# Combining Natural Language and Hypermedia as New Means for Information Access

Giuseppe Carenini, Marco Ponzi, Oliviero Stock
Istituto per la Ricerca Scientifica e Tecnologica
38050 Povo, Trento, Italy
e-mail: stock@irst.it

## 1. Introduction

Artificial intelligence approaches to information access through natural language (NL) have yielded a lot of insight to the issue and a number of specific techniques; yet, so far they have not been sufficiently creative in seeing in the computer an agent with a large bandwidth of communication with the human user. In our view, natural language can become a fundamental paradigm of communication between human and machine, provided it is integrated with other channels, some of which might also be unavailable in human-to-human communication. Just to make the claim more concrete with an example: if we want to refer in a dialog to something mentioned before we would use a pronoun or a definite noun phrase, causing a difficult and nondeterministic process in the hearer (and probably in the speaker too); if we have the references available on the screen, we could explicitly refer to them through pointing. We believe that in the long run a new modality of natural language will emerge, after the spoken and the written one: the perspectives we are exploring for information access are just a window into that.

In this paper we shall begin with discussing multimodality and hypermediality and how they fit into an artificial intelligence view of information access. After that, we are going to analyze some issues in the integration of images with natural language and in the production of language, combined with hypermedial capacity as a means to provide information to the user in a flexible, individual and rich way.

All these ideas have emerged while developing a prototype of a dialog system connected to a videodisc that gives information about Italian frescoes of the Fourteenth Century. In the paper we shall refer to this work. The system, beside using language, shows images of frescoes and combines film sequences. Images are active so that the user may refer to items by pointing, combined with the use of linguistic demonstratives. The dialog may cause zooming into details or changing the focus of attention into other frescoes. The aim is not only to provide information. The system also has the goal of promoting other masterpieces: in the art field the potential experiencer is likely to build an interest when exposed to images in a coherent way.

## 2. Multimediality and hypermediality

Recent advances in digital video and audio real-time processing suggest that we will soon face a new revolution in the design of interactive systems (Fox, 1989).

The first effect of these advances is the feasability of the implementation of multimodal systems, i.e. of systems that use traditionally independent communication media (such as pictures, animation, sound, speech, printed text etc.) in an integrated way.

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). Multimodality and hypermediality, more than other paradigms, put an emphasis on the possibility of realizing an intelligent interface that amplifies capabilities we have in nature (as opposed to trying to reproduce them). In fact, while it is true that people, in face-to-face communication, not only use spoken language but also take into account gestures, the visual environment and more, it is not so easy for people to reproduce or create images, and especially to make an integrated active medium out of the combination of images and language.

Intelligent multimodal systems are often used in interfacing a user with a dynamic process such as the simulation of the operations inside a factory (Cohen et al., 1989), a simulated battlefield (McLaughlin and Shaked, 1989) or the activities of U.S. Navy ships (Arens,1989). In such applications, the advantage of integrating multiple media in output is obvious, for instance, 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.

Hypermediality has recently opened interesting perspectives on the problem of accessing loosely structured information. Such 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. An hypermedia system can be used in different ways for different goals, just like a book (e.g. a manual) that can be read in a completely sequential way, quickly leafed through in order to get an idea of its content, or scanned in search of particular information.

## 3. How hypermedia and NLP can help each other

Many of the problems that prevent NL systems from providing useful interfaces stem from the current inability to handle the difficulties and misunderstandings that arise during dialogs. The study of actual human interactions suggests that the various forms of human dialogic communication can be seen as modifications to a single conversation's basic structure. The organization of conversation depends in essential ways upon the material and social circumstances of human interaction, i.e. upon those features that characterize dialog as an example of situated action. Suchman (1987) underlines the importance of local control for the success of face-to-face human interaction: we can remedy the uncountable troubles that occur during communication not because we can predict all the possible problems, but because the organization of conversation provides a very powerful mechanism for exploiting the contextual information shared by the people taking part to a dialog. Multimediality offers a way of enlarging the share of the interactional situation that is available both to the user and the system, allowing for a better understanding between them. If both the participants can refer to a rich range of contextual knowledge, having the opportunity of shifting initiative in order to signal and remedy interactional problems, it should be possible to achieve a satisfactory simulation of the recover mechanisms of human interaction. For instance, if the user fails in understanding a natural language description identifying a character represented in a picture, the system might generate a kind of deictic action such as a graphical highlightening of the area of the picture representing the character.

An example of synergetic use of language and other media is deictic pointing to graphical representations of objects by means of a mouse or a touch screen. Pointing to objects in the world is one of the many interactional devices that we use in face-to-face dialog and which are completely unavailable when communicating with a machine by a single communication channel.

Hypermediality provides a powerful mean for the organization of highly heterogeneous and unstructured information. This kind of knowledge is not easily handled by natural language systems (and AI systems in general). 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 could be the dynamic generation of hypertextual (maybe hypermedial) nodes pointing to a canned hypermedia network about those areas of the domain which have not been formally represented.

We now turn to discuss how NLP can solve a major hypermedia problem. One of the problems of hypermediality underlined by Conklin (1987) is disorientation: 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 s/he is and how to get to some other place that s/he knows (or thinks) exists in the network.

One of the proposed solutions to the second aspect of the disorientation problem is the integration of a query facility within the system. Such a facility should provide a way of jumping inside the network without having to follow the predefined paths through it.

Halasz (1988) underlines that it is possible to identify two different types of queries about an hypermedia network:

1) content queries allow to retrieve all objects (nodes or links) that satisfy some requirements. All objects in the network "are considered as independent entities and are examined individually for a match to a given query": content search ignores the structure of the network.

2) structure queries allow to retrieve a subnetwork matching a given pattern

NL can be a useful tool 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.

### 3.1. Towards a deeper complementarity

Both hypermedia technology and NLP are interesting paradigms for the development of intelligent interfaces of the future. Beyond the constructive local interactions just discussed, the two fields are starting to share common techniques and methodological foundations.

The necessity of an experimental and psychologically grounded approach is asserted in a large part of the literature about hypermedial systems. For instance, Nielsen (1990) describes the techniques adopted in the design of a HyperCard application and the research results extracted from the usability tests performed during the development phase. It would certainly be very useful if similar approaches were more common in the NL community. There are few exceptions to the general tendency of ignoring the experimental issues in the design of AI systems. In the field of NLP, one interesting possibility is the study of the practical consequences of Suchman's remarks about the relevance of conversational analysis for human-computer interaction. Frohlich and Luff (1989) have build an experimental system based on these ideas. Oviatt and Cohen (1989) have addressed the problem of evaluating the effectiveness of different forms of spoken and written linguistic communication and discussed the implications for the design of NL systems.

Many of the problems addressed in the research on the development of intelligent interfaces are connected to the need of tailoring the behaviour of the system to the characteristics of the particular user. User modeling is a fundamental area of research in dialog processing. The relevance of the field for NL understanding is connected to the need of representing the communicative context in which a sentence is uttered (Kass and Finin, 1988). In the field of NL generation, from content selection to lexical selection, from the choice of particular rhetorical structures to the choice of particular syntactic structure, a user model is a powerful knowledge source that drives the generation process. The generated text, tailored on the intended reader, is a more effective communicative act (Hovy 1987).

The hypertext research community begins 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 it is possible to follow from each node. He then suggests a number of possible solutions based on the clever use of various multimodal techniques (e.g. a one to three line pop-up explanation of the side reference, or a graphical browser which shows the subnetwork into which a link leads). 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 at all (or to display them with a lower degree of relevance). The development of powerful user modeling techniques could probably cause an impressing increase in the applicative usefulness of hypermedia systems. For instance, Fox (1989) claims that the coupling of multimediality and an accurate user model might dramatically improve the bandwidth and effectiveness of intelligent tutoring systems.

We believe that there is an important aspect that must not be underestimated in NLP sytems: the problem of providing to the user a model of the data available and possibly of the kind of queries that the user can eventually propose to the system. Most NL systems were proposing a paradigm for which NL was a front end to a DB. If the user was made familiar with the contents of the DB, the problem became unnaturaly limited.This is the case for instance of administrative DB's. The type of contents is generally well known to people that eventually had the possibility of using a NL front end through which they could propose the questions they would have had to express in SQL or similar query languages. If, on the other hand, one has a menu of choices that provide in every point the possible values of parameters in a request , the model is too restrict to accomodate a more personal view of the domain, and, obviously, the expressive power of requests is necessarily very limited.

The combination of the relative freedom provided by a natural language interface and a graphical presentation of some organized subdomains may be of potential impact. The suggestive power of a graph, or also of explicit promises of more information in the form of buttons may provide a bias toward guiding the view of the system and of the information it can provide, that the user is to adopt. And of course the user before making a precise request can explore “the surroundings”.

This particular aspect needs new studies on the cognitive side. See also [Slack and Conati in this volume].

## 4. First steps towards a new form of interface

The natural language processing community, naturally interested in applications in the information access area, so far has given too little attention to innovative situations in more complex computer-human interaction. Access to information within the NLP community has often been considered from two points of view:

- the data base tradition brought into an intelligent question answering system. In applications of this kind , there can be quite sophisticated input, but there is no specific user model and discourse management is accordingly limited;
- the dialog oriented approach, usually employed for expert systems, takes into account intention recognition, and more generally a user model in the context of some limited task.

We are building a hypermedial system with the goal of integrating these two AI approaches to NL interaction by dynamically switching from information seeking subdialogs to the role of an expert adviser. Kass and Finin (1988) point out that these two situations embody two symmetric aspects of natural communication: in the first case the user always has the initiative and s/he is responsible for the successful progress of the interaction; in the second the initiative will be handled by the system which will have to build a sophisticated user model in order to avoid possible misunderstandings. In our case, the user will have the responsibility for communication when exploring the domain, i.e. when asking for frescoes and other works of art with certain characteristics. The system will observe the behaviour of the user in order to build a model of his/her interest and of his/her degree of expertise in Italian art. It should take the initiative when it detects a user's goal which requires some kind of expert reasoning or when it detects a user's misconception.

Research on the design of interactive systems capable of extensively handling initiative shifts is of basic importance for the improvement of human-computer interaction. A system that assumes sole responsibility for communication will not give to the user the opportunity of interrupting it and try to recover from any trouble that could have arisen (Kass and Finin 1988). On the other hand, if the system just answers the queries of the user, without sharing the responsibility for communication, it will not even need a sophisticated user model and it will never attempt to solve any problem unless explicitly requested.

The idea of exploration dialogs fits well in this framework: if a user is facing a system that s/he has never used before it is very unlikely that s/he might have a clear idea of which formulation of his/her problem can be understood by the system. It is also possible that the user does not have enough information about the domain to be able to produce any clear formulation of his/her problem. The situation of the system is more or less symmetric: it does not know what the user is interested in and, if it knew, it does not know how to present relevant information in an optimal way, since it cannot yet consult any user model. So the interaction will initially be handled on a local basis; during this phase each participant will build a model of the other and will start setting up the goals that it/he/she will pursue during the progress of the dialog.

## 4.1 The AL Fresco System

In order to address the complexity of the interaction just outlined, our system has grown into a workbench where to integrate and test different communicative paradigms and media.

The AL Fresco system is an interface to information about Italian art of the 14th century. Such knowledge consists of:

- a KB expressed in a KLONE-like language called KRAPFEN (Franconi, 1990). It represents knowledge about features of the frescoes, painters, monuments etc.
- a NoteCards hypermedia network (see below) containing unformalized knowledge such as critical opinions on the paintings and their authors;
- a set of images and films stored in a videodisc.

We have considered the integration of NL and hypermedia 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 (Samek and Strapparava, 1990).

For instance the user can ask a question such as: "Who is this person?" while touching the image of a person appearing in the tcurrently 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.

NL and hypermediality are coupled in the output phase as well. The system's answer is an hypermedia node containing a (possibly generated) text with links to both the hypermedia 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 an hypertextual entry point to the underlying hypermedia network. Rethorical schemata promote the introduction of individual contents for which, in that situation, it is sensible to associate buttons pointing to hypermedial information.

The hypermedia module is based on the NoteCards system. NoteCards is a general hypermedia environment developed at Xerox PARC. It was originally designed as "an authoring system providing its users with facilities for creating and modifying hypermedia structures", but its powerful programmer's interface makes it easy to integrate Lisp programs into the NoteCards environment and/or to integrate NoteCards into another Lisp-based environment (Halasz, 1988). 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).

Our system presents generated text as a card, providing the user a dynamic entry-point to the static base of hypertextual information. Presenting the user a mix of generated and canned text can be misleading. In fact the user could: (a) make references to canned text (the system is unable to understand); (b) overestimate the system's capabilities. The use of graphics overcomes this problem, making clear the difference between the two kinds of text. Canned text can be presented in a radically different graphical format (e.g. a simulated open book). Also the integration between the NL generation component and the videodisc is achieved through hypertext technology. We have defined a new card type such that opening a card of that type causes the execution of a videodisc action: in order to access images or film sequences relative to an instance mentioned by the system, the user only needs to click on the link icon associated to its name.

## 5. Conclusions and future work

We have discussed multimodality and hypermediality within an artificial intelligence view of information access. Some issues in the integration of images with natural language and in the production of language, combined with hypermedial capacity, have then been analyzed. We have presented the experimental dialog system AL Fresco: an NL system connected to a videodisc that gives information about Italian frescoes of the Fourteenth Century, under development at IRST.

Up to now AL Fresco has been used as a workbench for the implementation of a bunch of ideas concerning the integration between NL and other communication modalities; in the near future we shall test the usefulness of such ideas for the design of systems that should operate in applied domains. In this perspective, psychologically founded analysis of explorative behaviour of users of an information retrieval system will become essential. Slack and Conati (1990) present some preliminary results. We think that the interaction of hypermediality with AI 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.

## References

Arens, Y., Miller, L., Shapiro, S.C., Sondheimer, N.K., “Automatic Construction of User Interface Displays”, Proc. of the 7th AAAI Conference, St. Paul, Minnesota, August 1988

Cohen, P.R. et al., "Synergetic Use of Direct Manipulation and Natural Language", Proc. CHI 89, Austin, Texas, May 1989

Slack, J., Conati, C., “A Framework for Exploring Exploration”, Proc. ECCE5, Urbino, 1990

Conklin, J., “Hypertext: An Introduction and Survey”, IEEE Computer, 20, 9, 1987

Fox, E.A., “The Coming Revolution in Interactive Digital Video”, Communications of the ACM, Volume 32, n. 7, July 1989

Franconi, E., “The YAK Manual: Yet Another KRAPFEN”, IRST Manual n. 9003-01, Trento, Italy, 1990

Frohlich, D.M., Luff, P., “Conversational Resources for Situated Action”, Proc. CHI 89, Austin, Texas, May 1989

Halasz, F.G., “Reflections on NoteCards, Seven Issues for the Next Generation of Hypermedia Systems”, Communications of the ACM, Volume 31, n. 7, July 1988

Hovy, E., “Generating Natural Language under Pragmatic Constraints”, PhD Thesis, Yale University, March 1987

Kass, R., and Finin, T. ,"Modeling the User in Natural Language Systems", Computational Linguistics, Volume 14, n. 3, September 1988

MacLaughlin, D.M., Shaked, V., “Natural Language Text Generation in Semi-Automated Forces”, BBN Report n. 7092, July 1989

Nielsen, J., "The Art of Navigating through Hypertext", Communications of the ACM, Volume 33, n. 3, July 1990

Samek-Lodovici, V., Strapparava, C., “Identifying Noun Phrase References, the Topic Module of the AL Fresco System”, Proc. ECAI 90, Stockholm, Sweden, 1990

Suchman, L.A., "Plans and Situated Action", Cambridge University Press, 1987