Evaluating Plan-Based Hypermedia Generation

Berardina De Carolis and Fiorella de Rosis
Department of Informatics, University of Bari
Via Orabona 4, 70 126 Bari, Italy
nadja@aos2.uniba.it, derosis@di.uniba.it

Dianne C Berry and Irene Michas
Department of Psychology, University of Reading
d.c.berry@reading.ac.uk

Abstract
This paper describes a algorithm for generating a hypermedia from a discourse plan and discusses results of an experimental evaluation study whose aim was to assess whether hypotheses behind the generation method were justified. A first aid hypermedia instruction manual is taken as a case study to illustrate the algorithm.

1. Introduction

Interest in hypermedia document generation is growing very fast. The main difference between generating a one-shot document and a hypermedia is that, in the first case, the algorithm has to set up the information content and media employed, render it as a text or as a multimedia and solve presentation issues such as formatting, style refinement, redundancy control, emphasis or de-emphasis of specific issues and so on. These are tasks of hypermedia generation too: in this case, however, the algorithm has to establish, as well, how information has to be distributed among the various nodes, which links have to be established among these nodes and which navigation paths and orientation supports have to be provided. This requires a knowledge base (KB) that includes two main components: a *domain* knowledge and a *generation strategy*. In electronic encyclopedias (which are, so far, the most popular application domain of hypermedia generation), the first component is a taxonomy of objects, concepts or terms to be explained. The generation strategy is employed to define the layout of every single node and its relationships with the remaining structure; in these 'dynamic hypermedia' (Dale et al, 1998) or 'virtual documents' (Gruber et al, 1997), the generation algorithm does not produce the whole document but only a portion of it: the set of nodes that the user needs to see during an interaction session. It is generally claimed that this approach corresponds to leaving to the user the 'high-level planning' task, and assigns to the system only the role of presenting the requested information in the best way: the system has some power in orienting the user goal in every phase of interaction, as it provides a list of nodes that might be relevant in that context; it leaves, however, to the user the last word in making a choice among alternatives. Building local portions of the network on request (as in dynamic hypermedia) is more convenient than building the whole structure (like in 'static' hypermedia) when some conditions hold: (i) the portion of the network that will be examined at each interaction is minimal, and/or (ii) the frequency with which the hypermedia has to be updated is high. In these cases, dynamic generation enables avoiding storage of documents that will be read only in part or will have to be updated frequently. When the hypermedia is produced dynamically, its structure is a local mapping of the underlying domain KB: this enables the structure of a hypermedia to clearly show to the reader the organisation of ideas in the expert's knowledge (Jonassen, 1993). When the domain KB is a taxonomy, a natural way of embedding a structure in the hypermedia is to make explicit the position of the displayed item in the hierarchy and its relations with other items, and to employ comparison or contrast as strategies to generate explanations about the item itself (Milosaljevic,1997): the purpose is, in this case, to provide context and user-customised information about items in the domain KB. What happens, though, if the purpose of the hypermedia is instruction or argumentation? In this case, the underlying KB is, typically, a plan: this may describe how the task being explained may be decomposed into a hierarchy of subtasks (like in explanation of software systems, in Johnson and Erdem, 1997) or may be a more general discourse plan that specifies how the intended mental state of the user can be achieved (Moore, 1995). In both cases, the main task of generation is to translate this plan into a usable hypermedia.

Elsewhere, we described a system that translates a discourse plan into a user-adapted hypermedia in HTML. The system is called GeNet and was implemented in C++ under SunOs2. In this paper, we discuss aspects of the system that concern the relationship between the underlying KB structure and the hypermedia structure in more depth. In particular, we examine the following open problems: (i) how much, of the goal/subgoal structure of the discourse plan, should be explicitly reflected into the hypermedia structure; (ii) how much freedom should be left to the user, in navigating in such a structure; (iii) which forms of orientation to the navigation should be provided. In GeNet, we made some specific choices about each of these aspects, which we wanted to verify using an experimental evaluation study. In the next Section, we will outline the algorithm employed in GeNet, then describe the evaluation study with its results (in Section 3) and finally discuss the implications of results on future developments of GeNet (in Section 4).

2. Outline of GeNet

Let us call Dplan the source discourse plan, and Hypm the page-oriented hypermedia document that GeNet generates from Dplan. The particular Dplan that we employ is produced by a hierarchical planning and is represented as a hierarchy of node-objects to which the following elements are attached:

Dplan:= <Go, RR, Ro, Co, In>

- discourse goal-subgoal Go (*goal structure*)
- rhetorical relations (RR) among children nodes (*rhetorical structure*)
- *role* Ro of the node in the RR attached to its mother (nucleus or satellite)
- *complexity* Co of the subtree originating from the node (H=high, M=medium, L=low)
- media employed and *information content* In (only for leaf nodes).

Figure 1 shows an example of a portion of Dplan about a first-aid manual: nucleus-nodes are denoted with a small circle; for space reasons, only goals attached to higher level nodes are displayed.

Hypm is an oriented graph in which *pages* are associated with nodes and links between pages with edges. Pages have a fixed structure that includes the following elements: ***Hypm : = <Ti, He, Bo, Fo>***

- a *Title* Ti, which identifies the page and orients the user in the navigation history
- a *Header* He, which cues the user to the page purpose
- a *Body* Bo, in which Dplan nodes are aggregated and structured
- a *Footer* Fo, which represents links, in whole or in part, depending on the navigation style.

A Hypm node is built from a Dplan subtree according to a strategy that aims at achieving three main purposes at the same time: (i) to avoid building too complex pages, (ii) to insure that strongly related information is not split into different pages and (iii) to insure that a logical order of presentation of information is respected[1]. The algorithm explores Dplan in a depth-first way; when examining a node, it decides whether information associated with the subtree that starts from this node should go into a unique Hypm page or should be subdivided into separate pages. In doing this, it considers (i) the node complexity, which is a function of the depth and breadth of the subtree and of the media employed in its leaf nodes, (ii) the RR that links the examined node to its sisters, and their focus. RRs are classified as in (Hovy, 1993); they are labelled as 'strong' (all Interpersonal relations and the Semantic Comparative), of medium strength (all the remaining Semantic relations) and weak (the Presentational relations). Information associated with the subtree starting from the examined node is introduced into the Body of a unique Hypm page if the node has a 'low' complexity or if, though being of 'medium' complexity, the RR that links it to its next sister is not weak and the two nodes have the same focus. In the remainder of cases, only the main subject of the discourse represented by the subtree is mentioned in the page, and a link to other pages providing more details about this subject is introduced in the page itself. We call *information- page* a page of Hypm in which a Dplan subtree has been translated entirely into the Body; *structure-page* a page whose Body includes only the list of subjects associated with daughter-nodes and *mixed-page* a page whose Body includes both types of information. Structure pages are generated from the highest part of Dplan: their role is to show the structure of the discourse in a clear and explicit way; descriptions of subjects in these pages are built from the discourse goals/subgoals attached to Dplan nodes. RRs help to make clear the underlying discourse structure

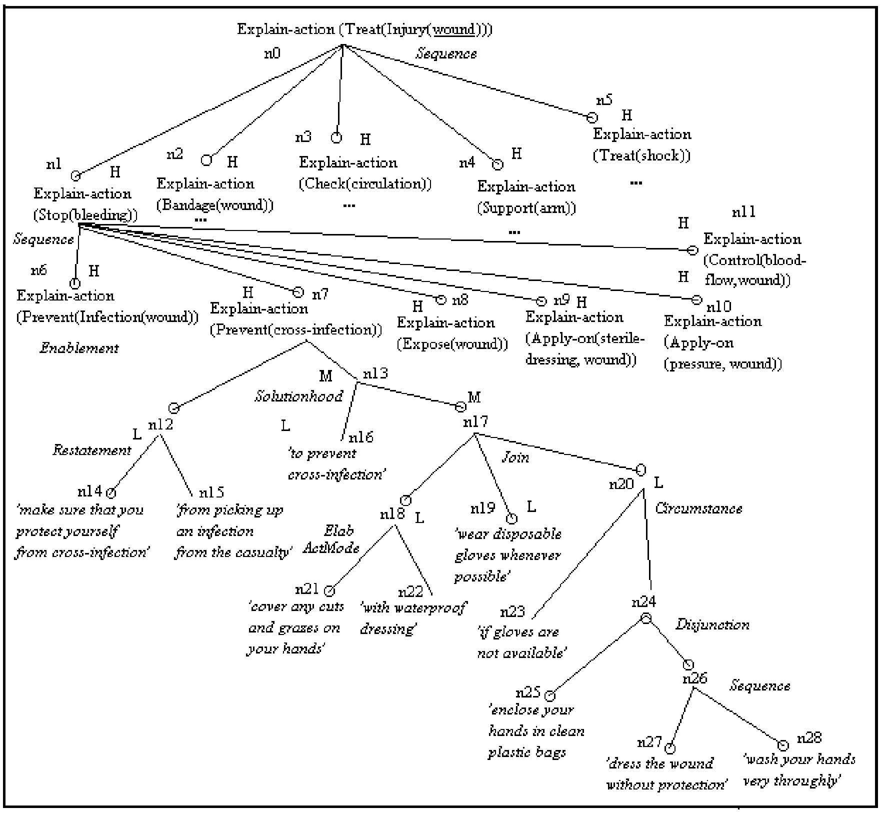

Figure 1 : a portion of discourse plan for a first aid instruction manual

[1] This purpose may be translated into the following rules: (i) never expand the satellite of a RR before expanding its nucleus and all previous satellites, if any and (ii) never expand a nucleus of a multinuclear RR before expanding the previous ones, if their order of presentation is important (for instance the *Sequence*).

by formatting the Body; as in all NLG systems, they are also employed to introduce linguistic markers linking related sentences and to establish the relative position of information items. The page Title is built from the goal attached to the examined node and the Footer always includes to-sister links.

In addition to these general criteria of Hypm generation, other criteria introduce the possibility of diversifying the resulting hypermedia:

- though Hypm structure is *hierarchical*, due to its correspondence with Dplan's goal-subgoal structure, it may be *circular* or *linear*; in the first case, each page includes links to all its daughters and sisters and to its mother; in the second case, it only includes one link to the next page (a daughter or a sister) and another one back to the previous page. We call *free* the navigation style enabled by the circular mode, as it leaves to the user more freedom in selecting the next page to visit; we call *fixed* the style corresponding to the linear mode, as the user is only enabled to follow the path that the system considers as 'preferable';
- links to other pages may be labelled according to the content of the destination-pages or to the relationship between origin and destination page; in the first case, the goal attached to the node from which the destination page is generated is employed to produce the label, and the link is represented as an anchor in the Body (*to-daughter-links-in-the-Body* mode); in the second case, the RR linking the two nodes in Dplan is employed instead, and the link is represented as a button in the Footer (*all-links-in-the-Footer* mode).

Hypermedia with different characteristics are presumed to be suited to different user categories (Wright and Lickorish, 1989); the user-adaptation component of GeNet reserves the 'hierarchical-circular, free, to-daughter-links -in-the-Body' mode to 'expert' users, while the 'hierarchical-linear, fixed, all-links-in-the-Footer' mode is activated for novice users. A third and intermediate type of hypermedia is generated to middle-experience users, in which the free mode is enriched with dynamically-generated navigation suggestions: these suggestions are produced when the user leaves the navigation path that is considered as 'optimal'. Suggestions generated in a specific context depend on the number of nodes that the user omits seing and on the strength of the RRs that link them. A 'light' suggestion is provided in case of medium-strength RRs, whose effect is exhalted from knowing all the elements (for instance, a Sequence): this suggestion includes only a recall of the jumped discourse segments through their communicative goal, with links to the corresponding pages. A 'heavy' suggestion is provided in case of strong RRs, whose effect is conditioned to knowing all elements (for instance, a Motivation); this suggestion includes the 'Comprehensive Locus of Effect' of the subtree originating from the jumped node (Mann, Matthiesen and Thompson, 1989). We show, in Figure 2, an example of the portion of Hypm that is produced from the Dplan in figure 1, in the case of expert users. This figure shows, on the top, an example of structure-page; its Title reflects the goal attached to the root of the subtree (n1); its Body is built from goals attached to high-complexity daughter nodes (from n6 to n11), and its formatting shows that these nodes are linked, in Dplan, by a RR of Sequence. Links in this page send to mother, sister and daughter pages; to-mother and to-sister links are grouped in the Footer, to-daughter links in the Body: they are all labelled according to the destination node's goal. The bottom page is an example of information-page that was generated from the Dplan subtree starting from n7; this subtree includes all medium or low-complexity nodes, linked by various RRs.

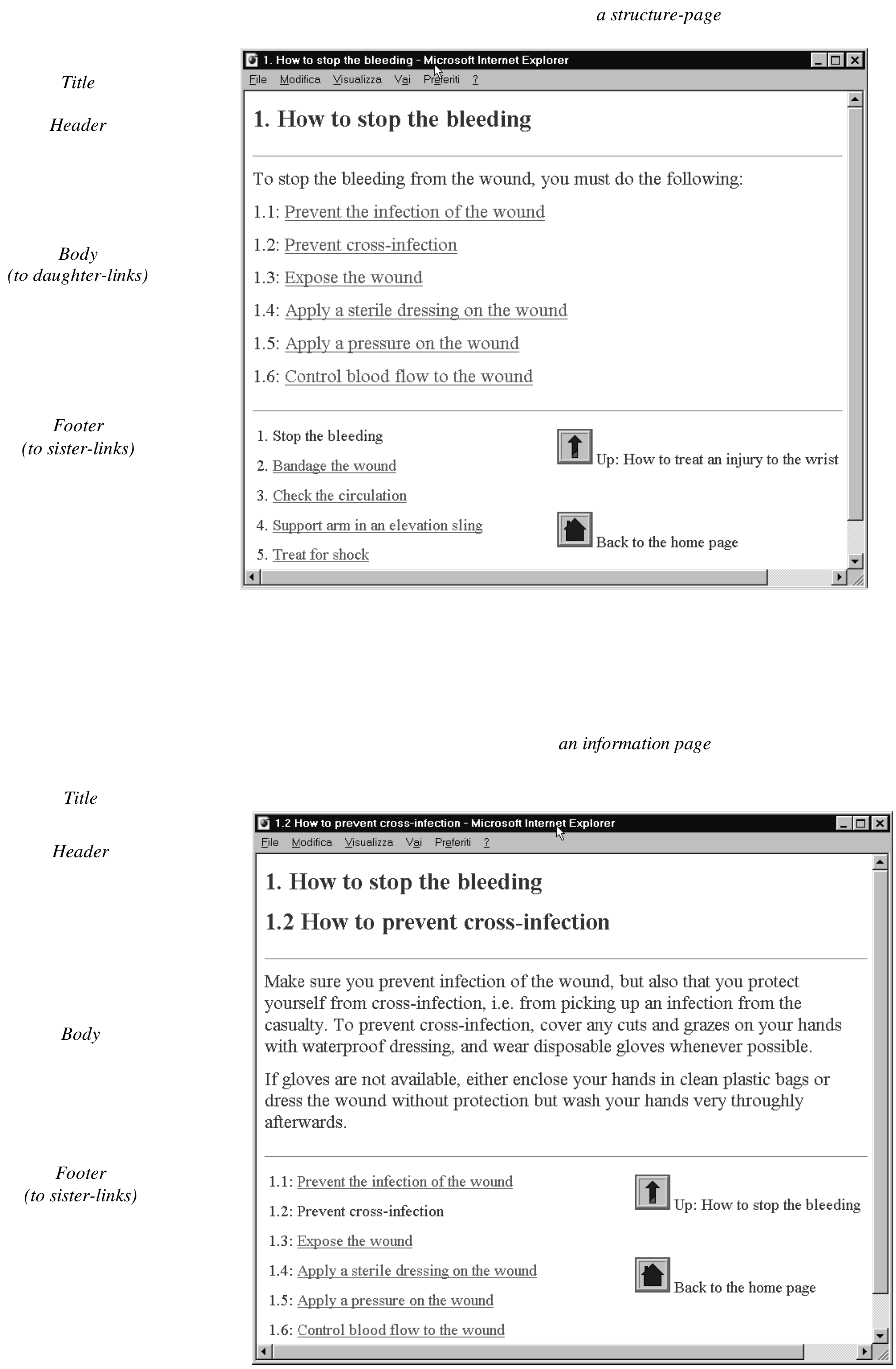

Figure2: two examples of pages generated from the discourse plan in Figure1, in the case of 'expert' users.

We applied GeNet to produce hypermedia in some application domains of limited size: explanations about prescriptions of drugs, a first-aid manual and a domestic blood pressure measurer manual. Before applying it to more complex domains, we wanted to verify, through a systematic evaluation study, whether the main assumptions of the algorithm are reasonable; as we could not test too many hypotheses in a single experiment, we had to select what we considered to be the most critical of them.

3. Evaluation study

Knowledge in the discourse plan may be exploited in different ways to produce a Hypermedia:

- *Body content*: as we said, the Body of information-pages contains information that is associated with a subtree of Dplan of a medium or low complexity; the higher level discourse structure produces, on the contrary, structure pages that are aimed at showing to the users the information items they may find in other pages. These items may be organised in alphabetical order or (as we do in GeNet) in an order that reflects Dplan's goals-subgoals organisation. The two approaches (that we call '*index*' vs '*goal*') reproduce the 'index' vs 'table of content' organisation of content lists in books.
- *Footer content*: the Footer's aim is twofold: it shows links to other pages and indicates how the page is positioned in the hypermedia. Parallel to the two options for the Body content, links in the Footer may be organised in a purely alphabetical order ('*unguided*' mode) or, again, according to Dplan's discourse logic ('*guided*' mode).
- *Navigation mode:* the user may be left totally free to choose the next page to visit ('*free'* mode) or may be enabled to select only one 'next' and only one 'previous' page ('*fixed*' mode).

The purpose of our study was to verify whether the three options (index vs goal, guided vs unguided, free vs fixed) influence the hypermedia usability, and how their effects combine. We measured usability in terms of user satisfaction with the hypermedia and of recall of its content. The study questions are therefore stated as follows:

a. does showing explicitly the *goal structure* of the discourse influence hypermedia usability?
b. is usability improved if *guidance* is provided in the links to follow next, by showing the content of destination pages in a order that reflects the discourse structure?
c. is it preferable to leave users free to navigate in the hypermedia or to show them a fixed path to follow?
d. is it preferable to employ the discourse structure as an orientation tool for navigation by showing it in the Body or in the Footer?

3.1. Method

The hypermedia that we employed in the evaluation study was obtained from part of a hypermedia produced by GeNet, by introducing a few changes aimed at controlling variables in the study; the main change is that all links to other pages are grouped in the Footer by separating to-sister from to-daughter links, and are denoted with anchors. The application domain was a section of a first-aid manual, concerning 'How to treat an injury to the wrist with severe bleeding'. The hypermedia included 24 pages overall, 18 being 'information-pages' and 6 'structure-pages'. Eight types of hypermedia were prepared, by combining the 'goal vs index', the 'guided vs unguided' and the 'free vs fixed' conditions. Structure pages in the goal&guided mode and in the index&unguided mode had a consistent layout of Body and Footer: in the first case, both components reflected the goal-subgoal structure, while in the second case they both showed the items in alphabetical order. The goal&unguided and index&guided modes had, on the contrary, a different organisation of Body and Footer: the goal-subgoal structure was reflected, in the first case, only in the Body, in the second case only in the Footer. These four modalities enabled us to test, at the same time, the effect of showing the discourse structure instead of a simple list of items, the preferred place in which to show it (the Body or the Footer) and the importance of organising consistently these two portions of the page. The free vs index option was added to each of the four mentioned modes, to check the effect of freedom in navigation. To record the path followed by each subject, we built a program that updates a user model every time a page is visited, by adding in this record the page identification number and the initial access time. The 120 subjects included in the study were assigned randomly to one of the eight conditions (15 subjects per condition) and were given 12 minutes maximum to visit the hypermedia; they were then required to compile a questionnaire which included the following categories of questions:

a. *Subjective evaluation of the presentation*, in a scale from 1 to 6:

How free did you free you were in browsing through pages of this hypertext?

How satisfied were you with the options for navigation you were given?

How clear did you find the order of presentation of the material?

How satisfied were you with the order of presentation of the material?

b. Recall of information in the hypermedia:

recall of specific information presented for the various topics (18 'content' questions)

recall of higher level organisation of material, that is of the relationships among topics (6 'procedural' questions)

c. *Background user knowledge:* experience of using computers and knowledge in the first aid domain.

Another source of data was given by analysis of logs, which enabled us to examine the path followed by each subject in the 'free' condition, their matching with the 'correct' path and/or with the path implicitly suggested in the Footer or in the Body, the number of pages omitted or examined several times, and so on.

3.2. Main results

Before presenting the main results, let us say that subjects in the 8 conditions did not differ in their experience of using computers and in their domain knowledge.

a. *Subjective evaluation of the presentation*: people prefer 'free' over 'fixed' navigation and 'guided' over 'unguided' mode: they feel more free to navigate and more satisfied with the 'free' option of navigation, and find that the order of presentation is more clear when links are organised, in the Footer, according to discourse goals rather than in alphabetical order. The organisation of the Body does not influence, on the contrary, their preferences (see **Table 1**).

	Free mode (60 subjects)	Fixed mode (60 subjects)	p
Freedom of navigation	5.2	3.3	.00001
Satisfaction with options for navigation	5.0	3.0	.00001
	Guided mode (60 subjects)	Unguided mode (60 subjects)	
Clearness of order of presentation	5.1	3.4	.00001
Satisfaction with order of presentation	5.0	3.7	.0001

Table 1: main results of the subjective evaluation of the presentation; numbers are averages of score, in a scale from 1 to 6.

b. *Recall of information*: overall, performance is better in the preferred navigation and presentation modes. In 'content' questions, there is no individual significant effect of the three examined conditions; in the 'free' condition, there is a positive effect of consistency in the organisation of Body and Footer (goal&guided or index&unguided are significantly better than goal&unguided and index&guided: 59.3 vs 48.3 percentage of correct answers); the same effect does not exist in the 'fixed' condition. In 'procedural' questions, on the contrary, recall of information is affected by the three examined conditions; freedom of navigation is the most important factor, but also Body and Footer organisations are (**Table 2**). The maximum rate of correct answers to these questions corresponds to the goal&guided&free (ggf) condition (78.9), the minimum to the index&unguided&fixed one (51.1). The ggf condition is always the one in which the maximum score is obtained, both in content and in procedural questions.

(30 subjects in each group)	% of correct answers	p
Goal Index	65.3 54.2	$P < 0.001$
Guided Unguided	62.8 56.7	$.05 < p < .1$
Free Fixed	62.5 56.9	$.05 < p < .1$

Table 2: percentage of correct answers to procedural questions in the 'free' condition

As far as analysis of log files is concerned, results of the 'free' condition show that subjects tend to follow the goal/subgoal structure, in selecting the next link, both when this is given in the Footer and not in the Body (as in the index&guided mode: 9 over 15 cases) and when it is given in the Body and not in the Footer (goal&unguided mode, 6 over 14 cases). However, following a 'correct' path does not influence the rate of correct answer to procedural questions. This analysis also shows that there are more omissions in the 'free' (47% of pages vs 12% in the 'fixed'), 'unguided' (43% vs 15% in the 'guided') and 'index' (40% vs 18% in the 'goal') conditions. These results show that analysis in the goal&guided&fixed mode is more systematic; the combination of the first two factors seems to affect positively learning of subjects, while it is still not clear why the same does not happen for the fixed condition (this is something we plan to investigate in the future).

4. Implications of the evaluation study results on GeNet's algorithm

We could evaluate, in our study, only some of the variables that affect generation of hypermedia from a discourse plan; we could not, for instance, check the criteria employed in GeNet to establish how much information of Dplan to compact in a unique page and how to format it, whether to represent links as anchors or as buttons, whether to label them according to the destination-page content or to the relationship between origin and destination page, and so on. In spite of these limits, the results we obtained enforce us in the main choices we made when designing GeNet. They prove that showing the goal-subgoal discourse structure affects positively both the subjective evaluation and the user performance. A notable result was that discourse structure improves learning not only of the relationships among discourse parts (what we call the 'procedural' knowledge) but also of the individual chunks of information that are displayed in an information-page (what we call 'content' knowledge). A possible interpretation of this finding is that the goal structure is more beneficial for constructing a mental model of the task as a whole, in addition to facilitating understanding and memory of the sequence of steps of the task. The importance of organising consistently the main page components (the Body and the Footer) was proved as well: as people tend to build their decision of how to navigate in the hypermedia from information that is displayed in a logical order (even when this is shown in the Body rather than in the Footer), inconsistency in the two parts' organisation risks to disorient them.

References

R Dale, J Oberlander, M Milosaljevic and A Knott: Integrating natural language generation and hypertext to produce dynamic documents. *Interacting with Computers*, in press.

B De Carolis, F de Rosis, F Grasso, A Rossiello, D C Berry and T Gillie: Generating recipient-centered explanations about drug presctiption. *Artificial Intelligence in Medicine*, 8, 1996.

F de Rosis and B De Carolis: Automatic hypermedia presentational planning. *Submitted for publication.*

T R Gruber, S Vemuri and J Rice: Model-based virtual document generation. *International Journal of Human-Computer Studies*, 46, 1997.

E H Hovy: Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63, 341-385, 1993

D J Jonassen: Effects of semantically structured hypertext knowledge bases on users' knowledge structures. In C Mcknight, A Dillon and J Richardson (eds): *Hypertext, a psychological perspective.* Ellis Horwood, 1993.

W L Johnson and A Erdem: Interactive explanation of software systems. *Automated Software Engineering*, 1997

W C Mann, CMIM Matthiesen and S Thompson: Rhetorical structure theory and text analysis. *ISI Research Report 89-242,* 1989.

M Milosaljevic: Content selection in comparison generation. *Proceedings of the 6th EWNLG*, Duisburg, 1997.

J Moore: *Participating in explanatory dialogues: interpreting and responding to questions in context*. The MIT Press, 1995.

P Wright and A Lickorish: The influence of discourse structure on display and navigation in hypertexts. In N W Williams and P Holt (eds): *Computers and writing*. Oxford Intellect Limited, 1989.