Natural Language and Exploration of an Information Space: the ALFresco Interactive System

Oliviero Stock*
IRST - Istituto per la Ricerca Scientifica e Tecnologica
38050 Povo, Trento, Italy
e-mail: stock@irst.it

Abstract

In traditional natural language systems, the channel of communication between the user and the system was a narrow and constraining device. In many areas natural language-based information access requires the possibility for the user of exploring the domain and a system's comfortable habitability. The present work presents a rationale for building intelligent interfaces that combine natural language and hypermedia as new means for human-computer interaction and in particular gives an outline of a prototype of this kind, built for the exploration of Italian frescoes: the ALFresco interactive system.

1 Introduction

Natural language interface design has for a long while been based on the teletype approach. The channel of communication between the user and the system was a narrow and constraining device. On the other hand, certainly not all the necessary techniques for understanding "pure" natural language, especially in the pragmatics field, seem to be advanced enough to reproduce this traditional way of communicating.

In recent times an innovative approach has emerged [Arens *et al.*, 1989], that acknowledges that a larger bandwidth communication between user and computer can be established, in particular by exploiting graphics. Intelligent multimodal systems have been developed for interfacing a user with a dynamic process such as the simulation of the operations inside a factory [Cohen *et al.*, 1989], a simulated battlefield [MacLaughlin and Shaked, 1989] or the activities of U.S. Navy ships [Arens *et al.*, 1988]. In such applications, the advantage of integrating multiple media in output is obvious, for example, to explain sequences of operations or to display the status of a complex process. Similarly for input, pointing to images on a screen may individuate the objects involved in some desired action [Wahlster, 1988; Hollan *et al.*, 1988]. Very recently some projects of multimodal information presentation have combined graphics and language in output in a new creative way [Feiner and McKeown, 1990; Wahlster *et al.*, 1991].

A different and nowadays quite popular technology is hypertext. The generalization of the idea of hypertext to multimediality gives rise to the concept of hypermedia:, "a style of building systems for information representation and management around a network of multimedia nodes connected together by typed links" [Halasz, 1988]. Hypermedial systems promote a navigational, explorative access to multimodal information: the user, browsing around the network, is at the same time both exploring the network and searching for useful information.

Multimodality and hypermediality provide the possibility of realizing intelligent interfaces that amplify capabilities we have in nature. The present work presents a rationale for building intelligent interfaces that combine natural language and hypermedia as a new means for human-computer interaction and gives an outline of a prototype of this kind, built for the exploration of Italian frescoes: the ALFresco interactive system.

2 Combining hypermedia and NLP

The main reason for trying to integrate two approaches that have up to now represented independent lines of research is that there are a number of advantages that accrue in both directions. Integrating NL with hypermedia facilities provides the following advantages: from the NL perspective: a means for organizing heterogeneous and unstructured information, for favouring the direct manipulation of all objects integrated with language, and for facilitating explorative behaviour; from the hypermedia perspective: a solution to the problem of disorientation and of the cognitive overhead of having too many links; from an unbiased point of view that looks at all this as an independent new approach: the offer of a high level of interactivity and system habitability in which each modality overcomes the constraints of the other, resulting in the whole being more than the sum of the parts.

We shall briefly discuss these points.

Many of the problems that prevent NL systems from providing useful interfaces stem from the current inability to handle the difficulties, lack of knowledge and mis-

*The various modules of ALFresco were realized by the NLP interface group at IRST. The group currently includes G. Carenini, E. Franconi, A. Lavelli, B. Magnini, F. Pianesi, M. Ponzi, C. Strapparava, and previously also F. Cecconi and V. Samek-Lodovici. Many of the ideas reported here have emerged working with them. G. Carenini, M. Ponzi and C. Strapparava have given a particular contribution to the present work.

understandings that arise during dialogs. Hypermediality has the virtue of offering a powerful means for organizing highly heterogeneous and unstructured information, a kind of knowledge not easily handled by natural language systems (and AI systems in general). For example, in situations in which an AI application has a formal representation of only a subpart of the relevant domain knowledge, it is not possible to generate natural language text about information not explicitly represented in the system. A possible solution is the dynamic generation of hypertextual (possibly hypermedial) nodes pointing to a canned hypermedia network about those areas of the domain that have not been formally represented. Direct manipulation of images, buttoned text etc., integrated with the natural language channel is a powerful concept. On the interaction strategy side the crucial point is to integrate an exploration modality in the environment. With this a user finds it easy to move around, see what is available here and there, possibly follow some exploration path, without being necessarily constrained by any definite goal.

Looking at interaction from the hypermediality point of view, one of the major problems is *disorientation*, as pointed in [Conklin, 1987]. Hypermedia offer more degrees of freedom, more dimensions in which one can move, and hence a greater potential for the user to become lost or disoriented; the user has the problem of having to know where she is and how to get to some other place that she knows (or thinks) exists in the network.

A solution is the integration of a query facility within the system, providing a way of jumping inside the network without having to follow the predefined paths through it. [Halasz, 1988] identifies two kinds of queries:

1. *content queries* allow retrieval of all objects (nodes or links) that satisfy some requirements;
2. *structure queries* allow retrieval of a subnetwork matching a given pattern.

NL can be the best way for handling such queries, if we assume that the system has some information about the knowledge presented by the different nodes and about the semantics of the structure of the network. But there is more to it than just adding queries to a hypertext system.

The hypertext research community has begun to acknowledge the importance of user modeling. [Conklin, 1987] points out that a common difficulty arising in interacting with a hypermedia system is the *cognitive overhead* caused by the number of links that may be followed from each node. It would be very useful to be able to tell which links the specific user is less likely to follow, in order not to display them all (or to display them with a lower degree of relevance). User modelling has a long history and great power in NLP, basically to represent the communicative context in which a sentence is uttered [Kass and Finin, 1988; Kobsa *et al.*, 1986]. So, for instance in generation, a text can be naturally tailored (at all levels, from the rhetorical to the lexical choices) for the intended reader, yielding an effective communicative act [Hovy, 1988; Paris, 1987].

The handling of initiative shifts and of different attitudes on the side of the user is of basic importance for the improvement of human-computer interaction. When a user encounters a system that she has never used before she is very unlikely to have a clear idea of how to formulate a problem so that the system understands it. It is also possible that the user does not have enough information about the domain to be able to produce *any* clear problem formulation. We believe that at least in typically "explorative" and "individually creative" domains, a substantial environment habitability is of the utmost importance, and does greatly benefit from a global, even if approximates user model rising through dialog. The combination of the relative freedom provided by a natural language interface (with the power of making complex and precise requests and answers) and a visual presentation (with direct manipulation possibility) of some organized subdomains has immense potential impact. And of course the user can interleave precise requests with concrete exploration of "the surroundings". This last aspect needs original studies on the Cognitive Science side. One step in this direction is [Slack and Conati, 1991].

3 A note on discovering art

Hypertextual systems for museums are presently quite common, but they provide limited possibilities. The main weakness lies in what we have emphasized before: the difficulty of expressing a precise complex request, the lack of possibility of pursuing a particular goal through conversation, etc. It is well known that even without entering in the complex and subtle world of art *understanding*, art curiosity develops as a function of two factors: a) the cumulative effect of information (especially information answering free requests by the person or giving some limited further hints, relevant to her inclination); b) fruition of an art work when the person experiencing it is prepared and anticipates it. Art curiosity grows slowly and autonomously and imposing more data or experiences than the person is prepared to cope with has no positive effect at all. Only by combining information and perception with a sensation of driving the game and enjoying circumscribed moments of mere exploration, can a user experience the possible development of an interest in art. A typical consequence might be that you take a trip just to go and see one painting that you saw before among others and did not even notice.

Now, in the design of an artificial system that provides an environment for a user interested in art, the same concepts must be exploited. Also to be noted, as far as the modality of language communication is concerned, is the fact that *written* language is acceptable even if not the most desirable solution: a user interested in this domain (as opposed to a generic user of an automatic counter at a train station) is likely to know how to use a keyboard.

4 ALFresco

ALFresco is an interactive system for a user interested in frescoes. It runs on a SUN 4 connected to a videodisc unit and a touchscreen. The particular videodisc in use includes images about Fourteenth Century Italian frescoes and relevant monuments[1]. The system, besides un-

[1]The videodisc "Italian Art History" was provided by Rizzoli SpA; we take this occasion to thank them.

derstanding and using language, shows images and combines film sequences. Images are active in that the user may refer to items by combining pointing with the use of linguistic demonstratives. The system's linguistic output includes buttons that allow the user to enter in an hypertextual modality. The dialog may cause zooming into details or changing the focus of attention onto other frescoes.

The overall aim is not only to provide information, but also to promote other masterpieces that may attract the user. ALFresco will eventually provide a proposal for a cultural tour appropriate for the particular user, through further negotiation with her.

The system is based upon domain knowledge represented in different ways:

- a KB expressed in a KLOne-like language called YAK (a.k.a. KRAPFEN) [Franconi, 1990]. YAK is a hybrid system, whose terminological component consists of a tangled hierarchy in which generic concepts and roles are defined. The assertional box consists of instances that are represented as frames (in the spirit of LOOM, [MacGregor and Bates, 1987]) connected to the terminological box by a realizer that builds in the terminological hierarchy the most specific generic concept to which the particular instance belongs. The KB is used for defining everything the system can reason about: frescoes, monuments, painters, contents of frescoes, towns etc. and provides the base for ALFresco's deductive inference capabilities;
- a NoteCards hypermedia network (see below) containing unformalized knowledge such as art critics' opinions on the paintings and their authors.

Monodirectional pointers link instances or features in the knowledge base to images and film fragments stored in the videodisc. Bidirectional pointers connect knowledge base entities to regions of images.

At this point an explanation is due. Fourteenth Century frescoes have a content that almost always is centered on a sacred scene. The scene includes an event that can be reasonably well described (for instance the event "annunciation" performed by the angel Gabriel to Mary, in which the contents of the message is another event, namely the forthcoming birth of Jesus) and includes a number of well identified recurring characters: humans, animals, saints, angels etc.

The contents that we represent are the contents of the foreground of the paintings, while nothing is said about the background where the artist could have expressed any real world scene. In this particular context the indexabilty of the objects and the related concept of the granularity of internal representation are quite easily made clear to the user at the beginning of the interaction.

NL and hypermedia are integrated both in input and output. In input, our efforts have been focussed on combining the interpretation of NL deictic references with pointing to images displayed on a touch screen. For instance, the user can ask a question such as: *"Who is this person?"* while touching the image of a person appearing in the currently displayed fresco. Both the linguistic and the graphical deixes may be ambiguous, but usually the correct integration of the information provided by the two channels allows the identification of the intended reference. In output, images and generated text with buttons are yielded that offer entry points for further hypertextual exploration. A sketch of the ALFresco system architecture is shown in Figure 1 and in the following subsections we shall give an outline of the main components of the system.

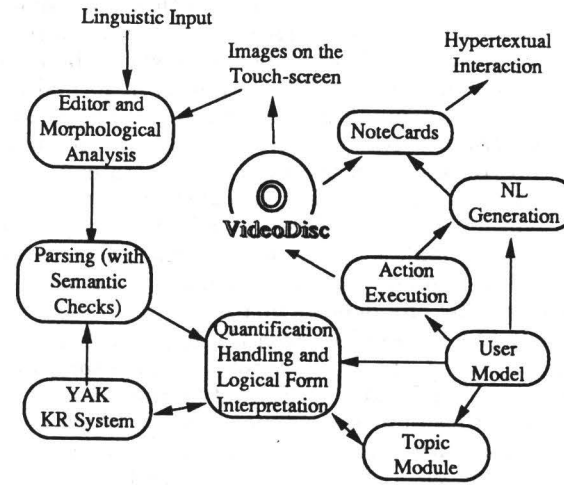

Figure 1: ALFresco architecture

An English version of a dialog fragment (taken from a real session with ALFresco) oriented to requests of information follows:

U: *I would like to know if A.L. ever painted a work that represents a sacred scene with an angel.*

S: *Yes, by example, A Lorenzetti, the Annunciation. The angel is the Arcangel Gabriel.*

U: *Who is this person?* [pointing at Mary on the touch screen]

S: *Mary.*

U: *Can you show me an enlargement containing it?*

S: [The system shows an enlargement]

U: *Was the author influenced by someone?*

S: *For example by Giotto*

U: *Where did this master see the light?* [meaning "where was he born"]

S: *Colle di Vespignano*

U: *Show me a detail with this scene.* [pointing to some place about the centre of the scene]

S: [The system shows the detail]

U: *In what town was A.L. born?*

S: *Siena*

U: *Show it to me.*

S: [the system shows a small film segment about Siena]

U: *I would like to see a fresco of the master in a monument of the town.*

S: [the system shows a film segment about a large fresco, "the Effects of Good Government", and outputs the name with a "description" button.]

At this point the user presses that button and a generated text is issued. Clicking on buttons the user can move inside an hypertext with comments written by art critics, or see related images, or explore the network. At any point it is possible to resume the dialog.

4.1 Parser

As far as sentence analysis is concerned, ALFresco is based on the WEDNESDAY 2 parser, (described in [Stock, 1989]). It is a chart-based parser able to cope with a large number of linguistic phenomena, including declarative, imperative and interrogative sentences, complex relative clauses (and other phenomena connected with so-called long distance dependencies), idiomatic expressions, various kinds of ellipses, constituent coordination, etc. One relevant capability of the parser is its ability to analyze idiomatic forms, with all the flexibility they may display in natural language. This characteristic is strictly integrated in the chart-based approach.

A morphological analyzer works on-line while the user types the sentence on the keyboard, thereby providing an immediate building of inactive lexical edges for the chart parser. The editor through which the sentence is input behaves transparently: for instance words can be deleted guaranteeing that the edges are reorganized accordingly. An important feature is that pointing to the currently displayed fresco can occur while the user is entering the sentence, in correspondence to demonstratives.

In general, the parser *per se* would output a large number of alternative interpretations of a sentence (this is particularly unfortunate for Italian, which is a freer word order language than English and in which even simple sentences can be ambiguous from a syntactic point of view). Through continuous interactions with the conceptual knowledge co mponent via a lexical semantic analysis component the possible functional relations are filtered; this permits greatly reducing the number of alternative representations of the sentence.

Another vital aspect for an interactive system that must allow good habitability and integration of modalities is the capability of treating elliptical forms. Also the solution to this problem is built into the chart approach (see [Lavelli and Stock, 1990]). The following dialogue fragment:

U: *Did Giotto portray St.Francis in a work located in Assisi?*
S: *Yes, for example in the Sermon to the birds.*
U: *Gioacchino in a fresco located in Padova?*

results in the last sentence being interpreted as *Did Giotto portray Gioacchino in a fresco in Padova?.*

4.2 Semantic interpretation

As far as semantic interpretation is concerned, multimediality made it necessary to have a layered and modular approach. In some recently developed systems a multilevel semantics approach has been proposed, in which various levels of formalization, each with its own particular functionality, correspond to successive levels of abstraction of logical-linguistic phenomena [Scha, 1983; Stallard, 1986]. A level bound to lexical semantics can be individuated (usually denoted as EFL, English-oriented formal language) and a level of meaning representation (usually denoted as WML, world model language; see Figure 2). Meaning representation is therefore as much as possible independent from the application domain. In ALFresco the interpretation modules' task is to present a meaning representation of the sentence that can be used by the various modules of the system.

The interpretative phase has several functions: in the first place it computes quantifiers' scopings. The algorithm is based on the concept of Cooper storage as in [Hobbs and Shieber, 1987] and on a) lexical classification of quantifiers; b) syntactic and surface order characteristics; c) presence of disambiguating expressions. It solves definite expressions such as determinative NP's, demonstratives (for example deriving from pointing actions), pronouns, by interacting with the topic module. It interprets the sentence dynamically, exploiting the semantics of operators, quantifiers, verbal modifiers, various levels of coordination and so on. The resulting representation is intensional, i.e. it is possible to abstract it in relation to time and context. The integration with temporal and contextual interpretation modules is currently under investigation.

```
(indef x (indef y
          (and (monument y)
           (has-place nil y (definite (town z))))
          (and (fresco x)
           (has-place nil x y)
           (made-by nil x (definite (painter w)))))
        (want nil speaker (see nil nil x)))
```

Figure 2: A WML form for *"I would like to see a fresco of the master in a monument of the town."*

In our case, without passing to a DBL form (Data Base Language in [Scha, 1983]) WML expressions are selectively and dynamically mapped into YAK's assertional language. In the process the topic module (see below) is called in.

4.3 The Topic Module and pointing

In a multimedial dialog system, the topic module must integrate global focus strategies, local focus approaches and deictic reference techniques. The main points of ALFresco's topic module [see [Samek-Lodovici and Strapparava, 1990] are the following.
ALFresco basically structures discourse in turns, with confirmations of referents from previous turns into the current one. The user normally refers to entities bound to particular frescoes. So the basic idea is to combine a) Grosz's idea of factoring the search for referents into topic-spaces [Grosz, 1977]: our topic-spaces are typically built around frescoes; and b) Hajicova's approach consisting in letting the entities that have been mentioned slowly fade away unless they are mentioned again with a certain functional role [Hajicova, 1987].

The system relies on two stacks: 1) a stack of turns, where each turn contains the referents inserted or confirmed by the user or by ALFresco , and 20 a stack of Topic Units, in which each Topic Unit contains all the referents inserted by the user or by the system while discussing a particular fresco.

As far as the deictic context is concerned, in our case it changes whenever a new fresco is displayed on the screen. Similarly to XTRA [Wahlster, 1988], [Kobsa *et al.*, 1986] we associate the accessible entities with the regions they occupy on the screen. Regions overlap significantly (e.g. the region of the "annunciation" and the region of the angel Gabriel in Giotto's Annunciation). As in CUBRICON [Neal and Shapiro, 1988], the ALFresco topic module permits both the use of linguistic input

to solve ambiguous touches and the use of touches to solve ambiguous linguistic input. More generally, both a touch and a linguistic expression may be ambiguous and yet yield a unique referent through mutual constraint.

4.4 The user's interest model

The user's interest model consists of an activation network, the nodes of which are associated with ordered sets of individual concepts. The grouping of the individuals is performed according to a measure of domain-dependent pragmatic closeness; whenever a set is activated, all the entities composing it are considered somehow relevant. Each set represents an *interest area* and is identified by particular concepts or individuals represented in YAK. The individuals that belong to an area and their relative relevance (represented by the ordering) are computed on the basis of their proximity to the identifier of the area (if it is an individual) or to its instances (if it is a terminological concept). The process that builds the interest areas interprets the A-box as a graph in which the nodes are the individual concepts and the edges connect an individual to those individuals filling one of its roles or having a role filled by it.

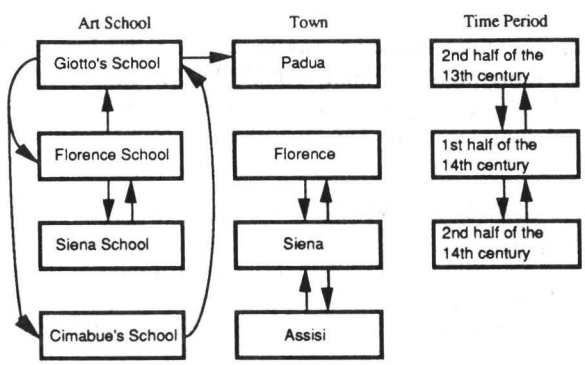

Figure 3: Some interest areas

The dynamic information represented by the model consists in the activation level associated with the interest areas. A node receives a certain amount of activation whenever one of the individuals it contains is referred to by the user or the system. The intensity of these activation impulses depends on how explicitly the individual is mentioned. After each user query, activation impulses are propagated through the network.

Figure 3 shows a simplified activation network relative to the domain of ALFresco. The areas are grouped in structures that represent interest dimensions, i.e. classes of entities that are likely to be at the centre of the user's interest. The dimensions currently used in the user model correspond to the T-box concepts defining *art schools, towns* and *time periods*. The weighted links connecting different areas can represent different kinds of pragmatic closeness: for instance, a link between two areas of the Towns dimension represents geographic closeness. "Geographic closeness" is intended as a pragmatic relation, with consequences relative for instance to the possibility that the user considers a trip from one town to another.

The generator makes use of the user's interest model. For instance, it asks the latter which is the most relevant area of a certain dimension of interest, and the model returns the main element associated with the most active area of that dimension.

4.5 Actions performed by the system

The actions performed by ALFresco are realized exploiting the media the system is based upon: a) the system can show one or more images or pieces of film, b) give some punctual answer to a question by replying with instances such as the title and location of a painting, or dates etc., or, finally, c) give a more complex description of a fresco or some other entities, through natural language. All three different media allow the user to interact with them by direct manipulation. A higher level, *ad hoc*, pragmatic component decides how to react in the given dialogic situation, considering the type of request, the context, the model of the interest of the user, the things already shown or said to the user and so on. We shall focus our attention on the language generator. There are three main aspects:
1) the semantic/pragmatic component and its interactions with the other modules of the dialog system (user model in particular); 2) the syntactic realization component; 3) the integration of NLG and hypertext techniques.

The communicative goal of the generation component, providing information about instances of the domain (i.e. frescoes), is achieved by execution of rhetorical schemata [McKeown, 1985].

The schemata's task is to select and order appropriate attributes, generic-classes, values, instances (Figure 4) from a set of possible alternatives; the process of selection is driven by both an *a priori*, strongly domain-dependent ordering between the KB items and a sequence of queries to the KB. For instance, the selection of the comparative-attribute for a fresco starts by asking the KB for other instances with a *similar* value for the *content* attribute (the first in an *a priori* ordering) and values for the other attributes subsumed by concepts stemming from the user modeling module; if the KB returns one instance, this is used as a comparative-instance and the selection process ends, if it returns more than one instance another *a priori* ordering (this time between the instances) is exploited; in both cases the content would be the comparative-attribute. When no instance is returned, the selection process iterates on other attributes.

The user modeling component provides a relevance criterion for operating selections between the individual concepts described in the assertional box of the knowledge representation system. After this phase of determination of the rethorical schemata and of the particular contents, the syntactic realization component is called in.The algorithm works on a unification based formalism in a bottom-up, head driven fashion [Pianesi, 1991] and yields complex natural language sentences.

4.6 Integrating NL Generation and hypertextual techniques

The system's answer is a hypermedia node containing a (possibly generated) text with links to both the hyper-

media network and images from the videodisc. In particular, the output of the NL generation component is passed to the hypermedia module which transforms the plain text into a hypertextual entry point to the underlying hypermedia network. Buttons pointing to hypermedial information are associated to individual contents selected by rethorical schemata.

The hypermedia module is based on the NoteCards system, a general hypermedia environment developed at Xerox PARC. We chose Notecards for its powerful programmer's interface to a Lisp environment and general tailorability. The system provides an extremely flexible environment in which it is possible to integrate many of the different kinds of interaction that can take place in the exploration of a complex set of knowledge.The primitive constructs of the system are *cards* and *links*:

- Cards are usually displayed on the screen as standard Xerox Lisp windows containing a piece of text. NoteCards allows for the creation of new card types.
- Links are used to interconnect individual cards into networks; they are displayed as icons inside the substance of a card. Clicking the mouse in such an icon retrieves the destination card and executes the associated actions (usually the card is displayed on the screen).

The tailorability of NoteCards has been exploited in ALFresco in the implementation of virtual (dynamically constructed) links that connect text cards to images and fragments of films stored in the videodisc.

ALFresco's generated text appears as a card providing the user with a dynamic entry-point both to images from the videodisc and to a static base of hypertextual information. Figure 4 shows the card produced in relation to the last point in the dialog of section 4. The canned texts accessible through buttons deal with particularly complex topics (such as comparisons between styles as they are elaborated by art critics) definitely outside the expressive scope of current KR systems.

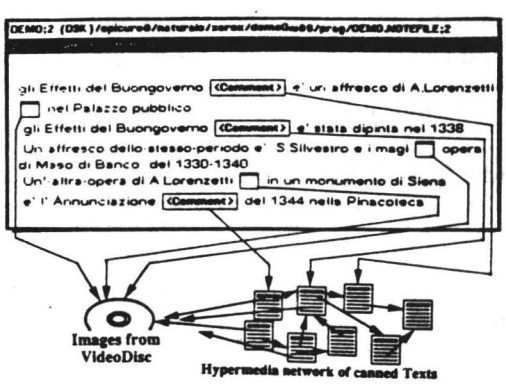

Figure 4: *"The Effects of Good Government is a fresco by A.L. in the Public Palace. The Effects of Good Government was painted in 1338. A fresco from the same period is S. Silvestro and the Holy Kings by Maso di Banco, painted in 1330-1340. Another work by A.L. in a monument of Siena is the Annunciation, of 1344 in the Pinacoteca".*

Presenting the user with a mix of generated and canned text in general can be misleading; in fact the unaware user might: a) make references to canned text (the system is unable to understand) and b) overestimate the system's capabilities. The use of graphic containers overcomes this problem making clear the difference between the two kinds of text: canned text has to be presented in an evocative graphical format, such as a simulated open book.

Another important aspect is the *Browser construct.* Browsers are constructs containing a structural diagram of a network of cards with different types of links visualized in different dashing styles. Browser cards can be automatically computed by the system and, when generated, have an active behaviour: i.e. it is possible to click on a node of the displayed network in order to access the corresponding card. The user can refer to this overall structure in her exploration.

Two general problems, connected with the rationale of section 2 are worth discussing:

1. User modeling connected to linguistic communication may suggest the possibility of applying the same model to the control of hypertextual communication. Anyway, it is not clear whether a simple user model (such as that of ALFresco) could really improve the efficiency of hypertextual interaction: it might be the case that only very sophisticated (and yet not understood) modeling techniques can help hypertext exploration when another powerful means for focusing a request (i.e. natural language) is available in the same environment.
2. Would it be advantageous that the model of interest (to be used in the language-centered modality) is changed as the user browses through the network? We feel that given a) the limited accuracy of current user models and b) the fact that hypertext modality on its own would only be used in a very limited way to explore the surroundings or have a general bird eye's view of the domain, it may be wrong to constrain the behaviour of the whole system in that way. It seems to us better not to modify the user model through the hypertext modality thereby minimizing any consequences for the natural language-centered interaction.

5 Conclusions

We have discussed the combination of natural language dialog and hypermediality within an artificial intelligence view of information exploration. The habitability of a system that provides an active integration of the two paradigms is greatly enhanced and a number of interesting issues that amplify more traditional possibilities arise. The exploratory dialog system ALFresco has been presented: an NL and hypermedial system connected to a videodisc that gives information about Italian frescoes of the Fourteenth Century. The system as described here is implemented in Xerox Common Lisp and runs on a Sun 4 connected to a videodisc unit.

Many points are to be explored: among the extensions we intend to pursue the idea that the system can in the end provide through further negotiation a suggestion for a personal cultural-touristic itinerary. We believe that this area is of great potential because cultural tourism is increasing material relevance, especially in Italy, but also

because it is an area that requires a shift from a mass-oriented, impersonal perspective toward an individual-oriented, creative opportunity for all. In general, we think that the integration of hypermediality with AI and NLP technology opens a wide range of new perspectives for the development of intelligent interfaces and that it will ultimately lead to an innovative paradigm for man-machine interaction.

Acknowledgements

We would like to thank Jon Slack for many helpful discussions and Luigi Stringa for encouraging an integrated view of AI at IRST.

References

[Arens *et al.*, 1988] Y. Arens, L. Miller, S.C. Shapiro, and N.K. Sondheimer. Automatic construction of user interface displays. In *Proc. AAAI-88 Conference*, St. Paul, Minnesota, 1988.

[Arens *et al.*, 1989] Y. Arens, S. Feiner, J. Hollan, and R. Neches. *Proc. of the IJCAI Workshop on a New Generation of Intelligent Interfaces.* Detroit, 1989.

[Cohen *et al.*, 1989] P.R. Cohen, M. Dalrymple, D.B. Moran, F.C.N. Pereira, , J.W. Sullivan, R.A. Gargan Jr, J.L. Schlossberg, and S.W. Tyler. Synergetic use of direct manipulation and natural language. In *Proc. of CHI'89 Conference*, Austin, Texas, 1989.

[Conklin, 1987] J. Conklin. Hypertext: An introduction and survey. *IEEE Computer*, 20(9), 1987.

[Feiner and McKeown, 1990] S. Feiner and K. McKeown. Coordinating text and graphics in explanation generation. In *Proc. AAAI-90 Conference*, Boston, Mass., 1990.

[Franconi, 1990] E. Franconi. The YAK manual: Yet Another KRAPFEN. Technical Report 9003-01, IRST, Trento, Italy, 1990.

[Grosz, 1977] B. Grosz. The representation and use of focus in dialogue understanding. Technical Report 151, SRI Project 5844, SRI, Menlo Park, CA, 1977.

[Hajicova, 1987] E. Hajicova. Focusing: a meeting point of linguistics and artificial intelligence. In P. Jorrand and V. Sgurev, editors, *Artificial Intelligence II: Methodology, Systems, Applications.* Elsevier Science Publishers, 1987.

[Halasz, 1988] F.G. Halasz. Reflections on NoteCards, seven issues for the next generation of hypermedia systems. *Communications of the ACM*, 13(7), 1988.

[Hobbs and Shieber, 1987] J.R. Hobbs and S.M. Shieber. An algorithm for generating quantifier scopings. *Computational Linguistics*, 13(1-2), 1987.

[Hollan *et al.*, 1988] J. Hollan, E. Rich, W. Hill, D. Wroblenski, W. Wilker, K. Wittenburg, and J. Grudin. An introduction to HITS: Human Interface Tool Suite. Technical Report ACA-HI-406-88, MCC, Austin, Texas, 1988.

[Hovy, 1988] E. Hovy. *Generating Natural Language under Pragmatic Constraints.* Erlbaum Lawrence Associates, 1988.

[Kass and Finin, 1988] R. Kass and T. Finin. Modeling the user in natural language systems. *Computational Linguistics*, 14(13), 1988.

[Kobsa *et al.*, 1986] A. Kobsa, J. Allagyer, C. Reddig, N. Reithinger, D. Schmauks, K. Harbusch, and W. Wahlster. Combining deictic gestures and natural language for referent identification. In *Proc. COLING 86*, Bonn, Germany, 1986.

[Lavelli and Stock, 1990] A. Lavelli and O. Stock. When something is missing: Ellipsis, coordination and the chart. In *Proc. COLING 90*, Helsinki, Finland, 1990.

[MacGregor and Bates, 1987] R.M. MacGregor and L. Bates. The LOOM Knowledge Representation Language. Technical Report ISI/RS-87-188, USC/Information Sciences Institute, Marina del Rey, CA, 1987.

[MacLaughlin and Shaked, 1989] D.M. MacLaughlin and V. Shaked. Natural language text generation in semi-automated forces. Technical Report 7092, BBN, Cambridge, Mass., 1989.

[McKeown, 1985] K.R. McKeown. *Text Generation.* Cambridge University Press, 1985.

[Neal and Shapiro, 1988] J.G. Neal and S.C. Shapiro. Intelligent multimedia interface technology. In *Proc. of the Workshop on Architectures for Intelligent Interface Technology*, Monterey, CA, 1988.

[Paris, 1987] C.L. Paris. Combining discourse strategies to generate descriptions to users along a naive/expert spectrum. In *Proc. of IJCAI 87*, Milano, Italy, 1987.

[Pianesi, 1991] F. Pianesi. X-bar theory and deep structure: a tabular bottom-up generator. Technical Report 9104-11, IRST, Trento, Italy, 1991.

[Samek-Lodovici and Strapparava, 1990] V. Samek-Lodovici and C. Strapparava. Identifying noun phrase references, the topic module of the ALFRESCO system. In *Proc. ECAI 90*, Stockholm, Sweden, 1990.

[Scha, 1983] R. Scha. Logic foundation for question answering. Technical Report MS 12.331, Philips Research Labs, Einhoven, The Netherlands, 1983.

[Slack and Conati, 1991] J.M. Slack and C. Conati. Modeling interest: exploration of an information space. Technical Report 9101-03, IRST, Trento, Italy, 1991.

[Stallard, 1986] D. Stallard. A terminological simplification transformation for nl question answering system. In *Proc. of the 24th Meeting of the ACL*, New York, 1986.

[Stock, 1989] O. Stock. Parsing with flexibility, dynamic strategies and idioms in mind. *Computational Linguistics*, 15(1), 1989.

[Wahlster *et al.*, 1991] W. Wahlster, E. Andre', W. Graf, and T. Rist. Designing illustrated texts: How language production is influenced by graphics generation. In *Proc. 5th Conf. of the European Chapter of ACL*, Berlin, 1991.

[Wahlster, 1988] W. Wahlster. User and discourse models for multimodal communication. In J.W. Sullivan and S.W. Tyler, editors, *Architectures for Intelligent Interfaces: Elements and Prototypes.* Addison-Wesley, 1988.