Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

PAUL PIWEK, RICHARD POWER, DONIA SCOTT
AND KEES VAN DEEMTER

GENERATING MULTIMEDIA PRESENTATIONS FROM PLAIN TEXT TO SCREEN PLAY

Abstract. In many Natural Language Generation (NLG) applications, the output is limited to plain text – i.e., a string of words with punctuation and paragraph breaks, but no indications for layout, or pictures, or dialogue. In several projects, we have begun to explore NLG applications in which these extra media are brought into play. This paper gives an informal account of what we have learned. For coherence, we focus on the domain of patient information leaflets, and follow an example in which the same content is expressed first in plain text, then in formatted text, then in text with pictures, and finally in a dialogue script that can be performed by two animated agents. We show how the same meaning can be mapped to realisation patterns in different media, and how the expanded options for expressing meaning are related to the perceived style and tone of the presentation. Throughout, we stress that the extra media are not simple *added* to plain text, but *integrated* with it: thus the use of formatting, or pictures, or dialogue, may require radical rewording of the text itself.

1. INTRODUCTION

Most research on Natural Language Generation (NLG) has focussed on the production of plain text, the typical output being one or two paragraphs of continuous prose. In a survey of NLG research pre-1998, Paiva (1998) describes 19 applications of which 12 generate plain text, 5 generate text with some simple layout (e.g., enumerated or bulleted lists), and 2 generate text that includes diagrams. This restriction may have been sensible during the early development of the field, but to address the typical requirements for commercial applications we need to widen the scope of generation to include more complex layout (e.g., lists, boxes, footnotes, emphasis), diagrams or pictures, and the new media (such as animations) that can now be incorporated into documents on the web.

In several projects during the last five years, we have focussed on the commercial requirements of technical documentation, such as instruction manuals (Paris et al., 1995), patient information leaflets (Bouayad-Agha et al., 2002), and product information in the car industry (Krenn et al., 2002). During the course of these projects, we have analysed corpora of existing documents, and interviewed the technical writers and translators who produce them. Several general points have emerged. First, the documents make rich use of layout in order to enhance clarity and navigability. Secondly, most documents of this type include diagrams, which may even be more important than the text (e.g., in instructions on how to assemble equipment or furniture). Finally, the companies issuing the documents are extremely sensitive to questions of style: a manual or instruction leaflet is seen partly as a

T.H.E. Editor(s) (ed.), Book title, 1—6.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

portrayal of the company's character and its attitude to its clients. The lesson to be drawn is clear*: for commercial applications of NLG, we must develop generators which integrate a range of media and which allow control over style.*

How can the scope of a generator be widened so that it can plan layout and the inclusion of illustrations? The simplest solution would be to *add* these graphical features to the text – in other words, we might hope to take the output of an existing generator, and to apply specialised modules for formatting and illustration. In most cases this method will not yield good results; in our interviews with document designers, we have also confirmed that it is not the method they use. Reformatting, or addition of pictures, means at the very least that the text must be revised; we have even encountered cases in which layout and pictures are drafted *before any text is written at all* (Bouayad-Agha et al., 2000). In short, text, layout, pictures, etc. have to be co-adapted, which means that they must be planned together. What is needed is a generator that can plan not simply a text but a multimedia document, while preferring patterns in all the various media that portray the desired personality or style.

Our aim in this paper is to survey our experiences in developing generators that are enhanced in this way. We contrast four presentations of the *same* semantic content. We begin by considering a presentation that is limited to plain text. We then consider presentations in which graphical formatting methods like vertical lists and emphasis are added to the text. Next, we discuss the implications of adding pictures. Finally, we jump to a much more advanced and recent medium: a computer animation of a situated dialogue, based on an automatically generated screenplay or 'dialogue script'. The semantic content in each case will be a section adapted from a patient information leaflet, which provides instruction to patients on how to take their tablets. In surveying presentations in the various media, we focus on three issues:

Meaning Graphical devices are at least partly a means of expressing semantic or rhetorical content. Just as a conventional generator needs rules linking semantic patterns to syntactic ones, an enhanced generator needs some way of representing the meanings of layout patterns and pictures, so that it can make appropriate choices. We have had to address questions like 'What are the semantic or rhetorical implications of a vertical list?' and 'How can the meaning of a picture be specified in an NLG application?'

Style Languages and related media provide an enormous stock of patterns for expressing the same semantic content; this allows authors the opportunity of signalling a desired personality or style by their preferences among these options. Some linguistic indicators of style are already well-known (e.g., sentence length, informal vs formal terminology); similar indicators can be found within the options for layout and illustration. In the case of scripted dialogue the situation is complicated by the presence of *two* characters with potentially different conversational styles.

Wording Since the presence of text is the common factor among the four presentations that we discuss, it is important to see how the nature of this text changes when combined with different media. We show, for example, that criteria of linguistic correctness change when a document uses graphical layout, so that stripping away the layout would yield an ill-formed plain text. Similarly, we will show that pictures and dialogue can have a profound effect on wording.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

As might be expected in such a wide-ranging survey, we are more interested in raising issues than in demonstrating specific results. Although our observations are backed up by study of corpora, they rely more on intuition than on firm empirical evidence. In fact, it is difficult to say what kind of empirical basis would be appropriate for a system that generates documents. Do we want to emulate average human behaviour? Presumably not, since many people express themselves ineffectively. We might prefer to emulate professional document designers – yet can we be sure their judgement is always reliable? Perhaps the best guide would be a study of the effect on readers, although this is not an easy thing to measure. Luckily, most of the points we make are relatively elementary, so we can reasonably hope that our intuitions coincide both with expert opinion and objective effectiveness.

2. GENERATING TEXT

Since the generation of text is the primary capability of NLG systems, and because all the multimodal systems that will be discussed later in this paper build on this capability, it will be useful to sketch informally what 'generation of text' amounts to in practice. Space will not allow a detailed discussion, and we refer to Reiter and Dale (2000) for elaboration.

2.1 Representing meaning

One usually speaks of NLG when an abstract representation of meaning is put into the words of a natural language. The representations that form the input to the system can be of various kinds. What is crucial is that they are not already the expressions of a natural language (in which case we speak of Machine Translation, not NLG). These representations may be created specifically for processing by an NLG system, or they may have led an independently useful life, for example as the output of a Question-Answering system, or as the facts in a database. In this article, we will focus on a *logical* representation language, of the kind that is created by the so-called WYSIWYM knowledge editing interface (Power and Scott, 1998). In addition to the usual logical operators, such as negations and conditionals, the logical language contains modal operators such as '(it is) obligatory (that)' and special constructs of the kind that are familiar from Artificial Intelligence planning. For concreteness, we will assume that the input to the system is equivalent to the conjunction of the following six formulas, which contain variables as 'handles' for expressing coreference.

```
(1) obligatory(s:suggest(y:doctor(z:patient),d:dose))
(2) obligatory(follow(z,s))
(3) [unsure-about(z,d) v
    unsure-about(z,timing(d))]-> obligatory(ask(z,y))
(4) procedure(take(z,t:tablet),
    [remove(z,t,foil,finger,back(t))&
    swallow(z,t,water)])
(5) [w:take(z,overdose)] -> obligatory([tell(z,y,w)
```

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer


```
      OR visit(z,casualty(hospital(z)))])
(6)   store(a:person,m:medicine) ->
      obligatory(storeawayfrom(a,m,children)
```

The first formula says that it is obligatory that a person who is the doctor of the patient (henceforth identified through the variables `y` and `z` respectively) suggests what dose `d` to take. The second formula uses the same variables to say that the patient `z` should follow this suggestion (identified as `s`). Consistent with conventions in Discourse Representation Theory (Kamp and Reyle, 1993), most variables (i.e., the ones that do not occur in special constructs such as the antecedent of a conditional) are interpreted as *existentially* quantified. Because they do occur in the antecedent of a conditional, the variables `a` and `m` in (6) are universally quantified: For every person `a` and medicine `m`, if `a` stores `m` then `a` has to store `m` away from children. Formulas of this kind leave a lot of information implicit – witness the complex relation `storeawayfrom`, for example – but they are precise enough to serve as a basis for an NLG system that is designed with these formulas in mind.

Some NLG systems use inputs that distinguish between different kinds of conjunctions, using the relations proposed by Rhetorical Structure Theory (RST, Mann and Thompson, 1987). In this way, it is possible to encode the fact that the information in (1) is more important than that in (2), for example, by saying that (2) is an *Elaboration* of (1). We will refer to rhetorical relations between propositions when and where this is appropriate, assuming that they are available in the input to the NLG system.

2.2 Planning document structure

In most NLG systems, the assignment of semantic material to large-scale textual units like sentences, paragraphs, and lists, is performed by relatively simple 'microplanning' or 'aggregation' rules (Reiter & Dale, 2000). Typically, the rhetorical structure in a text plan is realised by more or less 'reading it out' in a left to right fashion, making each leaf into a (grammatical) clause, and producing the text by simply realising these clauses as grammatical sentences, occasionally making more complex sentences with the help of discourse connectives or aggregation. The limitations of this approach for our purposes include the following:

- The structure of the resulting text will be isomorphic with the rhetorical structure. In other words, an RST analysis of the generated text will be exactly the same as the input RST structure. But rewriting even a short piece of text without removing any content can result in a range of different RST structures: e.g., clauses, sentences and whole paragraphs can be reordered; clauses and sentences can be made into adjectives or adverbs, or *vice versa*; material can be pulled apart and distributed between different sentences or paragraphs; horizontal lists can be made into indented bulleted lists etc.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

- The generated text will only include information that is present in the input rhetorical structure. However, texts that are more graphical than those considered by ordinary NLG systems – for example, patient information leaflets – often include additional information. For example, headings ("How to take your medicine"), references to other parts of the document ("See figure 2"), and figure captions.

To overcome these and other related limitations, we have introduced an additional level of representation – *document structure* – which is a realisation of rhetorical structure that is not necessary isomorphic to it (Power et al., 2003a). Document representations encode information about the graphical presentation of a document, such as textual level (paragraph, orthographic sentence etc.), layout (indentation, bulleted lists etc.) and their relative positions. Document structure can be seen as an extension to Nunberg's 'text-grammar' (Nunberg, 1990); it is also closely related to 'logical markup languages' like LaTeX and HTML. We are not able to describe document structure in much detail here, and refer the reader instead to (Power et al., 2003a). However, we think the reader will benefit from a small example, showing the transition from rhetorical structure to document structure and the eventual generated text. Consider the following text, of the kind that can be generated by the ICONOCLAST system (Power et al., 2003a):

> Your doctor should suggest a dose; follow your doctor's advice and do not change the dose. If you are unsure about the dose or if you are unsure of when to take it, ask your doctor. To take a tablet, remove it from the foil and swallow it with water. If you take an overdose, tell your doctor immediately or visit your hospital's casualty department.

Figure 1: One way of expressing the input of section 2.1

Consider the first sentence of this text, which is based on the first two formulas of the input introduced in section 2.1:

(1) `obligatory (s:suggest(y:doctor(z:patient),d:dose))`

(2) `obligatory(follow(z,s))`

(The text uses two different formulations to express the information in (2).) The rhetorical structure for this sentence is shown informally in Figure 2. Note that although the nucleus and satellite are identified along with the rhetorical relation BACKGROUND their ordering is not yet determined, since this is a task for the document structurer.

An example simple document representation is shown in Figure 3. Here there is a single document structure containing a paragraph at the top level, which comprises a sequence of text-sentences. Shown in more detail is the part of the document

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

structure representing the first sentence in Figure 1. This text-sentence is composed of two text-clauses, the second being further decomposed into a pair of text-phrases. Each node has the associated features of TEXT, INDENTATION, POSITION and MARKER, with appropriate values. Note that the *message* of the text, as described in Figure 2 has been expanded at the level of document structure to include a repetition of (2), and that the propositions are now ordered and their associated discourse markers have been assigned.

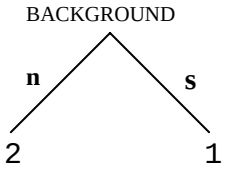


Figure 2: Example rhetorical representation (informal)

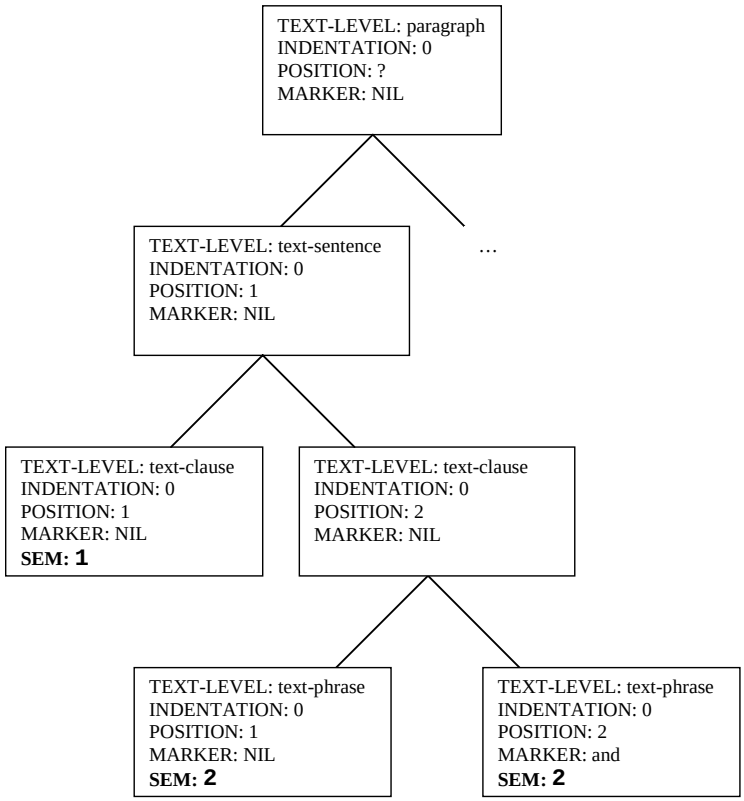


Figure 3: Document representation for first sentence of Fig.1

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

2.3 Controlling linguistic style

Patient information leaflets have been a rich source of material for us. They present information that is safety critical, and for this reason their content is strictly legislated by regulatory bodies (e.g., the EAMA in Europe). Nevertheless, they employ a range of different linguistic devices to convey the required information. This is because the same generic drug tends to be marketed by many companies; given this fierce competition, and given the constraints on what they *can say*, pharmaceutical companies try to make their mark by adopting an individual, often distinctive, style.

Even the most cursory perusal through a corpus of these leaflets (e.g., ABPI 1997) is revealing. At one end of the spectrum, there are those that favour a 'tabloid' style, with simple language (e.g., *throat* instead of *larynx*; *ask* instead *consult*), liberal use of formatting devices (e.g., font and face alternations, indented bulleted lists), pictures, and containing many single sentence paragraphs; they also tend to be rather 'personal' with references to "*your* doctor" and "*your* medicine". At the other end, there are those that favour a more serious 'broadsheet' style, sometimes at first glance looking more like an article in a medical journal than a leaflet for patients – with more technical language, fewer and longer paragraphs, in smaller font size and fewer pictures.

The following examples show how the semantic content defined in section 2.1, and shown in Figure 1 may be expressed linguistically in a range of styles. The problem of generating the same content using different styles has been studied by various authors, including Hovy (1988), Walker et al. (1996), Power et al., (2003b).

> **Version 1** Take your tablet by removing it from the foil and swallow it with water. Your doctor will tell you your correct dose; follow your doctor's advice and do not change the dose. Ask your doctor if you are unsure of the dosage or its timing. If you miss a dose, take another as soon as you remember or wait until it is time to take the next dose. Then proceed as before. However, If you take an overdose, you should immediately tell your doctor or go to the casualty department (emergency ward) at your hospital. Keep your medicine in a place where children cannot reach it.

> **Version 2** The method of delivery for tablets is the removal of the tablet from its foil, followed by its ingestion with water. Doctors must instruct their patient about the correct dosage, and patients must follow this instruction strictly. In the event that a patient forgets the required dosage or its timing, the patient must ask their doctor. In the event of an overdose, the doctor must be informed immediately by their patient, or else the patient must report immediately to the casualty department of their hospital. The medicine must be kept away from the reach of children.

> **Version 3** To take a tablet, you should first remove it from the foil and then swallow it with water. Your doctor will tell you the dosage. Follow his advice and do not change it. If you are unsure of your dosage or its timing, you should ask your doctor. If you take an overdose, you should inform your doctor. If you are unsure of your dosage or its timing, you should ask your doctor. If you take an overdose, you should inform your doctor immediately or go straight to your hospital's emergency ward. Store your tablets out of the reach of children.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

3. LAYOUT

Layout enhances plain text by introducing various graphical devices like indented lists, tables, boxes, footnotes, along with extra character formatting (italics, bold face, small type, etc.). There is not a sharp distinction between 'plain text' and 'text with layout', unless by 'plain text' we mean literally a string of words, with no punctuation at all. Devices like semi-colons, full stops, and parentheses serve as graphical aids as much as bulleted lists or bold face; the only difference is that they happen to be realised as characters, and can thus be specified in a plain text editor; owing to this contingency, 'layout' is typically used as a term for devices that cannot be expressed through a string of ASCII characters.

In languages like LaTeX and HTML, this essentially means anything for which a tag is needed. Here is a version of our sample text in which the features *indented lists* and *emphasis* are added, followed by its LaTeX source:

> Your doctor should suggest a dose. Follow your doctor's advice, and DO NOT CHANGE THE DOSE. Ask your doctor if
>
> - you are unsure about the dose
> - you are unsure when to take the dose
>
> To take a tablet:
>
> (1) Remove the tablet from the foil.
> (2) Swallow it with water
>
> If you have taken an overdose, tell your doctor *immediately* or visit your hospital's casualty department.

```
Your doctor should suggest a dose. Follow your doctor's advice, and
{\sc do not change the dose}. Ask your doctor if
\begin{itemize}
\item you are unsure about the dose
\item you are unsure when to take the dose
\end{itemize}
To take a tablet:
\begin{enumerate}
\item Remove the tablet from the foil
\item Swallow it with water
\end{enumerate}
If you taken an overdose, tell your doctor {\em immediately} or
visit your hospital's casualty department.
```

In HTML, essentially the same specifications would be given, except that the tag names are different and paragraphs have to be tagged explicitly rather than separated by an empty line.

3.1. Layout and Meaning

Some recent research in NLG has explored the relationship between the rhetorical structure of an argument, and the document structure that expresses this argument in

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

a text (Bateman et al., 2001; Power et al., 2003a). At a rhetorical level, our example includes *a procedure* for taking the medicine; this procedure comprises a *goal* (taking a tablet) and a *method*, which consists in turn of a sequence of *steps*. These are all rhetorical concepts, which are part of the message no matter how it is expressed. In the example, these rhetorical units have been mapped to units of abstract document structure: the procedure is realised by a text-sentence; the goal is realised by a text-clause; the method by an enumerated list; and the steps by list items, each also a text-sentence.

In instructional texts, this way of presenting a procedure has become almost standard – for instance, it is used consistently in software manuals for any product by Microsoft or Apple. Why is it considered a *good* format? Note that one cannot answer this question usefully by subjective reports like 'the list format is clearer' or 'it is easier to read' – in this case, why do we not use a list format for everything? Are there any general principles that would allow an NLG system to decide when an indented list is appropriate? Here are some relevant factors:

- *Space:* Indented lists introduce larger separations between segments of the text. Using an indented list therefore incurs a cost: either you must use more paper, or some other part of your message must be left out.
- *Navigability:* Readers often refer to instructions while performing the task (e.g., preparing a recipe, or assembling a bookcase), rather than memorising all the steps beforehand. This means that it should be easy to re-enter the text after looking away, in order to retrieve the next step.
- *Uniformity:* Indented lists carry the implication that the items play similar roles in the argument (e.g., in a procedure, they are all steps).
- *Size:* The value of an indented list is enhanced if the items are relatively large – up to a point. (If the list overflows several pages, confusion obviously results, since the eye can no scan across the bullets or numbers, or align them if the list has multiple levels.)
- *Length:* The value of an indented list is also enhanced if it has more items – again up to a point. With just one or two steps in a procedure, navigation is less of a problem, so the list formatting might be regarded as overkill.

In terms of Rhetorical Structure Theory (Mann and Thompson, 1987), the uniformity constraint might be stated as follows: use indented lists only for multinuclear relations, not for nucleus-satellite relations. By definition, a nucleus-satellite relation is asymmetric – the nucleus is more important than the satellite – and therefore does not have uniform constituents. It would be odd, for example, to present a whole procedure through a list in which the first item was the goal and the second item was the method:

- To take a tablet:
- Remove the tablet from the foil and swallow it with water.

Turning to the size constraint, the following presentation of the sentence 'the medicine contains terbinafine and gestodene', while not anomalous, is also strange,

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

because any gain in clarity seems poor compensation for the waste of space. This objection would no longer apply if the list were longer, or if each ingredient in the list was described more fully.

> The medicine contains:
>
> - terbinafine
> - gestodene

Ubiquitous use of indented lists diminishes their value. If there are bullets dotted around all over the page, it is no longer simple to look away from the instructions and then re-enter at the right point.

We have reviewed at some length the motives for using indented lists; we now consider, more briefly, the use of emphasis (shown, for example, by italics or small capitals). Here there are various possible rhetorical correlates, which have in common the notion that the word or phrase is important. In plain text, crucial warnings can be marked by words like 'vital', 'must', 'absolutely', and sentences can be arranged so that new (as opposed to given) information appears at the end. One sign of careless writing is the repeated use of emphasis in order to recover from poor sentence organisation:

> Consult your doctor if you have any *problems* concerning your treatment, or any *questions* about your treatment.

> Consult your doctor about your treatment if you have any problems or questions.

In the first sentence, *questions* is italicised because the next phrase 'about your treatment' presents old information; in the second sentence the new information comes at the end, so emphasis through formatting is unnecessary.

In written texts, emphasis is not only the counterpart of stress in speech; it also serves as a navigational aid. For instance, if important warnings are presented in bold face, they can be found at a glance, even if they appear in an inaccessible location such as the middle of a paragraph. As in the case of indented lists, this benefit depends on using the device sparingly: if each page has dozens of emphasized phrases, the warning becomes a needle in a haystack. At an absurd extreme one might imagine an author emphasizing the whole text on the grounds that every word is of vital importance.

3.2. Layout and Style

So far, we have discussed formatting devices as a means of conveying ideas and improving usability; however, authors also have to take account of another aspect, namely style. Different formatting styles will be interpreted as indicating different personalities or attitudes to the reader. Consider the eccentric author just mentioned, who emphasizes an entire leaflet by formatting it (let us say) in capital letters. The drawback is not only that the reader cannot distinguish degrees of importance; there is also an unpleasant tone, a feeling of being shouted at.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

The different textual styles that were discussed in section 2.3 tend to go together with systematic variations in formatting. In what we called the 'tabloid' approach, one sees frequent use of bullet points and emphasized words; in 'broadsheet' style, one sees long dense paragraphs and little use of indented lists. The reason, presumably, is that an indented list represents an exchange of clarity for depth. The crucial points can be found more easily, but since space is wasted, there is less room for giving additional explanation. Some readers will thank the author for easing their task; others will perceive the leaflet as an insult to their intelligence. Similar differences are probably found between academic fields: while common in scientific articles, bulleted lists are rare in humanistic fields like philosophy, literary criticism, and history, where the dignity of the material seems to demand long paragraphs of continuous prose and to preclude anything so vulgar as a list.

3.3. Layout and Wording

A potential misconception about formatting is that can be added to a text as a kind of adornment, like a cherry on top of a cake. In terms of an NLG system, this would be equivalent to an architecture in which the formatting module is invoked last, after the syntax and wording have already been fixed. In reality, layout and wording interact, especially when the layout includes indented structures like lists and quotes. As a simple example of such an interaction, consider the following sentence from our example:

> Ask your doctor if:
>
> - you are unsure about the dose
> - you are unsure when to take the dose

Suppose that, owing to a change of policy, this sentence had to be revised to take out the bulleted list. Obviously the author could not simply remove all the LaTeX tags, leaving the following non-sentence:

> Ask your doctor if: you are unsure about the dose you are unsure when to take the dose.

Two changes are needed. First, correct syntax now requires the introduction of a connective (i.e., 'or'). Second, the punctuation must be corrected – the colon after 'if' is no longer appropriate, and a full stop must be added at the end:

> Ask your doctor if you are unsure about the dose or you are unsure when to take the dose.

What seems to be happening here is that the extra separation and organisation provided by an indented list makes the full stop and the connective unnecessary (provided that the context makes it clear whether the relation is a conjunction or a disjunction), so licensing a small departure from normal syntax. However, the departure can be much more radical, as the following example shows:

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

> In rare cases, the treatment can be prolonged for another week; however, this is risky since:
>
> - The side-effects are likely to get worse. Some patients have reported severe headache and nausea.
> - Permanent damage to the liver may result.

Here the list items are paragraphs (the first has two full text-sentences), yet the whole list serves syntactically as the complement of the subordinating conjunction 'since'. As a result, removal of the indented list would require more far-reaching changes, as in the following version:

> In rare cases, the treatment can be prolonged for another week; however, this is risky for two reasons. First, the side-effects are likely to get worse; some patients have reported severe headache and nausea. Secondly, permanent damage to the liver may result.

4. PICTURES

Many documents, whether they are meant to be seen on paper or on a computer screen, contain more than just formatted text. In addition to formatting, they may contain such graphical elements as formulas, diagrams, and pictures. A considerable amount of research has been done on the meaning and use of diagrams (e.g., Kerpediev and Roth, 2000); here we will focus on pictures. Pictures are sometimes distinguished from other graphics by the fact that they are 'iconic': the meaning of a picture arises mainly by its similarity to what it depicts (Hartshorne and Weiss, 1958). Photographs are pictures, and so are the more stylised sketches that are often find in Patient Information Leaflets. In the present subsection we will ask (a) how pictures contribute to the meaning of a document, (b) how they affect the style of the document, and (c) how they can affect the wording of the document. In addition, we will discuss how documents may be generated that contain pictures as well as formatted text.

4.1. Pictures and Meaning

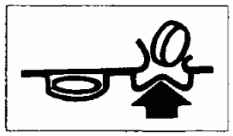
It is not easy to say in general terms what the meaning of a picture is. An interesting exploration of this question can be found in Levesque (2003), who claims that pictures tend to convey 'vivid' information: information that contains no logical structure beyond predication and conjunction. A picture might say, for example, that Carter shook hands with Castro: it cannot say that they *either* shook hands *or* beat each other up. The notion of vivid information is an important concept in artificial intelligence, and pictures are sometimes seen as a prime example. Here, we will adopt a variant of this view, viewing pictures as basically expressing existential information: A picture of two unrecognisable men shaking hands can be argued to mean that at *some* point in time, *two men* with certain outward appearances shook hands in a particular way. (If one of them is recognisable, then we can think of this as depicting the property of being equal to Carter.)

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

'Photographic' pictures of this kind convey a wealth of information, and would be difficult to generate automatically (unlike the more schematic ones generated in McKeown et al., 1992; Wahlster et al., 1993). The picture just discussed, for example, shows in detail *how* the handshake took place. This 'how' would be difficult to capture in words or mathematical symbols: any symbolic representation would tend to leave out something that the picture depicts. This is different with the kinds of pictures that are used in instructional texts like Patient Information Leaflets, where pictures are employed to convey 'discrete' information of the kind that might also have been conveyed by text. Consider the following illustrated version of the document, for example (version 3 in section 2.3).

To take a tablet, you should first remove it from the foil and then swallow it with water. Your doctor will tell you the dosage. Follow his advice and do not change it. If you are unsure of your dosage or when to take it, you should ask you doctor. If you take an overdose, you should inform your doctor immediately or go straight to your hospital's emergency ward. Store the medicine out of the reach of children.



Consider the picture showing someone storing away the medicine. The person is shown as a woman with longish hair; the cabinet has a specific size and a medicinal cross on each of its doors. Little of this has anything to do with the meaning of the picture in its current setting. In other respects, the picture is rather poor in information. For example, it does not show what the woman is doing; there is no medicine in sight! Clearly, the picture denotes both more and less than what is depicted: it denotes a person (whose gender and appearance are irrelevant) storing away a medicine in a place where children cannot reach it. This is the kind of information that is conveniently represented using a representation language that allows us to represent atomic propositions (involving one or more arguments), existential quantification, and conjunction. As before, we leave existential quantification implicit:

```
x:person & y:medicine & storeawayfrom(x,y,children)
```

('There is a person *x* and a medicine *y* such that *x* stores away *y* away from children.') Something similar is true for the other picture, which shows how to obtain the tablet. The only novel element here is the use of the term *back(y).*

```
x:person & y:tablet & remove(x,y,foil,finger,back(y))
```

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

('There is a person *x* and a tablet *y* such that *x* removes *y* from the foil by pushing a finger to the back of *y.*') Over the last few years at ITRI, we have explored the usefulness of an approach to picture meaning that uses logic to express their meaning, analogous to and inspired by the way in which logic is routinely used for capturing the meaning of texts. This has been done in the following way (van Deemter, 1999; Cahill et al., 2001; van Deemter and Power, to appear):

- We have built an interface that makes it easy to create formulas of the kind exemplified here, and to stipulate that a given picture is associated with a given meaning representation. In the case of the first picture, for example, we assume that the annotator knows that the picture will be used in the patient information leaflets, allowing her to associate it with the formula shown above. Crucially, the interface offers the annotator only a limited range of options; for example, it will not allow the annotator to specify the gender of the person depicted, since the system 'knows' that this is irrelevant. The picture annotation interface is a variant of existing WYSIWYM interfaces. As a result, the same interface can be used for creating the A-Box of section 2.1 and for annotating pictures.
- In earlier work, NLG tools have been constructed that allow WYSIWYM-created formulas to be expressed in a number of natural languages (Power and Scott, 1998). These tools allow a user to specify the content of a text at a high level of abstraction, leaving textual realisation to the system.
- Given an A-Box and an annotated library of pictures, a search algorithm finds those pictures in the library that *match* the information in the A-Box best. The best matching picture is included in the generated document, at a location that corresponds with their contribution to the text (Cahill et al., 2001)

This retrieval-based treatment of pictures is especially useful in applications like the Patient Information Leaflets, where there exists great variety between the types of pictures that are used, and where pictures tend to be reused heavily. Our work in this area has also been integrated into an interactive document authoring system, called ILLUSTRATE, which allows users to specify the semantic content of a document, leaving all details of graphical as well as textual expression to the system (Van Deemter and Power, to appear).

4.2. Pictures and Style

Even though pictures and text can express similar kinds of information, they do not have the same strengths and weaknesses. One strong point of pictures, for example, is the immediacy with which they tend to be understood (Pineda, 2000). Take a quick look at the leaflet under discussion, for example, in its illustrated version. Certain aspects stand out with much more clarity and immediacy than others: headers, for example, have a high perceptual salience, and the same is true for pictures. A related strength of pictures is that they are language-independent, making them especially suitable for conveying information to linguistic minorities.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

Another strong point of pictures, relating to their iconicity (see above), is their suitability for indicating information relating to the relative locations of objects. Consider the first of the two pictures in the leaflet, for example. It can be expressed textually, but to express everything that the picture conveys would tend to be cumbersome:

> Take your tablet by removing it from the foil by pressing your finger against the back of the tablet.

An informal study of the leaflets in APBI (1997) has shown that pictures are used in about 60% of the leaflets, and that they are used heavily to depict:

- complex pieces of equipment (anti-asthmatic inhalers, inoculators, etc.) whose spatial layout the reader of the document it is important to understand.
- actions, such as the steps that need to be taken to clean an inhaler. Often, entire sequences of actions are depicted.
- continuous quantities – e.g., when creams and ointments are used, one frequently sees depictions of the required quantity, sometimes positioned on a finger or juxtaposed to a coin to show the relative size of the blob.
- parts of the human anatomy.

Some companies do not use pictures at all. Those companies that avoid pictures appear to also favour a more formal textual style (section 2.3).

4.3. Pictures and Wording

Just like layout (see section 3), pictures are more than the cherry on the cake: when pictures are used well, they are an integral part of a document, which affects wording (and, obviously, layout). Although neither the RICHES system (Cahill et al., 2001) nor the ILLUSTRATE system (van Deemter and Power, to appear) have taken this fully into account, let us explore some of the ways in which pictures and wording can influence each other. First, we will discuss references to illustrations. Secondly, we will explain how texts may be reduced because of information expressed in the picture.

4.3.1. References to illustrations

References to illustrations can help to improve the readability of the text, by highlighting a connection between two or more of its parts. A sentence like ‘Compare Fig.7 of the previous subsection’, for example, makes crucial use of such a reference. Our semantic approach to pictures allows us to generate such references, whether they are based on the location of the picture (as in ‘See Fig.7’) or on its pictorial content (as in ‘See the picture *of the inhaler*’). Conversely, the fact that a picture depicts a certain domain object can also be exploited for identifying *that object*. An example would be the noun phrase ‘the red vial depicted in Fig.7’, which refers to a domain object via a document part. Descriptions of this kind can be

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

generated using a variant of standard generation algorithms (e.g., Dale and Reiter, 1995) provided document-related properties like 'being described by Fig.7' are treated on a par with other properties (such as being red, being a vial, etc.) This is only possible in a system that 'knows', of every picture in the document, which domain object it refers to. Algorithms for automatically generating 'document deictic' references of this kind are discussed in Paraboni and van Deemter (2002) and implemented in a small stand-alone generation program.

4.3.2. Reducing text

Something that is expressed through a picture may no longer have to be expressed textually. For example, let us return to the earlier-discussed removal of a tablet. If the picture is used then there is no obvious need for the text generator to explain in full detail how the tablet is to be removed, since this is already conveyed by the picture. Instead of (1) below, it is now sufficient to write (2), (3) or even (4):

(1) Take your tablet by removing it from the foil by pressing your against the back of the tablet
(2) Take your tablet by removing it from the foil (by pressing it through).
(3) Take your tablet by removing it from the foil.
(4) Take your tablet.

Mechanisms for achieving the kinds of reductions exemplified in (2)—(4) are described in van Deemter and Power (to appear). Note that the choice between these four possibilities is not trivial, since some degree of overlap between the information conveyed by text and pictures may be desirable.

5. FROM DOCUMENTS TO DIALOGUE VIA SCRIPTS

In this section, we move from documents to dialogue. The step from documents to dialogue might at first sight seem like a big one. Documents are objects, whereas dialogues are events. In order to bridge this gap, our initial concern will be not so much with dialogues as events, but rather with the written records of such events. In fact, our focus is first and foremost on records of fictitious, as opposed to real, dialogues. We use the term dialogue script to refer to a record of a dialogue, regardless of whether it is a real or fictitious dialogue.

5.1. Dialogue Scripts

We have defined dialogue scripts in terms of their origin, i.e., as records of dialogue events. Let us now move to a functional perspective on dialogue scripts. We want to discuss how they can be put to use. We focus on one specific type of exploitation: dialogue scripts used as *discourse*. From this perspective, dialogue scripts are on a par with, for instance, argumentative discourse and narrative: they are *used* for communicating with an audience. For instance, if a dialogue script is written entirely

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

by an author, this author is comparable to the author of any other type of discourse: she utilizes the discourse to communicate with the readers of that discourse. Note that even if the dialogue script is based on a real dialogue, it can still be used as a device for communication: the person(s) who selected the real dialogue and subsequently transcribed and adapted it might have done so with a particular audience and message in mind.

Philosophers of language have been occupied for a long time with the question what counts as a communicative act. Arguably, the most influential account is due to Grice (1957). Roughly speaking, Grice proposes that a speaker tries to communicate the message *m* by signal *s* to an audience if and only if the speaker intends that by presenting *s* the audience will recognize the message m partly by recognizing this very intention of the speaker.

It is characteristic of scripted dialogue that it involves communication on two layers. The dialogue script is not only a means for communication with an audience, but also a report of a communicative event – the dialogue – which has its own participants. These participants are usually different form the author and audience of the dialogue script. The presence of this second layer of reported communication provides some new opportunities for communicating information on the first layer. These opportunities are, however, orthogonal to the possibilities which text, layout and pictures provide. For instance, we have seen that, in text, rhetorical structure can be expressed by means of discourse markers. Since a dialogue script is a text, this device is still available as witnessed by the use of 'although' in the following fragment of a dialogue script.

> Although the patient asked when he should take the medicine, the pharmacist could only reply to him that his doctor would be able to tell him.

Layout can be used to enhance dialogue scripts in various ways. In the example above, the underlying dialogue structure is expressed through the wording ('the patient asked', 'the pharmacist could only reply'). Roughly the same meaning can be conveyed through layout:

Patient:	When should I take the medicine?
Pharmacist:	Your doctor will be able to tell you.

The choice between (in)direct speech and layout to convey dialogue is influenced by the same considerations which apply to the choice between plain text, enumerations and lists. For example, we pointed out that requirements concerning the navigability of a document can suggest a particular type of layout (e.g., a bulleted list). Now, consider a dialogue script, which is used as a direction to actors. During rehearsals, actors will often refer to individual turns while reciting them. Hence, like in instructional text, it becomes important that it is easy to re-enter the text after looking away, in order to retrieve the next turn. Character names are salient re-entry sites. A presentation of a dialogue script in the form given below can basically be seen as a *labelled* list, where labels occupy the place that is taken by bullets in a bulleted list.

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

Speaker *A*:	Text of turn.
Speaker *B*:	Text of turn.
…	…

As in previous sections, formatting is not just an embellishment of text. The differences between the pronouns in the two examples ('he' *vs* 'I', 'his' *vs* 'your' and 'him' *vs* 'you') show that wording interacts with layout. Finally, a scripted dialogue can also include pictorial illustrations. Again, reasons for adding pictures to (formatted) text can also be reasons for adding pictures to the (formatted) text of a dialogue script. This is not just a theoretical possibility. For instance, the website of the UK Department of Health (http://www.doh.gov.uk/adguid.htm.) presents some of its information in the form of question-answer sequences. Additionally, these are elaborated with illustrations.

A characteristic of scripted dialogue is that it introduces multiple voices into a text. Every text has an authorial voice. Through the text the author expresses ideas, emotions and opinions. Authors have, however, various devices at their disposal to introduce further voices that can articulate different or incompatible ideas, emotions, opinions, etc. In a scripted dialogue the voices are not only expressing further points of view, but also interacting with each other.

In plain text, different voices can be introduced through direct and indirect speech. We have seen that in dialogue scripts the same can also be achieved through layout. The move from text to layout has an interesting consequence. We move from an explicit report ('X said …') to one where the authorial voice is implicit in the layout, thus creating more distance between the reported dialogue and the reporter/author of the dialogue. Thus, the reported dialogue is dissociated from the reporter and associated more strongly with the characters involved in the dialogue. The reader is lured into viewing the reported utterances as genuine performances of these characters rather than creations or interpretations of the reporter/author for which the reporter/author can be held responsible.

A dialogue script with believable characters creates new opportunities for influencing its audience. The characters can be given certain traits that affect the interpretation by the audience of what they are saying. For instance, consider an author who is not a medical expert yet wants to communicate some piece of medical advice. One way to do this is to directly address the audience. Alternatively, s/he could report a (real or fictitious) utterance by a medical authority – say a pharmacist. The latter strategy would allow the author to imbue the communicated information with the authority of the character.

Scripted dialogue also provides new means for conveying rhetorical structure. Some researchers (e.g., van Kuppevelt, 1995) have suggested that the underlying rhetorical structure of discourse can be understood in terms of the (implicit) questions which the sentences in the discourse address. Let us go back to the meaning representation in section 2.1. The following dialogue script could be used to express the information it contains:

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

(1)	Pharmacist:	Here is your medicine.
(2)		Store it away from children.
(3)		Your doctor should suggest a dose.
(4)	Patient:	Can I change the dose?
(5)	Pharmacist:	No, you should follow your doctor's advice.
(6)	Patient:	When should I take the medicine?
(7)	Pharmacist:	Ask your doctor.
(8)	Patient:	How should I take the tablet?
(9)	Pharmacist:	Remove it from the foil by pressing
(10)		your finger against the back of the tablet.
(11)		Then swallow it with water.
(12)	Patient:	What if I take an overdose?
(13)	Pharmacist:	Tell your doctor or visit your
(14)		hospital's casualty department.

Here, (4) expresses a possible question which (5) answers. Thus (4) legitimizes (5) and reveals the purpose of (5). Complex rhetorical structures, such as those expressed by the following formula (number 4 of section 2.1) can also be expressed in a transparent way in a dialogue script.

```
procedure(take(z,t:tablet),
       [remove(z,t,foil,finger,back(t)) &
        swallow(z,t,water)])
```

This is achieved by distributing information amongst the interlocutors. In our script, the patient expresses the goal of the procedure (8) whereas it is the pharmacist who explains the steps (9—11).

Piwek and van Deemter (2002) provide examples of human-authored dialogue scripts with subdialogues which at first sight contribute no new information, for instance, subdialogues consisting of a clarification question and an answer. From the perspective of the interlocutors in the dialogue such subdialogues play an important communicative function. It is, however, less clear what function they play in the communication from the author of a dialogue script to the readers/audience. Why does did the author include these subdialogues? On the basis of the aforementioned examples Piwek and van Deemter (2002) hypothesize that subdialogues can serve the purpose of emphasising information. The author of our dialogue script between the pharmacist and the patient might have employed such a strategy if, for some reason, s/he thought that the information conveyed in 13. and 14. should be brought to the reader's attention with extra force. The idea would be to insert, for instance:

(15)	Patient	Do I really have to?
(16)	Pharmacist:	Absolutely, tell your doctor or
(17)		go to the casualty department.

This example illustrates that dialogue provides a natural medium for repeating information to achieve emphasis. Dialogue thus provides us with a new device for

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

emphasizing information in addition to the methods discussed earlier on, such as the use of typography.

5.2. Scripted Dialogue

We started this section by pointing out the gap between documents and dialogue. We have suggested that dialogue scripts might function as a stepping stone between the two. In the remainder of this section, we want to take the final step to dialogue. We already pointed out that dialogue scripts are written records. From here it is, however, a small step to scripted dialogue, i.e., the performance of a dialogue script by two or more actors. Thus, whereas a dialogue script is an object, a scripted dialogue is an event. The automated generation of scripted dialogue has been pioneered by André et al (2000).

Here we want to briefly describe an architecture for generating scripted dialogue which has been implemented in the NECA project (Krenn et al., 2002), automating both the generation of the dialogue script and the subsequent performance of that script. The input to the system consists of (a) a database or conjunction of logical formulae (as described in section 2.1), possibly annotated with further pragmatic information (e.g., which information is important) and (b) information about the characters (personality traits, role and interests). A pipeline architecture is employed: the system takes the input and puts it through the following modules:

1. A Dialogue Planner, which produces an abstract description of the dialogue (the dialogue plan).
2. A multi-modal Generator, which specifies linguistic and non-linguistic realizations for the dialogue acts in the dialogue plan.
3. A Speech Synthesis Module, which adds information for Speech.
4. A Gesture Assignment Module, which controls the temporal coordination of gestures and speech.
5. A player, which plays the animated characters and the corresponding speech sound files.

Each step in the pipeline adds more concrete information to the dialogue plan/script until, finally, a player can render it. A single XML compliant representation language, called RRL, has been developed for representing the dialogue script at its various stages of completion (Piwek et al., 2002).

In Piwek and van Deemter (2003) a revision-based approach to enforcing constraints on dialogue style (e.g., amount of emphasis by means of subdialogues) is presented. The revision takes place on the level of the dialogue planner. We assume that a first draft of the dialogue plan is generated, after which a number of revision operations can be applied (one of them involves inserting subdialogues) for optimising the dialogue with respect to a number of constraints. In particular, Piwek and van Deemter examine constraints on the length of the dialogue and the amount of emphasis expressed by means subdialogues. Specific attention is paid to the problem of determining how to simultaneously satisfy potentially conflicting constraints. In order to avoid situations where one constraint wins at the expense of

Published in: O. Stock and and M. Zancanara (Eds.), Intelligent Multimodal Information Presentation.

Text, Speech and Language Processing, Volume 27. Dordrecht: Springer

others, a number of game-theoretical solutions to this problem are examined (such as the Nash arbitration plan). This approach exemplifies a recent trend in formal and computational linguistics to put to use techniques from decision and game theory (Rubinstein (2000), but see also Jameson (1987)).

6. CONCLUSION

We have highlighted some issues in generating multimedia, using a thought experiment (based on several implemented generation systems) in which exactly the same semantic content is expressed (a) in plain text, (b) in text with layout, (c) in text with pictures, and (d) in a dialogue script that can be performed by two animated agents. The most general point to emerge is perhaps an obvious one*: a good presentation cannot be produced merely by adding the various media together, like ingredients mixed in a bowl*. Wording must be adapted to take account of layout and pictures; and in a dialogue, speech must be coordinated with gestures and other behaviour. What is not at all obvious is how this integration can be achieved in a multimedia generation system.

Our approach to multimedia generation in ICONOCLAST and NECA has been based on the notion of *constraint satisfaction*. The generated presentations in each medium have to conform to several kinds of constraint. First, they must be consistent with the semantic content (so ensuring that they 'sing from the same hymn sheet' – i.e., do not contradict one another). Secondly, they must respect any general requirements on style, so that for example a sober scientific article is not combined with frivolous pictures or layout. Thirdly, they must respect constraints on *cross-media realisation*, so that each is designed to take account of the others. This third category has been the main focus of our examples.

Recently, we have started to apply this approach to the generation of dialogues (Piwek and van Deemter, 2003). Here we concentrate on the generation of dialogues as presentations for an audience (i.e., dialogue in drama, comedy, feature films, commercials, infomercials, etc.). The focus is on the effect of the dialogue on the audience. This in contrast to traditional work on dialogue, which concentrates on the influence of dialogue (acts) on the participants of the dialogue.

The integrated approach contrasts with one in which the various media are considered in sequence. Imagine, for example, a system that first produces a plain text (perhaps a section for a patient information leaflet), then adds layout, then adds pictures, through three modules organised in a 'pipeline'. Such a system can achieve some degree of integration, because the layout and picture modules have access to the generated wording. However, the options open to the later modules will be restricted by decisions taken unilaterally by the text generator, so that the final result, even though well-formed, is likely to fall below the quality obtainable through integrated planning.

To state constraints across different media, it is useful to have a unified description language for the multimedia document. Our first step towards such a representation has been the notion of 'document structure', introduced in the ICONOCLAST and RAGS projects (Power, 2000; Power et al., 2003a; RAGS, 1999),