Dynamic versus static hypermedia in museum education: an evaluation of ILEX, the intelligent labelling explorer

Richard Cox

School of Cognitive & Computing Sciences
University of Sussex
Falmer, BN1 9QH, UK
richc@cogs.susx.ac.uk

Mick O'Donnell

Department of Artificial Intelligence
University of Edinburgh
80 South Bridge, Edinburgh, EH1 1HN, UK
micko@dai.ed.ac.uk

Jon Oberlander

Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place, Edinburgh, EH8 9LW, UK
jon@cogsci.ed.ac.uk

Abstract

This paper describes an evaluation of an intelligent labelling explorer (ILEX), a system that dynamically generates text labels for exhibits in a museum jewellery gallery. In the evaluation, learning outcomes in subjects who used the dynamic ILEX system were compared to those of subjects who used a static-hypermedia version of the system (i.e. a system more typical of current hypermedia systems). The aim was to attempt to isolate learning effects specifically due to dynamic hypertext generation. Several types of data were collected - user-system interaction logs of navigational and browser use, recordings of the type of information to which each subject was exposed, post-session tests of factual recall and a special 'curator' task in which the subjects were required to classify novel jewellery items. Results showed that performance measures (post-session tests) did not differ between subjects in the two conditions. However, the interaction-log data revealed that the two groups differed in terms of their navigational behaviour and in the type and amount of information to which they were exposed. These results are discussed in terms of the 'learning-performance' distinction often drawn in psychological accounts of learning. The paper concludes with an outline of further planned work.

1 Introduction

The term hypermedia refers to hypertext systems that include graphics, diagrams, photographs, movies, animations, etc. They are systems that allow non-linear access to multimedia resources.

The intelligent labelling explorer (ILEX) system produces descriptions of objects encountered during a guided tour of a museum gallery. ILEX seeks to automatically generate labels for items in an electronic catalogue (or museum gallery) in such a way as to reflect the interest of the user and also opportunistically to further certain educational (or other) aims. The ILEX domain is that of a 20th Century Jewellery Exhibit in the Royal Museum of Scotland.

At top level, the virtual gallery consists of a page ('virtual glass case') of 30 thumbnail images of jewels from the National Museum of Scotland collection. The jewels are quite varied - the exhibit features works by many designers (e.g. Gerda Flockinger, Jessie M. King) , in around 6 styles (art deco, arts & crafts, ...) from several periods. The exhibits are made from a wide variety of materials — from 'ephemeral' items in plastic such as a 'Beatles brooch' to items made in precious metal and precious stones. The user explores via a Web browser. The user begins by selecting an item from the jewellery 'case' - a larger image of the item is then presented and some explanatory text is generated by the system. An example of a generated label can be seen in Figure 1.

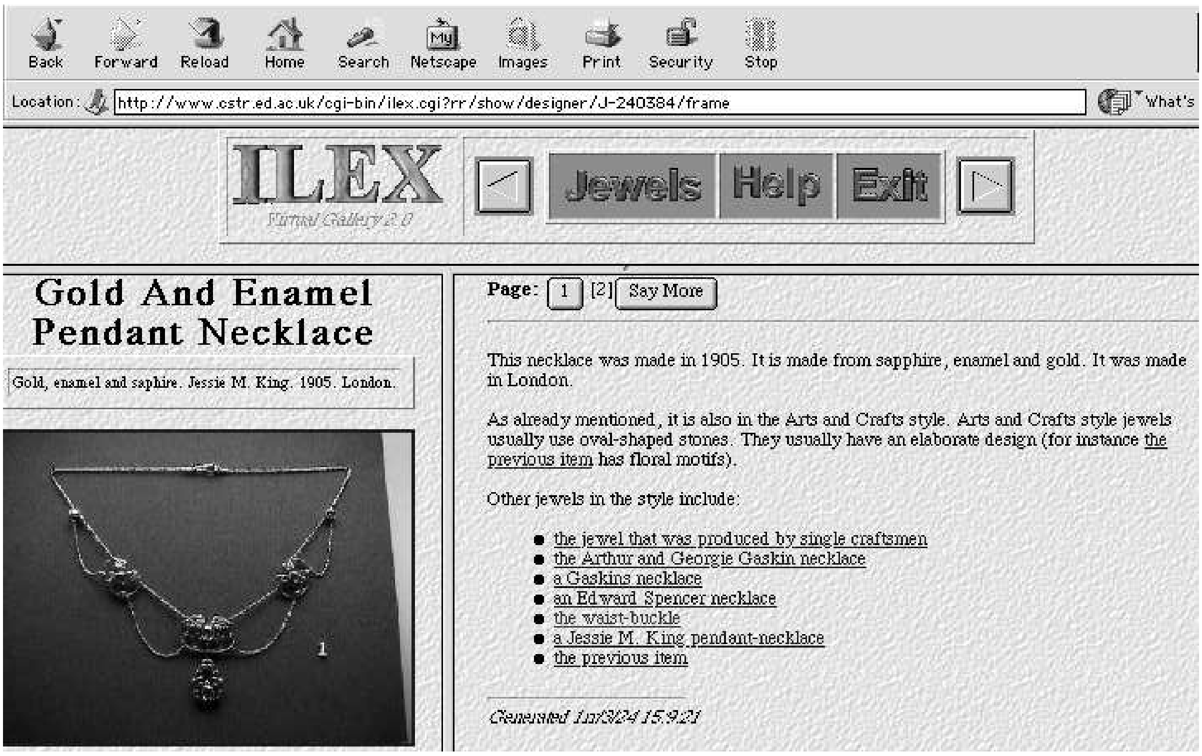

Figure 1: An example of a dynamically labelled exhibit. Note the referring expressions ('As already mentioned...') and user-modelling based on previous items seen ('for instance the previous item has floral motifs').

The pages produced by the system differ from conventional hypermedia pages in that they are generated dynamically—in other words, they are tailored to a particular user in a particular communicative situation. This flexibility has a number of advantages. For one thing, the discourse history of the user can be taken into account — the objects which the visitor has already seen — so that information the visitor has already assimilated can be taken into account in the current description. For instance, the description of the object currently being viewed can make use of comparisons and contrasts to previously-viewed objects, while omitting any background information that the visitor has already been told [6, 7, 8].

Dynamic hypertext makes it possible for the generation system to pursue its own agenda of educational and communicative goals, while allowing the user the freedom to browse the collection of objects in any order, as in a normal hypermedia system. The aim is to reproduce the kind of descriptions that a real curator might give, were the visitor to have one at their elbow.

Opportunistic text tailoring is achieved in ILEX via the use of referring expressions, comparison expressions, nominal anaphora and approaches derived from rhetorical structure theory [4, 6, 7][1]).

The aim of the evaluation was to attempt to assess the effect of intelligent label generation upon several types of learning outcome. Dynamic and static versions of the intelligent labelling explorer (ILEX) system were compared. The goal was to attempt to 'pin down' or isolate to some degree the specific effects of text which is tailored to the user and which takes into account his/her browse history.

Unlike typical hypermedia evaluation studies, the aim was not to compare hypermedia with traditional media, or to investigate aspects of hypermedia such as configurations of page links, but, rather, to compare two versions of hypermedia - a traditionally configured version (static pages, no user modelling) with the intelligent system (dynamically generated text containing referring expressions and comparisons based on a user-model).

[1] These papers and others are available from the project website http://cirrus.dai.ed.ac.uk:8000/ilex/

ILEX provides succinct and coherent information to the learner by relating information about a currently viewed object to previously viewed objects and thus, to some degree, organising, structuring and contextualising the material in a semantically coherent way. Thus a prediction was that learning outcomes in terms of factual recall would be greater from the dynamic system than from the static system. The ability of dyanamic-ILEX to draw out comparisons and make generalisations was also predicted to produce better outcomes on subjects' learning to classify novel, unseen, artifacts.

On the other hand, there is some evidence that unpredictable, varying, dynamic hypermedia do not always facilitate performance [5]. Hence, there was an alternative hypothesis that the dynamic system would produce poorer learning outcomes than the static system due to the variability and less-standardised and less-predictably-formatted nature of its output.

The aim, therefore, was to test these competing hypotheses by comparing learning outcomes from the two versions of ILEX.

2 The systems - two versions of ILEX

Two versions of the system were developed - 'dynamic ILEX' (the ILEX project system) and a comparison 'static ILEX' (described below). The two systems differed in terms of the format of the descriptions that they generated. Also, the response time of dynamic ILEX was somewhat slower than that of static ILEX, due to the computational overhead of dynamic label generation.

2.1 Dynamic ILEX

This system has been described in the introduction and more information can be obtained from the project Web site (see footnote 1).

2.2 Static ILEX

The static version of the ILEX system was prepared by generating all the available pages of each jewellery object (usually 4 or 5 per object) from dynamic-ILEX. The discourse history was maintained over the pages of an object, so that each sucessive page contained new information, and may have referred to entities introduced on prior pages. However, discourse history was erased before starting on the next jewel, so that the pages of the new object may repeat what was said for a previous object, and indefinite reference was used for first mentions of some objects. The generated system was 'frozen' and became the static ILEX system used in the study. Static-ILEX is more typical of current Web-based hypermedia systems than dynamic-ILEX.

2.3 Sample output from the two versions of ILEX

To illustrate how output from the two versions of the system differed, the same browsing sequence was conducted with each system in order to generate comparative output. The browse sequence in this exercise was as follows. The first jewel in the gallery 'case' was accessed, then a further 5 jewels in the gallery were viewed (the first row of jewels in the case). The first jewel was then revisited.

The jewel browse sequence was 1) A gold, enamel and saphire pendant necklace by Jessie M. King. 2) A gold, enamel and moonstone pendant-necklace by Jessie M. King. 3) A silver and enamel pendant necklace by Jessie M. King. 4) A waist buckle by Jesse M. King. 5) A gold, moonstone and opal necklace by Edward Spencer and 6) A silver metal, beryl and tourmalines necklace by Arthur and Georgie Gaskin. The first jewel in this sequence was then revisited.

The first three pages of output are shown below, for both versions of the system.

2.3.1 Output from dynamic ILEX

On the first viewing of the first jewel, dynamic ILEX produced the following labels (nb 'say more' shows output following click of 'say more' button (see Figure 1):

> *This jewel is a necklace and was made by a British designer called Jessie M. King. It is one of the four items in this case which were made for Liberty & Co. It is comprised of openwork gold floral sprays. It is set with jewels. It is in the Arts and Crafts style but it uses faceted stones. It has an elaborate design (specifically it has floral motifs). King was British. Other jewels designed by King include...(list of 3)*
>
> 'Say more' *This necklace was made in 1905. It is made from sapphire, enamel and gold. It was made in London. As already mentioned, it was made by Jessie M. King. King was not just a jewellery designer, she was an illustrator too. In fact, she did quite a lot of different types of creative work. Jewellery is just part of it. We have also mentioned that it is in the Arts and Crafts style. Arts*

and Crafts style jewels usually use oval-shaped stones and usually feature rounded stones. Other jewels in the style include...(list of 7)

'Say more' *This necklace has festoons and was produced in limited quantity. As already mentioned, it was made by Jessie M. King. King was important and lived in London. We have also mentioned that this necklace is in the Arts and Crafts style. Arts and Crafts style jewels usually have an elaborate design and are usually flexible. They usually demonstrate the artistic sensibilities of the wearer and were usually produced by single craftsmen. Other jewels in the style include: ... (list of 7)*

After visiting the next five jewels in the virtual 'case' the same jewel was revisited. The first page of the label now read:

Arts and Crafts style jewels usually feature rounded stones; for instance the previous item uses oval-shaped stones (in other words it features rounded stones). However this necklace does not featured rounded stones; rather it uses faceted stones.

Note that the label now contains a comparison between this jewel and the previous jewel viewed (a silver, beryl and tourmalines necklace by G & A Gaskins, also in the Arts & Carfts style). The phrase 'previous jewel viewed' was a hyperlink back to that item.

2.3.2 Output from static ILEX

On the first visit the label read:

Page 1 *This jewel is a necklace and is made by a designer called Jessie M. King. It is one of the four items in this case which were made for Liberty & Co. It is comprised of openwork gold floral sprays. It is in the Arts and Crafts style and is made in 1905. It is set with jewels, in that although Arts and Crafts style jewels usually use oval-shaped stones this jewel uses faceted stones. Other jewels designed by King include ... (list of three displayed). Other jewels in the style include: ... (list of 4 displayed)*

Page 2 *This necklace is made from sapphire, enamel and gold. It is made in London and has festoons. It is produced in limited quantity. It has an elaborate design (indeed Arts and Crafts style jewels usually have an elaborate design); specifically this necklace has floral motifs. Liberty and Co were at the interface between mass-produced jewellery and 'craft' jewellery – one-offs. They used the very best designers to design jewels for them, which were then produced in fairly limited quantity, but in quantity nevertheless. Not quite mass-produced, and not quite 'craft'. Liberty's designers were never credited, so a piece like this wasn't sold as a 'Jessie M. King (for Liberty)' necklace – it was simply sold as a Liberty necklace. But the design books survive, and we have a pretty good idea who designed what.*

Page 3 *As already mentioned, this necklace is in the Arts and Crafts style. Arts and Crafts style jewels usually feature rounded stones and is usually flexible. They usually demonstrate the artistic sensibilities of the wearer. This necklace is made for Liberty and Co.*

Since this version was a static hypermedia system, the output was identical when this item was revisited after the intervening five jewels.

3 Development of the outcome measures

Three instruments were devised for use in the evaluation. They consisted of 1. a recall test of factual knowledge about jewels in the exhibition, 2. a 'curator' task (to be described below) and 3) a useability questionnaire. The tests were administered to subjects on-line, as Web-forms linked to ILEX. Multiple choice check boxes and radio buttons were provided for subjects to indicate their response choices.

3.1 Factual recall test

This was a multiple choice test which was introduced to the subjects with the heading "What did you learn from the virtual exhibition?" Two examples (of 15 items in total) are shown below:

1. Clean lines and geometric forms are characteristics of which style(s)? [sixties, organic, scandinavian, machine age, art deco, arts & crafts]

15. Wendy Ramshaw and David Watkin made jewels in the 60's. They typically used which materials (choose up to 5 materials)? [paper, steel, perspex, wood ...(31 materials listed in total)]

3.2 Curator task

An educational aim valued by museum curators is the inculcation in visitors of a notion of *artifacts as evidence* - evidence of a particular time, place and set of beliefs [3]. Visitors are taught to look, describe, record, classify objects and artifacts and perhaps take away skills that may be useful for assessing the significance of (novel) artifacts outside of the museum context.

The second instrument therefore attempted to assess these kinds of skills. A typical item is shown below. It consisted of the presentation of a jewel not seen in the exhibition, with subjects instructed to 'Examine the photograph and then classify the jewel in terms of its Style'. Multiple choices options were [sixties, organic, scandinavian, machine-age, art deco, arts & crafts]. The complete test consisted of 15 items.

4 Evaluation study

4.1 Subjects

The subjects were University of Edinburgh students recruited via notices in academic departments and halls of residence. Thirty subjects were randomly assigned to one of two conditions - static ILEX or dynamic ILEX, with the constraint that more subjects (20) were allocated to the dynamic ILEX condition than to static ILEX (comparison) condition (10 subjects). Gender representation was as follows: static-ILEX 7 male, 3 female; dynamic-ILEX 11 male, 7 female. Data was lost for two of the subjects in the dynamic-ILEX condition due to equipment failure.

4.2 Methodology and procedure

The subjects were run over three, consecutive, 90 minute sessions. Each session consisted of exhibition browsing (45 minutes) followed by the post-tests and a questionnaire (45 minutes).

The experiment was conducted in one section of a computer laboratory. Subjects were seated at a row of Pentium PCs (one row each side of a central divide). In order to ensure adequate levels of performance, multiple servers were used to serve the ILEX web pages.

4.2.1 Subjects' task

Subjects were instructed to log on to the system and explore the ILEX virtual museum gallery. They were told that they would be required to answer quiz questions following the gallery browsing session. Subjects were not told which version of ILEX they were using, or that there were two versions of ILEX in use. Since static-ILEX subjects were seated on one side of a central divide and dynamic-ILEX users were on the other side, they were unaware of any differences between the systems.

4.2.2 User-system interaction logging

In order to track the domain content viewed by the users, and to relate hypermedia browsing and navigational behaviour to learning outcomes, a logging system was implemented on both static and dynamic ILEX systems. All subjects' button presses were recorded in time-stamped logs together with records of the pages visited and the information content of the pages.

5 Results

The recall test and curator task responses were forwarded electronically to the experimenter for marking. User-system interaction logs were automatically saved into separate directories for each subject.

5.1 Post test outcome

5.1.1 Factual recall test

Test 1 consisted of 15 items. Maximum possible score was 31. The test means for subjects in both groups are shown in Table 1.

5.1.2 Curator task

The second ('curator') task consisted of 15 items. Maximum possible score was 23. The test means for subjects on this test are also shown in Table 1.

As shown in Table 1, mean group scores were similar on both tests. However, the dynamic subjects showed considerably less variation in score, compared to static subjects, on both tests.

Although *performance* scores were similar in both groups, it was of interest to discover whether the underlying behaviour or learning *processes* by which they were achieved differed. To pursue this question, the log

Group		Factual recall test	Curator task
Static ILEX	mean(std dev)	18(4.2)	9(2.9)
Dynamic ILEX	mean(std dev)	18(3.0)	10(1.9)

Table 1: Mean scores and standard deviations for subjects in the static and dynamic ILEX conditions on the two tests.

data were analysed in order to determine whether the two systems produced differential effects in terms of a) browser/navigational behaviour and b) the amount and type of information browsed.

5.2 *Analysis of log data*

Two kinds of data were extracted from the user-system interaction logs — 1. the user's browser behaviour (navigational maneouvres) and 2. page content — jewellery facts presented to particular users in the course of their session.

5.2.1 *Browser/navigation events measures*

Several indexes of navigational behaviour were extracted from the logs. These consisted of: visits to the case of jewels (CASES); button clicks in the ILEX navigation bar provided at the top of each page (back, forward, go to jewellery case) (NAV, PNTR); requests for more information or another page of information about a jewel (NAV2); and total events (CASES, NAV, NAV2, PNTR plus other miscellaneous browser events).

Subjects using static-ILEX demonstrated approximately 60% more navigation-related button clicks than their dynamic-ILEX counterparts. However, this was an artifact of way static pages are generated by dynamic-ILEX – information about a particular jewel was distributed over several pages and contained more repetition of information compared to the dynamic-ILEX system. In order to standardise the comparison, therefore, the CASES, NAV, PNTR and NAV2 measures were expressed as proportions of total events to yield PCASES, PNAV, PPNTR and PNAV2 variables for each subject.

5.2.2 *Page content measures*

The following parameters were derived from the log data: the number of (different) jewels viewed (NODIFF); the mean number of pages browsed per jewel (MNPPO); the mean number of visits per page per jewel (MN-VPPO). This provided an index of the extent to which subjects revisited pages of information about a particular jewel; and the mean display time per page (MNDPP).

5.3 *Log analysis results*

5.3.1 *Browser/navigation events measures*

Statistically significant differences were found between the groups in the number of visits to the case of jewels (Static subjects PCASES=0.075, dynamic subjects PCASES=0.169, t=-5.26, $p < .0001$). Also in forward, back button, etc button clicks (ie. maneouvres using the ILEX navigation controls) – Static subjects PNAV=0.14, dynamic subjects PNAV = 0.24, t=-3.26, $p < .005$) and also the use of an alternative 'forward' button (Static subjects PPNTR = 0.15, dynamic subjects PPNTR=0.39, t=-3.94, $p < .005$)

5.3.2 *Page content measures*

Signicant differences between the groups were found for the following measures: number of jewels viewed (static subjects mean NODIFF = 29.8, dynamic subjects = 28.0; t=2.81, $p < .05$); mean number of pages viewed per jewel (static subjects MNPPO mean=2.96, dynamic subjects=2.19, t=-5.71, $p < .0001$); and mean number of visits per page, per jewel (static subjects MNVPPO mean=1.53, dynamic subjects=1.07; t=3.91,$p < .005$). There was no significant difference between the groups in mean display time per page (static subjects MNDTPP= 21.96 seconds, dynamic subjects = 21.13 seconds).

5.3.3 *Correlation between log measures and test performance*

Correlation of each of the browser event and page information parameters and the test scores were computed. The results showed that mean display time per page correlated significantly and positively with test 1 (factual recall) scores for dynamics (r=.61, $p < .01$) but not for static subjects (r=.05, n.s.). Another finding was that mean pages visited per jewel (MNPPO) correlated moderately positively with test 1 score for static subjects (r=.42, n.s.) but moderately negatively for dynamic subjects (r=-.53, n.s.). There were no striking differences in correlation pattern between the groups on test 2, the 'curator task'.

Table 2: Modal responses to questions by subjects in Static and Dynamic conditions.

Question	Static	Dynamic
1.I think I would like to use ILEX frequently.	disagree	neutral
2.I found ILEX unnecessarily complex.	disagree	disagree
3.I thought ILEX was easy to use.	agree	strongly agree
4.I think that I would need the support of a technical person to be able to use ILEX.	strongly disagree	strongly disagree
5.I found the various functions in ILEX were well integrated.	agree	agree
6.I thought there was too much inconsistency in ILEX.	disagree	agree
7.I would imagine that most people would learn to use ILEX very quickly.	strongly agree	agree
8.I found ILEX very cumbersome to use.	disagree	disagree
9.I felt very confident using ILEX.	agree	strongly agree
10.I need to learn a lot of things before I could get going with ILEX.	strongly disagree	strongly disagree

Useability questionnaire The results for each group are shown in Table 2.

There were no substantial differences between the two groups' perceptions. Most differences were small and in favour of dynamic ILEX, except in the case of the 'inconsistency' question 6.

Results summary To summarise, both groups scored similarly on the two tests in performance terms. Scores of subjects in the dynamic ILEX condition were less variable than those of subjects in the static-ILEX condition, for both tests. All subjects tended to score better on the factual recall test (58% accuracy) than on the 'curator' task (approx. 40% accuracy).

The log data revealed that dynamic subjects made more visits to the case of jewels than static subjects, and made proportionately more navigation-related button clicks than their static ILEX counterparts.

However, static ILEX subjects looked at more jewels, more pages per jewel[2], and made more repeat visits to pages.

6 Discussion

Subjects in the dynamic-ILEX condition did not score significantly better than subjects in the static-ILEX condition and so the hypothesis regarding the learning benefits of dynamically generated labels was not supported by the gross performance data.

However, the alternative hypothesis – that dynamic hypertext can be counter-productive due to its varying, non-standardised nature [5] – was also not supported. Both groups scored equivalently on the performance tests. But, despite similar performance scores, the process data showed up interesting differences between the systems.

Subjects in the dynamic-ILEX condition were exposed to fewer examples than static-ILEX subjects. A higher proportion of dynamic-ILEX subjects' button clicking activity was concerned with accessing further information and visiting the top-level jewellery case. But the information they received was 'denser' and spread over fewer pages. Dynamic-ILEX subjects made fewer repeat visits to pages, perhaps indicating that information was more easily aborbed 'first time around' from that system than from static-ILEX.

The correlation patterns (of log parameters with test scores) showed that the correlation between the time for which a page was displayed and test 1 (factual recall) score was higher in dynamic subjects than in static subjects. This was despite the fact that static-ILEX subjects looked at more jewels and more pages per jewel. This suggests that the dynamic text (with its unique characteristics — referring to previously seen items, use of comparison expressions, etc) may have produced more learning effect per unit of information displayed.

The difference between the groups in the sign of the correlation between repeat visits to pages and test 1 score is also interesting. It may suggest that static subjects were acquiring knowledge from the exhibits via a learning process akin to induction, whereas, in contrast, dynamic subjects were processing more declarative (and individualised) information from a 'cooperative' system.

[2] This particular index must be interpreted a degree of caution as the measure is unadjusted for the fact that, in the static-ILEX condition, on average, information tended to be spread over more pages, with more repetition of information, than was the case for the dynamic-ILEX condition.

One factor that requires discussion concerns the slower speed of the dynamic-ILEX system compared to static-ILEX. Despite the use of multiple servers, the dynamic text generation process is complex, and on some (worst case) occasions, took up to one or two minutes to produce. At such times, subjects were instructed: ‘This label is being dynamically generated. Please look at the photo in the meantime.’ Thus it is possible that dynamic subjects scrutinised the visual features of the jewels to a greater degree than the static subjects. This issue may be the focus of further work.

This study also showed that as in previous studies [1, 9], the collection of process data in addition to performance test data provides insights that cannot be achieved through the use of performance tests alone. The ‘learning-performance’ distinction is well known in psychology - performance may not necessarily reflect learning for a variety of reasons, often because the student motivation at the time of testing is reduced or because the test fails to ‘tap’ the particular information acquired. The collection and analysis of rich process data allows the distinction to be elucidated to some degree.

Further analyses of the user-system interaction logs are planned. This will be in order to compute, for each subject, how many ‘test-relevant’ (i.e. subsequently-tested at post-test) facts they were presented with by ILEX during their browse session. Subjects were free to browse all of the jewellery gallery but only a subset of facts were tested at post-test. The aim, then, is to elucidate further subtle differences between the systems.

Further, it is possible to divide the facts into two types - those that were explicitly presented complete statements of the form ‘Paula Dennet designed in the 60’s style’ or ‘arts and crafts style jewellery is usually produced by a single craftsman’ and, on the other hand, those that, taken together, would allow inference by induction eg. the repeated pairing of the terms ‘Scandinavian’; ’clean lines’ or ‘Sixties’ and ‘perspex’, within and across pages.

It may turn out to be the case that subjects differ considerably in the amount of test-relevant information they browsed. They may also differ in terms of the proportions of explicit and implicit information that they were exposed to prior to testing. Different navigational strategies and browser use may be associated with test scores and the modes in which information is acquired.

Acknowledgements Thanks to Chris Mellish and Alistair Knott for useful discussions in the course of this work. Thanks too to the anonymous reviewers for helpful suggestions. The ILEX project is supported by EPSRC grant GR/K53321 and the third author is supported by an EPSRC Advanced Fellowship. We acknowledge the valuable assistance of the National Museums of Scotland.

7 References

[1] Cox, R. & Brna, P. (1995) Supporting the use of external representations in problem solving: The need for flexible learning environments. *Journal of Artificial Intelligence in Education*, **6(2)**, 239–302.

[2] Dale, R., Oberlander, J., Milosavljevic, M., and Knott, A. (in press) Integrating natural language generation with hypertext to produce dynamic documents. Interacting with Computers, **9**.

[3] Durbin, G, Morris, S. & Wilkinson, S. (1996) Learning from objects: A teacher’s guide. English Heritage,

[4] Hitzeman, J., Mellish, C. and Oberlander, J. (1997) Dynamic generation of museum web pages: the intelligent labelling explorer. *Archives and Museum Informatics*, **11**, 107-115.

[5] Hook. K. & Svensson, M. (1999) Evaluating adaptive navigation support. *Proceedings of IUI’99*, Los Angeles, USA.

[6] Mellish, C., Knott, A., Oberlander, J. and O’Donnell, M. (1998a) Experiments using stochastic search for text planning. In *Proceedings of the 9th International Generation Workshop*, Niagra-on-the-Lake, Ontario, Canada.

[7] Mellish, C., O’Donnell, M., Oberlander, J. and Knott, A. (1998b) An architecture for opportunistic text generation. In *Proceedings of the 9th International Generation Workshop*, Niagra-on-the-Lake, Ontario, Canada.

[8] Oberlander, J., Mellish, C., O’Donnell, M. and Knott, A. (1997) Exploring a gallery with intelligent labels. *Proceedings of the Fourth International Conference on Hypermedia and Interactivity in Museums*, Paris, September, 153–161.

[9] Stenning, K., Cox, R., & Oberlander, J. (1995) Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, **10, 3/4**, 333-354.