The Web as Dialogue: the role of natural language generation in hypertext

Robert Dale†, Maria Milosavljevic† and Jon Oberlander‡
†Language Technology Group, Microsoft Research Institute
School of MPCE, Macquarie University, New South Wales 2109 Australia.
{mariam,rdale}@mpce.mq.edu.au
‡Human Communication Research Centre
University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland.
J.Oberlander@ed.ac.uk

Abstract

A very common metaphor used when discussing the World Wide Web, and hypertextual systems in general, is that of navigation within a space. An alternative metaphor is to view a session with any such system as a dialogue or conversation. This potentially richer metaphor acknowledges the interactive nature of the experience, and may open up new ways of thinking about how the medium can be used. Unfortunately, existing hypertext systems are not up to the task: a true dialogue requires a conversational participant who is able to take account of the content of the dialogue so far, and to respond appropriately in context. Existing systems are, by and large, providers of pre–written, canned material. By using natural language generation techniques, however, we can tailor the system's contributions to take account of the ongoing discourse context. In this paper, we look at how natural language generation techniques can be used to enhance the coherence of the system's turns in the conversation. We demonstrate the ideas by reference to two Web-based language generation systems we have constructed, the PEBA-II system and the ILEX system.

Introduction

The explosion of interest in the World Wide Web in the last few years has resulted in the availability of a vast amount of on-line textual material. Researchers in natural language processing now have an immense corpus at their finger–tips; not surprisingly, this has resulted in a variety of attempts to apply natural language analysis theories and techniques to some useful effect. Relatively little attention has been paid, however, to the role that natural language *generation* techniques can play in the context of the Web. There is tremendous scope for using these techniques for the automatic creation of new textual content for delivery on the Web. This paper looks at how natural language generation techniques allow us to view interactive sessions with the Web (or any other hypertext system) as dialogs between the user and the host system, resulting in a much richer interaction than is possible with static hypertext systems. They key element is that NLG techniques allow us to vary the text that is presented to the user on the basis of how they reached their current location in hypertextual space: by doing this, we can improve the coherence of the system's contributions to the ongoing conversation with the user.

The structure of the paper is as follows. In the first section, we begin by noting the benefits that flow from using natural language generation techniques in the construction of 'conventional' documents; we then indicate how the use of natural language generation in a hypertextual environment like the World Wide Web works to the mutual advantage of both technologies. The next section provides an overview of two working NLG systems which generate multi-paragraph descriptions of entities as WWW pages. We then focus on one of the characteristics of these systems, and argue that by viewing a user's interaction with the Web as a dialogue, we can utilise natural language generation techniques to enhance the coherence of the resulting conversation. We examine two aspects of coherence that can be addressed using NLG techniques, which we call TEXTUAL COHERENCE and CONCEPTUAL COHERENCE. We end the paper by pointing to some issues that arise as a result of viewing the Web as dialogue.

Natural language generation and hypertext

Below, we first review the benefits that follow from using natural language generation techniques to produce documents in general. We then go on to look at the benefits that follow from using natural language generation techniques in the construction of hypertextual documents.

Benefits of natural language generation

Any given document is written with a particular audience and a particular context of use in mind. Presenting the same information content, or some variation of it, to a different audience, or for use in a different

context, requires the writing of a new document. Natural language generation techniques make it possible to automatically create a variety of documents from one underlying source of information, thus making information provision more flexible, and even making possible some forms or frequency of document delivery that would otherwise be impractical or impossible. The potential benefits of NLG technology are well-rehearsed elsewhere, so we will only list them briefly here:

Reduced text construction costs: Writing can be a very expensive task. If the same information has to be presented in a variety of different documents, perhaps expressed slightly differently in each case, the cost is multiplied, even where much of the same material can be reused. Provided the number of different documents to be created is large enough, the expense of this task can be reduced by using a single knowledge base of facts which can be realised into multiple natural language texts using NLG techniques.

Variation according to purpose: Often the same essential information needs to be presented in different ways: for example, a reference manual for a piece of software may contain essentially the same information as a user guide or tutorial document for that piece of software, but that information will be organised differently, and different elements will be presented with different emphases. Natural language generation techniques can leverage off a representation of the common underlying knowledge to produce a variety of documents from the one source.

Variation according to user characteristics: Each user of a document will have a different set of background knowledge. Natural language generation techniques make it possible to countenance the creation of documents that are tailored to individual needs, provided of course some computationally tractable characterisation of the user can be obtained (see Paris 1987).

Variation according to target language: Natural language generation techniques make it possible to produce documents in different languages using the same underlying knowledge source. At the expense of creating an abstract underlying representation of the information to be conveyed, this approach avoids the problems inherent in achieving high quality machine translation.

Variation according to resource boundedness: Sometimes the most appropriate content of some linguistic interchange will depend on the time and space available. Natural language generation techniques enable the provision of different texts for users in different resource-bound situations.

Interpretation of unfriendly data: It is often said that a picture is worth a thousand words; but there are contexts where, faced with a large set of statistical or numerical data, it can be hard for a user to determine what is salient and what is not. In many such situations, graphical means may not help; often what is required is a concise summary of what is important. NLG techniques provide powerful mechanisms for generating natural language summaries of large datasets.

Benefits of integrating NLG and hypertext

Combining NLG techniques with hypertextual information delivery results in a dynamic environment in which the benefits of each technology is enhanced by the presence of the other. In fact, the two technologies compensate for each other's deficiencies. We call the result of this merging DYNAMIC HYPERTEXT.

NLG benefits hypertext Hypertext interfaces have been used for many information delivery systems, including help systems and on-line encyclopædias. The World Wide Web provides a familiar and consistent hypertext interface for a broad audience, and requires relatively little effort to utilise for an information delivery system. Not surprisingly, a substantial amount of online information has been adapted to take up this presentational opportunity.

Employing natural language generation techniques to automatically generate the textual content of such systems brings a number of advantages, in addition to those that hold for 'conventional' documents. These have been discussed at some length in the previous literature: see in particular (Reiter et al 1992), (Moore 1995) and (Milosavljevic et al 1996). For our present purposes, the most significant benefit that NLG brings to hypertext is the scope for using a discourse history to ease the user's navigational burden. In a hypertext environment, we can keep track of the nodes which the user has previously visited, as well as the content which was delivered to the user at those nodes. This allows us to adapt how the user sees the hypertext network and the content of subsequent nodes. By tailoring the presentation of information to take account of previous context in this way, we may be able to alleviate some of the effects of becoming 'lost in hyperspace' (see Conklin 1987).

Hypertext benefits NLG Of course, natural language generation techniques are not without their problems, and many issues in developing and using the

technologies remain. It turns out that the medium of hypertext can help overcome some of the outstanding problems in NLG. In particular, a big problem in NLG is the task of discourse planning: this involves working out what the structure and content of multi-paragraph texts should be, and often requires access to a detailed model of the user's needs and background. Obtaining and maintaining such a model can be anywhere from very difficult to effectively impossible. Here a real benefit is gained from using hypertext: the user of a dynamic hypertext system can be allowed the freedom of performing high-level discourse planning, thus removing some of this responsibility from the system. The result is an environment where the user has freedom to explore, but this freedom is constrained by the hypertext links which the system provides to the user.

As a particular hypertextual medium, the World Wide Web provides some additional, practical advantages:

- The number of users who can use the system without additional training is very large.
- A separate interface module does not have to be developed; a dynamic hypertext system must only produce the output in HTML format.
- The incorporation of multi-modal information is inexpensive.

In summary, the use of natural language generation techniques permit three things that would not be achievable otherwise:

- we can constantly update the text on the basis of an underlying knowledge source;
- we can vary the text on the basis of a model of the user; and
- we can vary the text on the basis of what has been shown to the user before.

In the next section, we provide an overview of two working dynamic hypertext systems, and the section that follows focuses on the last of the three benefits outlined above: the use of the discourse history, which most directly overcomes some of the limitations of static hypertext.

Two Web-based hypertext generation systems

The ILEX System

The ILEX system[1], developed at the University of Edinburgh, produces descriptions of the items displayed in the National Museums of Scotland's 20th Century Jewellery Gallery. An example description produced by the system is shown in Figure **??**.

The primary goal of the ILEX system is to produce accurate and interesting labels of museum items which convey the important aspects of the items to the reader. The system must therefore weigh these three criteria against each other in order to educate the visitor about the items held in the gallery. ILEX achieves these goals by using a system agenda to reason about its goals, the information space and the user's knowledge. It provides the user some freedom to explore the gallery, but only within the constraints it sets through its educational goals. For more details on the system, see (Knott et al 1996).

The PEBA-II System

The PEBA-II system[2], developed at Sydney's Macquarie University, produces descriptions and comparisons of animals. Figure **??** shows a description produced by the system.

The main goal of the PEBA-II system is to explore techniques for developing an intelligent encyclopædia, where different descriptions are provided to different users based on their knowledge, previous interaction with the system, or other contextual factors. When describing a new concept to a user, the system can make comparisons with known entities in order to tailor the description to the user's knowledge, or to ensure that the user does not confuse two similar entities. For example, the description of the alligator in Figure **??** contains a comparison with the crocodile in order to distinguish the two entities.

For more information about the PEBA-II system, see (Milosavljevic et al 1996).

The Web as Dialogue

The key shift in thinking that natural language generation brings to hypertext results from the fact that the system's output varies depending upon the previous interactional context. We often talk in terms of a navigational metaphor when we describe interactions with the Web: we talk of moving around in a space, of jumping from place to place, of being lost, and so on. The alternative metaphor that NLG ushers in is that of web sessions as dialogues or conversations, where the system is a conversational partner. This acknowledges the interactive nature of the experience, and may open up new ways of thinking about how the medium can be used.

[1] A version of the ILEX system is available on the WWW from URL: `http://cirrus.dai.ed.ac.uk:8000`

[2] A version of the PEBA-II system is available on the WWW at URL: `http://www.mri.mq.edu.au/ltg/peba/`

Figure 1: A description produced by ILEX

Figure 2: A description produced by PEBA-II

It is, of course, possible to view existing static hypertext systems in terms of this metaphor. The conversation consists of a sequence of turn-taking events, alternating between mouse clicks on the part of the user, and the delivery of the material at the end of a URL by the system. The semantics of the user's mouse clicks are effectively questions whose nature is determined by the answer received, or more specifically, by the intention of the author in writing the document returned. The lack of explicit typing of links in Web pages, and the fact that the same page will be returned in response to mouse clicks in potentially very different contexts, means that the user's question can only be characterised in the vaguest terms: at best, something along the lines of 'tell me something about x', where x is the textual manifestation of the link. This is a rather impoverished notion of dialogue.

Using NLG techniques to dynamically construct the response to a mouse click means that it can be tailored to the context in which the 'question' was asked, thus allowing for different contextually-defined questions to be answered in different ways. This is still not real dialogue, of course: if the only dialogue contributions available to the user are mouse clicks on links on the screen, then the system is in effective control of the dialogue because it determines what links will be made available.

Coherence in dialogue

One property that distinguishes a dialogue from a random collection of utterances is that, in a dialogue, each conversational participant's contribution is related to the previous utterance by the other conversational participant. This observation is what underlies the notions of question–answer pairs and turn-taking, and other related concepts, in work on conversational analysis.

Much of this essential property comes down to COHERENCE: the way that the dialogue contributions flow together and connect to make a seamless whole. A dialogue is not jerky or discontinuous; it is fluent. For a hypertext conversation to be coherent, a number of properties are required:

- Each 'utterance' should be coherently related to the previous one. Within a monologic context, a considerable amount of research in natural language generation has looked at the role of rhetorical relations in ensuring that this property holds. Within a hypertextual dialogue, this is controlled by only offering valid rhetorical moves.

- The conversation should flow in terms of thematic structure and focus-shifting. Ideas like those developed in Grosz, Joshi and Weinstein's centering theory may be useful here: we can think of the set of links from a web page as offering forward-looking centers, any one of which the user can choose to accept.

- Previously unseen concepts should be introduced appropriately, so that we do not assume knowledge that the user does not have.

In a dynamic hypertext system, the resource that makes coherence possible is a discourse history: a record of the content of the utterances the system has provided to the user. In PEBA-II, for example, we model this by allocating a unique user ID to each user at the start of a session, then including this as a parameter in all subsequent requests the user poses to the system. This enables the system to maintain an arbitrary number of discourse histories simultaneously. On the basis of this information, the system can construct subsequent texts taking into account the information that has already been presented to the user.

We can distinguish two kinds of coherence in text:

Textual Coherence: a textual contribution should use syntactic resources that are consistent with the thematic development of the text;

Conceptual Coherence: the content of a text should not make reference to concepts that are unknown to the reader except in order to introduce them to the reader.

These types of coherence matter both within web pages, and between them. Below, we show how these principles can be embodied in a hypertext system using NLG.

Textual coherence between pages

First, let us consider an aspect of *between-page coherence*. PEBA-II allows the user to explore the Linnaean taxonomy of animals. When we describe an animal, the text schema[3] we use always relates a class to its superordinate class and its subordinates. This means that any node in the taxonomy may be reached from one of two directions: either from the node immediately underneath, or from the node above. The user may also reach a node from a document describing a similar entity which has a link to the node, or she may commence interaction with the system at any given node, coming on from nowhere, so to speak. Different texts are produced in each case; see (Dale and Milosavljevic 1996).

To take the last of these possibilities first: suppose we are at the beginning of a session with the system where no previous nodes have been visited, and the user requests a description of the marsupial class; we might then generate the following text:[4]

> The Marsupial is a type of Mammal that carries its young in a pouch. The Marsupial has the following subtypes ...

However, suppose the user reaches the marsupials node from the mammals node, as would be the case after reading the following text and then clicking on the marsupials link:

> There are three kinds of mammals:
>
> - the monotremes
> - the placental mammals
> - the marsupials

This would result in the following text:

> The Marsupial differs from other Mammals in that it carries its young in a pouch. It has the following subtypes ...

On the other hand, suppose we reach the marsupial node from viewing a description of the kangaroo class; i.e., after reading the following text:

> The kangaroo is a kind of marsupial which has a powerful tail and back legs.

Here, if we click on marsupial, then we reach the marsupials node from below, with the resulting text as shown below:

> Apart from the Kangaroo, the class of Marsupials also contains the following subtypes ...

By varying the way in which the information is presented in this way, we can generate more fluent texts. In each case, the way in which the queried class is introduced is determined by how best this fits into the ongoing discourse.

We have only begun to explore the potential here for varying the text to take account of the discourse history: in the example just given, this is analogous to modifying the introductory section of a document for different audiences but leaving the body of the text substantially the same.

Textual coherence within pages

Secondly, let us turn to an aspect of *within-page coherence*. Consider the following two alternative ILEX texts, each of which describes the same necklace:

1. This necklace is in the arts-and-crafts style. ... It has very elaborate festoons. It has faceted stones.

2. This necklace is in the arts-and-crafts style. ... Arts-and-crafts jewels tend to be intricately worked; for instance, this piece has very elaborate festoons. However, unusually for arts-and-crafts jewellery, this piece has faceted stones. Most arts-and-crafts jewels (see for example the jewels in case 8) have cabochon stones.

Obviously, neither text is ideal; both might benefit from aggregation, and the need to establish conceptual coherence (see below) might require definitions for terms like *festoons* and *cabochon stones*. Nonetheless, the use of generalisations in (2) helps, in part because without them, the propositions describing the properties of the selected jewel do not in fact stand in obvious relationships to each other. Indeed, exploiting aggregation in (1) as its stands would tend to make matters worse.

To generate text like (2), ILEX's content selection algorithm collects all the simple facts involving the selected jewel. For each fact, a search is made of the defeasible and indefeasible rules which can be expressed in connection with these facts. For a rule to be expressible, the general class about which the rule holds must be introduced by a simple fact (here, *This necklace is in the arts-and-crafts style*); this prevents the expression of the rule from acting as an unexpected topic-shift. The introducing fact is then linked to the generalisation (here, *Arts-and-crafts jewels tend to be intricately worked*) via the coherence relation Definition. The generalisation is then linked back to another simple fact about the jewel, by an appropriate relation. Here, with

[3] See (McKeown 1985) for a discussion of schema–based text generation.

[4] Underlined words in these examples correspond to hypertext links.

for instance, this piece has very elaborate festoons, it is Exemplification, since the next fact accords with the rule); but equally, the relation could be Concession, if the fact located does not accord with the rule. Other aspects of (2), and the wider use of generalisations to improve textual coherence, are discussed in Knott et al (1997).

Conceptual coherence

More radical variations in the text are possible if we also strive for conceptual coherence. One way of doing this is to make use of the longer-term discourse history by making comparisons with those concepts which have been described to the user in the past discourse.

Milosavljevic and Dale (1996) describe the types of comparisons which are found in encyclopædia descriptions of entities. These comparison types are employed in the PEBA-II system in order to produce descriptions which aid the user's conceptualisation of the entities being described. The goal here is to describe new concepts by comparing them to concepts which the user is already familiar with. There are two main types of comparisons which are used in order to achieve this goal:

Clarificatory comparison: If there is another entity which shares several salient features with the *focussed entity*, and if the user is familiar with that similar entity, then a clarificatory comparison of the two enties will aid the user's understanding of the focussed entity. For example, the description of the Alligator in Figure **??** includes a clarificatory comparison with the Crocodile. If the user already has knowledge of the Crocodile, then this comparison would aid her understanding of the Alligator. Of course there are other reasons why clarificatory comparisons might be made, such as to distinguish potential confusors.

Illustrative comparison: If the focussed entity has a property which is shared by a commonly known or a recently described entity, then an illustrative comparison can be made with that entity in order to allow the user to more easily understand the property and also to make an association between the two concepts. For example, the PEBA-II system produces sentences such as *The platypus is about the same length as the domestic cat.*

The ILEX system produces illustrative comparisons with those items described in the discourse history. For example, the following sentence is produced in order to highlight the similarities of the current entity to a recently described entity:

> Like the earlier piece (Gold and enamel pendant necklace), this piece was designed by Jessie M. King (Glasgow) in c. 1905.

In the PEBA-II system, it is assumed that the user reads all the text displayed and, additionally, is a perfect learner—she will remember everything. By contrast, in ILEX, some pieces of information are re-iterated; how often depends both on the item's assimilation score, and on the user-type's assumed assimilation rate. It remains to be seen which strategy will prove to be the more effective; it is possible that different strategies will be appropriate in different domains.

Open questions

Obviously, the systems we have described require structured knowledge bases from which to generate text. This undoubtedly limits their generality, but there are two reasons for hope. First, the constraint does not mean that they cannot be used in conjunction with relatively unstructured text bases. Indeed, an explicit goal of the ILEX project is to permit the generator flexibly to combine both full-generated text with canned, human-generated text. Secondly, of course, the use of robust parsing techniques on existing text bases on the Web may allow us to represent existing texts at a level deep enough to allow NLG techniques to be re-applied to them. While our methods make room for this possibility, they do not presume it.

It should also be noted that there are some respects in which the metaphor of Web as dialogue does not quite fit as comfortably as we might hope. In particular:

- In a real conversation or dialogue between two human participants, we can usually assume that the other party is listening to what we are saying, and asking questions when they don't understand what has been said. The situation is somewhat different on the Web: from introspection alone, it is obvious that people do not read Web pages from beginning to end. Rather, they skip around, skim a page, very quickly scan the material to see if there is something of interest. This suggests that, in generating text for the Web, we should assume skimmers rather than good listeners. There is a fine balance to be achieved here: on the one hand, NLG techniques give us a way of avoiding saying things we have said before on some previous page; on the other hand, we have no way of knowing whether the reader even noticed that something was said on a previous page.
- Real conversations only go forwards. The Web, however, allows us to go backwards. This raises the ques-

tion of what should happen when the user asks a second time for a description of some entity. In an NLG-based system, this corresponds to re-accomplishing a goal that has already been accomplished; and so the the system has two choices, which we term RESTATEMENT and REPETITION. In restatement, the reposting of goal leads to a new realisation of the content, where the interim discourse history—all the things that have been said between the first realisation of this goal and the second realisation—makes a difference. In the case of repetition, we have what amounts to a request for verbatim re-realisation, so that the interim discourse history is effectively ignored for this realisation. Which is the best strategy? This is not clear: on the one hand, users expect things to be pretty much as they were last time they saw them, thus favouring repetition; on the other hand, restatement is closer to what happens in 'real' conversation.

Conclusions

In this paper, we have suggested that a shift of metaphors—from navigation to conversation—allows us to think about interactions with the Web, or any hypertextual system, in different ways; and we have argued that natural language generation technologies make this shift possible. We have focussed in particular on what is required to give interactions with the Web a dialogic character, and we have suggested that a key notion is that of coherence. We have shown how the notions of textual and conceptual coherence play a role in two functioning Web-based natural language generation systems.

Ultimately, the answers to the open questions depend on the expectations that users develop with regard to this new medium. The use of NLG techniques allow an escape from the restricted interaction of static hypertext, but it is too early to determine how this freedom should be balanced with the responsibility of minimising the disorientation experienced by users.

Acknowledgements

The third author is an EPSRC Research Fellow, and the Human Communication Research Centre is supported by ESRC. Thanks to Alistair Knott, Chris Mellish and Mick O'Donnell, my collaborators on the ILEX project, which is supported by EPSRC, under project grant GR/K53321.

References

Conklin, J. 1987. Hypertext: An Introduction and Survey. *IEEE Computer* 20(9):17–41.

Dale, R. and Milosavljevic, M. 1996. Authoring on Demand: Natural Language Generation of Hypermedia Documents. In Proceedings of the First Australian Document Computing Symposium (ADCS'96). Melbourne, Australia.

Knott, A.; Mellish, C.; Oberlander, J.; and O'Donnell, M. 1996. Sources of flexibility in dynamic hypertext generation. In Proceedings of the 8th International Workshop on Natural Language Generation. Herstmonceux, Sussex, UK.

Knott, A.; O'Donnell; Oberlander, J.; and Mellish, C. 1997. Defeasible rules in content selection and text structuring. Submitted manuscript.

McKeown, K. R. 1985. Discourse strategies for generating natural-language text. *Artificial Intelligence* 27:1–41.

Milosavljevic, M. and Dale, R. 1996. Strategies for comparison in encyclopædia descriptions. In Proceedings of the 8th International Workshop on Natural Language Generation. Herstmonceux, Sussex, UK.

Milosavljevic, M.; Tulloch, A.; and Dale, R. 1996. Text generation in a dynamic hypertext environment. In Proceedings of the Nineteenth Australasian Computer Science Conference. Melbourne, Australia.

Moore, J. D. 1989. A Reactive Approach to Explanation in Expert and Advice-Giving Systems. Ph.D. diss., University of California, Los Angeles.

Paris, C. L. 1987. The Use of Explicit User Models in Text Generation: Tailoring to a User's Level of Expertise. Ph.D. diss., Columbia University.

Reiter, E.; Mellish, C.; and Levine, J. 1992. Automatic Generation of On-Line Documentation in the IDAS Project. In Proceedings of the Third Conference on Applied Natural Language Processing, 64–71. Trento, Italy.