Exploring a gallery with intelligent labels*

Jon Oberlander Chris Mellish Mick O'Donnell Alistair Knott

University of Edinburgh
http://www.cogsci.ed.ac.uk/people/jon/

Abstract

Ordinary labels on museum or gallery objects compromise the competing demands of differing visitors, curators and educators. Although adaptive hypermedia systems offer a degree of personalisation to the visitor, they do not go as far as they might. Intelligent labelling can achieve higher levels of personalisation, and this is desirable because it leads to a more coherent and educational visit, and because the overall interaction with the hypermedia resource becomes more like an active conversation, and less like reading a static book. We use natural language generation techniques, from the field of artificial intelligence, to generate personalised labels on demand. Our system can deliver these labels as Web pages, in an electronic gallery, or as synthesised speech in a physical gallery. We discuss aspects of coherence and conversationality, and illustrate them with a simple case.

1 Introduction

Ordinary labels on museum or gallery objects represent a compromise between the competing demands of differing visitors, curators and educators. There are various ways of tailoring them more effectively to different classes of visitor; for example, currently available audio guides can be configured to deliver information expressed as captured speech, pre-recorded at varying levels of detail.

*The first and last authors are at the Human Communication Research Centre. The second and third authors are at the Department of Artificial Intelligence. This work was supported by the UK Engineering and Physical Sciences Research Council, through grant GR/K53321. The support of the Economic and Social Research Council for HCRC is also gratefully acknowledged. Our thanks to our collaborators in the Education Department at the National Museums of Scotland, and also to Robert Dale, Janet Hitzeman, and Maria Milosavljevic.

An analogue of this technique can be adopted in hypermedia galleries, too, by confining a user to a sub-part of a larger hypermedia system. This is, effectively, what most current approaches to 'adaptive' hypermedia must do (cf. Brusilovsky [1996] for a good survey of such approaches).

Obviously, this is a step in the right direction, in that descriptions of museum artefacts are targeted more directly at their intended readers. But there are drawbacks: many (often near-identical) texts have to be authored in parallel, for the different user groups. Doing this properly will drive up the cost of the system; failing to do it properly will lead to only coarse personalisation of the visit. A still more serious drawback, however, is that most approaches to adaptive hypermedia cannot actually take into account what else the system has already presented to the visitor (cf. Mathe and Chen [1994]). That is, although the system might have a model of *who* it is talking to, it has no model of *what* it's already told them. Because of this, the system's presentation cannot really be tailored to an individual's history of interaction with the system: the hypermedia experience will not be genuinely personalised.

Personalisation is not merely a luxury: there is at least two good reasons to prefer it, hinging on the issues of *coherence* and *conversation*. The rest of this paper outlines these two reasons in a little more detail, after introducing the methods and system we have been developing at the University of Edinburgh, in collaboration with the National Museums of Scotland. With the arguments in favour of personalisation in hand, we look at a simple case which allows us to say in a little more detail how our system achieves an appropriate degree of tailoring. We then mention some current (and future) related work, and conclude that the prospects for genuine personalisation—both in galleries, and on screens—are good.

2 The Intelligent Labelling Explorer

We have been working on an approach, which is related to adaptive hypermedia, but uses somewhat different techniques to dynamically synthesise personalised, educationally-slanted museum commentaries. These can take into account: who the visitor is; where they've already been; what they've seen; and what the curator believes they *really* ought to know about the gallery they're in now. With the National Museums of Scotland, we have been developing the Intelligent Labelling Explorer (ILEX), an opportunistic natural language generation system, which uses techniques and tools developed within natural language processing, a sub-field of artificial intelligence.

An early version of the system, and its overall architecture, has been described for a museums audience by Hitzeman, Mellish and Oberlander [1997], and a more specific discussion of recent developments, directed at computational linguists, can be found in Knott, O'Donnell, Oberlander and Mellish [1997].

For current purposes, it suffices to note that ILEX exploits a knowledge base, a user model, a discourse history and a system agenda to generate personalised descriptions of artefacts in a gallery. The knowledge base is populated partly by importing and converting records from the Museum's existing database, and partly by eliciting further curator knowledge through interview. Once populated, the automatic generation system takes over, and the simplest version of the system then outputs, on demand, dynamically built HTML pages, for web browsing. Figure 1 is a snapshot of the entry screen to the Web version.

The same main generator core can be interfaced to a speech synthesis system. Instead of delivering HTML, the system outputs text marked up in STML, a mark-up language for speech synthesis, and this can be used by any STML-aware synthesiser (Taylor and Isard [1997]). We use Festival, developed by colleagues at the University of Edinburgh. With this change, descriptions can be delivered direct to visitors in a gallery, using digital cordless phones. Thus, while an 'electronic' visitor browses web pages, and selects objects for description by clicking on links, the 'physical' visitor browses a real gallery, and selects objects for description by keying numerical codes on the cordless handset. Thus, the output device in the second version of the system superficially resembles an audio guide wand.

3 Reasons to be personal

3.1 Coherence is good

Knowing what has already been said in an interaction with a visitor allows a system to produce much more *coherent* descriptions: comments that links the current object to ones already seen, and to important messages about the gallery as a whole.

Obviously, coherence works at a number of levels. At the micro-level, text tends to be incoherent if, for example, its author keeps using a long description to refer to some entity (such as an artefact's designer), instead of using pronouns for subsequent references.[1] However, at larger scales, coher-

[1] Hitzeman et al. [1997] describe how ILEX achieves this kind of coherence.

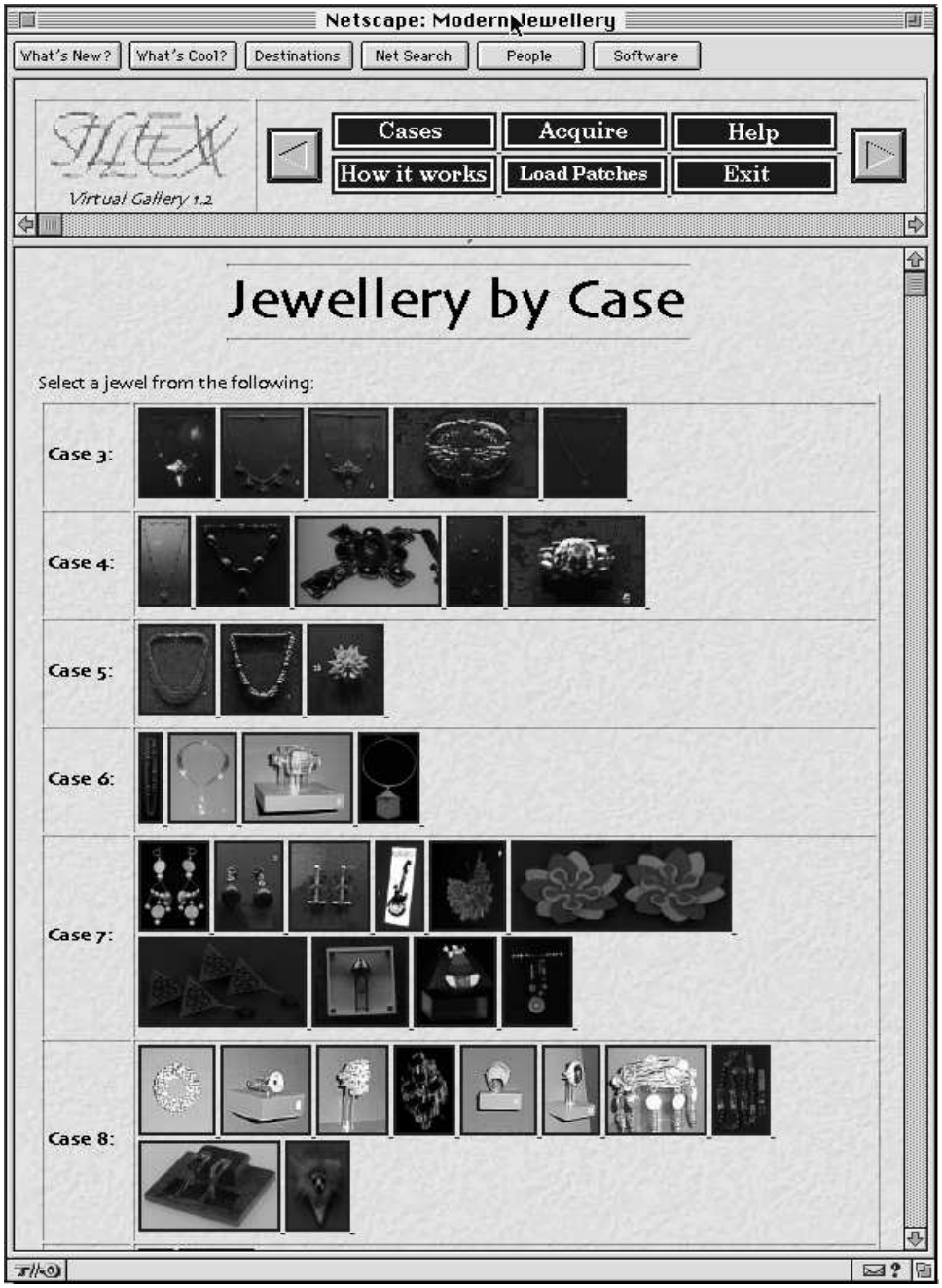

Figure 1: Intelligent Labelling Explorer. Entry screen to the 20th Century Jewellery Gallery demonstration.

ence arises from the effectiveness with which individual pieces of information are related to one another, and in particular, to the 'higher-level' knowledge that helps one understand the overall structure of some field.

If a system fails to produce descriptions that ensure this kind of coherence, then it will tend to output strings of apparently unrelated facts. These will then be relatively hard to recall. By contrast, relating particular facts to generalisations will tend to make it easier for people to remember to relevant points. Choosing which generalisations to use, and when, requires personal information about the visitor: both their general type (child/adult, expert/novice), the type of their visit (causal/educational, quick/extended), and also, more specifically, which individual facts and generalisations they've already been exposed to, or can otherwise be assumed to know.

3.2 Conversation is natural

A more general argument in favour of personalisation of descriptions runs as follows. There is persuasive evidence that people tend to treat computers in the same way as they treat people (cf. Reeves and Nass [1996]). For example, if a computer uses polite language in its help messages, people (unconsciously) judge it to be more co-operative than otherwise—just as they would judge a person. Or, if a system uses language which is associated with a character trait which matches that of its user, the user will like the system more than otherwise. For instance, if the system uses language which is associated with dominance (for example, by avoiding hedge expressions like *perhaps*, and *could*), then dominant users will like the system more, and submissive users will like it less.

It has been argued that what this body of research shows is that people cannot help responding to computers in the same ways they respond to people. Thus, good design must take this expectation into account.

Now, consider a visitor interacting with a hypermedia system. They visit a page about a necklace, then go to another, about a brooch, and then return to the first page again. What should the system do? Current design lore is that, to avoid the visitor getting 'lost in hyperspace', the system should treat pages just like pages from a book: everything should be repeated, all over again. But what do people really expect in an interaction with an intelligent system? It's true that they usually have some experience with static, unchanging books. But they have even more experience with *talking to people*. Suppose I walk from one object to another, and back again, with a human curator commenting on the objects. Would they repeat everything they said about the first object, on the previous visit, all over again? Surely

not.

Thus, while people expect books to be unchanging, they expect people—and hence perhaps also hypermedia systems—to change what they say from time to time. To do this properly, a system must take into account the history of the previous interaction with the user: once again, it must personalise its descriptions.

4 Coherence and conversation in ILEX

The abilities to generate coherent objects descriptions, and to do so in a conversationally natural fashion, are therefore very important. In this section, we discuss these issues in more detail, and in the next section, we sum up by discussing a simple case, which concisely illustrates the consequences of coherence and conversational behaviour.

4.1 Coherence in ILEX

Consider the following three alternative ILEX texts, each of which describes the same necklace:

1. This necklace is in the arts-and-crafts style. It is made of silver, amethysts and pearls. It has very elaborate festoons. It has faceted stones.

2. This necklace is in the arts-and-crafts style. It is made of silver, amethysts and pearls. Arts-and-crafts jewels tend to be intricately worked; for instance, this piece has very elaborate festoons. Arts-and-crafts jewels also tend to have cabochon stones. However, this piece has faceted stones.

3. This necklace is in the arts-and-crafts style and is made of silver, amethysts and pearls. It has very elaborate festoons and faceted stones.

Obviously, neither (1) nor (2) is ideal, since aggregation of sentences could improve them (by the joining together of information that here is presented in separate sentences), and the need to establish higher-level coherence might also require definitions for terms like *festoons* and *cabochon stones*. (2) is preferable for at least two reasons: Firstly, (2) contains more information than (1): it informs about the class of arts-and-crafts jewels as well as about the particular jewel being described. This is likely to be the sort of

material that the museum curator is really trying to get across. Secondly, this additional material allows a greater degree of *cohesion* in (2) than is possible in (1). (2) is the kind of text ILEX can produce, and the use of generalisations in it 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 be actively misleading—as in (3).

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 generalisations which can be expressed in connection with these facts. For a generalisation to be expressible, the general class it relates to 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 a discourse coherence relation of DEFINITION. The generalisation is then linked back to another simple fact about the jewel, by an appropriate discourse coherence relation. Here, with *for instance, this piece has very elaborate festoons*, it is EXEMPLIFICATION, since the next fact accords with the rule. Note that facts about an individual jewel can also conflict with a generalisation, in which case they are linked to the generalisation by the relation of CONCESSION signalled by *however*, as illustrated in the final two sentences of (2). Other aspects of (2), and the wider use of generalisations to improve textual coherence, are discussed in Knott, O'Donnell, Oberlander and Mellish [1997].

4.2 Conversation in ILEX

First, 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.[2] The situation is somewhat different in hypermedia systems, such as the World Wide Web: it is obvious that people do not read Web pages from beginning to end. Rather, they skip around, skim a page, and scan the material to see if there is something of interest. This suggests that, in generating text for the Web, we should assume that visitors are skimmers rather than good listeners.

[2]This section leans heavily on discussion in Dale, Oberlander, Milosavljevic and Knott [in press].

The obvious engineering solution is to *keep pages short*: this should increase the probability that individual words are read. Indeed, keeping it short is probably even more important in the speech-delivery version of our system: however good the speech synthesis, people don't want to stand around listening to paragraphs of speech; it is more forgiving to deliver short segments, and leave it up to the user to request more, if they want it.

Secondly, real conversations only go forwards. Hypermedia navigation facilities, however, allow us to go backwards. This raises the question of what should happen when the user asks a second time for a description of some entity. In a system like ILEX, this corresponds to re-accomplishing a goal that has already been accomplished; and so the system has two choices, which we term RESTATEMENT and REPETITION. In restatement, the reposting of a 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, designers thinks that 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. We thus prefer the latter option at present.

5 An illustration

We have tried to build ILEX so that it exhibits coherence, and conversational behaviour in some measure. To see a specific illustration of these properties, let us consider two descriptions of one object, generated by the Web version of ILEX, reproduced in Figures 2 and 3.

The second description (in Figure 3) exhibits several kinds of coherence: it uses pronouns correctly; and it refers back to Page's brooch, which was visited in the interim. Most importantly, for the current discussion, the second description introduces generalisations relating to Organic-style jewelry, and relates various features of the object to these generalisations. Thus, specific facts are related to a larger-scale knowledge structure. The curator's own educational goals are thus more likely to be achieved, than by simply listing the facts linked together here.

In addition, the difference between Figure 2 and Figure 3 exhibits the conversational nature of ILEX. Firstly, both descriptions are short—certainly,

Figure 2: First visit to the Flockinger necklace.

Figure 3: Second visit to the Flockinger necklace (after an excursion to a brooch by Page).

shorter than a single, more comprehensive description would have been. They are thus more likely to be read in their entirety. Secondly, different things are discussed on the two occasions of visiting the same necklace. It should be observed, of course, that the decision to discuss generalisations only on the second visit may not be ideal—but it was taken automatically by ILEX, which ranked the relevant facts by their user-interest and educational-importance of the relevant facts, before delivering the finished pages.

6 Related and future work

Obviously, there is other work connected to our own. To pick two examples: first, in museums, we see the closest related work as being that pursued by Stock's group at IRST (Not et al. [1997]). Secondly, and more generally, the work on dynamic encyclopedia descriptions, by Milosavljevic and Dale, at Macquarie, takes a similar line to our own. We are now collaborating with both groups, and with IRST and other partners, we are embarking on a new European project, on Hyper Interaction in Physical Spaces.

Closer to home, our next steps will involve completion of an initial speech-output version of our system, and comparative field-evaluation of a fully functional ILEX, and a 'de-personalised' version. Currently, a number of small demonstration versions are still available on the Web.[3]

7 Summary

Museum educationalists can use the objects in their collections as evidence for the main educational points they wish to communicate. Thus, describing a particular object is not simply a recital of the set of true statements about the object. Rather, the description will relate facts about the object to generalisations in the domain—and to misconceptions that the visitor may be assumed to possess. So, personalised descriptions help to sustain arguments, by supporting examples, contrasts, exceptions, defeasible and indefeasible generalisations. Our system is designed to emulate this behaviour, and it makes for a coherent, smoother tour than is otherwise possible.

Secondly, and more simply, people treat computers the same way they treat people. Thus, while hypermedia systems can be thought of as fancy books, they can also be thought of conversational partners. Our system aims to be conversational, rather than bookish.

[3] `http://cirrus.dai.ed.ac.uk:8000`

8 References

Brusilovsky, P. [1996] Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction* **6(2-3)**.

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.

Dale, R., Oberlander, J., Milosavljevic, M. and Knott, A. [in press] Conversing with dynamic documents: natural language generation meets hypertext. *Interacting with Computers*, to appear.

Festival: `http://www.cstr.ed.ac.uk/projects/festival.html`

Hitzeman, J., Mellish, C. and Oberlander, J. [1997] Dynamic generation of museum web pages: the intelligent labelling explorer. Presented at Museums and the Web, Los Angeles, March 1997.

ILEX: `http://cirrus.dai.ed.ac.uk:8000`

Knott, A., Mellish, C., Oberlander, J. and O'Donnell, M. [1996] Sources of flexibility in dynamic hypertext generation. In *Proceedings of the 8th International Natural Language Generation Workshop*, pp151–160. Herstmonceaux, Sussex, June 1996.

Knott, A., Oberlander, J., O'Donnell, M. and Mellish, C., [1997] Defeasible rules in content selection and text structuring. In *Proceedings of the European Workshop on Natural Language Generation*, pp50-60. Duisburg, Germany, March 1997.

Mathe, N. and Chen, J. [1994] Adaptive dynamic hypertext based on paths of traversal.
`http://mac-30.aded.uts.edu.au/projects/ah/Mathe.html`

Not, E., Petrelli, D., Stock, O., Strapparava, C. and Zancanaro, M. [1997] Augmented space: bringing the physical dimension into play. Presented at the International Workshop on Flexible Hypertext, Southampton, April 1997.
`http://ecate.itc.it:1024/petrelli/HT97/HT97-final.html`

Reeves, B. and Nass, C. [1996] *The Media Equation: How people treat computers, television, and new media like real people and places*. Stanford: CSLI Publications.

Taylor, P. A. and Isard, A. [1997] SSML: A speech synthesis markup language. *Speech Communication*, **18**, in press.