

Constraints on Layout in Multimodal Document Generation

John Bateman

University of Bremen
Bremen, Germany
bateman@uni-bremen.de

Judy Delin

University of Stirling
Stirling, Scotland
j.l.delin@stir.ac.uk

Patrick Allen

University of Bradford
Bradford, England
p.t.allen@bradford.ac.uk

Abstract*

A model of layout in illustrated documents must be informed by the rhetorical structure of the content. However, real-life production constraints often force an 'imperfect' realisation of the rhetorical structure on the page. Navigational elements must consequently be introduced into the layout to support the reading path. Layout structures, including the placement of the navigational elements, are then created based on the content segmentation and importance assignments suggested both by the rhetorical structure and by generic constraints for the document genre and production media.

1 Overview

In this paper, we analyse the sources of constraint on generation of documents that include text and graphics. We argue for a particular decomposition of the representation of multimodal documents that allows us to describe both similarities and differences in information presentation that arise when moving across presentation genres. The presentation genres we illustrate in the current paper are the printed and electronic versions of daily newspapers. Central to our approach is the recognition of constraints that are due to the nature of the medium being used for information presentation and to the processes of production employed.

Space will preclude detailed analyses of all of the presentations that we show here; the reader is therefore referred both to the oral presentation that this paper is to accompany

and to longer versions in preparation. Space similarly precludes adequate discussion of previous work in the area; this is in no sense to be taken as an indication either of lack of interest or awareness on our part. The present paper serves only to show some types of analysis that we consider crucial for multimodal presentation; we introduce these representations and discuss some of their interactions in order to provide a basis for further, more detailed, discussion.

2 Levels of representation for multimodal generation

Our representation for multimodal generation assumes five levels of structure: content structure captures the structure of the information to be communicated, rhetorical structure represents the rhetorical relationships between content elements, navigation structure captures the ways in which reader's pathways through the document are supported, layout structure describes the nature, appearance and position of communicative elements on the page, and linguistic structure captures the structure of the language used to realise the layout elements. Our scheme combines significant input from two main areas of research: first, the use of rhetorical organisation for multimodal and hypermedia presentations as adopted in a number of generation systems (e.g., Wip, PPP, Comet, GeNet, ARIANNA); and second, a revision and extension of the levels suggested by Waller (1987) in his already highly elaborated account of genre in document design. Waller's work has been particularly influential for us in its recognition of the crucial role that production constraints come to play in all real-life document design. This aspect has not yet received sufficient attention in approaches to the generation of multimedia documents.

* The work reported here was undertaken within the UK ESRC project GEM : Genre and Multimodality. Website : '<http://www.gem.stir.ac.uk>'.

3 The role of rhetorical structure in document generation (and where it ends)

Rhetorical structure in the form of Rhetorical Structure Theory (RST: Mann and Thompson 1988) has been adopted for multimodal information presentation in a number of systems and has served as one important motivator for layout. However, its relationship to the final layout of realistic documents remains unclear. The most common approaches are to take an RST structure of the intended communication and to ‘prune’ this in particular ways: for example, by leaving lower salience elaborations as hyperlinked material (possibly unexpanded unless called for), or by deciding a particular degree of granularity and collapsing all RST subtrees from a given depth to form a single ‘page’ content. In a further exploratory study of the relationship between RST-style analyses and professionally produced page layouts, Bateman et al. (subm.) found substantially more complex relationships between RST and ‘layout’. They show that any given rhetorical structure supports a host of very diverse multimodal documents and that some substantial ‘rearrangements’ of an hypothesized RST must be entertained to account for this.

Finding further constraints on this process forms the main goal of the GEM project, some results of which we report here. We have found constraints to be of several kinds, and our knowledge of them is derived principally from consultation with designers and from studying document production in real-life situations: constraints arise from the nature of the artefact being produced, from the production technology, and from the proposed purpose of the document. These practical constraints may not merely temper the realisation of the rhetorical structure, they may actually subvert it. We propose that this dislocation of rhetorical structures is then mediated by means of *navigation* structures. Navigation structures are typically expressed by the introduction of navigation-specific layout elements that are not

directly motivated by a consideration of the assumed RST structures.¹

4 An Illustration: *The Herald*

We begin the illustration of our approach by way of a news story from a Scottish daily newspaper, *The Herald*. The story is from the front page, and describes a proposed sale of the Cuillin hills on the Isle of Skye. As indicated above, each such page receives in our account an analysis at several levels of description. Of these, that nearest the material as it appears on the page is the layout structure (cf. Reichenberger *et al.*, 1995). We indicate the layout structure of the Cuillins story in Figure 1 (from page 1 of the newspaper) and Figure 2 (from the continuation of the story on page 2).

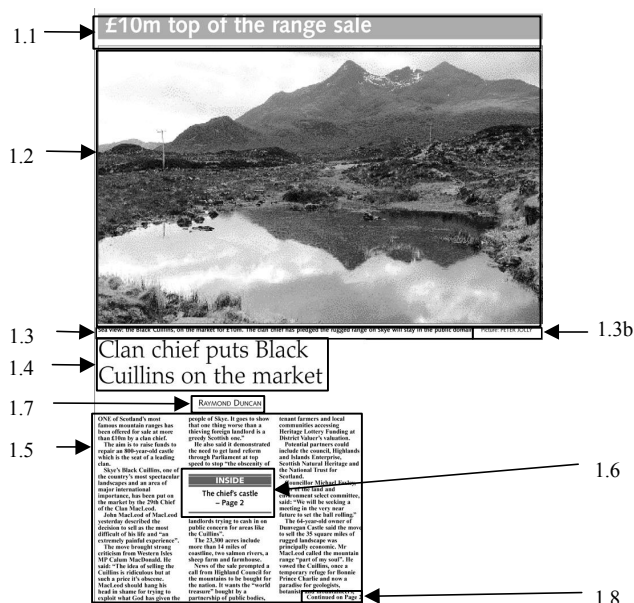


Figure 1 Cuillins story, page 1, with terminal layout elements

A full description of layout structure includes the nature of each element, its hierarchical relationship with other elements, its typography, size, design, weight and so on

¹ Since we are using an RST-style of representation for both textual and graphical material, this requires some extensions to the basic RST account along similar lines to those argued by, e.g., André (1995). We relate this to similar concerns in semiotics (e.g. Barthes 1977) and professional document design (e.g. Schriver 1997).



Figure 2: Cuillins, page 2, with terminal layout elements

but, for the purposes of this paper, we must generally content ourselves with simply labelling the leaves of the structure and showing their location on the page.

The rhetorical structure of the Cuillins story is signalled by the two lead paragraphs: 'One of Scotland's most famous mountain ranges has been offered for sale at more than £10m by a clan chief. The aim is to raise funds to repair an 800-year-old castle which is the seat of a leading clan.' The article is structured around this central nucleus—itsself composed of a *purpose* relation between the sale of the hills (nucleus) and the repair of the castle (satellite)—and the remaining elaborations involving the decision to sell, economic consequences, more about the castle, etc. Note that our RST structure is neutral regarding the mode of the 'text spans' it describes: in this case, they can be photos or text. A sketch of the RST analysis is given in Figure 3; this analysis supports standard results concerning the analysis of the structuring of news articles (cf. White, 1997, Delin, 2000).

Perhaps the first thing to note here is the nature of the split between pages (indicated in

the figure by a dotted grey vertical line). Whereas it is usually of no concern in the rhetorical analysis of a story how that story happens to be printed, in the genre of newspapers page-selection is of course highly significant: 'front page news' is precisely that. In the current story, there has been a decision to 'turn' or continue the story on page 2, not on rhetorical grounds, but because it was too long to fit on page 1. The competing top story, advertisements, and elements of navigation structure such as the index, 'teasers' about other articles, and the masthead of the paper all serve to limit the remaining available space. The decision to extract parts of the rhetorical structure for pages 1 and 2 is therefore practical and generic: there are never four competing photographs on the front page, so some must be relegated. The reason page 2, rather than some other page, is chosen, is also practical: page 2 is printed just before page 1 last thing at night, and stories from page 1 are therefore often turned to page 2, rather than to pages later in the paper.

In moving, then, from the rhetorical level of analysis to the layout structure, the original

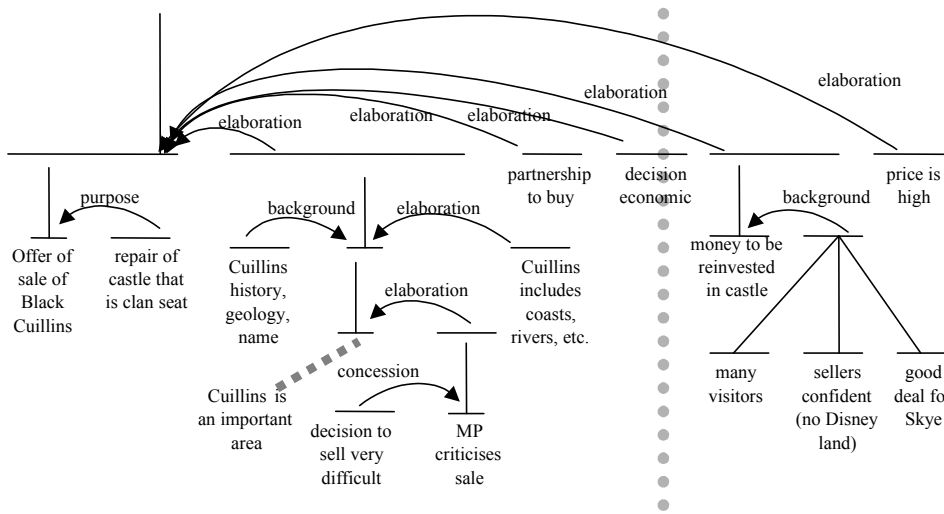


Figure 3: RST analysis of Cuillins article

units of the rhetorical structure cannot be respected. For this reason, it is usual (i.e., for this genre) to introduce further elements into the final page which serve to manage the reading path. There are several of these on both pages. First, the basic page split is signalled by the matching pair of forward looking element 1.8 ('Continued on Page 2') and backward looking element 2.11 ('Continued from Page 1'). More interestingly, the entire element 1.6 points towards page 2 as a news element in its own right: "Inside: The Chief's Castle—Page 2" as an inset box. This prominence of the castle goes back to its status as part of the nucleus of the article as a whole. This also motivates the prominence of the pictures of the castle chosen for page 2: that part of the rhetorical structure is in effect predominantly realised by photographs, the importance of the castle in the RST being realised by the prominence of these pictures.² Both of these navigational elements are generically constrained, in the sense that pointers of this kind and form are expected for newspapers and related genres where information may be divided across pages and it may also be necessary to 'push' the readers to continue reading. This RST analysis, however, does not and should not predict these elements. The structuring that is imposed on content in this

² André (1995) argues that graphical elements must themselves be deconstructed for analysis, a point also clear in the semiotics literature and with which we entirely agree.

medium does not respect any natural organisational structure inherent in the content.

In addition to these page navigation elements, there are additional navigation elements with a variety of functions. We discuss two here. First, it is common to 'break out' particular chunks of information as individual textually expressed layout elements. The motivations for such chunks are generally found in the 'background' nature of the information and the fact that it may potentially be linked to several points in an RST structure. An example is element 2.3 (realized as a bullet list) containing historical and geological information: when information of this kind is presented, it is common for it also to be attached to the navigational device of a 'header' (cf. element 2.4: 'The Cuillins'). Second, the fact that the article has been split over two pages also motivates an indication of the unity of the story: so here we have the story identification 1.1 ('£10m top of the range sale') picked up again on page 2 as 2.1 ('Skye's £10m top of the range sale').

Noteworthy here is the inherently nonlinear structure of even traditional print publications. In our model, we treat all such navigational elements as being generated independently of the content and RST analysis and as a result of the three sources of constraint given above. It is then perhaps surprising to find that the actual electronic versions of newspapers often do *not* have the same degree of nonlinearity: they still generally lag behind the print versions. This is due to a mixture of reasons, both technical and economic.

The structure of the *Herald's* website, for example, is quite straightforward. The frame is divided into three vertically: on the left are general or news item related links, on the right are links to particular sections of the electronic newspaper, and in the centre is a vertically scrollable region with the currently 'in focus' article. The text is unchanged from the newspaper version. Many of the elements present in the newspaper do not appear however. These missing elements include not only the no longer necessary navigational elements (there is no split across pages), but also the background information about the Cuillins, the photographs, and the story identifier 'top of the range sale' (perhaps providing further support that this is really just a navigational element). As is also common with the newspaper electronic versions, there is also a bushier hierarchy of news items: the newspaper homepage immediately divides the content into three broad categories: *News*, *Sport*, and *Inside*. The Cuillins story is then only reached as the second in a list of articles under the *News* link.

5 A further illustration: *The Guardian*

For a further illustration of the complex relations between RST and layout, as well as a far more complex web example, we now consider the print and electronic versions of *The Guardian* from April 11th, 2000. We will again focus on a single 'story', that concerning events in Zimbabwe, although in this case the story is presented as several more or less independent articles. The Zimbabwe coverage in the newspaper is split into two main parts: the front page information on how police action has been ruled out against black farm squatters, and a page 3 story about white reactions to the situation and white involvement in the Movement for Democratic Change (MDC). It is this involvement that Robert Mugabe cites as the reason that whites are culpable in the crisis and not deserving of police protection, since the right of whites to continue to own land after the dissolution of Rhodesia was based on their abstention from political influence.

A schematic representation of the front page and its terminal layout elements is shown in Figure 4. As with all newspapers, there is a clear ordering in terms of 'news value': the main story of the day selected is a murder story involving a booby trapped farm: this story receives two pictures, 3.14 and 3.17: the former being large and central. The Zimbabwe article on this page is self-contained and was written by two journalists in Brussels. It is presented at the bottom of the page and is composed of elements 3.19a, 3.19b, 3.20, 3.21 and 3.22. This article is presented with the standard Guardian 8-column per page format, spreading from the left over 6 columns. There are no text divisions apart from paragraphs and columns.

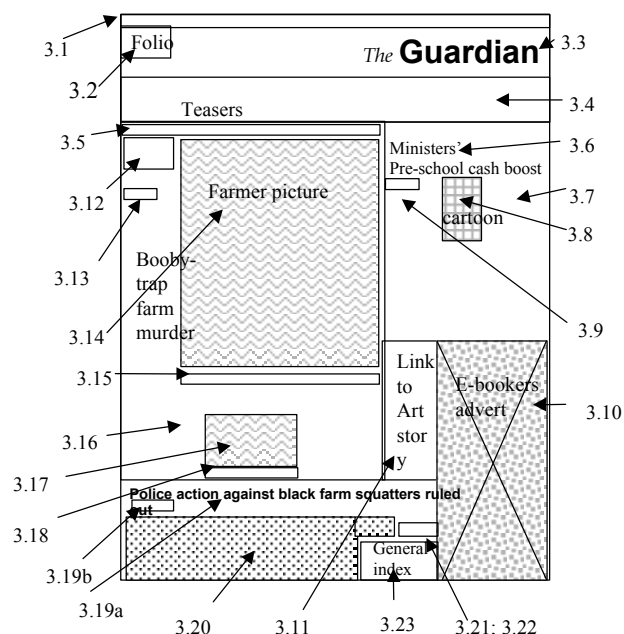


Figure 4. Terminal layout elements of the *Guardian* Front page, April 11th, 2000.

Of more interest to us here, however, are the layout elements immediately to the right of the text block: 3.21 and 3.22. These are pointers on to further articles as shown in Figure 5. The pointers are to three distinct further sources of information: an article in the printed newspaper on page 3, a 'special report' on the newspaper's website, to which we return in a moment, and to a particular 'audio dispatch' present on that website. Note again the explicit use of *both* story identification

(‘On the brink’) and physical page pointers (page 3); this is quite common in both newspapers (e.g., “Leader comment Page 18”, “Scottish bid Page 20”).

On the brink, page 3
 Special report and audio dispatch
 from Chris McGreal in Harare at
www.newsunlimited.co.uk/zimbabwe

Figure 5. Linking material concerning Zimbabwe on *Guardian* front page

Page 3 of this edition of *The Guardian* is devoted entirely (apart from a large advertisement) to the Zimbabwe story. Its layout and division into terminal layout elements is shown in Figure 6. To note here is that the page is dominated by a single photograph that would easily have been capable of standing as a front page picture: it is

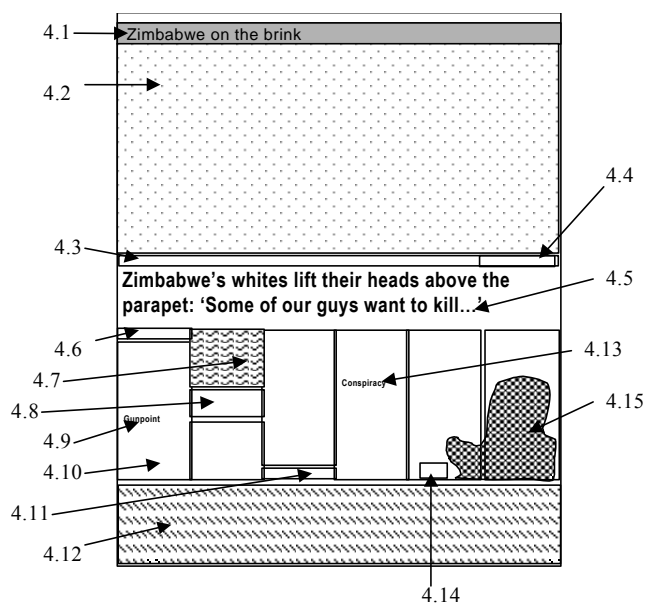


Figure 6. Guardian printed edition, page 3 coverage of events in Zimbabwe

likely that had there not been a suitably striking photograph available for the main lead on page 1, then the Zimbabwe picture of page 3 would have taken its place (as it indeed does on the web version of the newspaper). Such minor disassociations of the news hierarchies as presented in text and in pictures are common—that is, it is not necessary, although

clearly preferred, that the main page 1 picture is related to the main page 1 news story.

The page 3 coverage is a typically complex piece of newspaper layout: it includes the story identifier introduced in the pointer 3.21 (Figure 5) set out across the top of the page (4.1), a large photograph (4.2), a story headline (4.5), a smaller map of Zimbabwe and surrounding countries (4.7), a broken-out quotation in larger font (4.8), an inset portrait (4.15), captions and web addresses. The text of the article itself (4.10) is in addition broken down into three typographically signalled sections by the ‘crossheads’ *Gunpoint* (4.9) and *Conspiracy* (4.14). These headers do not in fact represent particularly important components of their typographical sections, which distinguishes them in function quite clearly from content-oriented section headings in other genres. Taken by themselves they provide another focusing of information that is carried more by the page layout as a whole as by the

Key:

- 4.1 Page caption
- 4.2 Photo of ‘poor blacks’
- 4.3 Caption: ‘Dozens of white-owned farms in Zimbabwe have been occupied by poor blacks claiming the land as their own, with the support of Robert Mugabe’s government, as the country’s political crisis deepens’
- 4.4 Photograph: Marion Herud
- 4.5 Headline
- 4.6 By-line: ‘Chris McGreal in Wedza, Zimbabwe’
- 4.7 Map of Zimbabwe in context of bordering countries, inset of position of Harare
- 4.8 Quotation: ‘I am not a party political person but things are so bad I had to get involved. If democracy does not return, the economy of this country will collapse’. ‘Monty’ Montgomery, white ex-policeman
- 4.9 Crosshead: ‘Gunpoint’
- 4.10 Text of article
- 4.11 ‘Links: <http://www.mdc.co.zw> Movement for Democratic Change’
- 4.12 Advert for charity
- 4.13 Crosshead: ‘Conspiracy’
- 4.14 Photo caption: ‘Morgan Tsvangirai, head of the opposition MDC: white support is crucial to his attempt to topple Robert Mugabe’
- 4.15 Inset photo of Morgan Tsvangirai

content of the text: that is, ‘gunpoint’ is in tune with a motif of violence and ‘conspiracy’ flavours the interaction of Morgan Tsvangirai and the Zimbabwean whites. The contents of the text are in contrast focused more on an account of the political situation and the views and reactions of those involved. We show an RST analysis of the material and presentation of pages 1 and 3 combined in Figure 7. The

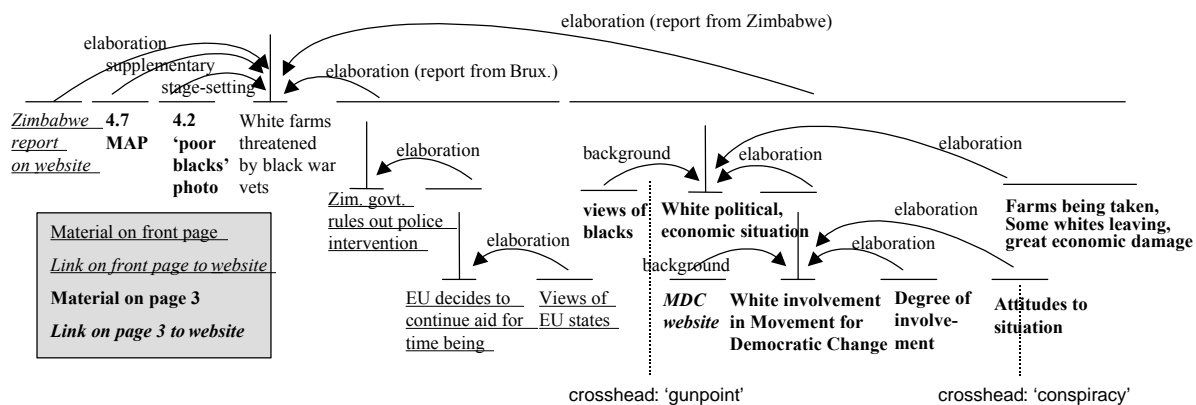


Figure 7: Rhetorical Structure Analysis: Zimbabwe material, *The Guardian*, 11 April 2000

more typographical than content related nature of the crossheads is also indicated in this figure, where it can be seen that they cross-cut the RST organisation somewhat arbitrarily.

In order to discuss the relationship between rhetorical organisation and layout decisions in more detail and to relate more readily to the rather complex *Guardian* website, we need to set out the layout structure of page 3 in more detail. This is shown in Figure 8 and depicts the hierarchical decomposition of the elements of the page, their graphical dependencies (dashed arcs), and relative assignment of importance (thick connecting lines). This should also emphasise again the non-trivial relationship between the construction of a page and its rhetorical organisation.

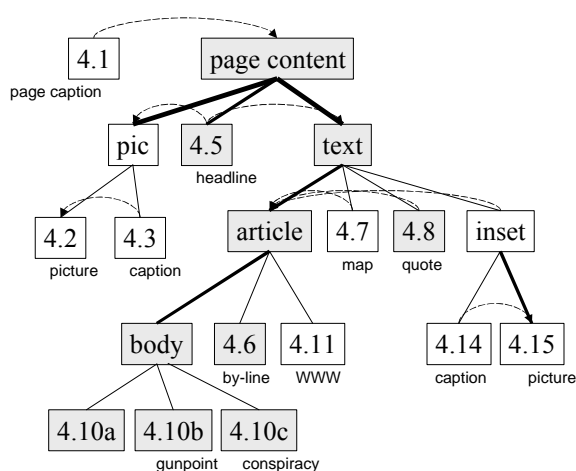


Figure 8. Layout structure for page 3 of *Guardian*

When we turn to consider the structure of the corresponding edition of the electronic *Guardian*, an interesting NLG-relevant question is ‘where should we start?’ If we start from the RST, then we shall see that certain of the decisions that appear in the layout structure have not been made and need to be repeated. However, if we start from the layout structure, then this overcommits because some of the elements introduced are motivated by the artefact constraints of the print medium. The electronic *Guardian* in fact shows some simplifications similar to those of the *Herald*: in particular, it has been decided (largely for practical reasons) that there should be no inclusion of photographs within stories—although there are some pictures and graphics used as the larger icon-graphics that identify particular stories, themes, and sections.

The elements of the layout structure that find their way to the web version are indicated in Figure 8 by shading. Predictably, the purely print-oriented navigational elements have been omitted (4.1). The textually-realized element 4.8 (a quotation extracted from the body of the text) has, however, *not*, been omitted. This is present as a second-level headline in the web version. Furthermore, just as is the case with the print version, the layout structure related to the content is itself embedded within a generically more stable layout structure that sets the overall style of the page—be that a webpage or a printed page. The webpage style of *The Guardian* is more sophisticated than that of *The Herald* particularly in its grouping

of stories together thematically and its inclusion of numerous off-site links that provide substantial sources of background information concerning the themes addressed in its articles. However, these links are clearly added by hand on the basis of topics addressed in the stories: there is no linking *within* stories such as might be supported by using an RST-like mark-up of those stories. The standard newspaper-genre style of presenting the main news item in the initial paragraph is also used extensively for producing the information-bearing links out of which the website is constructed.

The basic style of *The Guardian* web pages is to have a banner across the top, a horizontal navigational menu, a generally graphic or photographic horizontal block in colour presenting either teasers or section headings with a photograph, and then two columns: the left providing extensive links and navigation via searching, the right providing the body of the page. The body is either a list of links to articles, grouped by topic area (UK, Politics, International), or individual articles. The body scrolls vertically and is kept narrow, presumably for ease of reading. It is in the body of the article pages that we find again the elements of the layout structure as sketched in Figure 8 above.

The linking within the site is quite ambitious, however. The individual Zimbabwe articles can be reached directly from the International News link, but also collectively as contributions to a *Zimbabe: special report*. This report includes not only both articles mentioned above but also a third article from the day before. In addition, linking back to earlier articles in order to promote examination of the development of an issue across time is strongly supported. This is sufficient to show a very different *function* for the electronic version: it is not simply a web version of the newspaper as is the case with *The Herald*. There is a serious attempt to provide more of a database of articles and information with the necessary navigational interlinking that this requires. In providing an analysis of the website, therefore, this difference in genre needs to be explicitly considered, for without it

the motivation for design decisions will not be given justice.

6 Conclusions

In this short paper, we have been able to do little more than sketch a selection of some particular kinds of representations and issues that will need to be addressed when considering the generation of realistic multimodal documents. The relationship between an RST structure and their presentation is more complex than has generally been assumed; often stories are realised differently due to the differing practical constraints inherent in the genre and its technology. Navigation elements are then used to manage a disposition of elements that may not be ideally supportive of rhetorical structure.

The study of medium and production constraints are also an essential precondition for empirical work: it does not make sense to try and model particular styles of layout if the determinative constraints for those layouts have been left out of the picture.

7 References

- André, E. (1995). *Ein planbasierter Ansatz zur Generierung multimedialer Präsentationen*, DISKI 108, Sankt Augustin: Infix.
- Barthes, R. (1977). *Image, Music, Text*. London: Fontana.
- Bateman, J., T. Kamps, J. Kleinz and K. Reichenberger (subm.). The Dart_{bio} system: constructive text, diagram and layout generation for information presentation. Submitted.
- Delin, J. (2000). *The Language of Everyday Life*. London: Sage.
- Mann, W.C. and S.A. Thompson (1988). "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text* 8 (3): 243-281.
- Reichenberger, K., K.-J. Rondhuis, J. Kleinz, and J. Bateman (1995). Effective presentation of information through page layout: a linguistically-based approach. (<http://www.cs.tufts.edu/~isabel/reichen/page-layout.html>)
- Schrifer, K. (1997). *Dynamics in Document Design*. New York: Wiley.
- Waller, R. (1987). *The typographical contribution to genre*. Unpublished PhD dissertation, University of Reading.
- White, P. (1997). Death, disruption and the moral order: the narrative impulse in mass-media 'hard-news' reporting. In: F. Christie and J.R. Martin (eds.) *Genre and Institutions: social processes in the workplace and school*. London: Cassell. 101-133.