Automatic Generation of Formatted Text

Eduard H. Hovy* and Yigal Arens†

Information Sciences Institute of USC
4676 Admiralty Way
Marina del Rey, CA 90292-6695
hovy@isi.edu, arens@isi.edu

Abstract

Few texts longer than a paragraph are written without appropriate formatting. To ensure readability, automated text generation programs must not only plan and generate their texts but be able to format them as well. We describe how work on the automated planning of multisentence text and on the display of information in a multimedia system led to the insight that text formatting devices such as footnotes, italicized regions, enumerations, etc., can be planned automatically by a text structure planning process. This is achieved by recognizing that each formatting device fulfills a specific communicative function in a text, and that such functions can be defined in terms of the text structure relations used as plans in a text planning system. An example is presented in which a text is planned from a semantic representation to a final form that includes English sentences and LaTeX formatting commands, intermingled as appropriate.

The Problem: Text Layout

No paper is submitted to this conference without a heading, section titles, and occasional italicized text; and most of them contain itemized lists, footnotes, indented quotations, boldfaced terms, and other formatting devices.

Why? The reason is clear: each such formatting device carries its own idiosyncratic meaning, and writers select the device that best serves their communicative intent at each point in the text.

A more interesting question is: How? That is, how do writers know what device to use at each point? How is device selection integrated with the text production process in general? Can the two processes be automated — can a text production system be made to plan not only the content and structure of the text but also the appropriate textual formatting for it?

The answer is yes, and this paper describes an experiment that demonstrates this ability.

Though manuals of style (such as [CMS 82, APA 83, Van Leunen 79]) may seem relevant, they contain little more than precise descriptions of the preferred forms of textual devices in fact. Their recommendations of use, when provided, are not detailed enough to help in the selection process itself. Instead, our approach to this problem builds upon ideas and techniques proven useful for multisentence text generation (specifically, text planning) and multimedia communication planning (specifically, the problem of best integrating different media, such as language, diagrams, tables, maps, etc., into a single coherent display. In research on these two questions, we came to realize that the problem of text formatting forms a natural point midway between pure text planning on the one hand and pure multimedia display planning on the other: though text formatting devices have features that make them resemble different media (two-dimensional offset in some cases, highlighting in others, etc.), they remain essentially textual. This similarity enabled us to extend some of the previous work on text planning done at USC/ISI (see [Hovy 88, Moore & Paris 89]) over the past several years to perform not only standard text structure planning but also text formatting.

The next section describes our characterization of text formatting devices. The section that follows it describes how we extended our text planner to handle the requirements of text layout, including an example of the generation, from a semantic representation, of a formatted segment of an Air Traffic Control manual — the type of document which would be difficult to comprehend without appropriate textual devices. The final section provides a description of the definition of additional text formatting devices and discusses shortcomings in the relevant theories.

*This author was supported in part by the Rome Air Development Center under RADC contract FQ7619-89-03326-0001.

†This author was supported by the Defense Advanced Research Projects Agency under Strategic Computing program contract no. MDA-903-87-C-0641.

Textual Devices

In the course of our work on automated modality selection (described in some detail in [Arens & Hovy 90]), we noticed an interesting fact: not only are the different text layouts and styles (plain text, itemized lists, enumerations, italicized text, inserts, etc., which we here call **Textual Devices**) used systematically in order to convey information, but *it is possible to define their communicative semantics precisely enough for them to be used in a text planner.* What's more, the systematicity holds across various types of texts, genres, and registers of formality. It is found in books, articles, advertisements, papers, letters, and even memos. The information these devices convey supplements the primary content of the text.

As a result, we believe that one can treat the different textual devices as different communication modalities. That is to say, the same type of reasoning that goes into the central data-to-modality allocation problem (the problem of deciding when and how to choose between using a picture, a table, or a sentence, etc.), goes into deciding whether to generate a straight paragraph or to use an enumerated list or a table or an insert. The reasoning is based, in large measure, on the contents to be conveyed to the reader. The process of device selection, like that of modality selection in general, consists of choosing one or more devices whose characteristics are suited for conveniently expressing essential portions of the contents.

As is the case with modalities in general, we find that textual devices are distinguished along several independent dimensions; we identify three broad classes: *Depiction*, *Position*, and *Composition*. In all three cases, the communicative function of text devices is to delimit a portion of text for which certain exceptional conditions of interpretation hold.

1. Depiction:

Depiction involves selecting an appropriate letter string to express the text. Examples of different depictions are:

Parentheses: Usually indicates that the parenthesized text is tangential to the main text. Used mainly for clause-sized text bodies.

Font switching: Indicates special importance of the delimited text, either that it is a new term being introduced, that it is of central importance, or that it is a foreign expression. This includes the use of boldface, underlining, and doubled emphasis (italicization *when the surrounding text is* not *italicized*).

Capitalization: Indicates that the text string names (identifies) a particular entity. Used mainly for single-word text bodies.

Quotation marks: Usually used to signal that the text body was written by another author, and occasionally used to indicate that the meaning of the text body is different than a standard interpretation would yield.

2. Position:

Repositioning involves moving the text block relative to the surrounding text on the page. Examples:

Inline: The non-distinguished normal case.

Offset (horizontal repositioning): Indicates either that the text block is authored by someone else (e.g., a long quoted paragraph, indented) or that it summarizes material that is especially relevant, as with an indented paragraph.

Separation (vertical repositioning): Indicates that the text block addresses a single point (as with paragraphs) or that it identifies or summarizes the subsequent text (as with or headings or titles).

Offpage: Offpage text provides material (usually of explanatory import) that is tangential to the main text, as with an appendix, footnote, or sidebar.

3. Composition:

Compositional devices impose an internal structure on the affected text body. Examples are:

Itemized list: A set of entities/discourse objects on the same level of specificity with respect to the subject domain, but that (in general) contain too much material to be expressed as a simple list within a single sentence. (This list is an example.)

Enumerated list: A set of entities/discourse objects on the same level of specificity with respect to the domain, which are, furthermore, ordered along some underlying dimension, such as time, distance, importance.

Term definition: A pair of texts separated by a colon or similar delimiter, in which the first text is the name of a discourse object and the second defines it or expresses some other fact related to it. (The typical form is "Term: Text string"; each entry in this itemization of Composition types is an example of Term Definition.)

From these definitions, it is clear that selecting an appropriate textual device (or combination of such) relies in large measure on the author's ability to accurately characterize the meaning expressed by the specific portion of text as well as its relationship to the surrounding text. That is, in order to know when to use a footnote, an itemized list, or an enumeration, for example, the author must be able to match the communicative function of the block of text *in its current context* against the communicative semantics of the textual devices so as to select the appropriate one. After all, the same sentence can properly be a footnote in one text and a parenthesized part of the text proper in another: the difference depends on how the text as a whole is organized to achieve its communicative purpose.

Thus (ignoring until Section such issues as textual prominence and style), there are three parts to the prob-

lem: the underlying semantic content to be communicated, the textual structure in which this meaning is couched, and the textual devices available for expression. With respect to semantics, we take a standard approach (namely, using frame-like representation structures that contain terms from a well-specified ontology). In order to define the communicative semantics of textual devices, we employ a theory of text structure that describes how coherent texts achieve their communicative purposes. We turn next to recent work on text structure planning.

Background: Text Planning

There is more to building coherent text than the mere generation of single sentences. At the very least, one has to delimit the content of each sentence and specify their order of appearance. Texts only achieve their communicative functions when they are coherent — when the reader is able to build up an understanding not only of each individual sentence but also of how each sentence relates to the whole. In order to produce coherent paragraphs, one requires an understanding of the interrelationships between the parts of a paragraph. For example, the following paragraph is simply not coherent, because the logical interrelationships between the sentences are not respected rhetorically:

> At the very least, one has to delimit the content of each sentence. Texts only achieve their communicative functions when they are coherent. One has to specify the order of appearance of each sentence. There is more to building coherent text than the mere generation of single sentences.

The question "What makes text coherent?" has a long history, going back at least to [Aristotle 54]. A number of researchers have recognized that in coherent text successive pieces of text are related in particular ways, and have provided different sets of interclause relations (see, for example, [Hobbs 79, Grimes 75, Shepherd 26, Reichman 78]). After an extensive study of several hundred texts of different types and genres,[Mann & Thompson 88] identified 25 basic rhetorical relations, which they claimed suffice to represent all intersentential relations that hold within normal English texts. Some relations are PURPOSE, ELABORATE, SEQUENCE, and SOLUTIONHOOD, the first three of which are typically signaled by "in order to", "for example", and "then" respectively (the last has no cue phrase). Their theory, called Rhetorical Structure Theory (RST), holds that the relations are used recursively, relating ever smaller blocks of adjacent text, down to the single clause level; it holds that a paragraph is only coherent if all its parts can eventually be made to fit under one overarching relation. Thus each coherent paragraph can be described by a tree structure that captures the rhetorical dependencies between adjacent clauses and blocks of clauses. The RST relations subsume most of the rhetorical relations proposed by previous researchers; recent attempts at more encompassing taxonomies that synthesize several hundred relations from linguists, computational linguists, philosophers, and other interested parties appear in [Hovy 90].

Within the past few years, a number of computational research projects have addressed problems that involve automatically generating coherent multisentence paragraphs that achieve a given communicative goal. Almost all of these use a tree of some kind to represent the structure of the paragraph. An ongoing effort at ISI, involving one of the authors, uses RST relations (and extensions of them) which are represented and formalized as plans in a top-down hierarchical planning system reminiscent of the Artificial Intelligence planning system NOAH [Sacerdoti 77]. The structure planner functions between some application program (such as an expert system) and the sentence generator Penman [Penman 89, Mann & Matthiessen 83]. From the application program, the planner accepts one or more communicative goals along with a set of semantic representations of relevant material which can be used to form the text. During the planning process, the structurer assembles the input entities into a tree that embodies the paragraph structure, in which nonterminals are RST relations and terminal nodes contain the input material. It then traverses the tree, noting the linking phrases at tree branches and submitting the leaves to Penman to be generated in English. The planning process is described in much more detail in [Hovy 88].

Extending the Planner: An Example of Layout Planning

The RST text structure planner (linked to Penman) has been used in several domains, most recently to plan and generate paragraphs of text about the procedures to be followed by air traffic controllers. The host system, ARIES [Johnson & Harris 90], is being developed in the context of an automatic programming project, and is intended to perform many air traffic control operations automatically.

In our example, the structurer is activated with the goal to describe the procedure to be followed by an air traffic controller when an aircraft is "handed over" from one region to the next. The underlying representation for this example consists of a semantic network of 18 instances, defined in terms of 27 air traffic do-

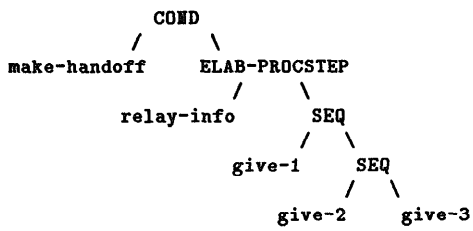

Figure 1: RST tree for ATC domain.

main concepts and 8 domain relations, implemented as frames in the Loom knowledge representation system [MacGregor 88].

The planner finds the RST-based plans CONDITION, ELABORATE-PROCESSSTEP, and SEQUENCE, and builds the paragraph tree shown in Figure 1. The leaves of this tree are reformulated in Penman's input language and the text generated is:

```
When making a handoff, the transferring
controller relays information to the
receiving controller in the following
order. He gives the target's position.
He gives the aircraft's identification.
He gives the assigned altitude and
appropriate restrictions.
```

Though the text closely mirrors that of the actual Air Traffic Control Manual [ASA 89], the differences in formatting are significant; and these differences make the manual much more readable. The manual contains headings, term definitions signaled by italicized terms, enumerated lists, etc.

We have recently embarked on a study of several instructional texts, including recipes, school textbooks, and manuals for cars, sewing machines, and video players. An early conclusion is that certain textual formatting devices are highly correlated with specific configurations of the underlying text structure tree. For example, a series of nested SEQUENCES, such as appears in Figure 1, is usually realized in the text as an enumerated list. Exceptions occur (in general) only when the individual items enumerated are single words (in which case the whole list is realized in a single sentence) or when there are few enough of them to place in a paragraph in-line (though usually in this case the keywords *first*, *second*, etc., are added).

On the assumption that we can capture most of the reasons for using such formatting devices as enumerations on the basis of RST alone, we augmented the text plan SEQUENCE in order to include explicit formatting commands and adapted the structure planner accordingly (namely, to ignore formatting commands when building the sentence generator inputs and finally to append the formatting commands and the sentence generator's output in the order mandated by the tree). For the formatting command we used LaTeX forms [Lamport 86] such as `\begin{enumerate}` `\item` `\end{enumerate}`. Although our implementation was done within the framework of our specific generation technology, we believe a similar augmentation could be performed with most if not all the text planners being developed at this time. The resulting tree (with formatting commands indicated) is shown in Figure 2. The text generated is:

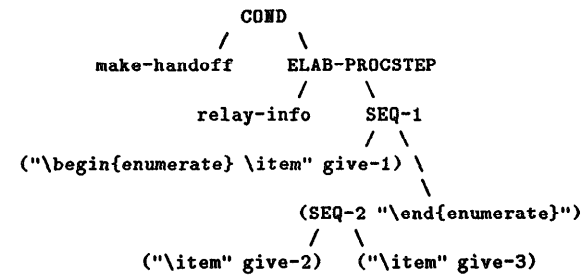

Figure 2: Augmented RST tree for ATC domain.

```
When making a handoff, the transferring
controller relays information to the
receiving controller in the following
order. \begin{enumerate} \item He gives
the target's position. \item He gives
the aircraft's identification. \item He
gives the assigned altitude and appro-
priate restrictions. \end{enumerate}
```

Note that due to the asymmetry of the LaTeX formatting command (i.e., the `\begin` and `\end` statements at either end), we had to create two SEQUENCE text plans, one for the top of a sequence and one for its recursive continuation. The relevant parts of the two SEQUENCES are shown in Figure 3. The asterisk in the Nucleus and Satellite fields indicates to the planner the position in which to include the semantic content which eventually becomes the text.

The final text, as planned from the semantic representation, is produced and formatted directly by the system as follows:

When making a handoff, the transferring controller relays information to the receiving controller in the following order.

1. He gives the target's position.
2. He gives the aircraft's identification.
3. He gives the assigned altitude and appropriate restrictions.

This example illustrates the planning of only a single formatting command. Despite its rather extreme

```
Name: SEQUENCE-1
Results:
 ((BMB SPEAKER HEARER (SEQUENCE-OF ?PART ?NEXT)))

Nucleus: ("\begin{enumerate} \item" *)     <-----
Satellite: (* "\end{enumerate}")           <-----

Nucleus requirements/subgoals: (...)
Satellite requirements/subgoals: (...)
Nucleus+Satellite requirements/subgoals: (...)
Nucleus growth points: (...)
Satellite growth points:
 (...
  (BMB SPEAKER HEARER (SEQUENCE-OF-2 ?NEXT ?FOLL)))

Order: (NUCLEUS SATELLITE)
Relation-phrases: ("" "then" "next")
Activation-question: "..."
```

```
Name: SEQUENCE-2
Results:
 ((BMB SPEAKER HEARER (SEQUENCE-OF-2 ?PART ?NEXT)))

Nucleus: ("\item" *)                       <-----
Satellite: ("\item" *)                     <-----

Nucleus requirements/subgoals: (...)
Satellite requirements/subgoals: (...)
Nucleus+Satellite requirements/subgoals: (...)
Nucleus growth points: (...)
Satellite growth points: (...)

Order: (NUCLEUS SATELLITE)
Relation-phrases: ("" "then" "next")
Activation-question: "..."
```

Figure 3: Top and recursive SEQUENCE text plans. (Details elided due to space limitations.)

simplicity, however, the example is sufficient to demonstrate that as long as one can characterize textual formatting devices in terms of particular structural configurations of the text tree, one can include appropriate formatting commands of several types into the generated output.

We next outline the definition of other textual devices in terms of text structure theories such as RST and point out some limitations of this idea.

Semantics of Textual Devices

The insight that the communicative semantics of text formatting devices can to a significant extent be stated in terms of text structure is a powerful one. Two major limitations should however be borne in mind: first, that there are additional factors that determine the use of most formatting devices, and second, that the representational power of current theories of text structure is still very limited.

Various types of additional factors affect the use of formatting devices. These range from the style of the author and the amount of visual prominence the author wishes to accord a text block, all the way to limitations on the text length. In current work at USC/ISI and the University of Nijmegen, we are attempting to systematize the various factors that play a role and to represent them in a single formalism that makes clear their nature and interrelationships [Vossers, Arens, & Hovy 91]. This work studies four major groups of factors — the nature of the communication, the goals of the author, the essential features of the information to be conveyed, and the capabilities of the reader — which together control the actual types of expressive media used (including normal text, formatting devices, and line diagrams).

With respect to defining how footnotes, sidebars, italicized regions, etc., really function in a text, we are well aware that precise definitions of the various devices are limited by the descriptive power of the particular theory of text structure employed. Such theories as RST, among the best available at the present time, still do not provide a great deal of detail and descriptive adequacy. But they do at least enable one to capture the essential functionality of the following textual devices:

- **Enumeration:** As described in the example above, the text structure relation SEQUENCE can generally be formatted as an enumerated list. The enumeration follows the sequence of the relation, which is planned in expression of some underlying semantic ordering of the items involved, for example time, location, etc.
- **Itemization:** The textual structure that relates a number of items without any underlying order is the RST relation JOINT, which can be realized by an itemized list (unless the items are small enough to be placed into a single sentence).
- **Appendix, footnote, and parentheses:** These are three devices that realize the same textual relation, namely BACKGROUND. They differ in the amount of material included in the relation's Satellite.
- **Section title or heading:** This device realizes the textual relation IDENTIFICATION, which links an identifier with the body of material it heads. A section or subsection is appropriate when the IDENTIFICATION is combined with a SEQUENCE chain that governs the overall presentation of the text.

For some textual devices, no text structure relation

has been identified by discourse linguists. For example, the Quotation device realizes the linguistic relation Projection (that which links a sayer to what is said; see [Halliday 85]), which is not included in RST, Conjunctive Relations, or Hobbs's or Dahlgren's theories. Other textual devices work on a level too detailed for text coherence theories, since they operate on individual words within a clause; text structure theories typically relate independent clauses only and provide no intraclausal links. And unlike the case for Projection, no purely linguistic constructs exist to handle them either. Thus text formatting devices such as italicization and capitalization for word definition or emphasis cannot at this time be represented. (Members of the Penman and EES projects at USC/ISI are currently building a new text planner with a network of considerably more detailed plans, in order to help address this problem of lack of expressive delicacy, a problem that also impinges on the syntactic realization of text.)

An additional shortcoming with our approach is the fact that we embed LaTeX commands literally into the RST plans. The text structure planner thus has no ability to reason about the implications of its formatting. Better would be to develop an abstract representation of textual devices which, when included in the text plans, would be realized into LaTeX (or Scribe, or any other formatting language) commands at the time the content is realized into English.

However, despite the problems with definitional delicacy, we are able to represent many of the textual devices listed above within the existing text structuring relations, which are all related to more or less well-understood text plans and communicative goals. To this extent, then, the insight that text formatting devices can be defined in terms of text structure relations (as described here) enables the automatic planning of appropriately formatted multisentence texts, a new and very useful capability.

Acknowledgments

Thanks to Bill Swartout for the idea of using LaTeX to illustrate the point, to Richard Whitney for implementation help, and to Lewis Johnson for ARIES.

References

[APA 83] *Publication Manual of the American Psychological Association* (third edition). 1983. Washington: American Psychological Association.

[Arens & Hovy 90] Arens Y., and Hovy, E.H. 1990. How to Describe What? Towards a Theory of Modality Utilization. In Proceedings of the 12th Annual Conference of the Cognitive Science Society, 487-494.

[Aristotle 54] Aristotle. The Rhetoric. In *The Rhetoric and the Poetics of Aristotle*, W. Rhys Roberts (trans). New York: Random House.

[ASA 89] ASA, Inc. 1989. *Airman's Information Manual.* Aviation Supplies and Academics.

[CMS 82] *The Chicago Manual of Style* (thirteenth edition). 1982. Chicago: University of Chicago Press.

[Grimes 75] Grimes, J.E. 1975. *The Thread of Discourse.* The Hague: Mouton.

[Halliday 85] Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* Baltimore: Edward Arnold Press.

[Hobbs 79] Hobbs, J.R. 1979. Coherence and Coreference. *Cognitive Science* 3(1):67-90.

[Hovy 88] Hovy, E.H. 1988. Planning Coherent Multisentential Text. In Proceedings of the 26th ACL Conference, 163-169.

[Hovy 90] Hovy, E.H. 1990. Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. Presented at the 5th International Workshop on Text Generation, Pittsburgh.

[Johnson & Harris 90] Johnson, W.L. and Harris, D. 1990. Requirements Analysis Using ARIES: Themes and Examples. In Proceedings of the 5th Knowledge Based Software Engineering Conference, 79-84. Tech Report no. RL-TR-91-11, Rome Laboratory.

[Lamport 86] Lamport, L. 1986. LaTeX *User's Guide and Reference Manual.* Reading: Addison-Wesley.

[MacGregor 88] MacGregor, R. 1988. A Deductive Pattern Matcher. In Proceedings of the 6th National Conference on Artificial Intelligence (AAAI-88), 696-701.

[Mann & Matthiessen 83] Mann, W.C. and Matthiessen, C.M.I.M. 1983. Nigel: A Systemic Grammar for Text Generation. Research Report RR-83-105, USC/ISI.

[Mann & Thompson 88] Mann, W.C. and Thompson, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3):243-281.

[Moore & Paris 89] Moore, J.D. and Paris, C.L. 1989. Planning Text for Advisory Dialogues. In Proceedings of the 27th ACL Conference, 67-75.

[Penman 89] *The Penman Documentation.* 5 unpublished volumes, USC/ISI.

[Reichman 78] Reichman, R. 1978. Conversational Coherency. *Cognitive Science* 2(1):283-327.

[Sacerdoti 77] Sacerdoti, E. 1977. *A Structure for Plans and Behavior.* Amsterdam: North-Holland.

[Shepherd 26] Shepherd, H.R. 1928. *The Fine Art of Writing.* New York: The Macmillan Co.

[Van Leunen 79] Van Leunen, M.-C. 1979. *A Handbook for Scholars.* New York: Alfred A. Knopf.

[Vossers, Arens, & Hovy 91] Vossers, M., Arens, Y., and Hovy, E.H. 1991. On the Automated Control of Communications Involving Text, Formatting, and Line Diagrams. Master's thesis, University of Nijmegen. Forthcoming.