

# From Genre to Text Critiquing in Multimodal Documents<sup>1</sup>

John Bateman<sup>2</sup> and Judy Delin<sup>3</sup>

**Abstract:** We suggest that the genre of multimodal documents resides in the way each document harmonises five levels of structure: content structure, rhetorical structure, layout structure, navigation structure, and linguistic structure. In addition, a further determinant of genre is the way in which documents satisfy three kinds of constraints: canvas constraints, production constraints, and consumption constraints. Document genres are then conceptualised as clusters of specifications composed of descriptions at each of the five levels that conform in characteristic ways to the three kinds of constraint. Using a worked example, we show how documents can be analysed in terms of these goals and constraints and ways in which the framework may be used as a means of critiquing document design, producing predictions for document usability.

## 1. Introduction

Document genres are not distinct entities: they are interrelated. We propose that eight parameters<sup>4</sup> together define a ‘space’ of possible identities for documents, electronic or paper, in which these movements and relationships take place. Building on and extending the work of Waller (1987), we suggest that the genre identity of a document arises out of the way in which the document satisfies communicative goals at five levels:

<i>Content structure</i>	the structure of the information to be communicated;
<i>Rhetorical structure</i>	the rhetorical relationships between content elements; how the content is ‘argued’;
<i>Layout structure</i>	the nature, appearance and position of communicative elements on the page;
<i>Navigation structure</i>	the ways in which the intended mode(s) of consumption of the document is/are supported; and
<i>Linguistic structure</i>	the structure of the language used to realise the layout elements.

In addition, however, the final appearance of a document rests on how the document satisfies a number of potentially competing and/or overlapping constraints:

---

<sup>1</sup> This research is funded by ESRC grant no. R000238063: ‘Genre and Multimodality (GeM): A Computer Model of Genre in Document Layout’.

<sup>2</sup> FB10, Sprach- und Literaturwissenschaften, Universität Bremen, 28334 Bremen, Germany. Email: bateman@uni-bremen.de

<sup>3</sup> Department of English Studies, University of Stirling, Stirling FK9 4LA. Email: j.l.delin@stir.ac.uk and Enterprise IDU, 79 High Street, Newport Pagnell, Bucks. MK16 8AB. Email: judy.delin@enterpriseidu.com

<sup>4</sup> A more detailed discussion of these levels is presented in Delin *et al.* (subm.) For a further discussion of the role of rhetorical structure in layout and the effect of production constraints, see Allen *et al.* (1999) and Bateman *et al.* (2000).

<b><i>Canvas constraints</i></b>	Constraints arising out of the physical nature of the object being produced;
<b><i>Production constraints</i></b>	Constraints arising out of the production technology;
<b><i>Consumption constraints</i></b>	Constraints arising out of the time, place, and manner of acquiring and consuming the document.

In our view, document genre can be described by analysing the way in which the goals and constraints are satisfied in every document. Genres are identified by looking at how documents ‘cluster’ in terms of the way they do this. We will illustrate the basic approach by means of analysing the text in Figure 1, which we call ‘the tiger page’. First, we look at the five levels of structure in turn, and then we examine how the document conforms to the three sources of constraint.

## 2. The Tiger Page: Analysis

### 2.1 Content Structure

The content of the tiger page can be organised into five main segments: an overview (given in bold type just under the title), the appearance and functional description of the Bengal tiger, its size, mating habits, etc, physical comparisons with other tigers, and related species. Figure 2 represents the hierarchical relationships between the content segments of the tiger page. This hierarchical representation makes clear what parts of the content are related, and does not discriminate between what is presented linguistically and what graphically. We can now see that there are two main sets of facts provided in the document: a larger set relates to the Bengal tiger, and the other relates to other tigers and other animals.

The content representation does not seek to capture how the content is argued, nor how it is represented on the page. Statements, for example, about the Siberian tiger, other big cats, or about the height of the average human, are treated in the content analysis as factual propositions. It is not the job of the content representation to make clear *why* this page features such content, or what its role is in the argumentation presented (for example, for comparison with other animals or tigers). It is important to allow the analysis to reflect the fact that content and rhetorical presentation can vary independently of one another.

### 2.2 Rhetorical Structure

The *rhetorical structure* of the text (language and pictorial elements) is the way in which the content is argued and the various segments interrelated textually or graphico-textually. In order to analyse rhetorical structure, we use a framework known as Rhetorical Structure Theory or RST (Mann and Thompson 1988). RST provides a set of concepts and a notation to express the way in which segments of text are hierarchically related to one another in the presentation of a coherent text. We must, of necessity, adapt it and allow it to represent graphics as well. For the purposes of explication, we will restrict the analysis we present here to the main blocks of the page: the large tiger picture and accompanying text and inset pictures, and the ‘vital statistics’ panel. What is obvious here is the centrality of the tiger picture, and the role which smaller elements play (descriptions of the functions of its attributes, such as coat, teeth, claws) in

elaborating upon that central image. Within the smaller elements there are also notable rhetorical relationships: there is a description of the tiger's coat and an explanation of its function (in RST terms, a *purpose* relationship between those two parts of the argument), and there is a description of its eyes and how they work (a *means* relationship), for example. RST distinguishes between the part of the rhetorical relationship that is core, central or *nuclear*, and the peripheral, additional part that is referred to as a *satellite*. For example, in the segment describing the tiger's coat, we can further analyse the text into a nucleus that states that every tiger has a unique pattern of stripes and a satellite that provides the purpose of these stripes.



## PROFILE BENGAL TIGER

### VITAL STATISTICS

WEIGHT	180–265kg
LENGTH Head & Body	1.9–2.2m
Tail	80–90cm
SHOULDER HEIGHT	90–95cm
SEXUAL MATURITY	Female 3–4 years; male 4–5 years
MATING SEASON	Winter to spring
GESTATION PERIOD	95–112 days
NUMBER OF YOUNG	2 to 4
TYPICAL DIET	Sambar deer, chital deer, buffalo, wild pigs, gaur and monkeys
LIFESPAN	Up to 26 years in the wild

**The tiger's powerful muscles and massive build are the keys to its hunting success: in seconds, it can knock down and kill prey that weighs nearly a tonne.**

**EYES**

The tiger's night vision is six times better than our own, aided by a mirror-like layer at the back of the eye that reflects extra light.

**EARS**

Hearing is the tiger's sharpest sense. The white spots behind the ears help tigers to identify one another in the dark jungle.



**COAT**

Every tiger has a unique pattern of black stripes on a deep-orange coat. This breaks up the outline of the body in dense cover.



**TEETH**

The tiger uses its long canine teeth to stab and kill prey. The molars behind them act like scissors, slicing strips of flesh from a carcass.



**CLAWS**

The claws are used to grip prey and to scratch trees. They retract when the tiger walks, to remain sharp and to allow it to stalk prey silently.



**CREATURE COMPARISONS**

The Siberian tiger is the world's largest cat: adult males may reach more than 3.6m in length. The coat is shaggier and paler than that of the Bengal tiger, equipping the tiger for its icy northeast Asian habitat. The Siberian tiger occupies a vast territory and preys mainly on wild pig. There may be no more than 150 left in the wild.

Bengal tiger



Siberian tiger





**RELATED SPECIES**

- There were once eight subspecies of tiger. Today, only five remain: the Bengal, the Siberian, (below left), the Sumatran, (below right), the Indochinese and the Caspian. The tiger's closest relatives are the other 'big cats' in the genus *Panthera* — the lion, leopard, jaguar and snow leopard.




Figure 1: 'The Tiger Page'



Figure 2: Hierarchical representation of the tiger page showing content relations.

The diagram in Figure 3 represents this part of the tiger document in RST. The curved lines point from satellites to nuclei, so that a concentration of lines converging on one segment shows the more 'central' elements of the document. In some list-like relationships, there is no nucleus-satellite relationship, and the elements are referred to as being in a 'multinuclear' relation, designated by straight lines (as is the case with the tiger's attributes in the second elaboration segment). The segments with vertical lines above represent nuclei.

The main structure of the document as exposed by the RST analysis is based on two main elaborative segments around the nuclear tiger picture: the tiger's body parts, and other attributes. The relationships with other tigers and other animals are presented as background. At this stage, we can see how the rhetorical structure draws upon the content structure to make an argument. Hierarchy is clear within it: we know, for example, how all the elements belonging to the functional description of body parts relate to one another, and that those elements are themselves complex. For example, since the segment describing the function of the tiger's stripes is a 'purpose' satellite of the nucleus that first mentions the stripes, the nucleus should be presented first and the two should not be split up. Similarly, the three elements describing the teeth have an internal hierarchy, with the two statements about canines and molars subordinate to the picture showing the tiger's mouth.

### 2.3 Layout Structure

The next stage of the analysis provides a detailed characterisation of the concrete layout decisions that have been made in a given document. Layout is described in terms of a hierarchically organised set of layout elements and the properties (graphical, typographical) of those elements. Just as different rhetorical organisations can be selected for communicating a given content structure, diverse layout structures can be selected for any given rhetorical organisation (cf. Bateman et al. (subm.) for extended discussion). Again, this is the primary motivation for maintaining these as distinct descriptive levels.

The tiger page falls broadly into three blocks: the top left block (we will call this block A), with labels organised around the central illustration of the tiger, the right-hand column (B), consisting of two panels of information, and a third block (C) consisting of text and illustrations below block A. Each block has a different organising principle. In block A, labelled text entries and detailed illustrative call-outs are ranged around the central illustration, as closely as possible to the relevant feature described, although positioning is clearly also determined by available space.

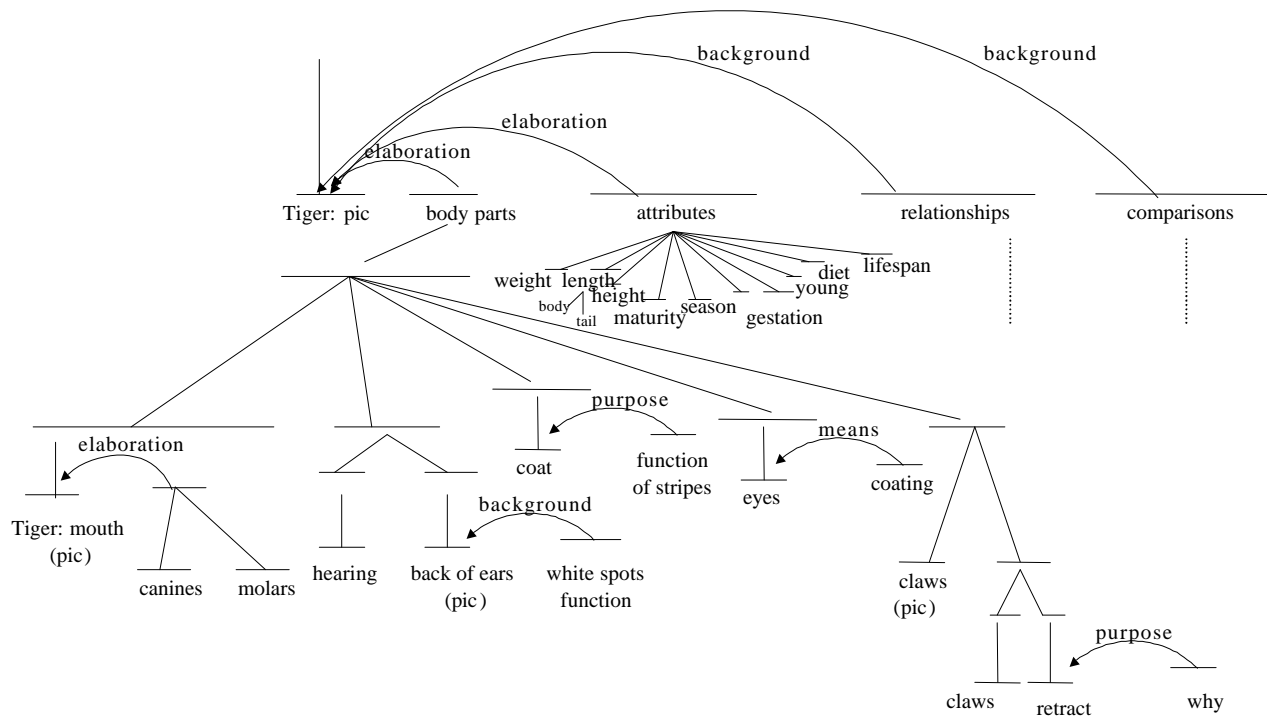


Figure 3: Full-blown RST tree showing purpose, elaboration, background and means relations between content nuclei and their associated satellites.

The text around the tiger is in sans serif 10pt mixed case, ranged left, ragged right, while the labels are in the same typeface, 12pt small caps, large initial cap, ranged left, ragged right (note that the document is reduced in reproduction). It is organised to preserve alignment along the top and bottom of the text and to keep white space around the large illustration. The illustrations are full-colour artwork, the smaller ones with a 1pt border. In block B, the first element is another labelled list except this time more simply organised in a column, entry labels are in a serif font, small caps, matching the title of the whole page (although smaller) rather than the labels of the adjoining block B. Spacing of the entries is tight against the red title 'vital statistics' but then generous, including an extra large space between 'typical diet' and 'lifespan' to allow space for the intruding tail of the tiger. The lower panel of block B has a darker red title and a bright red bullet before the bold, sans serif text, and two square photographs. In block C, the title 'Creature Comparisons' is in black with a small graphic, with light sans serif text ranged left and the full-colour and line drawing graphic to the right. The whole is bordered in dark red.

The breaking-down of the layout structure into distinct elements, and elements within those elements, therefore allows us a clear hierarchical analysis for the page, and an association with typographical characteristics with each of the elements.

## 2.4 Navigation Structure

The navigation structure consists of the 'signposts' that enable the rhetorical structure as it has been disposed upon the page to be understood. Given the rhetorical decomposition above, we would expect major segments to be given a label of some kind: this is the case with 'vital statistics', 'related species' and 'creature comparisons'. The dominant status of the information in block A is indicated by the greater space it occupies on the page, its top-left positioning, and its lack of a title: it seems to share the title of the whole page. The text labels and detailed pictures are linked to the illustration by lines, reinforcing their related status.

As we noted in the layout structure, page elements are differentiated by size, positioning, boxes, and panels. The fact that the two panels we are referring to as block B somehow 'go together' is indicated by the fact that they are the same (sandy) colour, although the different colours of titles (bright red top, dark red bottom) have the effect of dissociating them. The overlap of the tiger's tail with the upper panel in block B has the effect of linking that panel with block A, and integrating the page as a whole.

'Creature comparisons' as a unit diverges to quite an extent from content structure, which simply recorded the existence of statements about the Siberian tiger and about man, but as unrelated to one another. In the representation on the page, however, the relationship between the facts about height and size of the different animals is clearly intended to be one of comparison and contrast, centimetre rule at the left.

The two uncaptioned pictures at the bottom of block B are referred to in the text ('below left' and 'below right'), while, in Block A, lines connect the text about the tiger both to the smaller detail pictures and to the main illustration. This explicit labelling to direct the reader within a document is common to several of our genres, such as the newspapers' need to refer readers of a story to more content 'continued on page 34'. The navigation structure is where such internal referencing or 'document deixis' is naturally captured (for a discussion of the phenomenon of document deixis, see Paraboni, 2000).

## 2.5 Linguistic Structure

Language on this page can be divided into several types. Titles and labels consist only of nouns (*eyes, claws*) and noun phrases (*vital statistics, related species*). These elements are selected for word-play: the double-meaning of 'vital statistics', for example, or the alliteration of 'creature comparison'. The vital statistics list contains a variety of grammatical forms but none of them complete clauses, and the complete sense of them is only inferrable in the context of the labels. Body text around the main illustration, and in the introductory paragraph at the top of the page, is evaluative: the tiger is *powerful, massive*, and it has *success* in hunting, for example. The structure of each text entry around the tiger is to evaluate positively each of the highlighted

characteristics: the tiger's night vision is *six times better than our own*, it has a *unique* pattern of black stripes, and it stalks prey *silently*. The text in box C, 'Creature Comparisons', and in the 'related species' panel, is more factual than evaluative, although signs it has been written for young readers exist in the Siberian tiger's coat being described as *shaggier* and its habitat *icy*.

## 2.6 Constraints

We noted at the outset that a description of genre does not only consist of analysis at the five levels proposed, but of observing constraints on production. Here, an obvious constraint is the necessity of fitting all the information on one side of paper, given the canvas constraint presented by the chosen encyclopedia format: one page per entry. This constraint would have an effect in the navigation structure: the status of the new entry is made clear by a title. Although this is a page with several blocks of information, it appears that the user is expected to traverse the page from the top left, a conventional consumption strategy. This consumption constraint is reflected in the layout structure by the positioning of the main graphical and textual element (block A) in the top left of the page. A consumption constraint also exists in terms of the expectation of the readership: young readers require simple, engaging text (reflected in the linguistic structure) comparing familiar concepts with new ones, and prefer enticing and often highly-coloured productions (reflections in the layout structure) with simple navigation.

## 3. Discussion: from Text Description to Text Critiquing

We have now described the characteristics of the tiger page on all five of the levels we propose in our framework, and briefly sketched the primary additional constraints that play a role in the construction of a complete document. Our research, based on a corpus of four different document types, is aimed at elucidating systematic relationships between the different levels of description we describe. In the tiger example, the layout structure informs us that block A (the tiger) is presented as being more important. A crucial question then is whether the chosen layout structure reflects the RST analysis, and whether it helps or hinders the reader in recognising rhetorical relationships. The descriptive approach therefore offers tools for designing more usable documents.

The tiger page illustrates several aspects of this issue very well. The content of the first elaboration in the rhetorical structure (cf. Figure 3), the physical, functional characteristics of the tiger, are kept together in the layout structure: they are arranged around the main tiger picture upon which they elaborate. There is no nucleus-satellite relation between elements in this list: they are instead arranged where they would fit around the tiger, as closely as possible to the points they refer to, but kept separate. Furthermore, the rhetorical distinctions within these text labels have been realised to a particular level of delicacy. The text entries for 'eyes', 'ears', 'coat' etc. do not visually distinguish 'means' relations (how the eyes work) and 'purpose' relations (what the stripes are for): all the content is realised as plain text labels. The second elaboration, again unordered, contains the 'vital statistics' information. This is separated in a panel, although the list elements are labelled. In keeping with a left-to-right reading strategy, the large central segment is placed top left.

While the layout structure selected in the tiger page thus broadly observes the rhetorical and content distinctions, there is slippage between layout and rhetorical or even content structure. The selections of light and dark type, and between serif and sans fonts, do not appear to be functionally motivated from the rhetorical structure and so we need to ask whether this differentiation is helpful or distracting for the reader of the page. There are therefore clear indications that our approach can help us ask useful questions about usability. Similarly, the alignment of panels on the right also appears not to carry any functional load. We would also want to account for the relative size of the three panels in relation to their function, their placement on the page, and the rationale for using a colour background in only two of the panels.



Figure 5: Bank letter: actual linguistic form and content, constructed ‘neutral’ layout

Similarly, the back view of the tiger's ears is in fact an equal partner in the 'ears' list element that elaborates on the main picture, but is not clearly connected to either the text that is its sister or the



picture that is its parent. Indeed, the proximity of the picture of the tiger's ears is closer to the description of its coat, and is thus a stronger cue than the line from the text to the main image: this does not coincide with the rhetorical intent. Finally, the circular picture border for the tiger's mouth lacks coherence with the rectangular borders around the other inset pictures, given that they perform equivalent rhetorical functions.

Taking an example from a commercial context, we can see how layout structure that observes rhetorical structure arguably results in a better design. Figure 5 is an example of the content of a bank letter written in a clear style, and presented in a fairly ordinary form<sup>5</sup>.

There are five key elements of the content of this letter: the **intro**, consisting of the recipient's address, a salutation, and a date; the central **message**, which thanks the recipient for returning a mortgage deed, tells her it is filled out incorrectly, asks her to fix the problems and return it, and tells her who to contact in case of a problem, and how; the **outro**, which consists of a valediction, signature, and name of signatory; a **cross-marketing** element, which attempts to sell her something else (loans and savings) and tells her who to contact if she's interested; and finally the **watermark** content, which is the company address, registration details, and logo which are part

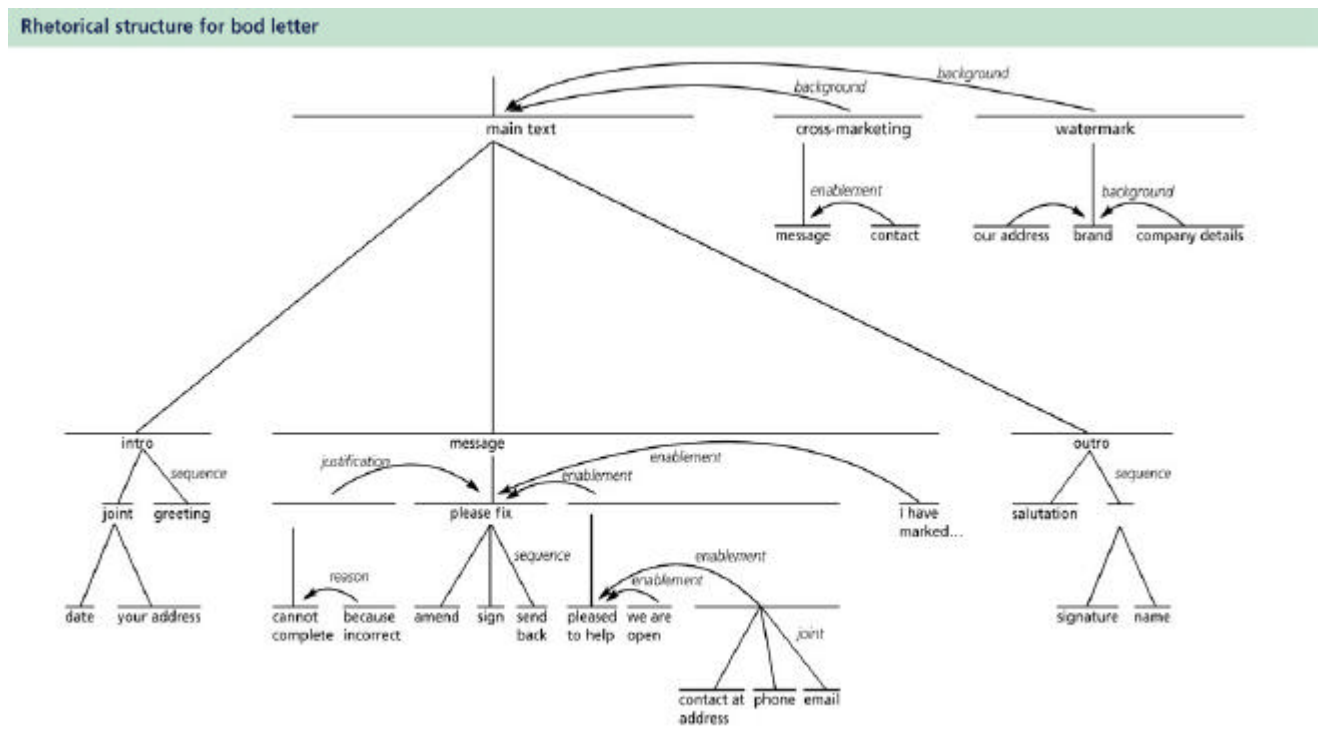


Figure 6: Rhetorical structure for Bod letter content

<sup>5</sup> This letter is 'back-constructed' from the actual example (bank name excised) given in Figure 7. The linguistic structure and content are therefore exactly the same as the letter actually used by the bank. Thanks to Robert Hemsall for the artwork on these examples.

of the stationery. The rhetorical structure that is suggested by the content and the way it is argued is illustrated in Figure 6.

We can see from the rhetorical structure that there is a great deal of information in the letter that focuses on enabling the reader to do as she is asked in filling out and returning the form: contact details and offers of help. The key function of the letter is to ask her to perform the required actions and return a form, and one segment of the text is devoted to justifying why this needs to be done (it was incorrectly filled out). Because it is not the central function of the letter, the cross-marketing information is cast as background, as is the watermark information. The 'body text', therefore, is made up of intro, message, and outro, with the message being the most central element.

Figure 7 shows the letter as redesigned by Enterprise IDU, using their usual expertise with no knowledge of the rhetorical analysis just presented. This is the design that was used by the bank.

The improvements represented by the redesign can be explained to a large extent in terms of operations on the rhetorical structure. These changes not only respect rhetorical segments, they go some way to highlight rhetorical relationships. First, the whole cross-marketing segment, represented in RST terms as background information, has its background status reflected in its placement in a side-bar. Watermark content is left where it is, since it was already presented as peripheral; the date element is singled out, however, and expanded. All the information represented as enabling segments (how to contact the company by phone, email, etc.) is distinguished in placement from the main message element. This group of satellites is given a title of its own ('Contacting us') that highlights their similarity in function with respect to the main purpose of the letter. Likewise, each separated enablement element is given a label in medium, rather than light font weight. These titles not only highlight the content of what follows, but point to the fact that each segment stands in a similar relation to the title ('Contacting us') and to the text as a whole.

The 'gist' of the central and most nuclear component of the text, the request to sign and send back the form, not only appears in the main body of the text, but is excised and given in a different font, enlarged and medium weight, as the title of the whole letter. No transformation is made that 'ungroups' the main message component, except for a title insertion {'What to do now') that introduces the three-step sequence of amending, signing, and returning the deed, arguably the central multinuclear element of the whole letter. Although it is clear that much further study is required to investigate the relationship between rhetorical structure preservation and document usability, it is interesting that the changes that were made by the designer, and adopted by the client as improvements, adhere to and promote the understanding of rhetorical structure in this way.

Both in modelling for computer generation and in trouble-shooting for real design applications, we are seeking to motivate layout decisions in terms of the functional discriminations needed to communicate the rhetorical structure. When there are commonalities in layout decisions across distinct elements (be they of size, colouring, type face selections, alignment, etc.), then there is a



Figure 7: Bod letter, as redesigned

natural tendency for the reader to interpret those elements as being in some way rhetorically related. Conversely, when there are distinct layout decisions made, then there is a tendency for rhetorical discriminations to be perceived. Furthermore, insertion of text and typography that emphasise rhetorical nuclei and topic information is perceived, in this case at least, to make the information easier to understand. The detailed analysis afforded by our levels of description now allow us to state and probe these putative generalisations with a far higher degree of precision than has been possible up to now.

## References

Allen, P., Bateman, J., and Delin, J. (1999). Genre and layout design in multimodal documents: towards an empirical account. *American Association for Artificial Intelligence Fall Symposium*

*on Using Layout for the Generation, Analysis, or Retrieval of Documents*, Cape Cod, Autumn 1999.

Bateman, John, Kamps, Thomas, Klein, Jörg, and Reichenberger, Klaus. (subm.). The Dartbio system: constructive text, diagram and layout generation for information presentation. Submitted to *Computational Linguistics*.

Bateman, J., Delin, J., and Allen, P. (2000). Constraints on Layout in Multimodal Document Generation. *INLG2000: The First International Natural Language Generation Conference, Workshop on Coherence in Generated Multimedia*. Mitzpe Ramon, Israel, July 12 2000. 7-14.

Delin, J., Bateman, J., and Allen, P. (submitted). A Model of Genre in Document Layout. Submitted to *Information Design Journal*.

Mann, William, C. & Thompson, Sandra A. (1988). "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text* 8 (3): 243-281.

Paraboni, Ivandré (2000). Describing Document Parts. *Proceedings of the 3rd UK Special Interest Group for Computational Linguistics (CLUK-3)*, University of Brighton, UK. 34-41.

Waller, Robert. (1987). *The Typographic Contribution to Language*. Unpublished PhD Dissertation, Department of Typography and Graphic Communication, University of Reading.