of complex information graphics, and

(3) an application-independent approach to intelligent graphic design based upon communicative goals.

Acknowledgements

The design and implementation of AutoBrief was conducted under the sponsorhip of ARPA Grant DAA-1593K0005 while N. Green, S. Kerpedjiev, J. Mattis, and S. Roth were at the School of Computer Science at Carnegie Mellon University, and while G. Carenini and J. Moore were at the Intelligent Systems Program at the University of Pittsburgh. This paper was written and revised with the support of the authors' current employers. Also, we thank the reviewers of this journal for many helpful suggestions.

References

André, E., Rist, T., 1994. Generating coherent presentations employing textual and visual material. Artificial Intelligence Review 9, 147–165.

André, E., Rist, T., Mueller, J., 1999. Employing AI methods to control the behavior of animated interface agents. Applied AI 13, 4–5, 415–448.

Bertin, J., 1983. Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, Madison, WI.

Beshers, C., Feiner, S., 1993. AutoVisual: rule-based design of interactive multivariate visualizations. IEEE Computer Graphics and Applications, 41–49.

Card, S.K., Mackinlay, J., Shneiderman, B. (Eds.), 1999. Readings in Information Visualization: Using Vision to Think. Morgan-Kaufmann Publishers, Inc., Los Altos, CA (Chapter 1).

Carenini, G., Moore, D.J., 2001. An empirical study of the influence of user tailoring on evaluative argument effectiveness, Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI 2001).

Casner, S.M., 1991. A task-analytic approach to the automated design of information graphic presentations. ACM Transactions on Graphics 10 (2), 111–151.

Cassell, J., 2000. More than just another pretty face: embodied conversational interface agents. Communications of the ACM 43 (4), 70–78.

Chuah, M.C., Roth, S.F., 2003. Visualizing common ground. In: Proceedings of the Seventh International Conference on Information Visualization (IV'03), London, July, 16–18, 2003.

Cleveland, W.S., 1994. The Elements of Graphing Data Revised Edition. Hobart Press, Summit, NJ.

Cleveland, W.S., McGill, R., 1985. Graphical perception and graphical methods for analyzing scientific data. Science 229, 828–833.

Dalal, M., Feiner, S.McKeown, K., Pan, S., Zhou, M., Hollerer, T., Shaw, J., Feng, Y., Fromer, J., 1996. Negotiation for automated generation of temporal multimedia presentations. In: Proceedings of ACM Multimedia, Boston, MA, November 1996, pp. 55–64.

Dale, R., Reiter, E., 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. Cognitive Science 18, 233–263.

Donnellan, K., 1966. Reference and definite descriptions. Philosophy Review LXXV, 281–304.

Elhadad, M., McKeown, K., Robin, J., 1997. Floating constraints in lexical choice. Computational Linguistics 23 (2), 195–240.

Elzer, S., Green, N., Carberry, S., 2003a. Exploiting cognitive psychology research for recognizing intention in information graphics. Proceedings of the Annual Conference of the Cognitive Science Society.

Elzer, S., Green, N., Carberry, S., McCoy, K., 2003b. Extending plan inference techniques to recognize intentions in information graphics. Proceedings of User Modeling 2003.

Fasciano, M., Lapalme, G., 1999. Intentions in the coordinated generation of graphics and text from tabular data. Knowledge and Information Systems 1999.

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

Goldstein, J., Roth, S.F., Kolojejchick, J., Mattis, J., 1994. A framework for knowledge-based, interactive data exploration. Journal of Visual Languages and Computing 5, 339–363.

Green, N., 1999. Some layout issues for multimedia argument generation. In: Proceedings of the AAAI Fall Symposium on Using Layout for the Generation, Understanding or Retrieval of Documents, November 5–7, 1999. AAAI Technical Report, FS-99, pp. 47–51.

Green, N., 2001. An empirical study of multimedia argumentation. Proceedings of the International Conference on Computational Systems, Workshop on Computational Models of Natural Language Arguments, San Francisco, CA, May 2001. Springer Lecture Notes in Computer Science, Vol. 2073. Springer, Berlin, pp. 1009–1018.

Green, N., 2002. An experiment to evaluate the effectiveness of cross-media cues in computer media. Annual Meeting of Assoc. for Computational Linguistics Special Interest Group on Discourse and Dialogue (SIGDIAL 2002), Philadelphia, PA.

Green, N., 2003. Towards an Empirical Model of Argumentation in Medical Genetics. Proceedings of CMNA 2003 (IJCAI-03 Workshop on Computer Models of Natural Argumentation) Acapulco.

Green, N., Kerpedjiev, S., Roth, S., Carenini, G., Moore, J., 1998a. Generating visual arguments: a media-independent approach. In: AAAI98 Workshop on Representations for Multi-modal Human–Computer Interaction. Madison, Wisconsin, July 26–27, 1998. Technical Report WS-98-09.

Green, N., Carenini, G., Moore, J., 1998b. A principled representation of attributive descriptions for integrated text and information graphics presentations. In: Proceedings of the Ninth International Workshop on Natural Language Generation, Niagara-on-the-Lake, Ont., Canada, August 5–7, 1998.

Green, N., Carenini, G., Kerpedjiev, S., Roth, S., Moore, J., 1998c. 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).

Grice, H.P., 1957. Meaning. Philosophical Review 67, 377–388.

Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J. (Eds.), Syntax and Semantics, Speech Acts, Vol. 3. Academic Press, New York.

Iverson, G.R., Gurgen, M., 1997. Statistics: The Conceptual Approach. Springer, New York.

Kerpedjiev, S., Roth, S.F., 2000. Mapping communicative goals into conceptual tasks to generate graphics in discourse. In: Proceedings of Intelligent User Interfaces (IUI2000), New Orleans, pp. 60–67.

Kerpedjiev, S., Carenini, G., Green, N., Moore, J., Roth, S., 1998a. Saying it in graphics: from intentions to visualizations. In: Proceedings of the IEEE Symposium on Information Visualization Research Triangle Park, NC, (InfoVis '98), pp. 97–101.

Kerpedjiev, S., Carenini, G., Roth, S.F., Moore, J.D., 1998b. AutoBrief: a multimedia presentation system for assisting data analysis. Computer Standards and Interfaces 18, 583–593.

Kerpedjiev, S., Roth, S., Mattis, J., 2000. Functional unification approach to automated visualization design. In: AAAI Spring Symposium on Smart Graphics, Stanford, pp. 101–108.

Kolojejchick, J.A., Roth, S.F., Lucas, P., 1997. Information appliances and tools in Visage. IEEE Computer Graphics and Applications 17 (4), 32–41.

Kosslyn, S.M., 1989. Understanding charts and graphs. Applied Cognitive Psychology 3, 185–226.

Levinson, S., 1983. Pragmatics. Cambridge University Press, Cambridge, UK.

Mackinlay, J., 1986. Automating the design of graphical presentations of relational information. ACM Transactions on Graphics 5 (2), 110–141.

Mann, W.C., Thompson, S.A., 1988. Rhetorical structure theory: toward a functional theory of text organization. Text 8 (3), 243–281.

Marks, J., Reiter, E., 1990. Avoiding unwanted conversational implicatures in text and graphics. In: Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-1990), pp. 450–456.

Maybury, M., 1991. Planning multimedia explanations using communicative acts. In: Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 61–66.

Mayer, R.E., 2001. Multimedia Learning. Cambridge University Press, Cambridge, UK.

Mittal, V., 1997. Visual prompts and graphical design: a framework for exploring the design space of 2-D charts and graphs. In: Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-1997), pp. 57–63.

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

Moore, D., 1997. Statistics: Concepts and Controversies. W.H. Freeman and Company, New York.

Moore, J.D., 1995. Participating in Explanatory Dialogues. MIT Press, Cambridge, MA.

Oberlander, J., 1996. Grice for graphics: pragmatic implicature in network diagrams. Information Design Journal 8, 163–179.

Pinker, S.A., 1990. A theory of graph comprehension. In: Freedle, R. (Ed.), Artificial Intelligence and the Future of Testing. Laurence Erlbaum Associates, Hillsdale, NJ, pp. 73–126.

Reed, C.A., Long, D.P., 1998. Generating the structure of argument. Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL98), pp. 1091–1097.

Reiter, E., Dale, R., 2000. Building Natural Language Generation Systems. Cambridge University Press, Cambridge, UK.

Rickel, J., Johnson, L., 2000. Task-oriented collaboration with embodied agents in virtual worlds. In: J. Cassell, et al. (Eds.), Embodied Conversational Agents. MIT Press, Cambridge, MA, pp. 95–122.

Roth, S.F., Hefley, W.E., 1993. Intelligent multimedia presentation systems: research and principles. In: Maybury, M. (Ed.), Intelligent Multimedia Interfaces. Morgan Kaufmann, San Francisco, CA, pp. 13–58.

Roth, S.F., Mattis, J., 1990. Data characterization for intelligent graphics presentation. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 193–200.

Roth, S.F., Kolojejchick, J., Mattis, J., Goldstein, J., 1994. Interactive graphic design using automatic presentation knowledge. In: Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI '94), Boston, MA, pp. 112–117.

Roth, S.F., Chuah, M.C., Kerpedjiev, S., Kolojejchick, J.A., Lucas, P., 1997. Towards an information visualization workspace: combining multiple means of expression. Human–Computer Interaction Journal 12 (1 & 2), 131–185.

Seligman, D., Feiner, S., 1991. Automated generation of intent-based 3D illustrations. Computer Graphics 25 (4), 123–132.

Shneiderman, B., 1994. Dynamic queries for visual information seeking. Technical Report CAR-TR-655, University of Maryland.

Shneiderman, B., 1998. Designing the User Interface: Strategies for Effective Human–Computer Interaction, 3rd Edition. Addison-Wesley, Reading, MA.

Smith, S.F., Lassila, O., Becker, M., 1996. Configurable, mixed-initiative systems for planning and scheduling. In: Tate, A. (Ed.), Advanced Planning Technology. AAAI Press, Menlo Park, CA.

Stone, B.A., Lester, J.C., 1996. Dynamically sequencing an animated pedagogical agent. In: Proceedings of the 13th National Conference on Artificial Intelligence, pp. 424–431.

Stone, M., Doran, C., 1996. Paying Heed to Collocations. Proceedings of INLG 1996, pp. 91–100.

Toulmin, S., 1969. The Uses of Argument. Cambridge University Press, Cambridge, UK.

Tufte, E., 1983. The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT.

Vallduvi, E., Engdahl, E., 1996. The linguistic realization of information packaging. Linguistics 34 (3), 459–519.

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

Webber, B.L., 1983. So what can we talk about now? In: Grosz, B., Jones, K.S., Webber, B.L. (Eds.), Readings in Natural Language Processing. Morgan Kaufmann, Los Altos, CA.

Woods, W., 1983. Semantics and quantification in natural language question answering. In: Grosz, B., Jones, K.S., Webber, B.L. (Eds.), Readings in Natural Language Processing. Morgan Kaufmann, Los Altos, CA.

Young, M.R., Moore, J.D., Pollack, M.E., 1994. Towards a principled representation of discourse plans. In: Proceedings of the 16th Conference of the Cognitive Science Society, pp. 946–951.

Zhou, M., Feiner, S., 1997. Top-down hierarchical planning of coherent visual discourse. In: Proceedings of the Intelligent User Interfaces (IUI-97), pp. 129–136.