Museum audio guides which adapt to the user and context

Mick O’Donnell

Istituto Trentino di Cultura, Italy

The technology of Natural Language Generation (NLG) has reached the point where real applications are becoming viable. One such application involves the dynamic generation of descriptions of museum exhibits. As a visitor walks around the museum, the system selects the information about the exhibits most relevant to the visitor, and generates an audio presentation expressing this information. The presentation needs to be coherent and interesting, and not repeat information the user has already been presented with. Comparisons may be made to exhibits already seen by the user, so tracking the visitor's movements is important.

This paper presents three systems which the author has worked on, all of which take different approach to this task: ILEX (full text generation, mapping database-represented information onto semantic and syntactic structures, generating audio via speech synthesis), HIPS (template-based generation, assembling audio fragments), and MPIRO (a new project, which integrates the above two approaches). The advantages and disadvantages of the approaches will be compared.

1. Introduction

Audio guides for museums provide spoken descriptions of exhibits as a visitor walks around a museum. The descriptions are usually *static,* in that the audio is pre-recorded. Different users will receive the same descriptions of exhibits, regardless of age and interest. Static audio guides may also repeat information at several points, as the guide cannot know if the information has been given already in relation to some other exhibit.

New technology is evolving using *dynamic* audio guides: audio guides which adapt presentations to the type of visitor, and also take into account what the visitor has already been told. These systems, using Natural Language Generation (NLG), compose exhibit descriptions on the fly, taking into account the interests of the visitor, what they are assumed already to know, what they have already been told, what they can see, etc. The expression of the information can also be tailored to the visitor, using simpler expressions for children, and more complex for adult visitors, or domain experts.

This paper reports on three inter-related projects which the author has worked on, all of which take different approaches to this task. These projects are:

- **ILEX** (Intelligent Labelling Explorer): an EPSRC (UK) funded project carried out at the University of Edinburgh, from 1996-1998.
- **HIPS** (Hyper-interaction in Physical Space): an Esprit (European) funded project, with partners including University of Edinburgh, University of Siena, Alcatel (Italy), Istituto

Trentino di Cultura (Italy), University College of Dublin, GMD (Germany), CB&J (France) and SINTEF (Norway), from 1998-2000.

- **M-PIRO**: (Multi-lingual Personalised Information Objects) : another European project (IST programme), involving University of Edinburgh, Istituto Trentino di Cultura (Italy), National Centre for Scientific Research (Athens), Knowledge Engineering Laboratory (Athens), National and Kapodistrian University of Athens, Foundation of the Hellenic World (Athens), and System Simulation Ltd (UK). From 2000 - 2002.

I will start by giving some surface details on these systems, and then compare them on three aspects: i) the means of generating presentations; ii) the methods of adapting the text to the user; and iii) the delivery systems which interface generation systems to users.

2. Overview of the three systems

ILEX

ILEX was not initially conceived of as an audio guide, but rather as a general exhibit label generator, supporting both hypertext access, as well as for audio-based applications. ILEX was developed using an exhibit database from the National Museum of Scotland (NMS), including most of the exhibits contained in the Twentieth Century Jewellery exhibition. This information was received in a database format, such as in figure 1. We implemented a routine to automatically import the museum database into our own knowledge representation system. The information was then used as the basis for generating descriptions of exhibits, their designers, styles, etc. For example, see figure 2. As will be discussed below, the text in the second and third paragraphs of the text are generated using additional techniques, allowing canned texts and generalisations to be used. To see

ILEX in operation, please visit: http://cirrus.dai.ed.ac.uk:8000/ilex/.

item:	j-990656
class:	pendant-necklace
designer:	King01
date:	1905
style:	"Arts and Crafts"
material:	silver
material:	enamel

Figure 1: Database information used by ILEX

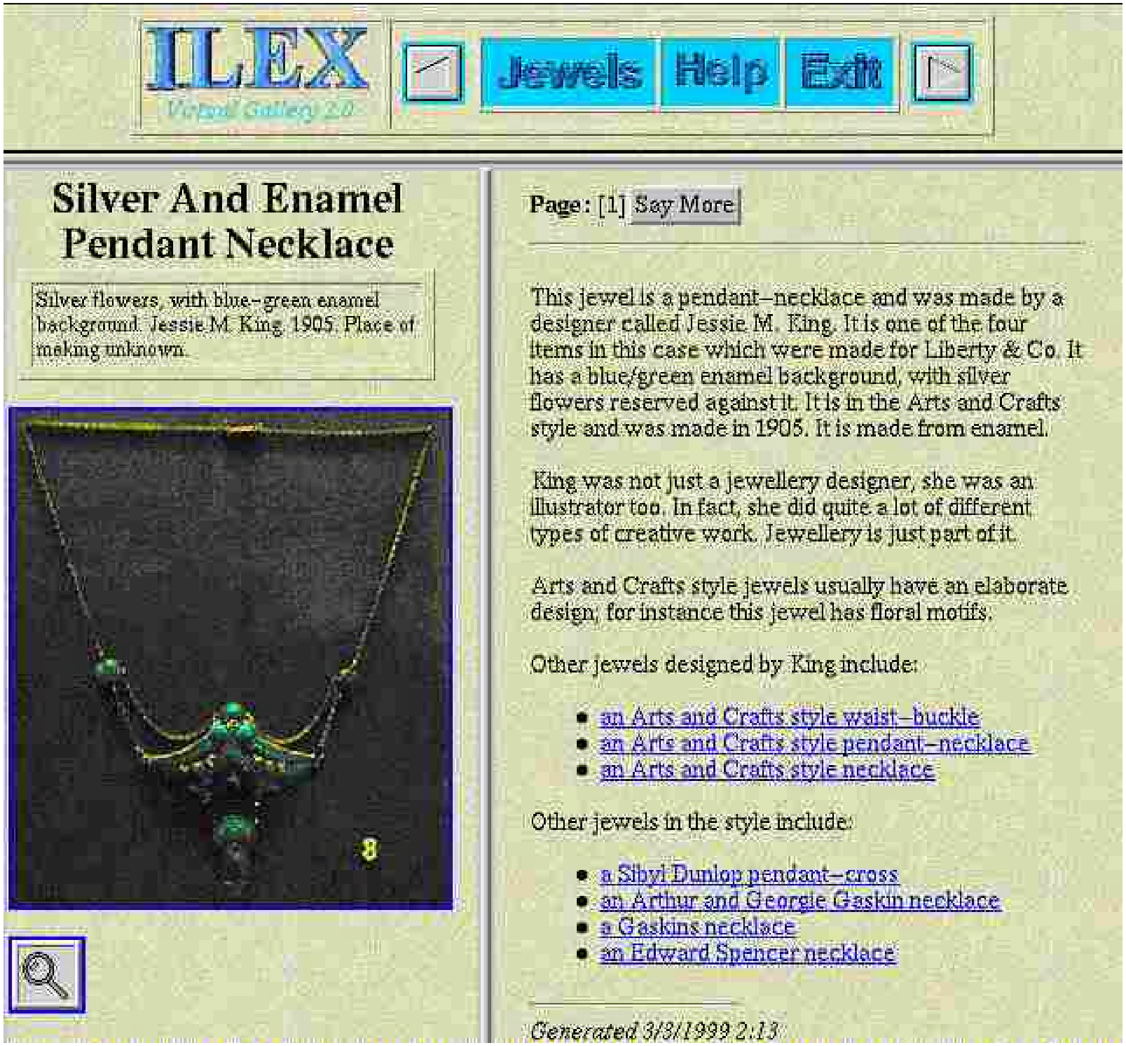

Figure 2: A typical ILEX page

HIPS

The HIPS system was from the start conceived of as an audio guide for museums and similar environments (e.g., city guides). The main medium is audio, the user wearing headphones, while the user controls the system in two ways:

Physical movement: HIPS uses location hardware (infrared, GPS, electronic compass) to track the visitor's movements through the museum, and uses this information to select the exhibits to describe.

Button or pen interaction with a handheld device: The user uses a GameBoy like device to receive basic information about the exhibit in front of them, and to select sub-topics to follow. Other interfaces are also available, providing additional information for the visitor such as maps, indexes to exhibits, etc.

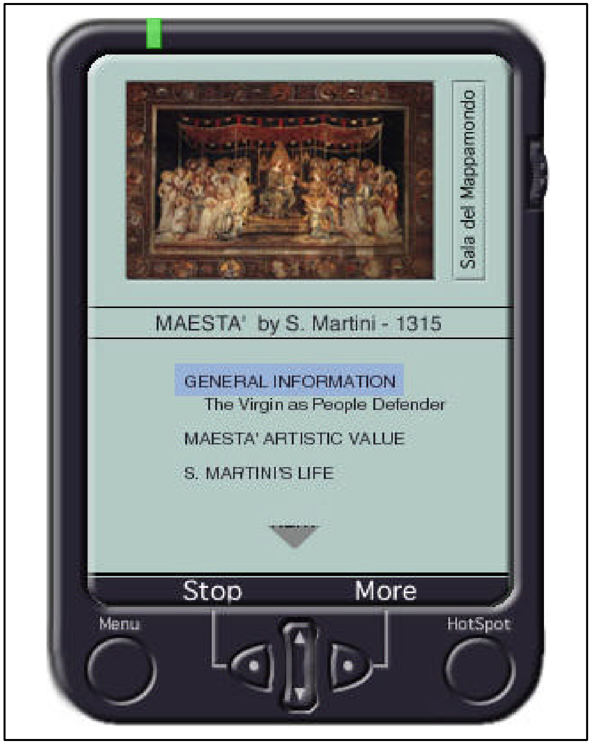

Figure 2: A typical HIPS display

M-PIRO

M-PIRO as a project is still in its first year, so is still taking shape. The project, like ILEX, is not intended to produce a system for a single media, but rather focuses on the natural language engineering aspects of the task. The project is pre-supposing the hardware components (localisation, audio delivery, communication, etc.) developed by prior projects and is instead focusing on producing a natural language generation system which deals with multiple languages, and multiple modalities (output either as text, or using speech synthesis). The system will also focus on strengthening the user-based adaptation of generated texts.

The rest of this document will explore three areas of difference between these three systems.

3. Granularity of text generation

Text generation systems vary as to the amount of informational detail they require to work. For some applications, chunks of text are assembled together to create larger texts. These “chunks” could be paragraphs, sentences or even phrases. These systems are “generating text” in that they can assemble the chunks in different ways, choosing a sequence of chunks which in some way suits the current user and context of interaction.

In other systems, the information the system starts with is not expressed as text, but in some database form. This information is packaged into sentences using either templates (e.g., *This $CLASS was made by $ARTIST*), or by using full NLG, using linguistic knowledge to map the domain information onto lexified syntactic structures, and thus producing text.

We can thus talk about two dimensions of granularity here. In terms of the inputs to the system, the domain information can be large-grain (paragraphs of text), medium grain (a simple database specification), or fine-grain (database specification using type hierarchies to organise entities, etc.). In terms of the outputs of the system, granularity can range from large-grain (fixed paragraphs), medium-grain (templates), to small-grain (using full NLG). While large-grain input necessarily uses large-grain output (paragraphs in, paragraphs out), note that small-grained input can be expressed using templates or full NLG.

Below we will consider the three systems discussed here in these terms.

3.1. ILEX

ILEX was designed principally to express information contained in existing museum exhibit databases, typically, a relational database, with a record for each exhibit, defining artist, place of origin, materials, style, techniques, etc. Each of these attributes may in itself be a key to another database record, providing details of that entity, for instance, details of the artist. We also assume that that database is organised in terms of a *type hierarchy*, for instance, a hierarchy of exhibit types: *artwork: painting: fresco*, or *craftwork: jewellery: wristwear: bracelet*. The database we obtained from the NMS was of this form, with an implicit type hierarchy.

Data in this form would support either template-based generation, or full NLG. We started with template-based generation, but found that the sheer variety of different ways we needed to express each different fact-type actually called for a more robust approach. For instance, the database fact: designer($Jewel, $Designer) could be expressed using a template: *This $CLASS was designed by $DESIGNER*, but if the exhibit is multiple (e.g., a pair of bangles), then we need another template: *These $CLASS were designed by $DESIGNER*. If our starting point is in fact the designer, we need a template for the active voice: *$DESIGNER designed this $CLASS*, and so on. Once we introduced the need for negation, generalisation, etc., we found the template approach was too cumbersome for our needs. For this reason, we connected the ILEX system to a robust sentence generation system WAG (see O'Donnell,1996) to handle the complexity of sentence realisation.

On the other hand, we found that database information by itself produced very uninteresting labels, so we allowed our system to handle information from two other sources to produce better quality text:

Rules representing expert knowledge

Because information in ILEX, is formally represented, the program can perform logical reasoning over data. On request, curators provided generalisations about the domain, such as *Most art-deco jewellery is made using enamel*. From such rules, we can automatically generate complex text, as shown in the third paragraph of figure 2.

Canned text

Where information from the curator or books was not easily represented in terms of our knowledge representation system, we allow for canned "stories" to be associated with entities, or classes of entities. The text can be included in generated text to improve the fluency of the page. The second paragraph of figure 2 is an example. ILEX thus allows for varying degrees of granularity in regards to text generation.

The various facts contained in the relational database, and the additional information from generalisations and canned text, are organised into a coherent text using the ILEX text planner (see Oberlander *et al*.,1998). Nominal references in the text are generated to be contextually appropriate (see O'Donnell *et al*., 1999).

3.2. HIPS

In regards to inputs, HIPS starts from a different position. In HIPS, we argued that there are existing text repositories, descriptions of exhibits, both in the museum labels themselves, and in books and guides written about exhibitions. The goal of HIPS was to reuse existing textual repositories, but re-represented in such a way as to allow its presentation to be adapted to the type of user, and to the context of interaction.

The starting point of HIPS data preparation is the use of this material to construct ideal tour descriptions for each exhibit, along with stated variations, for different user types (e.g., some information for children only, others only for experts), and also for different visit contexts (e.g., returning to an exhibit will produce *You are back at the Guidoriccio* rather than an initial introduction). These transcripts and variants are then recorded by professional speakers, and the resulting audio stream is segmented into smaller segments, which will be reassembled on demand as suitable to the context of interaction.

The resulting datastructure is a network of *macronodes*, where each macronode corresponds roughly to a paragraph of text. The HIPS text planner traverses this network to select a sequence of macronodes for a given presentation. Once the sequence is selected, further contextual adaptation occurs in the realisation of a macronode. A macronode itself is represented by a graph, this time a *transition network*. It represents alternative ways of presenting the paragraph, each alternative path appropriate to a different context. See figure 4 for an example. This is, in effect, a template-based approach, of medium grain.

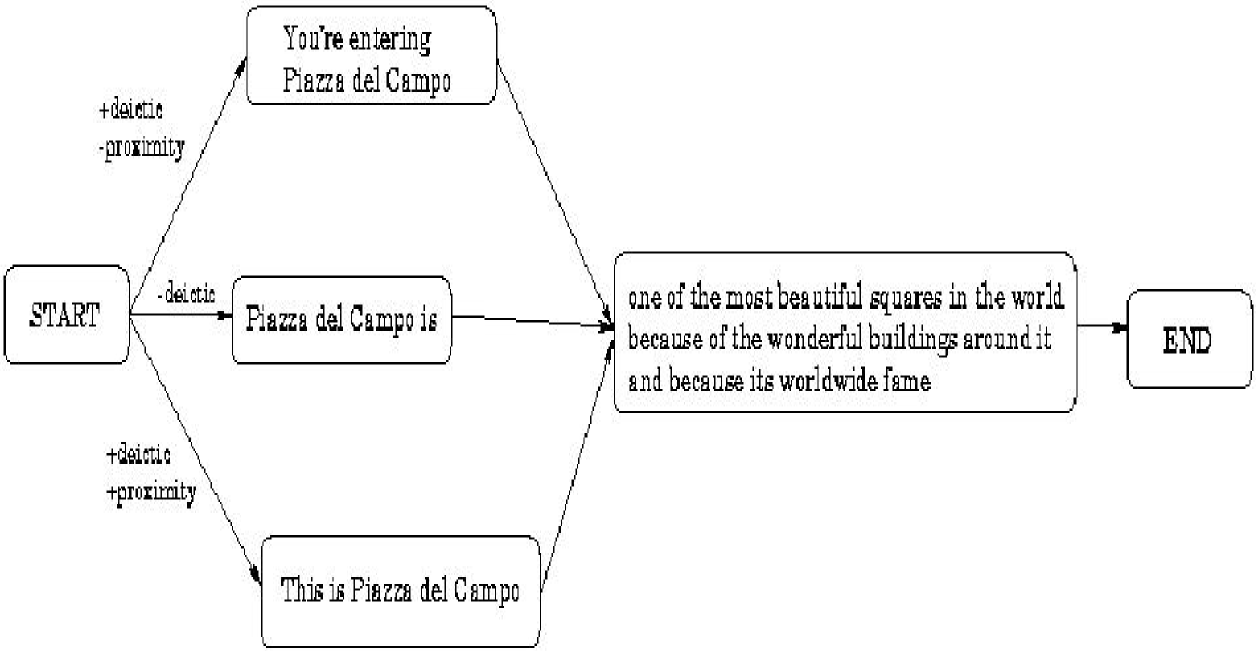

Figure 4: A sample macronode

3.3. M-PIRO

In **M-PIRO**, we opted for a hybrid approach, allowing for the high-grain input from museum databases, but also allowing for the reuse of textual resources via the macronode approach. The M-PIRO system is using a re-implementation of ILEX in Java. Like the original ILEX, it accepts database-defined inputs, and generates text using full NLG. However, we also wanted to re-use existing textual resources, so, in place of the canned stories of ILEX, we are using macronode-represented text. The schema-based text planner from HIPS is being reworked to handle both ILEX-style database information and HIPS-style macronodes. In this way, we hope to achieve the best of both worlds, reusing all available information.

M-PIRO also extends on the NLG architecture of ILEX by being multi-lingual. The database represented information will be expressed in English, Greek or Italian, using language-specific lexicon, inflectional morphology, syntax and rhetorical strategies.

3.3. Summary

Systems with high-grain input and output have more possibilities for adapting the text to the context of interaction, and to the user. On the other hand, machine-produced text is rarely as good as human-written text, so the low-grain systems often produce better quality (albeit less adapted) texts.

Another advantage of using fine-grain input is that existing museum databases can be used for generation, without the need of human-authoring of web-pages or audio-guides. Changes or additions to the database will be realised automatically in the produced text/audio. This approach can thus save significantly in authoring costs. However, databases do not always come in the desired format, or with all required information, so some data preparation might be necessary.

4. Adapting content to the user

The three systems take different approaches to adaptation to users and context.

In **ILEX**, adaptation is in regards to two main aspects:

- *Discourse history*: the system keeps track of what has been told to the user, so does not repeat what it has already told them (except for rhetorical purposes). The system also selects referring expressions using this resource.
- *User model*: the system is given basic user models of expected user types (child-visitor, adult-visitor, curator, etc.). This model details what facts the user should already know, what they will find interesting, how often information needs to be repeated to be assimilated, etc. This information is taken into account when deciding which of the available information to present to a user at each point.

HIPS takes a stronger approach to user modelling than ILEX. Firstly, the modelling of interest is *dynamic*, while ILEX's is *static.* In ILEX, the model of user interests is set at startup, while in HIPS, the system watches the choices the user's make, and modifies the perceived interest in each exhibit (work by GMD, Germany, see Specht & Oppermann, 1999).

HIPS also models the user by observing the way the user moves through he exhibition space, and uses this to predict how long they will typically stay in front of an exhibit (how much opportunity will the system have to say what it wants to say), which exhibit will they move to next (following walls, or jumping at random), do they stay in the centre of a room to get an overview and thus they may benefit from comparisons between artworks, etc. (work by University of Siena, see Marti *et al*., 1999).

M-PIRO intends to enrich the approach to user modelling begun in ILEX. In ILEX, adaptivity of the text is largely in regards to which facts are included. M-PIRO aims to extend this to allow for: variations of lexical selection for different user types; or in phrasing (telegraphic vs. elaborated, etc.)

5. Delivery systems

HIPS and ILEX have each explore different delivery systems – how the presented information is delivered to the user. The main variables of difference in delivery are *mode* (text vs. speech) and *selection* (how is the exhibit of relevance selected). M-PIRO will not be discussed here, as its focus is on the NLG technology, not the delivery.

5.1. Mode

ILEX

The main focus of ILEX was on an HTML-based interface, delivering textual descriptions of exhibits on demand. The presentation of the descriptions is multi-modal, in that the layout intermixes text and graphics.

A related project, called SOLE (see Hitzeman *et al*., 1998), explored linking ILEX to the Festival speech synthesis system (Taylor *et al*., 1998) to produce synthesised versions of exhibit descriptions. Festival was extended to take into account the rhetorical structure of the text in synthesis, for instance, delivering examples in a distinct intonation than for contrasts, etc.

Using speech synthesis, we explored the use of ILEX over a cell-phone. However, voice quality at present was not good enough to make this approach viable.

HIPS

HIPS chose the audio mode for delivery. We rejected speech synthesis on the basis that synthesised speech tends to irritate users, and HIPS aimed at commercialisation. The template-based generation method used in HIPS supports a system whereby audio fragments are assembled into audio presentations. The fragments are of human voice, spoken by professionals. They range in size from nominal referents (“it”, “the painting in front of you”), to whole paragraphs. When delivered in sequence, the fragments produce reasonable quality flow of speech. See Not *et al* (1999) for more detail.

5.2. Selection

ILEX

ILEX was intended for use in a museum, and we explored various mechanisms for use of the system in situ. Ideally, this involved a device that users carried around with them, and either:

The device registers their location and thus delivers descriptions of nearby objects. We considered mounting an infrared emitter over relevant exhibits, with an infrared receiver mounted on the handheld device. However, where many exhibits are very close together (as in the jewellery gallery), this technique is impractical.

The user indicates to the device the object of interest themselves. This could involve i) the visitor uses a card-swipe device which notifies the central server of the exhibit of interest; or ii) the user selecting an exhibit using the interface of the device they carry.

Using the cell-phone approach mentioned above, the user can select exhibits by entering exhibit codes into the phone keypad.

We also have a demo system using a handheld Libretto PC, which allows the user to select exhibits as they walk around, by clicking on the picture of the exhibit. Descriptions are delivered in text.

However, in the end, we decided that the focus of ILEX was the NLG technology, not the delivery mechanism. We implemented a web interface for ILEX. This could be set up in a gallery as a resource for visitors.

HIPS

The starting idea in HIPS was to use localisation technology (infra-red, GPS) to track the visitor's movements in physical space. The system would then generate descriptions of exhibits in front of them. We explored two scenarios: indoors, we used a system of infrared emitters mounted on walls and an infra-red receiver carried by the user. These were configured in such a way that the system could locate the visitor's general position in the room. Outdoors, we depended on GPS (using satellites for localisation). In both cases, an electronic compass provided orientation.

This approach, in general, worked, except that it is only useful where the exhibits are large. Where exhibits are arranged closely, the system cannot determine automatically which exhibit the user is looking at. However, for selected scenarios, this approach works.

6. Summary

It should be clear that HIPS paid more attention to the delivery system than ILEX. ILEX focused on text generation, leaving the delivery system to be resolved with improved technology. HIPS on the other hand was more oriented towards producing a commercial system, and thus spent a lot of effort on the delivery system, in terms of selection of the handheld device, localisation technology, audio streaming, linking of hand-helds to a central server via a radio-based LAN, etc.

7. Conclusions

This paper has described three different approaches to adaptive museum guides. The three systems have been compared in regards to means of text generation, approach to user and context adaptivity, and the delivery mechanisms used.

In regards to text generation, we concluded that finer grain input, with more robust generation methods, allowed for better adaptation of the text to the user and context, but often produced worse text than methods with larger grain input. ILEX, accepting input from relational databases, allows re-use of already existing databases, while HIPS allows re-use of existing textual resources.

In regards to user modelling, HIPS improves on the ILEX approach by introducing dynamic modelling of visitor interests, and also modelling the way visitors move through the exhibition space. M-PIRO is also intended to improve on ILEX by using the user-model to adapt expression choices within sentences.

In regards to delivery mechanisms, all systems have experimented with alternative approaches, both in spoken and written mode. HIPS has gone furthest, with actual installations using localisation technology, radio-based LANs, audio streaming, etc.

REFERENCES

Hitzeman, J., A. Black, P. Taylor, C. Mellish & J. Oberlander. On the use of automatically generated discourse-level information in a concept-to-speech synthesis system. In

Proceedings of the 5th International Conference on Speech and Language Processing, 2763-2766, 1998.

Marti, P., A. Rizzo, L. Petroni, G. Tozzi & M. Diligenti. Adapting the museum: a non-intrusive user modeling approach. In *Proceedings of UM99*, 1999.

Oberlander, J., M. O'Donnell, A. Knott & C. Mellish. Conversation in the museum: experiments in dynamic hypermedia with the intelligent labelling explorer. In *New Review of Hypermedia and Multimedia*, 4, 11-32, 1998.

O'Donnell, M. Input Specification in the WAG Sentence Generation System. In *Proceedings of the 8th International Workshop on Natural Language Generation*, 13-5, 1996.

O'Donnell, M. Intermixing multiple discourse strategies for automatic text composition. In *Revista Canaria de Estudios Ingleses* (RCEI), 40, Special Issue on Intercultural and Textual Approaches to Systemic-Functional Linguistics, 2000.

O'Donnell, M., H. Cheng & J. Hitzeman. Integrating referring and informing in NP planning. In *Proceedings of the Coling-ACL'98 Workshop on the Computational Treatment of Nominals*, 46-55, 1998.

O'Donnell, M., A. Knott, J. Oberlander & C. Mellish. Optimising text quality in generation from relational databases. In *Proceedings of the First International Natural Language Generation Conference*, 133-140, 2000.

Not, E. & M. Zancanaro. Reusing information repositories for flexibly generating adaptive presentations. In *Proceedings of IEEE International Conference on Information, Intelligence and Systems*, 1999.

Not, E. & M. Zancanaro. The MacroNode approach: mediating between adaptive and dynamic hypermedia. In *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, 2000.

Specht, M. & R. Oppermann. User modeling and adaptivity in nomadic information systems. In Mimo Caenepeel, David Benyon & Duncan Smith (eds.), *Proceedings of the i3 Annual Conference: Community of the Future*, 65-68, 1999.

Taylor, P., A. Black & R. Caley. The architecture of the Festival Speech Synthesis System. In

Proceedings of the 3rd ESCA Workshop on Speech Synthesis, 147-151, 1998.

Veron, E. & M. Levasseur. *Ethnographie de l'exposition.* Paris, Bibliothèque publique d'Information, Centre Georges Pompidou, 1983.