Conversation in the museum: experiments in dynamic hypermedia with the intelligent labelling explorer

Jon Oberlander* Mick O'Donnell* Alistair Knott† Chris Mellish*

Affiliations

* Division of Informatics, University of Edinburgh
† Department of Computer Science, University of Otago

Email addresses

J.Oberlander@ed.ac.uk

M.O'Donnell@ed.ac.uk

C.Mellish@ed.ac.uk

alik@atlas.otago.ac.nz

Abstract

We outline experience with the Intelligent Labelling Explorer, a dynamic hypertext system developed at the University of Edinburgh, in collaboration with the National Museums of Scotland. First, we indicate a number of ways in which labels on museum objects ought to be tuned to take into account types of visit, the interests of visitors, and their evolving knowledge during a visit. Secondly, we sketch the general architecture of our system, and then focus on the conversational effects which the system can create. We then briefly indicate future directions of research, before critically discussing the applicability (or otherwise) of the spatial metaphor to flexible hypertexts.

Keywords: adaptive museum hypermedia, natural language techniques for dynamic hypertext generation, content adaptation in hypertext and hypermedia, web-based museum hypermedia, navigation design

1. Introduction

We take dynamic hypertext to be that variety of flexible hypertext which relies upon techniques from the field of natural language generation (NLG) (1). A survey and comparison of existing dynamic hypertext systems can be found in (2), and further discussion on the advantages of such systems is provided elsewhere (3,4). The key feature that NLG brings is the dynamic construction of textual content on demand, at viewing time, as foreseen by Halasz (5). It has been argued that flexible hypertext systems are particularly useful for relatively formal educational purposes, since they can tailor content to particular individuals, allowing them to progress at their own rate; see Brusilovsky (6) which includes a survey of the range of educational adaptive hypermedia systems. Such systems also offer considerable potential in less formal educational contexts, such as those found in museums and galleries.

Here, we outline experience with the Intelligent Labelling Explorer (ILEX), a system developed at the University of Edinburgh, in collaboration with the National Museums of Scotland. First, we indicate a number of ways in which labels on museum objects ought to be tuned to take into account types of visit, the interests of visitors, and their evolving knowledge during a visit. Secondly, we sketch the general architecture of our system, and then focus on the conversational, or textual, effects which the system can create. We then briefly indicate future directions of research, before critically discussing the applicability (or otherwise) of the spatial metaphor to flexible hypertexts.

2. What would make a label intelligent?

A number of different types of labels can be isolated in the physical museum (7). A gallery or exhibition may have an overall introductory label; groups of objects may possess a sectional label, providing background rationale for their display; objects may possess individual labels or captions, as well as follow-up labels providing interpretation or anecdote; and there may also be orientation signage, directing the visitor through the gallery or towards specific targets.

These labels may be realised purely conventionally—printed and mounted on walls and in cases—or they may exploit hypermedia in the gallery, via information kiosks of various kinds. In addition, of course, but still in the gallery, the labels may be realised via the current generation of commercially available audio guides. Beyond the gallery, and particularly on the World Wide Web, hypermedia can play a more extensive role, allowing virtual galleries to reproduce some, but not all, of a physical exhibition's characteristics.

However, considering first the most traditional forms of printed labels, it is clear that their static nature severely limits their utility. First, if an object has a single individual label, it must be written for a general audience, rather than targeted at specific visitor interests. Secondly, a single label will have a fixed number of words, which must be carefully controlled to maximise the probability that the general reader will read to its end; yet different visitors have differing amounts of time available. Most significantly, however, a single individual label cannot make effective links to other labels on other objects. If labels are designed to be read in any order, they cannot presume that any other labels have been read. Thus, a label may contain redundant information which effectively conceals the 'new' information about its object. Worse, by failing to link back to previous related objects, the labels offer an effectively incoherent commentary on the sequence of objects in the gallery. Static labels—unlike human curators and guides—fail to provide a coherent narrative.

Obviously, even with static individual labels, there are solutions. First, sectional labels provide background and reduce redundancy. Unfortunately, they may not be read at all if they are too far from the objects of interest. Secondly, linearisation of a gallery forces visitors to move past labels in a specific sequence. Unfortunately, visitors thus lose freedom of choice, and there is still no guarantee that labels will not be skipped.

Let us focus on these four specific issues: visitor interest, visit type, history and repetition, and history and linking. First, it would be preferable if labels could intelligently take into account a visitor's primary interest. For example, in a gallery of 20th Century jewellery, some visitors may be more interested in styles, and their evolution, while others may be more interested in the individual designers. Thus, for the former, we may prefer a label like [1]; for the latter, a label like [2]:

[1] This jewel is a necklace and is in the Organic style. It was made in 1976. It is made from opals, diamonds and pearls.

[2] This jewel is a necklace and was made by Gerda Flockinger, who was a designer and was English. The jewel, which is in the Organic style, was made in 1976.

Secondly, the amount of attention a visitor will devote to an individual object is strongly influenced by the overall amount of time they are currently intending to spend in the museum building. In our intelligent gallery, for brief visitors, we may prefer a label like [3]; for more leisurely visitors, we may prefer a label like [4]:

[3] This jewel is a finger ring and is a remarkably fluid piece. It is rather reminiscent of molten metal and was made by Frances Beck.

[4] This jewel is a finger ring and is a remarkably fluid piece. It is rather reminiscent of molten metal and was made by Frances Beck. It is also in the Organic style. It was made in 1969. It is also made from diamonds. It is made from tourmaline and 18-carat gold. It was made in Buckingham.

Thirdly, when a human guide describes a sequence of objects, they will intelligently suppress the re-expression of information they know they have already successfully conveyed. This allows them both to make individual descriptions shorter, but also to include new information in the time or space freed up by eliminating the redundancy. The effect is most dramatically seen when a visitor returns to a previously viewed object. An intelligent labelling system might describe an object when first seen by using [5]; but on returning to the object, it would not repeat [5], but instead, it should produce a label such as [6]:

[5] This jewel is a necklace and was made by a designer called Gerda Flockinger. It consists of pairs of rectangular panels separated by groups of pearls. The panels are decorated with openwork and applied silver wire. It also includes ...

[6] This necklace was made in 1976. It is made from opals, diamonds, pearls, gold and silver metal. It was made in London. It draws on natural themes for inspiration; indeed Organic style jewels usually draw on natural themes for inspiration.

Finally, as well as taking the history of the interaction into account to eliminate redundancy, and include new information about the individual item, an intelligent labelling system ought to make comparisons back to previously seen objects, or types of object, in order to maximise the coherence of the visitor's experience (8). Suppose the visitor saw an Art Deco necklace with the label in [7], and then saw several items of Scandinavian style jewellery, before returning to the necklace. Then with a conventional label, they would see [7] again; however, [8] would be a more intelligent way of making coherent links between the objects that have been seen:

[7] This jewel is a necklace and is in the Art Deco style. It was made in 1920. It is made from moonstone and silver rock-crystal. In colour, it is coral. It differs from the previous item, in that whereas that was made by Arthur & Georgie Gaskin, this was made by H.G.Murphy. <p> It has clean lines; indeed Art Deco style jewels usually have clean lines. They usually use geometric forms.

[8] As already mentioned, this necklace is in the Art Deco style. Like most of the Scandinavian style jewels, Art Deco-style jewels usually have clean lines.

Now adaptive hypermedia—both in physical galleries, and in virtual ones—offers potential solutions to these difficulties. Such systems can, in principle, alter the labels on objects, to take into account visitor interest, visit type, visitor knowledge, and the history of the interaction. Various adaptive techniques are available. We could maintain a larger set of static labels, and exploit dynamic linking to render only subsets available at any one time, depending on user type or session history (9, 10, 6). Or we could do away with fixed labels, and instead generate them completely dynamically. It has been argued elsewhere that natural language generation (NLG) techniques from the field of artificial intelligence are particularly well-suited to the latter task (3, 4); indeed, De Bra and Calvi (11) have recently observed that techniques which avoid the need for NLG—such as their use of conditional fragments—may simply not be practical for the task of imposing a coherent order on the adaptively assembled information. Thus, we have been investigating methods for accomplishing the dynamic generation task in museums (2, 12, 13).

All the examples in this section were in fact generated on-line by the current version of our system, the Intelligent Labelling Explorer (ILEX). As should be clear, the texts which the system produces are not always optimally tailored; however, they should serve to illustrate the type of conversational behaviour we are aiming at.

3. The intelligent labelling explorer

The Intelligent Labelling Explorer system uses NLG technology to generate descriptions of objects displayed in a museum gallery. Versions of the ILEX system are available on the WWW (14); the system can therefore be used over the web, or as a standalone interactive within a museum gallery. To date, three versions have been implemented (ILEX-0, -1, and -2); these systems describe objects in the National Museums of Scotland's 20th Century Jewellery Gallery. Both field and laboratory evaluations are currently under way.

3.1 The interface

Visitors start from a visual index, composed of thumbnail images. Clicking an image causes a description of the relevant object to be generated. There is no separate introductory article, since background material is incorporated into the descriptions generated on demand (on the relative usability of indices and introductions in more conventional museum hypertexts, see Shneiderman et al. (15)). Figures 1 and 2 show the initial page, and an example output from ILEX-2.0. It will be seen that no thematic organisation is imposed on the set of images. They are intended to compete for the visitor’s attention just as physical objects in a real gallery compete on the basis of their appearance. In principle, however, it is perfectly feasible to sort them, for instance, by period, style or designer, or to embed them in a more elaborate simulation of a physical gallery layout.

— Insert Figures 1 and 2 around here —

On ILEX's internal button bar, the 'Jewels' control is provided so that although a generated description contains suggested onward links, the visitor can at any time return to the visual index and choose a new object from there. The left-arrow button takes the visitor to the previously viewed object, but generates a new description, which can take into account what has subsequently been seen—so part of the page (the image and the name of the object) is as it was before, and part of it (the textual commentary) changes; the right-arrow button takes the visitor forward again, and again generates a new context-sensitive description. The 'Help' button provides basic assistance, and the 'Exit' button allows graceful termination of the user session.

Within a generated page, one frame contains a short label and image of the object; neither of these changes between visits to that object. The other frame contains the dynamically generated description. The 'Say more' button on the page requests a new follow-up description (16), similar in effect to returning to the object from another object. Invocations of 'Say more', or multiple visits to the object, will build up a sequence of description-versions. All previously-generated descriptions remain accessible, through the use of the numerically-coded buttons to the left of the 'Say more' button.

3.2 ILEX's knowledge representation and algorithm

ILEX uses a fairly traditional architecture for generation, which separates out the stages of choosing what to say from choosing how to say it. However, we view the user as being in control of the highest-level decisions about what is to be described, and so we have designed our algorithm and knowledge representation to support a kind of co-operation between user and system.

In outline, when an object is selected by the user, the system consults its knowledge base of facts, entities and relations. It first decides on a pool of relevant facts to express. Relevance depends on a number of parameters, including a model of the assumed knowledge and interests of the user. The system then organises these facts into a coherent structure. It then converts this structure into a text. Finally it presents this text, in written or spoken form, and updates its model of the assumed knowledge of the user. Here, we focus specifically on the knowledge representation, and then on the first two stages of content selection and structuring—content selection being the stage at which most user-adaptation occurs. By comparison, (13) provides less detail on the first two stages, and more detail on the third stage, of textual realisation.

3.2.1 Knowledge representation

ILEX begins by consulting its knowledge base of facts, to choose which to present. The knowledge base has two main sources; firstly information parsed straight from the museum's own database, and secondly, information gathered during a number of interviews with the gallery's curator. Knowledge is represented in a structure called the content potential, which is basically a graph containing entity-nodes (which represent objects), fact-nodes (which represent facts about objects) and relation-nodes (which represent relations between facts).

Entity nodes represent the participants in facts, and may be of two kinds: <u>specific</u> (such as an individual jewel or person); and <u>generic</u>, representing some class of entities (such as Scottish jewellers, or Art Deco brooches). The two kinds are treated alike for purposes such as tracking which entities are under discussion, and how to refer to them in the text.

Fact nodes represent the relations between entities, in both events (such as X made Y), and states (such as X owns Y). We have assumed that all facts involve binary relations between two entities. For instance, <u>made-by(J-9999, King01)</u> represents the fact that the designer King made item J-9999. This binary assumption simplifies our architecture, but at a later stage, we may allow more complex fact-representation. Complex sentences can be formed through aggregating together these binary facts. Each binary fact has the following fields:

Pred: The name of the predicate connecting the two entities.
Arg1: The entity in the relationship which the fact is primarily about. For instance, 'J-999 was designed by Jessie King' is primarily about J-999, not about King.
Arg2: The other entity in the relationship. This is sometimes another thing (such as 'Jessie King') and sometimes a quality (such as 'blue').

Various other fields exist which detail the polarity, defeasibility, interest, importance and assimilation of the fact. Facts representing general principles or negations of general misconceptions are expressed using generic entities.

Relation nodes represent relations between facts. Relations include <u>Example</u>, <u>Concession</u>, <u>Amplification</u>, <u>Similarity</u>, <u>Contrast</u>, <u>'In that'</u>, <u>'In other words'</u>, and

Specification. Each relation has a nucleus and satellite (as in Rhetorical Structure Theory (17)) as well as a set of precondition facts, which must be assimilated before the relation can be understood. There are no relations between relation-nodes in the content potential at present; relation-nodes only link fact-nodes. Figure 3 shows a small subgraph of the content-potential, showing two Concession relations between facts.

— *Insert Figure 3 around here* —

3.2.2 Content selection

In the content-selection process, this graph of content potential is traversed, beginning from the entity-node corresponding to the object to be described, and a set of fact-nodes to be described is thus collected. The decision about which fact-nodes are selected is determined by three factors. Firstly, facts are weighted according to the chain of relations back to the initial entity-node of the current page (18). This is a way of preventing lengthy digressions from the supposed topic of the text. Secondly, each fact is associated with numbers (between 0 and 1) which represent the opportunity 'value' of the fact. These opportunity values reflect both long-term and session-based user models. The key values are of two kinds:

Interest the estimated value of the fact to the user; for instance, being made of plastic or paper are more interesting (to the user), because they are unusual in jewellery. Canned anecdotes about a piece of jewellery may also have high interest values. Clearly, different types of users will find certain categories of information more interesting than others. Thus, the interest values of classes of facts are derived from our simple long-term user models, which we describe shortly.

Importance the value of the fact as regards the system's educational agenda; for instance, the museum considers it important to educate on stylistic development, so facts about styles are rated highly. Importance values of classes of facts are independent from user models or session histories.

Interest values thus reflect an aspect of long-term user modelling. By contrast, importance values are more like ‘system models’; they allow the orderly pursuit of an educational agenda, achieving similar goals to those De Bra and Calvi (11) accomplish via their rules for modifying the presentation of content fragments. Notice that neither interest not importance changes during a user session: they are not dynamically adapted. However, these opportunity values are moderated by two further fact annotations, both of which are involved in dynamic, session-level user modelling:

Assimilation the degree to which the fact is assumed known to the user, either from general knowledge, or through prior mentions in the web interaction (these values change dynamically). Facts also have an assimilation rate, depending on their assumed degree of difficulty.

Assimilation-rate the amount by which the assimilation of a fact changes each time it is presented. It is normally set to 1, but for some facts (in particular generalisations) it is less than 1. A fact's assimilation rate does not itself change dynamically.

The first three values—interest, importance and (1 – assimilation)—are multiplied together to calculate the local score of each fact. The overall opportunity value of a fact is the product of its local score, the overall opportunity value of the parent (the node through which it was reached), and a weight for the relation between them. It is the overall opportunity values that are used to select which items of content will be included in the generated description. When all the stages of selection, structuring, realisation and presentation have been achieved, the assimilation values for expressed facts are updated, in accordance with their assimilation-rates.

So, if we distinguish long-term user modelling from short-term, session-based user modelling, we can see that ILEX2.0 is simplistic regarding the former, but quite sophisticated regarding the latter. On the long-term side, ILEX2.0 in principle allows for three main types of user:

Expert: someone with general expertise about 20th century jewellery and 20th century styles.

Child: a school-child of age 12 or so.

General: a literate, non-specialist adult.

While eventually we would like the language produced for different user types to be different, at present the only differences are to do with how facts of different kinds are annotated—and hence, which will be selected. For instance, for user type expert, generalisations about jewel styles are assumed to be assimilated at the outset, while for user type child, they are assumed to be unassimilated, but of only moderate interest. In each case, these annotations have the effect of dispreferring such generalisations. For user type general, these generalisations are assumed to be both unassimilated and of high interest, which makes them very good candidates for selection. Notice that in each case, interest values are fixed, but assimilation will be dynamically updated, depending on the objects the user chooses to view, thanks to the differing facts the system selects for presentation. However, it is important to note that in the publicly accessible implementation of ILEX2.0, only the general user model is available, and the user cannot alter this.

On the short-term side, we can consider ILEX2.0 to be modelling the individual user via the changes it makes to the assimilation values of facts which happen to be presented. Which facts have their values updated will depend on the user's precise trajectory through the virtual gallery. If we consider the session-history to help define fine-grained user models, we end up with vast numbers of distinguishable user models. Consider ILEX2.0 with 40 objects loaded; suppose a user looks at just 10 of these. There are around 10^{12} different sequences in which this might be done, and differing sequences can result in different assimilation values. It is true that many sequences will have equivalent effects on the assimilation values stored in the content potential. Furthermore, as with De Bra and Calvi's AHA! (11), the interdependence between facts will help to ensure that the actual number of distinguishable session-level user models is much smaller than this. However, it will still be very large by most standards.

The content selection process is almost finished when it has found a sufficient number of facts, ranked by opportunity value. However, if it cannot find enough, it enters a ‘barrel-scraping’ phase, to locate new content. It is at this point that the algorithm finds negative facts, and locates other entities, to which comparisons can be made. Negative facts, expressed for instance, as <u>This item is not made of enamel,</u> are not stored separately, but are generated from existing positive polarity facts only in contexts where the reader has grounds for believing the proposition which is negated. The choice of appropriate comparisons is determined via a measure of the similarity between pairs of fact-nodes. The metric used for similarity is that developed by Milosavljevic (19); it is outlined only briefly here. Facts of different types require different notions of similarity to be invoked. As well as the annotations for interest, importance and so on, there are two further annotations on each fact node: scale, and variation. The former represents the type of scale that the attribute is associated with. There are four types of scale:

Nominal (such as place or material) is not ordered.
Interval (such as date) is ordered, but there is no notion of ‘the lowest possible value’.

Ratio (including dimensions) is ordered, and there is some notion of 'the lowest possible value'.

Ordinal (such as the case in which a jewel is displayed) is ordered, but just by convention, rather than by anything intrinsic to the different values. (In this case, a third annotation order is given, specifying what the ordering is.)

Variation, meanwhile, is captured by a number representing how much variation there is between items with this attribute; the larger the number, the more variation. For some scales, the variation can be given in advance, while for others it must be calculated by looking at the values for the collection of objects that the system knows about. In ILEX2.0, this calculation is carried out off-line, when the content potential is built. The similarities between the facts associated with candidate objects are determined by checking the values they take on the various scales, and the amount of variation possible on those scales. The most similar object will be found, together with the facts that make it similar to, and different from, the current object.

It is worth noting that a pair of global variables controls the system's overall tendency to choose comparisons as part of the content to be realised. The first determines the number of barrel-scrapes which are automatically made—that is, irrespective of whether the system has run out of facts to express. If the value of this variable is set high, comparisons (and also negations) will be sought even before the content-selection routine has run out of things to say. If it is set to 0, these extra facts will only be sought if there is nothing left to say. The second variable helps choose which method of barrel scraping—comparison or negation— is preferred, and to what extent.

3.2.3 Content Structuring

When all the facts to be expressed have been selected, a suitable discourse structure for expressing them is determined. The model of discourse structure used in ILEX is two-level. We assume that (descriptive) text is organised at a high level into collections of facts with the same Arg1—facts about the same entity—called entity-chains. The initial entity-chain is a collection of facts about the object to be described; these facts can mention other entities in Arg2 position, such as the designer or style of a jewel, which can serve to introduce subsequent entity-chains about these entities. The first phase of content structuring is to arrange the selected facts into a legitimate structure of entity-chains.

Within an entity chain, facts can be organised into rhetorical structure (RS) trees, using a number of coherence relations from the content potential, such as Example, Concession, Specification and so on. RS trees have a hierarchical structure; one fact (the top nucleus) dominates the others. In the second phase of content structuring, we find all possible RS trees that can be created from the selected facts, and then choose the best of these that can be legally added to one of the entity-chains. This process is then repeated until no more RS trees can be added.

3.2.4 Sentence Realisation

The third phase of generation is to decide on the surface form of the linguistic expressions which realise the facts to be generated. This process is driven by several different and interacting modules. For a clause expressing a fact, the decisions about the tense (for instance is versus was), mood (for instance can versus must), and surface polarity (for instance, worked versus did not work) are taken by the fact expression module. The RS tree realisation module determines how to express the relations between facts in a RS tree using sentence and clause conjunctions (for instance but or rather). The aggregation module determines when a group of facts can be combined into a single sentence, by co-ordination, by modification (for instance, British modifies jeweller in British jeweller) or by the inclusion of relative clauses (introduced, for instance, by who or which). The decision about how to refer to entities in noun phrases is determined by the NP planning module, which chooses between full descriptions (for instance, the Art Deco necklace), reduced descriptions (for instance, the necklace), and various types of pronominal expression (for instance, this piece or it).

It should be noted that ILEX2.0 intermixes deep and shallow representations in sentence realisation; that is, it produces both fully machine-generated sentences, and ones which include 'canned', human-written text, or phrasal templates. It therefore resembles in certain respects Geldof's navigation assistant (21), which makes more extensive use of templates. While templates and fixed content structures are well-suited to the production of Geldof's 'meta-texts', ILEX2.0 requires finer control of the form of linguistic expressions. Thus, when canning is used, it can in fact occur at different levels: at any level of syntactic constituency, or at any text node in discourse structure. To indicate roughly how much of the surface text is canned, and how much is generated word-by-word,the following examples use brackets to indicate pre-compiled phrasal units larger than single words:

[9] This jewel is also a necklace. It is in the machine-age style, and was made in 1930. It is made from chrome. It has [no fear of pattern], [in that] it incorporates [patterns with a repetitive element]. <p> Machine-age style jewels usually have [regularly repeated forms]. [To take an example]: this jewel is [made up of a pattern of interlocking rods]...

[10] This jewel resembles the waist-buckle, [in that] like most machine-age jewels it has [regularly repeated forms]. However, it differs from the waist-buckle, [in that] it is made from chrome, whereas the waist-buckle was made from enamel.

This engineering compromise improves the quality of the text somewhat, while allowing us to avoid coding every surface predicate or connective at a deep level; nonetheless, the examples here should make it obvious that, like most NLG systems, ILEX’s grasp of English falls in various ways rather short of that of a native speaker.

3.2.5 Text Presentation

Finally, the generated text is delivered to the user. Currently, texts are presented in written form. However, we are developing the option to present text as speech, via a speech synthesiser which uses the discourse structure of the text to generate appropriate intonation and emphasis. For more on this topic, see (20).

4. What difference does intelligent labelling make?

The structures and processes used by ILEX2.0 have specific conversational or textual effects. These can be seen both in the texts discussed in section 2, and by examining the difference between a dynamic version of the system, and a more conventional 'static' version.

4.1 Four aspects of intelligent labelling

Consider again the examples we used to introduce the four specific issues, of visitor interest, visit type, history and repetition, and history and linking. The examples are repeated here for convenience.

[1] This jewel is a necklace and is in the Organic style. It was made in 1976. It is made from opals, diamonds and pearls.

[2] This jewel is a necklace and was made by Gerda Flockinger, who was a designer and was English. The jewel, which is in the Organic style, was made in 1976.

The difference between [1] and [2] results from changing the user-interest values associated with facts about designers and facts about styles. In [1], style facts had value 0.9, and designer facts value 0.2; in [2], these values are swapped. This reflects a method of tailoring for visitors which, as we mentioned above, is not accessible in ILEX2.0, but which will be added in a later version.

[3] This jewel is a finger ring and is a remarkably fluid piece. It is rather reminiscent of molten metal and was made by Frances Beck.

[4] This jewel is a finger ring and is a remarkably fluid piece. It is rather reminiscent of molten metal and was made by Frances Beck. It is also in the Organic style. It was made in 1969. It is also made from diamonds. It is made from tourmaline and 18-carat gold. It was made in Buckingham.

The difference between [3] and [4] depends simply on a global parameter regarding the number of facts to be expressed. Again, this is not currently under user control, but this facility is simple to add.

[5] This jewel is a necklace and was made by a designer called Gerda Flockinger. It consists of pairs of rectangular panels separated by groups of pearls. The panels are decorated with openwork and applied silver wire. It also includes ...

[6] This necklace was made in 1976. It is made from opals, diamonds, pearls, gold and silver metal. It was made in London. It draws on natural themes for inspiration; indeed Organic style jewels usually draw on natural themes for inspiration.

The difference between [5] and [6] arises because facts which have been expressed gain assimilation value, and as a result are less likely to be re-expressed. They may well appear again, because they allow another unexpressed fact to be linked into the discourse, or because the system does not consider them fully assimilated; thus repetition remains possible, whenever it serves to improve coherence, or promote pedagogical goals.

[7] This jewel is a necklace and is in the Art Deco style. It was made in 1920. It is made from moonstone and silver rock-crystal. In colour, it is coral. It differs from the previous item, in that whereas that was made by Arthur & Georgie Gaskin, this was made by H.G.Murphy. <p> It has clean lines; indeed Art Deco style jewels usually have clean lines. They usually use geometric forms.

[8] As already mentioned, this necklace is in the Art Deco style. Like most of the Scandinavian style jewels, Art Deco-style jewels usually have clean lines.

The difference between [7] and [8] arises firstly because repetition is suppressed, as in [6], and secondly, because the system dynamically computes which comparisons can be made to specific or generic entities mentioned in the discourse so far.

4.2 Dynamic vs static hypertext: an example

For a slightly more extended comparison, Figure 4 contains texts [11] and [12]. The former is generated by ILEX2.0 in dynamic mode, the latter in static mode. The standard behaviour of the system, as described above, is dynamic. However, by re-setting assimilation values to their initial values after each description is generated, we can simulate a more standard hypertext system, in which each object label is essentially generated in isolation from any other object label, and discourse history is therefore not taken into account.

— Insert Figure 4 around here —

There are several notable differences between the dynamic [11] and the static [12]. First, dynamic generation allows the system to respond to user's move (a) by making a relatively short description containing a comparison with a previous item (a silver bracelet). Secondly, although the responses to user's move (b) are similar, the dynamic system registers the fact that it has been asked once more to describe a necklace (<u>This jewel is also a necklace</u>), and this small change has ramifications for the sentence structure for the rest of the paragraph in which it appears. Most obviously, however, the system's response to user move (c) in the dynamic system is much richer than in the static system. Comparisons are made back to two previously-seen items (<u>the waist-buckle</u>, and <u>the previous item</u>), and between styles (<u>machine-age</u> versus <u>organic</u>).

5. Related and future work

We noted above that Geldof's system (21) exploits template-based NLG techniques to generate dynamically its navigation-node meta-texts. While in some ways less flexible than the system we have described here, her system includes automatic adaptation to user interests on the basis of navigation behaviour, which ILEX2.0 does not. In fact the most closely related system of which we are aware is Milosavljevic's PEBA-II system (19, 22), for generating encyclopaedic descriptions of, and comparisons between, animals. Already, our approach to comparisons is based on development carried out under the PEBA-II project, and we foresee possibilities for further fruitful co-operation.

In the meantime, the Intelligent Labelling Explorer system will have some further features added to it in due course. First, to improve adaptability, user control over interest-values will be added, so as to allow adaptation of the simple long-term user-type models; similarly, user control over the number of facts to be expressed will allow simple visit-type control. As we have seen, ILEX2.0 already possesses adaptivity to changing (assumed) user knowledge. However, in contrast to Nill et al. (10) and Geldof (21), automatic adaptivity to model specific user interests, on the basis of observed behaviour, is not envisaged. Further work is planned on making the system support more varied user input, so that more interesting dialogue actions can be accomplished than at present. More immediately, concept-to-speech output is being developed, under the auspices of the SOLE project (20), and in another project, this speech output is being channelled via a wearable computer device (23, 24).

However, our current goals are to complete the analysis of the laboratory user trials of the system, and to carry out field testing in the National Museums of Scotland. The laboratory trials are comparative in design, contrasting user behaviour, subjective evaluation, and learning outcome, on the dynamic and static versions of the system, as illustrated above. In the meantime, however, our experience with the system leads to some general questions about the relation between conversational systems like ILEX, and PEBA-II (19), and the navigational paradigm underlying most hypertext systems.

6. Discussion: conversation versus navigation

In previous work, it has been asserted, more or less without supporting argument, that dynamic hypertexts are conversational (4). The foregoing account indicates in a general way how this claim can be substantiated by systems like ILEX2.0 and PEBA-II. Arguably, of course, other adaptive hypermedia systems are conversational too: the only thing that distinguishes dynamic systems might be the level of granularity at which content can be adapted. Be that as it may, we can go on to argue that a conversational approach is not only feasible for dynamic hypertext, but is actually preferable to the conventional spatial metaphor.

First, we have taken dynamic hypertext to be that variety of flexible hypertext which relies upon techniques from NLG to construct textual content on demand, at viewing time. In fact, as yet, there is relatively little evidence as to whether flexible hypertext systems—of whatever stripe—are actually acceptable to their projected users, and in particular, whether they are actually more acceptable than conventional, rigid hypertext. This is a genuine concern. For instance, Carter's work on automatic text-to-hypertext translation supplies reasons for doubt (25). He combines information retrieval and natural language techniques, and finds that his context-sensitive hypertext are not only unacceptable to users, but also give poor results on objective measures, such as hyperpath recall. Obviously, we expect the outcomes of our usability trials to address this concern.

Now, a possible reason for usability difficulties—if and when these are found—may lie in the fact that flexible hypertexts pose fundamental problems for the spatial metaphor underlying classical hypertext. According to the spatial metaphor, we visit nodes, navigate around, explore, get lost in hyperspace, and so on. These ideas fit quite well the notion that a document is a fixed artefact, whose parts (in this case, hypertext nodes) look the same from whatever direction they are approached. Indeed, part of the interest in hypertexts comes precisely from the fact that a given part can have many paths leading from and to it. Given all this, it is appropriate to require that the appearance of any given part of the document not vary from time to time, except perhaps through some wear-and-tear mechanism which indicates how popular the place is—as in visit-counters on Web pages.

But what happens in dynamic hypertexts? Consider the famous analogy in quantum mechanics: Schrödinger's cat is neither dead nor alive, so long as it remains unobserved. But as soon as it is observed, the very act of observation causes it to fall into one state or the other (26). Dynamic hypertext can be taken to offer a hypertextual parallel: a link's destination need have no determinate informational contents, until it is visited. As soon as it is visited, the very act of visitation causes Schrödinger's node to contain a determinate content. So flexible hypertext is to rigid hypertext as the quantum world is to the macroscopic world: it is 'non-classical'.

Now, many people have real problems understanding Schrödinger's cat, and a quantum world in which objects change their properties, depending on where we've been before we look at them—in which objects that used to be visible from a certain place may become invisible when we return. It should therefore be no real surprise that people have trouble with non-classical Schrödinger's nodes, too. Usability difficulties may thus stem from the fact that dynamic hypertexts are simply not classical, fixed artefacts, and thus do not fulfil the expectations engendered by the spatial, object-based metaphor.

There are thus at least two responses to this problem. First, we could design our dynamic systems so that they ensure that their users are fully aware that they are navigating in a bizarre quantum hyperspace. Secondly, we could design them so that their users do not believe that they are navigating in space at all, but instead believe that they are participating in a conversation evolving through time. On the latter view, Landow's rhetoric of departure and arrival (27) is replaced by pure rhetoric: there need be no spatial travel at all. As Milosavljevic and Oberlander say, 'for us, navigation really is conversation' (8). By replacing space with time, and establishing a vivid conversational metaphor, we would ensure that people carry neither false expectations about classical spatial navigation, nor half-understood ideas about a non-classical quantum space.

The ‘navigation is conversation’ view has an interesting relationship to Dillon and Vaughan’s recent discussion of studies on navigation in traditional, rigid hypertexts (28). They conclude that ‘the physical navigation framework fails to tackle the issue of semantic space’, and that it is not sensible to separate out the navigation process from the process of comprehending the content that the reader encounters on their journey. As an alternative to space, they propose ‘shape’, which they acknowledge to be ‘a somewhat elusive concept’, but which can be thought of as a property of information conveyed by both the physical form of a digital document, and its information content. We obviously agree with the need to consider the hypertext as a kind of discourse, but we must acknowledge that this focus should not lead us wholly to neglect the effects of the layout and physical appearance of the text and the interface in which it appears. However, the latter point does not dictate a return to navigation. Rather, it suggests that evaluation of hypertexts must consider the effects on discourse comprehension of both the content and the form of the documents being used.

Perhaps general recognition of the discursive nature of hypertext has been hindered by the fact that rigid hypertexts offer deviant, amnesic conversations, in which the system acts as if it had no memory of the previous discourse. As we have suggested, with only these as models, the spatial model appears relatively appropriate. Either way, both of the metaphorical responses to non-classical, flexible hypertexts—quantum space and conversational time—should be investigated further via engineering and empirical evaluation, to determine whether they help avoid any usability problems associated with applying the standard spatial metaphor to dynamic hypertexts. This paper has discussed a particular project aimed at making the conversational metaphor vivid, via the notion of intelligent labelling. Our intention has been to try to make the conversation coherent, and thereby to contribute to achieving the goals of museum curators and educationalists.

Acknowledgments

The first author is an Engineering and Physical Sciences Research Council (EPSRC) Advanced Fellow, and this work was carried out thanks to the EPSRC, through grant GR/K53321. Further development is supported by the European Commission, through Esprit Long Term Research grant PL25574. The support of the Economic and Social Research Council for HCRC is also gratefully acknowledged. Our thanks to our collaborators at the National Museums of Scotland, to Maria Milosavljevic and Peter Brusilovsky, to our anonymous referees, and to audiences who have responded to presentations of material reported here. Special thanks to the audience at the international workshop on flexible hypertext held in Southampton in April 1997, where some of the ideas described in section 6 were first presented.

References

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

(2) KNOTT, A., MELLISH, C., OBERLANDER, J. and O'DONNELL, M. Sources of flexibility in dynamic hypertext generation. In Proceedings of the 8th International Workshop on Natural Language Generation, pp. 151–160. Herstmonceux, Sussex, UK, 1996.

(3) LEVINE, J., CAWSEY, A., MELLISH, C., POYNTER, L., REITER, E., TYSON, P. and WALKER, J. IDAS: Combining hypertext and natural language generation. In Proceedings of the Third European Workshop on Natural Language Generation, pp. 55–62. Innsbruck, Austria, 1991.

(4) DALE, R., OBERLANDER, J., MILOSAVLJEVIC, M. and KNOTT, A. Integrating natural language generation and hypertext to produce dynamic documents. Interacting with Computers, **11**(2), 1998, 109–135.

(5) HALASZ, F. G. Reflections on NoteCards: Seven issues for the next generation of hypermedia systems. Communications of the ACM, **31**, 1988, 836–852.

(6) BRUSILOVSKY, P. Methods and techniques of adaptive hypermedia. User Modeling and User Adapted Interaction, **6**(2–3), 1996, 87–129.

(7) SERRELL, S. Exhibit Labels. AltaMira Press, 1996.

(8) MILOSAVLJEVIC, M. and OBERLANDER, J. Dynamic electronic catalogues: helping users to help themselves. In Proceedings of Hypertext 98, pp. 123–131. June 1998.

(9) STOTTS, D. and FURUTA, R. Petri–net based hypertext: document structure with browsing semantics. ACM Transactions on Information Systems, **7**, 1989, 3–29.

(10) NILL, A et al.. This Volume.

(11) DE BRA, P. and CALVI, L. This Volume.

(12) OBERLANDER, J., MELLISH, C., O'DONNELL, M. and KNOTT, A. Exploring a gallery with intelligent labels. In Proceedings of the Fourth International Conference on Hypermedia and Interactivity in Museums, pp153–161. Paris, September, 1997.

(13) HITZEMAN, J., MELLISH, C. and OBERLANDER, J. Dynamic generation of museum web pages: The intelligent labelling explorer. Archives and Museum Informatics, **11**, 1997, 105–112.

(14) Intelligent Labelling Explorer: (a) current version of the system at: http://cirrus.dai.ed.ac.uk:8000 (b) ILEX2.0 also available at: http://www.cstr.ed.ac.uk/cgi-bin/ilex.cgi

(15) SHNEIDERMAN, B., BRETHAUER, D., PLAISANT, C. and POTTER, R. The Hyperties electronic encyclopedia: An evaluation based on three museum installations. Journal of the American Society for Information Science, **40**, 1989, 172–182.

(16) MOORE, J. D. and SWARTOUT, W. R. Pointing: a way toward explanation dialogue. In Proceedings of AAAI90, 1990.

(17) MANN, W. C. and THOMPSON, S. A. Rhetorical structure theory: A theory of text organization. Text, **8**(3), 1988, 243–281.

(18) O'DONNELL, M. Variable-length on-line document generation. In Proceedings of the 6th European Workshop on Natural Language Generation Gerhard-Mercator University, Duisburg, Germany, 1997.

(19) MILOSAVLJEVIC, M. Maximising the Coherence of Descriptions via Comparison. PhD Thesis, Macquarie University. In press.

(20) Speech Output Labelling Explorer: http://www.cstr.ed.ac.uk/projects/sole.html

(21) GELDOF, S. This Volume.

(22) MILOSAVLJEVIC, M., TULLOCH, A. and DALE, R. Text generation in a dynamic hypertext environment. In Proceedings of the Nineteenth Australasian Computer Science Conference, pp. 417–426. Melbourne, Australia, 1996.

(23) Hyper-Interaction within Physical Space: http://www.ing.unisi.it/lab_tel/hips/Newhips.htm

(24) NOT, E., PETRELLI, D., STOCK, O., STRAPPARAVA, C., and ZANCANARO, M. Person-oriented guided visits in a physical museum. In Proceedings of the Fourth International Conference on Hypermedia and Interactivity in Museums, pp162–172. Paris, September, 1997.

(25) CARTER, E. Quantitative Analysis of Hypertext Generation and Organisation Techniques. PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.

(26) GRIBBIN, J. In Search of Schrödinger's Cat. London: Wild Wood House, 1984. (2nd edition, 1991, London: Black Swan.)

(27) LANDOW, G. The rhetoric of hypermedia: Some rules for authors. Journal of Computing in Higher Education, **1**, 1989, 39–64. Reprinted in P. Delany and G. Landow (eds.) Hypermedia and Literary Studies, pp. 81–104. Cambridge, MA: MIT Press, 1991.

(28) DILLON, A. and VAUGHAN, M. "It's the journey and the destination": Shape and the emergent property of genre in evaluating digital documents. New Review of Multimedia and Hypermedia, **3**, 1997, 91-106.

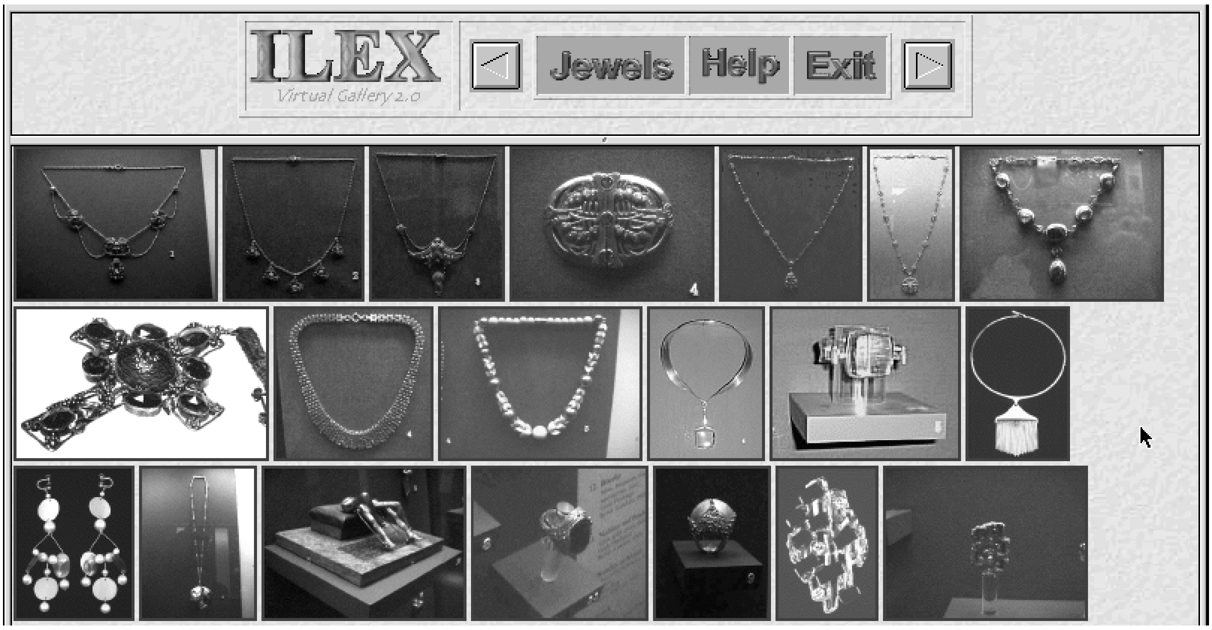

Figure 1: the front page of ILEX2.0

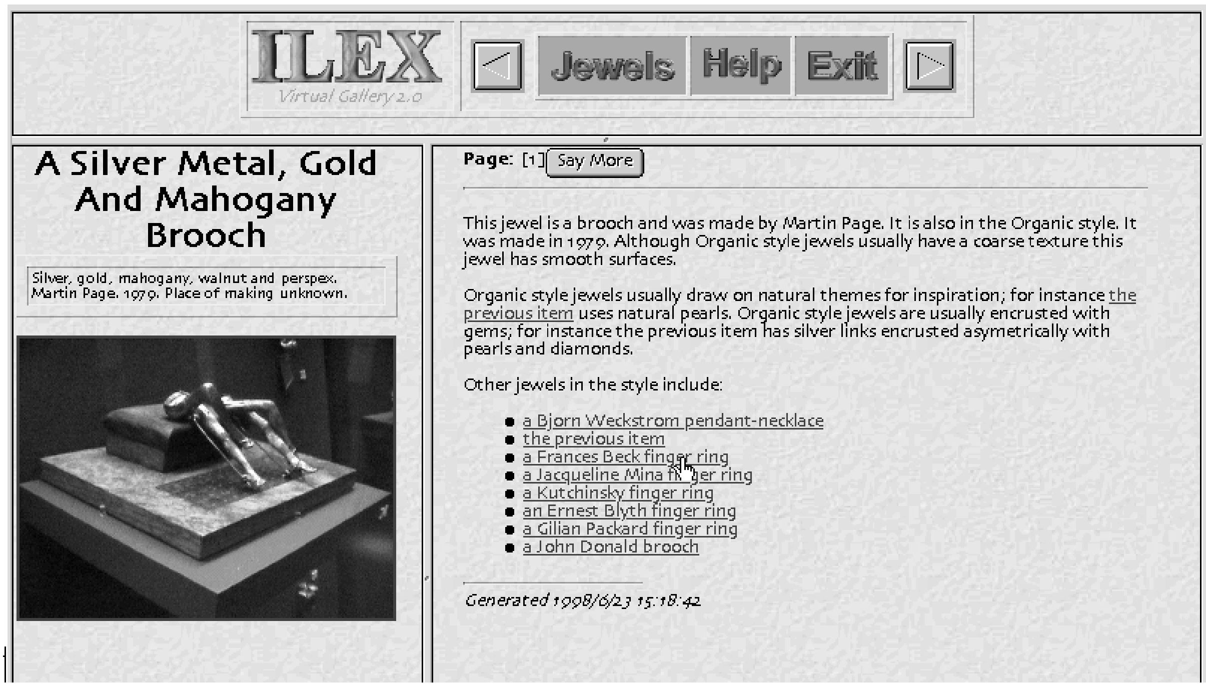

Figure 2: a page generated on demand by ILEX2.0

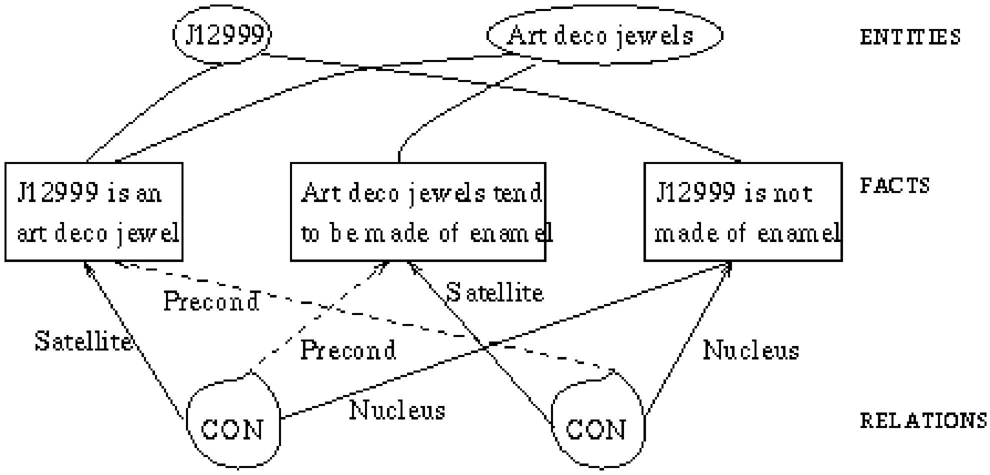

Figure 3: a fragment of ILEX2.0's content potential, showing concession relations between facts, and the entities which they involve. Notice that negative facts, like J1299 is not made of enamel, are not stored separately, but are generated from existing positive polarity facts only in contexts where the reader has grounds for believing the proposition which is negated, as here.

[11] <a: in move 8 of interaction, user re-selects Bülow-Hübe necklace>

This necklace was made for Georg Jensen. It resembles the bracelet, in that like the bracelet, it is made from silver metal.

<b: user selects machine-age necklace>

This jewel is also a necklace. It is in the machine-age style and was made in 1930. It is made from chrome. It has no fear of pattern, in that it incorporates patterns with a repetitive element.
Machine-age style jewels usually have regularly repeated forms. To take an example: this jewel is made up of a pattern of interlocking rods; indeed machine-age style jewels usually incorporate patterns with a repetitive element (for instance this jewel has regularly repeated forms).
Other jewels in the style include: ...

<c: user selects follow-up>

This necklace resembles the waist-buckle, in that like most machine-age style jewels, it has regularly repeated forms. However it differs from the waist-buckle, in that it is made from chrome, whereas the waist-buckle was made from enamel.
As already mentioned, this necklace is in the machine-age style. Machine-age style jewels differ from Organic style jewels. Organic style jewels usually draw on natural themes for inspiration; for instance the previous item looks crystalline (specifically it looks rather like a section of crystal).
Other jewels in the machine-age style include: ...

[12] <a: in move 8 of interaction, user re-selects Bülow-Hübe necklace>

This jewel is a necklace and was made by a designer called Torun Bülow-Hübe. It is in the Scandinavian style and was made in 1960. It is made from rock-crystal and silver metal. It was made for a client called Georg Jensen. It has clean lines (indeed Scandinavian style jewels usually have clean lines).
Other jewels designed by Bülow-Hübe include: ...

<b: user selects machine-age necklace>

This jewel is a necklace and is in the machine-age style. It was made in 1930. It has no fear of pattern, in that it incorporates patterns with a repetitive element.

Machine-age style jewels usually have regularly repeated forms. To take an example: this jewel is made up of a pattern of interlocking rods; indeed machine-age style jewels usually incorporate patterns with a repetitive element (for instance this jewel has regularly repeated forms).
Other jewels in the style include:

<c: user selects follow-up>

This necklace is made from chrome and draws for inspiration on machines and their components.

Figure 4: Comparison of [11] dynamic and [12] static conversations with ILEX.

[Note to Editors: ideally, [11] and [12] should be displayed side-by-side]