A Media-Independent Content Language for Integrated Text and Graphics Generation

Nancy Green*, Giuseppe Carenini, Stephan Kerpedjiev*, Steven Roth*, Johanna Moore****

*Carnegie Mellon University, Pittsburgh, PA 15213 USA

**University of Pittsburgh, Pittsburgh, PA 15260 USA

Abstract

This paper describes a media-independent knowledge representation scheme, or *content language*, for describing the content of communicative goals and actions. The language is used within an intelligent system for automatically generating integrated text and information graphics presentations about complex, quantitative information. The language is designed to satisfy four requirements: to represent information about complex quantitative relations and aggregate properties; compositionality; to represent certain pragmatic distinctions needed for satisfying communicative goals; and to be usable as input by the media-specific generators in our system.

1 Introduction

This paper describes a media-independent knowledge representation scheme, or *content language*, for describing the content of communicative goals and actions. The language is used within an intelligent system for automatically generating integrated text and information graphics[1] presentations about complex, quantitative information. The goal of the current implementation of the system is to produce analyses and summarizations of the quantitative data output by a transportation scheduling program.

In our approach [Kerpedjiev et al.1997a, Kerpedjiev et al.1997b, Green et al.1998, Kerpedjiev et al.1998], the content and organization of a presentation is first planned at a media-independent level using a hierarchical planner [Young1994]. In this way, a high-level presentation goal, such as to assist the user to evaluate a transportation schedule created by the scheduling program, is ultimately decomposed into media-independent subgoals, whose content is represented in the content language. The content language also is used to represent the content of the media-independent communicative acts, e.g., *Assert* and *Recommend*, selected by the planner to satisfy these subgoals.[2] Content language expressions are constructed by the plan constraint functions of the presentation plan operators during planning.

The content language in the presentation plan is used by the system's two media-specific generators, one for text and one for information graphics. A media allocation component decides which parts of the plan shall be realized by each generator. The text generator transforms its assigned parts to sentence specifications, for realization by a general-purpose sentence generator (SURGE) [Elhadad and Robin1996]. The graphics generator transforms its assigned parts of the plan to a sequence of user tasks which a graphic must support in order to satisfy the presentation goals. The tasks are then input to a graphic design system (SAGE)

This work was supported by DARPA contract number DAA-1593K0005.

[1] e.g., charts, tables, maps, rather than pictorial forms of representation.

[2] In other words, the content language describes that which is to be asserted, recommended, believed, etc., rather than the types of communicative acts to be performed or propositional attitudes which the acts are intended to achieve.

[Roth and Mattis1990, Roth et al.1994] which automatically designs and realizes a graphic supporting the tasks.

One of the requirements for our content language is the ability to represent complex descriptions of quantitative database attributes, such as *total port capacity of all ports* and *90% of the total weight of the cargo arriving by day 25.* In addition to application-specific concepts such as *port capacity*, such descriptions involve the specification of application-independent quantitative relations (e.g., *90% of ...*), aggregate properties of sets (e.g., *total ... of all ...*), and time-dependent relations (e.g., *increase from ... to ... during the interval ...*). Thus, we would like for the language to be able to express a wide range of quantitative and temporal relations and aggregate properties, rather than just those required for the current domain of transportation scheduling.

Another requirement is for the content language to represent these descriptions compositionally. A compositional representation should facilitate the work of the text and graphics generators, as well as media coordination.

A third requirement for the content language is the ability to represent subtle differences in communicative intention with respect to the same data. To give an example in the domain which will be used for illustration in the rest of the paper, the same data[3] could underly either the assertion that *Three newspapers that are circulated in Pittsburgh carry only national news* or the assertion that *Three newspapers that carry only national news are circulated in Pittsburgh.* However, while conveying the same facts about the three newspapers, the two assertions are not interchangeable. The first assertion would be more effective than the second in an argument such as

> Be careful which newspaper you read to find out what is going on locally. The Post-Gazette covers both national and local news, but *three newspapers that are circulated in Pittsburgh carry only national news.*

while the second would be more effective than the first in

> Pittsburghers are interested in national affairs. In fact, *three newspapers that carry only national news are circulated in Pittsburgh.*

As will be shown later in the paper, the content language enables related assertions such as these to be differentiated.

A final requirement is for the representation scheme to be media-independent in order to provide a common input language for the media-specific generators. We assume that such a common language will facilitate the difficult problem of media coordination. On the other hand, the language must satisfy the needs of both the text and information graphics generators.

In the rest of the paper, first we describe the content language, focusing on aspects of the content language which are applicable to other domains. Next, we illustrate how subtle variations in communicative intention can be represented in the content language, and give examples of how they can be expressed in text and information graphics. Finally, we describe some related work.

2 Content Language

In order to ensure that the language would be applicable to a variety of quantitative domains, we first performed a corpus analysis, the results of which are summarized in the next section. Then we describe the syntax we adopted to satisfy the requirements given in the introduction.

2.1 Corpus Analysis

We have collected samples of presentations with integrated natural language and graphics in order to describe and analyze the vocabulary and

[3] All data used in the paper is fictitious. However, many of the examples were inspired by a naturally occuring example about the numbers of readers of newspapers read in Pittsburgh. We have selected this domain for illustration because it requires minimal background knowledge.

structure of such presentations. To ensure generality, the corpus includes presentations from different disciplines (Economics and Medicine) and intended for different audiences.[4] It also includes samples from collections of presentations compiled by others, such as [Tufte1983, Tufte1990, Tufte1997, Kosslyn1994], and prescriptive examples found in books on how to design effective presentations [Zelazny1996, Kosslyn1994].

The analysis of this corpus contributed directly to the development of a vocabulary for the content language. To describe the content of the presentations in the corpus, we distinguish three different sets of predicates with associated modifiers, as follows:[5]

- **Comparison Predicates**: [*much*] Greater, Lower | Highest, Lowest | [*very*] Far-from, Close-to | [*almost* | *exactly*] Equal, *n*-Times.Comparison Predicates apply to any quantitative attribute of individuals or sets, e.g., *On this measure Central Europe's stock-markets are still puny compared with those of fast-growing Asian countries.*
- **Global Predicates**: [*widely* | *slightly*] Vary [*:from :to*], Constant.
  Global Predicates apply to quantitative attributes of sets, e.g., *Sales representative performance is uneven.*
- **Trend Predicates**: Remain-constant | [*considerably* | *slightly*] Increase, Decrease [*:from :to*] | Drop, Fall, Rise [*:from :to*] | Reach-a-Plateau | Fluctuate.
  Trend Predicates apply only to time series (a set of data ordered chronologically), e.g., *Production of television sets in Russia fell from* 4.5*m units in 1991 to fewer than* 1*m in 1995.*

2.2 Syntax

The first three requirements described in the Introduction (representing quantitative and temporal relations and aggregate properties, compositionality, and representing certain pragmatic distinctions) led us to make use of a first-order logic with restricted quantification (RQFOL), which has been used for representing the meaning of natural language queries involving complex referring expressions [Woods1983, Webber1983]. The features of RQFOL most useful for our purposes are (i) that it permits pragmatic distinctions to be made among expressions which are semantically equivalent, and (ii) that it supports the compositional specification of complex descriptions of discourse entities [Webber1983].

A pragmatic distinction supported in RQFOL and our content language is the distinction between the main predication of an expression and information to be conveyed about the objects of the main predication. For example, although (1a) and (1b) are semantically equivalent with (1c), they are not interchangeable in their effectiveness for achieving different communicative intentions (as was demonstrated in the Introduction.) In (1a) the main predication is about news coverage, whereas in (1b) it is about newspaper circulation.

(1a) Three newspapers that are circulated in Pittsburgh carry only national news.
(1b) Three newspapers that carry only national news are circulated in Pittsburgh.
(1c) There is a set of three newspapers such that for every newspaper in the set, it is circulated in Pittsburgh and carries only national news.

To represent this distinction in the content language, a communicative act has the form, *(Act Proposition Referents)*, where *Act* specifies the type of action (such as *Assert*), *Proposition* is a quantifier-free FOL formula describing the main predication, and *Referents* is a list describing the arguments of the main predication. (It is assumed that the agent performing a communicative action is the system, and that the audience is the user.) For example, (1a) and (1b) can be analyzed as realizing the assertions (2a) and (2b), respectively. In (2a), the main predication is *(has-coverage ?d1 National-only)*;

[4] Economics: *The Economist* (March-August 1996). Medicine: *UC Berkeley Wellness Letter* (June 1993 and September 1996), *Scientific American* (September 1996), *New England Journal of Medicine* (April-August 1996).

[5] Square brackets indicate optionality, the bar exclusive-or, and commas separate variants with differing orientations.

the variable ?$d1$ is further described as *three newspapers that are circulated in Pittsburgh.*[6] In (2b), the main predication is *(has-circulation ?d1 Pittsburgh)*; the variable ?$d1$ is further described as *three newspapers whose coverage is national news only.*

(2a) (Assert (has-coverage ?d1 National-only)
((?d1 (for (cardinal 3) ?x newspaper
(has-circulation ?x Pittsburgh)))))
(2b) (Assert (has-circulation ?d1 Pittsburgh)
((?d1 (for (cardinal 3) ?x newspaper
(has-coverage ?x National-only)))))

In general, each element of the Referents list has the form *(term description)*, where *term* is a variable or a database object identifier; and *term* denotes a discourse entity. If provided, *description* specifies information about *term* that is required to achieve the goal(s) of the communicative act, as opposed to information whose only function is to enable the audience to identify the entity. Only descriptions with an attributive function are specified in the presentation plan. Referential descriptions, whose function is only to enable the audience to identify an entity, are constructed by the media-specific generators. (For information about the different roles of attributive and referential descriptions in our system, see [Green et al.1998].) In general, *description* is of the form *(for quantifier variable class restriction)*. (In (2a) and (2b), *quantifier* is the cardinal 3, the class is *newspaper*, and the restriction is *(has-circulation ?x Pittsburgh)* and *(has-coverage ?x National-only)*, respectively.)

Complex descriptions can easily be expressed in a compositional manner in the content language. For example, (3a) is a possible realization in text of the assertion given in (3b). (A graphic realizing (3b) is shown in (3c) of Figure 1.) In (3b), the main predication, *(gt ?d1 ?d2)*, is that ?d1 is greater than ?$d2$. ?$d1$ is to be described as the unique integer ?x such that ?x is the number of readers of $\$PPG$. ($\$PPG$ is a database object denoting the Post-Gazette.) ?$d2$ is described as the unique integer ?x such that ?x is the total of ?$d3$; ?$d3$ is described as the unique set of integers ?y such that ?y is the number of readers of ?$d4$; and ?$d4$ is described as the elements of the set ($\$WSJ$, $\$NYT$, and $\$USA$), (whose elements are database objects denoting the Wall Street Journal, the New York Times, and USA Today, respectively).

(3a) The number of readers of the Post-Gazette is greater than the number of Pittsburgh readers of the New York Times, the Wall Street Journal, and USA Today combined.
(3b) (Assert (gt ?d1 ?d2)
((?d1 (for the ?x integer
(has-number-of-readers $PPG ?x)))
($PPG ())
(?d2 (for the ?x integer (total ?x ?d3)))
(?d3 (for the ?w set (lambda ?y integer
(has-number-of-Pitts-readers ?d4 ?y)))
(?d4 (for all ?z newspaper
(in-set ?z ($WSJ $NYT $USA))))
($WSJ ()) ($NYT ()) ($USA ())))

3 Examples

In this section we illustrate how different communicative intentions about the same data can be represented in the content language, and how these intentions can be expressed in text and information graphics. One goal of this exercise is to illustrate what distinctions *can* be expressed graphically, but not what information *should* be expressed in graphics. (The problem of deciding which media to use, *media allocation*, is beyond the scope of this paper.) Thus, the examples of graphics are minimal in the sense that they have been designed to convey the information to be asserted and as little as possible other information. However, in some cases it is not possible *not* to convey *more* in graphics than was intended.

For example in (3c) in Figure 1, which realizes (3b), the graphic also conveys information about relative numbers of readers of each of the newspapers, e.g., that the Post-Gazette has about one-third more than the sum of the others, and that the others have about the same number of readers each. Note that although it is not the communicative intention in (3b)

[6] By convention, symbols prefixed with ? are variables, and symbols prefixed with $ are database identifiers.

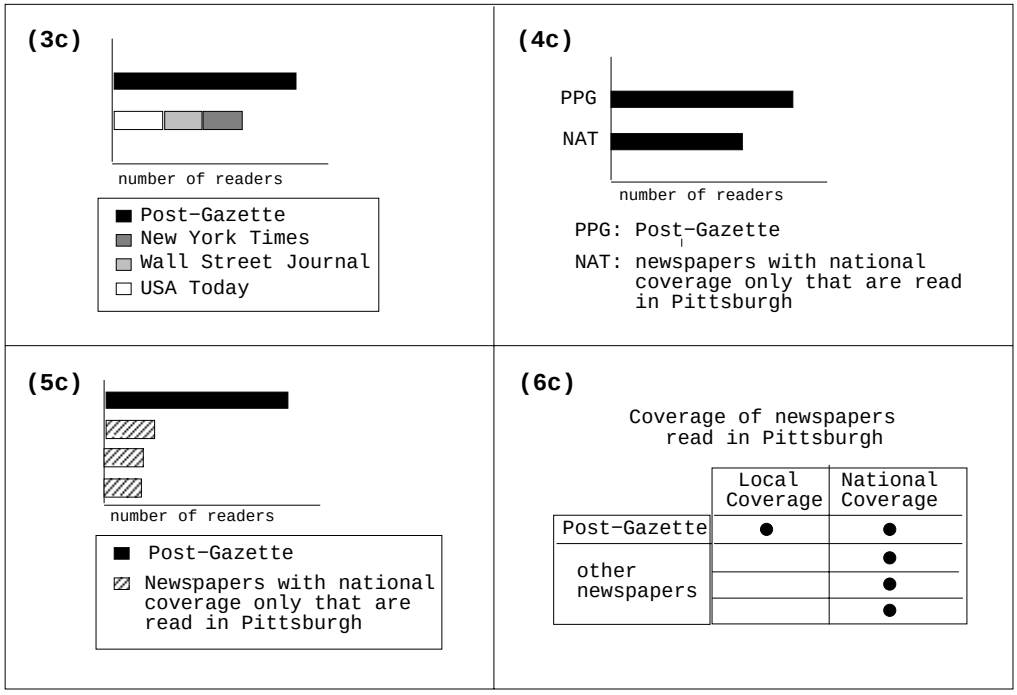

Figure 1: Assertions expressed in graphics

to convey the particular numbers of readers of each newspaper (hence the x-axis does not show actual numbers), information about the actual numbers of readers of each newspaper is needed during graphics generation to design (3c). (If the presentation's intention was to convey the particular numbers of readers of the newspapers, then different assertions specifying the actual numbers would be planned.)

Whereas in (3b), four newspapers are individuated, it is possible to make an assertion such as (4b) in which the members of the set ($\$NAT$) of newspapers with only national coverage are not individuated. The assertion in (4b) could be expressed in text as (4a), or in graphics in (4c) in Figure 1. However, this graphic still expresses more than (4b), e.g., that the number of PPG readers is about one-third more than the number of NAT readers (even though the x-axis does not show the actual numbers of readers).

(4a) The number of readers of the Post-Gazette is greater than the total number of readers of the newspapers read in Pittsburgh with national coverage only.

```
(4b) (Assert (gt ?d1 ?d2)
       ((?d1 (for the ?x integer
                (has-number-of-readers $PPG ?x)))
        ($PPG ())
        (?d2 (for the ?y integer
                (has-total-number-of-Pitts-readers
                     $NAT ?y)))
        ($NAT (for the ?w set (lambda ?z newspaper
                (and (has-coverage ?z National-only)
                     (has-circulation ?z Pittsburgh))))))
```

In contrast to (3b), (5b) differentiates the members of NAT, but does not identify or otherwise describe them. (5b) could be expressed in text as (5a), and in graphics as in (5c) in Figure 1. Once again, the graphic has side-effects. In this case, it conveys additional information about the relative numbers of readers among the newspapers with national coverage only, and the fact that there are three of those newspapers. Comparing (5c) to (3c), in (3c) the total number of readers of the three other newspapers is expressed by concatenating segments of bars representing the three newspapers into a single bar whose length represents the total number of readers of the three newspapers. Although this information can be *computed* from (5c), it is not directly realized in the graphic.

(5a) The number of readers of the Post-Gazette is greater than the number of readers in Pittsburgh of any newspaper with national coverage only.

(5b) (Assert (gt ?d1 ?d2)
((?d1 (for the ?x integer
(has-number-of-readers $PPG ?x)))
($PPG ())
(?d2 (for the ?y integer
(has-number-of-Pitts-readers ?d3 ?y)))
(?d3 (for each ?z newspaper
(has-coverage ?z National))))

In contrast to the preceding examples, (6b) illustrates a communicative intention (about the same data as in the other examples) with a different main predication. In text, (6b) could be expressed as in (6a); the main predication is about the coverage of the Post-Gazette rather than about the number of readers. This difference in main predication results in a graphic such as (6c) in Figure 1 with a different structure than those of the preceding examples.

(6a) Only 1 of the newspapers read in Pittsburgh, the Post-Gazette, has both national and local coverage.

(6b) (Assert (has-coverage $PPG Local-National)
(($PPG (for (only 1) ?x newspaper
(in-set ?x ?d1)))
(?d1 (for the ?w set (lambda ?x newspaper
(has-circulation ?x Pittsburgh)))))

4 Related Work

Several projects have studied the problem of media-independent knowledge representation schemes for automatic generation of multimedia presentations. The COMET [Feiner and McKeown1991] and WIP [Wahlster et al.1993] systems generate instructions for operating physical devices, and [Maybury1991] describes a system that designs narrated or animated route directions in a cartographic information system. These systems represent content about complex sequences of actions the user can perform on the physical device and their effects, as well as spatial concepts. However, this work is not relevant to information graphics generation.

The multimedia system whose focus is closest to ours, PostGraphe [Fasciano and Lapalme1996], is a system that generates multimedia statistical reports consisting of graphics and text. However, there are some fundamental differences with our approach. First, in Postgraphe it is assumed that a presentation is about the entire dataset, whereas our content language can be used to describe subsets and individuals in the dataset. Second, in Postgraphe graphics are generated directly from its knowledge representation language; then text is generated based upon the graphics. Thus, it is not clear whether the language is truly media-independent, i.e., whether it could be used to generate text directly. Also, Postgraphe's language of intentions is less general than our approach of generating presentation plans for achieving communicative goals. For example, in Postgraphe the language can be used to specify the intention to compare two variables of a dataset in a way that emphasizes an increase. In our approach, complex arguments can be planned.

5 Conclusions

This paper describes a media-independent knowledge representation scheme, or *content language*, for describing the content of communicative goals and actions. The language is used within an intelligent system for automatically generating integrated text and information graphics presentations about complex, quantitative information. To ensure that the language will be applicable to a variety of quantitative domains, it is based upon a corpus analysis of integrated natural language and graphics presentations. The language is designed to satisfy four requirements: to represent information about complex quantitative relations and aggregate propertiess; compositionality; to represent certain pragmatic distinctions needed for satisfying communicative goals; and to be usable as part of the input to the media-specific (text and

graphics) generators.

References

[Elhadad and Robin1996] M. Elhadad and J. Robin. 1996. An overview of SURGE: A reusable comprehensive syntactic realization component. Technical Report Technical Report 96-03, Dept of Mathematics and Computer Science, Ben Gurion University, Beer Sheva, Israel.

[Fasciano and Lapalme1996] M. Fasciano and G. Lapalme. 1996. PostGraphe: a System for the Generation of Statistical Graphics and Text. In *Proceedings of the 8th International Natural Language Generation Workshop*, pages 51–60, Sussex, UK, June.

[Feiner and McKeown1991] S. Feiner and K. McKeown. 1991. Automating the generation of coordinated multimedia explanations. *IEEE Computer*, 24(10):33–40, October.

[Green et al.1998] Nancy Green, Giuseppe Carenini, and Johanna Moore. 1998. A principled representation of attributive descriptions for integrated text and information graphics presentations. In *Proceedings of the Ninth International Workshop on Natural Language Generation*, Niagara-on-the-Lake, Ontario, Canada. To appear.

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

[Kerpedjiev et al.1997b] Stephan Kerpedjiev, Giuseppe Carenini, Steven F. Roth, and Johanna D. Moore. 1997b. Integrating planning and task-based design for multimedia presentation. In *International Conference on Intelligent User Interfaces (IUI '97)*, pages 145–152. Association for Computing Machinery.

[Kerpedjiev et al.1998] Stephan Kerpedjiev, Giuseppe Carenini, Nancy Green, Steven F. Roth, and Johanna D. Moore. 1998. Saying it in graphics: from intentions to visualizations. In *Proceedings of the Symposium on Information Visualization (InfoVis '98)*. IEEE Computer Society Technical Committee on Computer Graphics. To appear.

[Kosslyn1994] Stephen M. Kosslyn. 1994. *Elements of Graph design*. W.H. Freeman and Company.

[Maybury1991] Mark T. Maybury. 1991. Planning multimedia explanations using communicative acts. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 61–66, July.

[Roth and Mattis1990] S.F. Roth and J. Mattis. 1990. Data characterization for intelligent graphics presentation. In *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI '90)*, pages 193–200.

[Roth et al.1994] Steven F. Roth, John Kolojejchick, Joe Mattis, and Jade Goldstein. 1994. Interactive graphic design using automatic presentation knowledge. In *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI '94)*, pages 112–117.

[Tufte1983] Edward R. Tufte. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Conn.

[Tufte1990] Edward R. Tufte. 1990. *Envisioning information*. Graphics Press, Cheshire, Conn.

[Tufte1997] Edward R. Tufte. 1997. *Visual Explanations*. Graphics Press, Cheshire, Conn.

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

[Webber1983] Bonnie L. Webber. 1983. So what can we talk about now? In B. Grosz, K. S. Jones, and B. L. Webber, editors, *Readings in Natural Language Processing*. Morgan Kaufmann, Los Altos, California.

[Woods1983] W. Woods. 1983. Semantics and quantification in natural language question answering. In B. Grosz, K. S. Jones, and B. L. Webber, editors, *Readings in Natural Language Processing*. Morgan Kaufmann, Los Altos, California.

[Young1994] Michael R. Young. 1994. A developer's guide to the Longbow discourse planning system. Technical Report ISP TR Number: 94-4, University of Pittsburgh, Intelligent Systems Program.

[Zelazny1996] Gene Zelazny. 1996. *Say it with charts: the executive's guide to visual communication*. IRWIN Professional Publishing.