ALFRESCO: Enjoying The Combination of NLP and Hypermedia for Information Exploration

Oliviero Stock and the ALFRESCO Project Team*
IRST - Istituto per la Ricerca Scientifica e Tecnologica
38050 Povo, Trento, Italy

Integration of natural language with other communicative modalities has a number of different motivations. One is that there are things that in absolute terms, "in nature", humans best communicate and understand if transmitted non linguistically. This is the case for the analogical information we give when we transmit spatial information through maps, or through the act of pointing for indicating a direction. It would be nice to reproduce automatically our capacity for exploiting and choosing among different modalities, including providing graphics (maps, diagrams, etc.).
Another motivation is that the research community has recognized that the presently unsolved problems in natural language processing (NLP) hinder delivering a sufficiently sophisticated system that can be realistically offered as the basis for complex real-world interfaces with computers. We should then integrate NLP with other well developed techniques, so that problems for the language processing part can be circumscribed and solved in practical terms, while technology helps for the rest. In a way this happens with various kinds of menu driven systems integrated with NLP.
Yet another motivation is that there is a potentiality in computers that goes beyond communication as we know it in human-human interaction, but that still seems to be cognitively realisable; this potentiality relies on the technological possibility of moving in a rich information space with rapid interleaving of different information and different media, and on the high-level integration of the different communication modalities. The role of language seems to be central even here but in a novel way (see also [Stock, 1992]).
All these different motivations share the view that the so called teletype approach is too narrow for NLP applications and envision a larger bandwidth of communication.
In recent years there have been various projects sharing all or part of the above motivations and the chapters in the present volume give an excellent overview of the problems in designing a multimedial integrated system and of the proposed solutions.
A central problem concerns the integration, in the system output, of text generation with forms of graphics. For instance the combination of analogical images and text has been developed in several systems, notably in Comet [Feiner and McKeown, this volume], where the role of a coordinator that has to negotiate with the different mode specialists is emphasized. Particularly interesting is the case in which instructions are given so that temporal relations among actions are specified in a multimedial format [Feiner et al., this

* The current or previous members of the ALFRESCO Project Team are: G. Carenini, F. Cecconi, E. Franconi, A. Lavelli, B. Magnini, F. Pianesi, M. Ponzi, V. Samek-Lodovici, C. Strapparava.

volume]
In the WIP system [André et al., this volume], concerned with the presentation of information for the use of a technical device, the multimodal system starts from a formal representation of the system's goal (for instance the mutual belief between system and audience that the system has the goal of a certain action being performed by the audience) as defined by the back-end application. This is common to other generation systems, but here the main question is then how best to decompose this goal into subgoals to be realized by the mode-specific generators [André and Rist, this volume], so that they can complement each other. The interesting consequence is that semantic and pragmatic concepts such as coherence, focus, communicative act, implicature, discourse model, reference etc. acquire an extended meaning. An analogical representation of the technical device is also produced from an internal format. Another relevant aspect of the project concerns the capacity of incremental generation, i.e. of starting to produce output even when not all the information to be conveyed is yet available.
One of the main problems shared by different researchers is how to represent the actions that the system may perform in a coherent and flexible way so that the overall behaviour of the system can be built to realize the goals.
[Maybury, this volume] offers a very interesting proposal for starting the process of presentation from a communicative act perspective. What is proposed is an extension of the classical, linguistic oriented representation to include visual acts, and the integration of rethorical acts. This is then used by a planning system that organizes the overall multimedia presentation.
The potential of the perspective described above is very high for applications to the design of training systems [Goodman, this volume]. In particular the coherent integration of visual dynamics (videoclips, animation) becomes crucial.

Joint planning of text and diagrams has been addressed also by using presentation rules that hold for several media at once [Arens and Hovy, 1990]. Rules are generalized to take into account the system's communicative goals, a model of the reader, features characterizing the information to be displayed and features characterizing the media available to the system. In [Arens et al., this volume] the parameters that influence what media to use and how to combine them are analyzed in detail.

All these aspects are of central importance for the planning of multimedial presentation. Other work puts the emphasis on the potential bidirectionality of interaction: for instance pointing to images on a screen may individuate the objects involved in some direct action.
Systems of this type 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]

Still other research has been carried out, with the aim of putting intelligence and dynamicity into a "static" multimedial interaction pardigm [see, for instance, Woolf, Suthers andCornell, this volume], sometimes specifically inspired by research developed so far in the NLP area [Burger and Marshall, this volume].

The present work has a slightly different emphasis. It is about combining NLP with "hypermedia". Hypermedia is the generalization of the idea of hypertext to multimediality or as Halasz put it, "a style of building systems for information representation and management around a network of multimedia nodes connected together by typed links" [Halasz, 1988]. Hypermediality has opened interesting perspectives on the problem of accessing loosely structured information. 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. We think that often the user wants exactly this double capability from the system, of dealing with goal-oriented behaviour, typical of language and exploration-oriented behaviour. The most ambitious motivation of this work is of the third type specified at the beginning (amplifying natural communication potential). In addition we propose a view that makes these concepts of practical use.

This paper describes in particular the ALFRESCO interactive system [Stock, 1991], a prototype built for accessing images and information about Italian Fourteenth Century frescoes. We believe that the approach taken here is of particular interest for access to cultural information; an area where the shift from a mass-oriented to an individual-oriented creative offering, in which the user can "find his way", is really needed.

2. Hypermedia and NLP combined

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 to both:

(a) 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;

(b) From the hypermedia perspective: a solution to the problems of disorientation and of the cognitive overhead of having too many links.

(c) From an unbiased point of view that looks at all this as an independent 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 inhability to handle the difficulties, lack of knowledge and misunderstandings 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. 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 out in Conklin [1987]. Hypermedia offers more freedom, more dimensions in which to move, and hence a greater potential for the user to become lost or disoriented; the user problem is 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 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 an 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, so as to display them all (or to display them with a lower degree of relevance). User modelling has a longer history and great power in NLP, basically representing the communicative context in which a sentence is uttered [Kass & Finin, 1988; Kobsa, 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, 1987, 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 global environment habitability is of the utmost importance, and does greatly benefit from a global, even if approximate user model arising 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 side (one step in this direction is [Slack and Conati, 1991]), while as far as work on the evaluation of the effectiveness of different modalities is concerned, the two fields are starting to address some of the related cognitive and methodological problems (see for instance [Nielsen, 1990] on the hypermedia side and [Oviatt and Cohen, 1989] and [Whittaker and Stenton, 1989] on the NLP side).

It may be useful to add here some notes on discovering art and the role of computer-based systems. Hypertextual systems for museums are presently quite common, but they provide a very limited possibility. 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 experiencer 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 you did not even notice. Or else, in Italy, you learn of a minor fresco in a church in a remote village and go there because you know you will find pleasure in seeing it and, indeed, in "discovering" it.
Now, in the design of an artificial system that provides an environment for a user interested in art, the same concepts must be exploited. An interesting fact is that, as far as the modality of language communication is concerned, *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 station) is likely to know how to use a keyboard.

3. ALFRESCO

ALFRESCO is an interactive system for a user interested in frescoes. It is implemented in Common Lisp and Interlisp and 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 monuments*.
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

* The videodisc "Italian Art History" was provided by Rizzoli; we take this occasion to thank them.

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, the actions performed by ALFRESCO are realized exploiting the media the system is based upon: a) the system can show images , 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 [see Carenini et al 1992]. 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 user's interest, the things already shown or said to the user and so on. The system's linguistic output includes buttons that allow the user to enter 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, though not presently part of it, a proposal for a cultural tour appropriate for the particular user, through further negotiation with her.

A dialog fragment (taken from a real session with ALFRESCO) oriented to requests of information follows:

U: Vorrei sapere se Ambrogio Lorenzetti ha dipinto un'opera che raffigura una scena sacra con un angelo.
I would like to know if A.L. ever painted a work that represents a sacred scene with an angel.

S: Sì. Per esempio: A. Lorenzetti , l'Annunciazione. L'angelo é l'Arcangelo Gabriele.
Yes, for example, A Lorenzetti, the annunciation. The angel is the Arcangel Gabriel.

U: Chi è questa persona? [pointing at Mary on the touch screen]
Who is this person?

S: La Madonna.
Mary.

U: Puoi mostrarmi un ingrandimento che la contiene?
Can you show me an enlargement containing her?

S: [The system shows an enlargement]

U: C'è un pittore che ha influenzato l'autore?
Was the author influenced by some other painter?

S: Sì. Per esempio da Giotto.
Yes. For example by Giotto

U: Dove vide la luce questo maestro?
Where did this master see the light? [meaning "where was he born"]

S: Colle di Vespignano.

U: C'è un'opera di Ambrogio Lorenzetti a Trento?
Is there a work by A.L. in Trento?

S: No.

U: Un affresco di Giotto a Assisi?
A fresco by Giotto in Assisi?

S Sì. Per esempio la Predica agli uccelli.
Yes. For example the sermon to the birds

U: Mostrami un dettaglio con questa scena. [pointing to some place about the centre of the scene]
Show me a detail with this scene.

S: [The system shows the detail]

U: In quale città è nato Ambrogio Lorenzetti?
In what town was A.L. born?

S: Siena.

U: Mostramela.
Show it to me.

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

U: Vorrei vedere un affresco del maestro in un monumento della città.
I would like to see a fresco of the master in a monument of the town.

S: [the system shows a video segment about a large fresco, “the Effects of Good Governement,” 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.

A sketch of the ALFRESCO system architecture is shown in Figure 1 and in the following we shall give an outline of some components of the system.

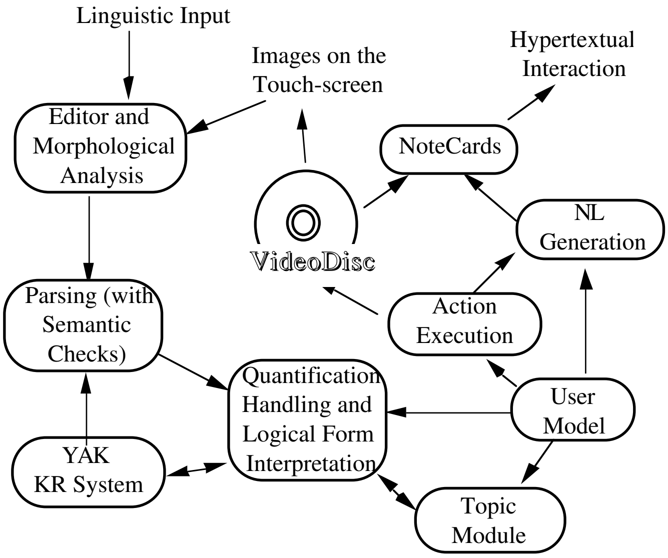

Figure 1

3.1 Knowledge Representation in AlFresco

The underlying Knowledge Base in AlFresco is expressed in YAK (Franconi, 1990), a descendant of Kl-One (Brachman, 1985). YAK is a hybrid system, whose terminological component (T-box) consists of a tangled hierarchy in which generic concepts and attributes (roles) are defined. The assertional component (A-box) consists of instances that are represented as frames 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.

In Figure 2 a view of a fragment of the ALFRESCO KB is sketched.

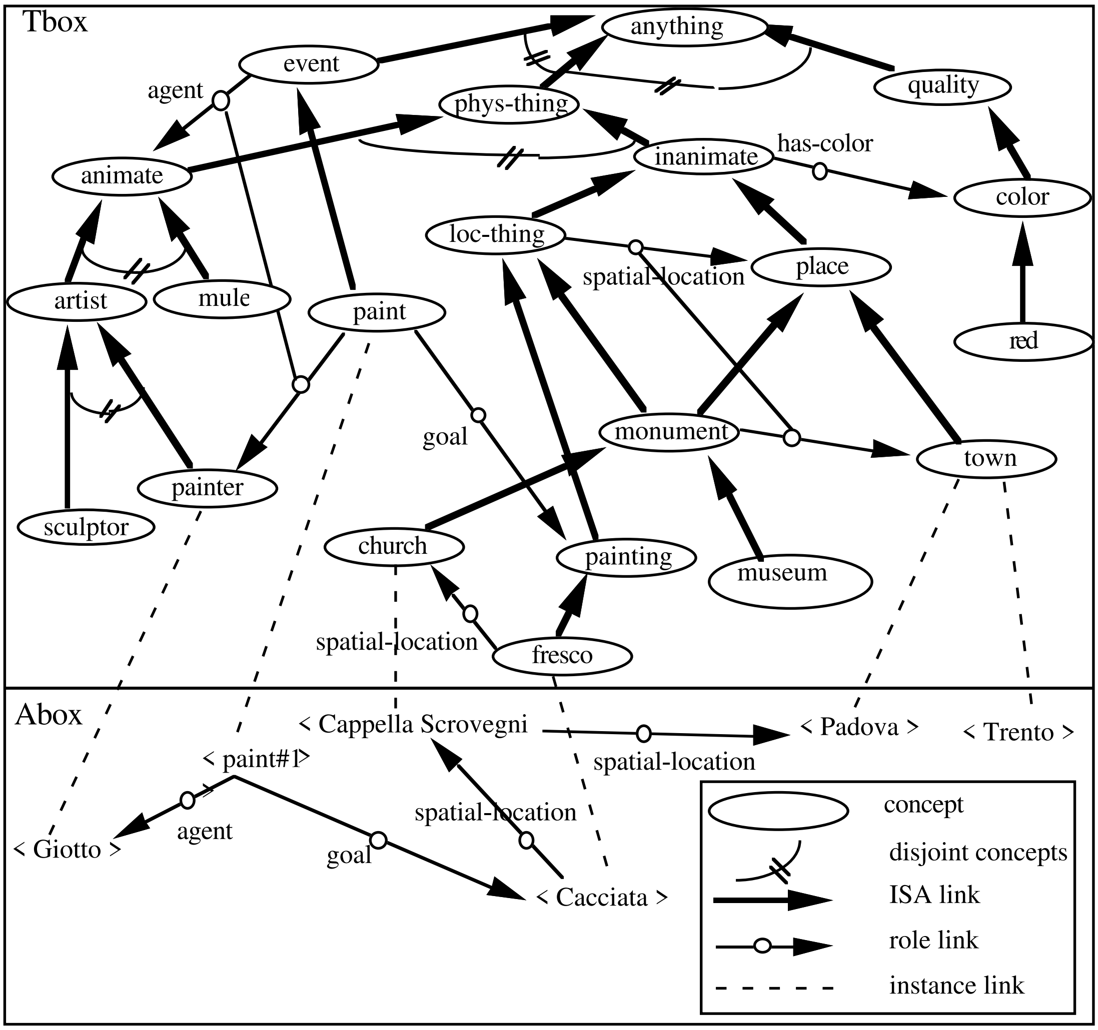

Figure 2

A NoteCards hypermedia network is included in the system, 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, where 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.

3.2 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 [Kay 1980] parser that can 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. The parser bases its work largely on information stored in the lexicon. The only centralized data used by the parser - in the form of a (non-augmented) recursive transition network - deals with the distribution of constituents. 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. If a word is mispelled the error is signalled and either it is removed or the user enters a modality in which she can specify a new lexical entry. The editor through which the sentence is input provides a transparent behaviour: 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 component via a lexical semantic analysis component the possible functional relations are filtered; this permits a great reduction in 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 & Stock 1990]. The following dialogue fragment:

U: Giotto ritrasse S.Francesco in un'opera di Assisi?
Did Giotto portray St.Francis in a work located in Assisi?

S: Sì, ad esempio nella Predica agli uccelli.
Yes, for example in the Sermon to the birds.

U: Gioacchino in un affresco di Padova?
Gioacchino in a fresco located in Padova?

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

3.3 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 formal language) and a level of meaning representation (usually denoted as WML, world model language; see Figure 3). 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's [1987] and on a) lexical classification of quantifiers; b) syntactic and surface order charcteristics; 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 [Westerstahl, 1986]. 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 3 A WML form for "Vorrei vedere un affresco del maestro in un monumento della città"
I would like to see a fresco by the painter in a monument of the town.

In our case, without passing to a DBL form (Data Base Language in Scha [1983]) WML expressions are dynamically mapped into YAK's assertional language. In the process the topic module (see below) is called in. Translation is started, for example when interpretation needs the extension of a certain concept or the truth value of a certain fact. For a description of the whole semantic interpretation process within the ALFRESCO system see [Strapparava, 1991] and [Lavelli et al., 1992].

3.4 The Topic Module and pointing

In a multimedial dialog system, the topic module must integrate global focus strategies [Grosz and Sidner, 1986], [Grosz, 1977], local focus approaches [Hajcova, 1987]) and deictic reference techniques [Wahlster 1988]. The main points of ALFRESCO's topic module [see Samek 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 of letting the entities that have been mentioned slowly fade away unless they are mentioned again with a certain functional role [Haijcova, 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 2) a stack of Topic Units, where 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 or the region of the mule and the region of Mary in "The Flight to Egypt"), differently from the one-to-one mapping of previous systems between the pointed region and the intended one [e.g. Brown et al., 1979] . As in CUBRICON [Neal et al., 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 in general both a touch and a linguistic expression may be ambiguous and yet yield a unique referent through mutual constraining.

3.5 The user's interest model

Presently, the most original part of the user model is the interest model. In ALFRESCO, in particular, interest for a given topic is assumed to be either a consequence of the activation of that topic in context (e.g. a specific request by the user) or of the anticipated exposition of the user to an art work (remember that the system includes a videodisc from which images of frescoes are retrieved). The interest model develops and becomes more focused in the course of the interaction between the user and the system.

The model consists of an activation/inhibition network whose nodes 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 to be somehow relevant. Each set represents an *interest area* and is identified by a particular individual or concept represented in YAK.

The top level organization of the model is a set of *dimensions*: each dimension is defined by a set of

(assertional or terminological) concepts from the KB. Such concepts are the heads of the interest areas associated to the dimension.
An area is defined by an ordered set of A-Box individuals that are considered to be "connected" to the head of the area in the particular domain. The individuals are listed in order of decreasing relevance as interest indicators for the head concept.
A dimension will typically represent a class that is particularly relevant for the current domain and its areas will be identified by (a subset of) its subclasses or individual instances. The sets associated to areas belonging to the same dimension are disjoint.

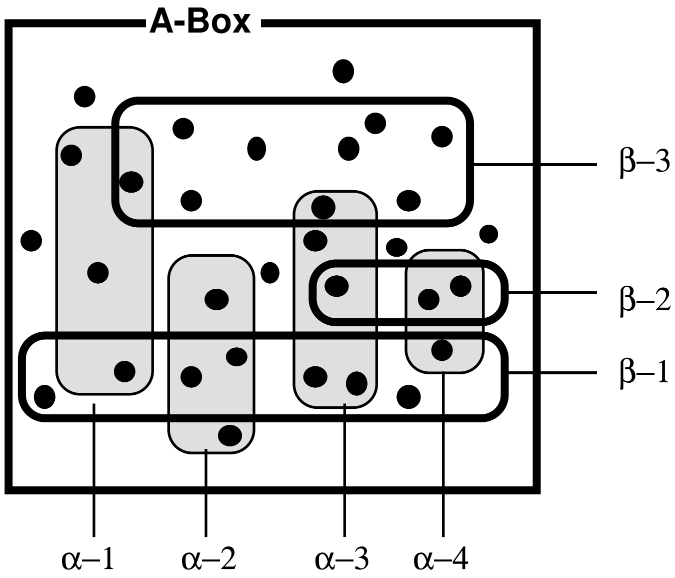

Figure 4

The A-Box shown in Figure 4 is structured along two dimensions: A (consisting of areas a-1, a-2, a-3 and a-4) and B (consisting of b-1, b-2 and b-3).

The areas of a particular model are all nodes of a single activation/inhibition network that connects them independently from the dimension they belong to. The links between the areas may be used to represent any kind of semantic/pragmatic proximity relevant for the domain. In the ALFRESCO system, the dimensions of the model correspond to painting schools, periods of time and towns. We have drawn activation links between areas corresponding to towns that are geographically close, time periods that are temporally close, but also (for instance) between a painting school and the town in which the members of the school painted most of their works.

Each time that a certain individual is designated by the user, or is selected as a possible answer to a query, the areas to which that individual is associated receive an activation impulse proportional to the explicitness of the reference. Such impulses are then propagated through the network, decreasing in intensity according to the weights of the links traversed .

The current implementation supports both activation and inhibition links. In traversing an inhibition link, an activation (i.e. positive) impulse is transformed into an inhibition (negative) one. Such links should be used to connect interest areas that are mutually incompatible: they represent the fact that interest in certain topics generally implies a lack of interest for other aspects of the domain. Inhibition impulses are not propagated through inhibition links: a negative impulse at a certain node also inhibits those areas that are reachable from it through activation links; but such an impulse does not cause the activation of those areas that are reachable from that node through inhibition links.

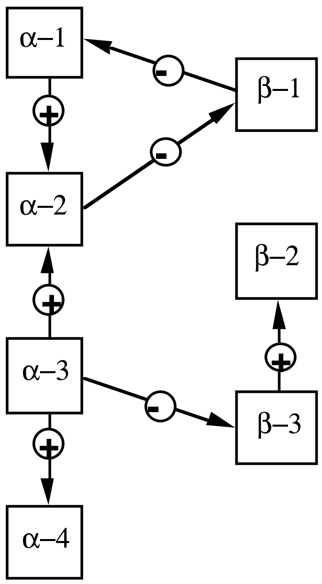

Figure 5

For instance, if the a-3 area in the network in Figure 5 receives a positive impulse, it will propagate to a-2 and a-3; b-1 and b-3 will receive an inhibition signal that will propagate to b-2 from b-3, but not to a-1 from b-1.

In the networks we currently use, the individuals associated with a particular head, as well as their static ordering, are computed on the basis of the content of the A-Box. In general, each individual *I* is assigned to an area for each dimension. The area is that for which *I* scores the highest relevance level. Such level is computed on the basis of the number and the length of the minimal paths that connect *I* to the head of the area in the graph whose nodes are the individuals and whose edges are the roles of the KB. Individuals that are judged to be equally relevant for two or more areas of a dimension are not assigned to any area.

The selection of the concepts to be used as heads of the interest areas and the links which constitute the network are not automatically computed and have to be explicity defined. This is unavoidable, since the information embodied in the structure of the activation network implicitly contains a set of hypotheses about the relevance of the different entities of the domain and some relations between them that are not necessarily represented in the KB.

The status of the interest model provides a criterion for the computation of the relevance of the A-Box individuals with respect to the current interactional context: the interest areas are dynamically sorted in order of decreasing activation level. The selection between individuals associated to the same area depends on the static ordering specified when the areas are built. Such relevance criterion is exploited in choosing among different possible individuals to be introduced by the natural language generator and has a consequence in the hypertext access.

3.6 Generation in ALFRESCO

The generator has the purpose of expressing linguistically statements that are in an internal alinguistic format. The strategic task is actually to select the appropriate contents and to plan the structure of the resulting text . In pursuing these goals the generator must take into account the situation (see Figure 6): the user model, the discourse context, the characteristics of the particular user's request, and finally the configuration of the underlying hypertext**.** The strategic component then defines the input for the tactical component, which has the task of expressing the planned text in the best possible linguistic manner.

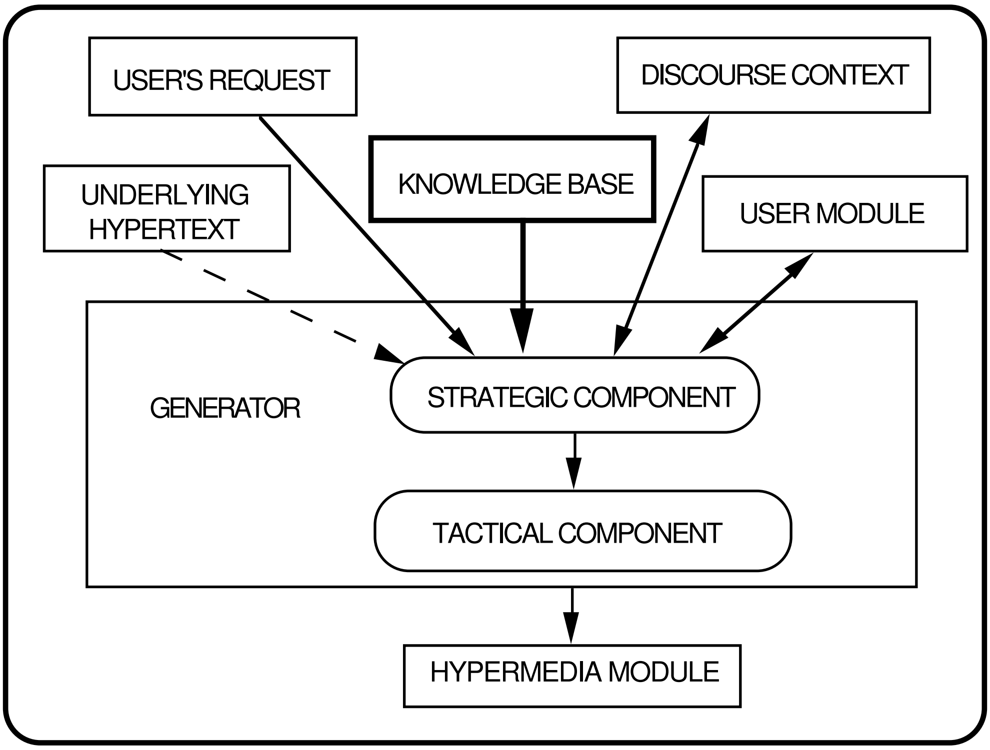

Figure 6 Generator architecture and its knowledge sources

The communicative goal of the generation component, providing information about instances of the domain (i.e. frescoes), is achieved by execution of rhetorical schemata (Figure 7) [McKeown 1985]. Briefly, a rhetorical schema is an abstract description of a portion of coherent text coupled with processes that select and order the relevant knowledge.

```
SCHEMA  S1
 (1) [(instance1 ( generic-description (attribute valueM)*))
 (2)  ((comparative-attribute not already used in (1) valueM)
            (S1 instance)*)
 (3)  ((generic-description
          (recovered-attribute already used in (1) value)*)
            (S1 instance)*)]

(2) introduces a new attribute of instance1 and makes a comparison with
other instances
(3) uses again some of the attributes already mentioned in (1), possibly
with a generalized value in order to introduce new istances
 * = 0 or more iterations
 M = possibility to add further information about the value
```

Figure 7

The schemata's task is to select and order appropriate attributes, generic-classes, values, instances (Figure 5) 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 (Figure 7) 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. In the case that no instance is returned, the selection process tries the steps described above for *content* with the second attribute in the ordering i.e. *date-of-realization* and goes on in this way until either it finds a comparative-instance (and so a comparative-attribute) or finishes the possible comparative attributes.

As mentioned before, the dynamic knowledge source most involved in the construction of the queries to the KB is the user modeling component. It 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 1992] and yields complex natural language sentences.

4 Integrating NL Generation and hypertextual techniques

The system's output is a hypermedia node containing a generated text with links to the hypermedia network and possibly to images (retrieved from the videodisc in the case of ALFRESCO). In particular, the output of the tactical generator 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 the strategic generator.

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 (Trigg et al., 1987). Independence from the target task domain allows NoteCards to be successfully used in a generation system. 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. The 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 8 shows the card obtained applying the Similarity Schema Instantiator after the request: " Can you show and describe to me a fresco by Ambrogio Lorenzetti in Siena?"

Figure8 *“The Effects of the Good Government is a fresco by A.L. in the Public Palace. The Effects of the 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 amonument of Siena is the Announcement, of 1344 in the Pinacoteca”*

The canned texts accessable 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.

Presenting the user with a mix of generated and canned text in general can be misleading; in fact the unaware user might: a) make reference to canned text (the system is unable to understand) and b) overestimate the system's capabilities. The use of graphic containers overcomes this problem by 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 contain 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.

The integration of NL generation with the NoteCards network is complete, since we have both links pointing to the cards containing the generated text and links pointing from such cards to "static" items of the network. NL generation provides a promising approach to the solution of some of the classical problems of hypertextual interaction, such as the lack of sensitivity to the overall context of the interaction and the scarce adaptation to the characteristics of the particular user.

Dynamically generated cards (being generated on the basis of specific queries and taking into account the status of the user model) result in often providing links between cards corresponding to individuals that are not connected in the knowledge base: such a selective and structured access to a specific partition of the network could also help in avoiding the disorientation of the user.

In order to support the user in the exploration of the information stored in NoteCards, our system also makes use of the NoteCards "Browser" cards. The graphs (and the underlying hierarchy) are dynamically updated in order to take into account the cards generated during the interaction.

For instance, after the generation of the card displayed in Figure 8, the graph in Figure 9.a is modified as shown in Figure 9.b.

Figure 9.a

Figure 9.b

The rationale of the different links corresponds to user interests. Different graphical outlooks are used for lines that denote different interest dimensions. This helps make browsing less confusing: the cognitive overload of the user (see [Conklin, 1987]) is reduced.

Two general problems 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. "Coloured" links are provided so that exploration can be carried on preferentially along different dimensions. A more personalized view of the options may be studied and indeed there are works in bringing in user modelling in hypertext (see for instance [Stein et al., 1991]).

Anyway, it is not clear whether a simple user model (such as that of ALFRESCO) could really improve the efficiency of hypertextual interaction: it may be 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 better to us not to build a user model through the hypertext modality thereby minimizing any consequences for the natural language-centered interaction. In practice, the only real consequence is at the discourse level, where it is taken into account

that the user has been provided with an hypertext access card that she has used.

On the User Interface aspect, a possibility is to use an analogical device, a gauge, that can be set by the user to determine the degree of desired "intrusion" of the interest model in the explorative interaction, in both ways.[1]

But let us reiterate that for the moment we think that the interest model, at least with current understanding, should only determine the way in which one can begin exploration through hypertext, yielding two complementary yet integrated modalities.

Among various further research themes, we are investigating the possible extension of the interest model to include T-Box (as well as A-Box) concepts as members of interest areas. Such an extension would provide a criterion to operate selections among classes of individuals and for subparts of generic descriptions. In order to make use of T-Box items also during the activation phase, it will be necessary to take into account the concepts that appear in the logical-form representing the input request sentence. Of course, also the automatic generation of the contents of the areas will have to be modified. For instance, the generator will take into account a criterion for choosing dynamically a certain description of an individual (among various possibilities), and therefore relate it to other individuals on the basis of that particular conceptual framework. Also the selection of attributes and the focussing on certain subspecifications of a structure of an individual would be dynamically performed for the particular interest of the user. This feature will provide both a text better tailored for the circumstance, focusing on what the user may be really looking for, and a better focus on the access to the underlying hypertext. A more limited number of relevant links will appear in the generated text, but more importantly, these links can be semantically qualified in a sense that the user is supposed to understand.

Recently there has been other work [Reiter et al., 1992] aimed at combining text generation and hypertexts. Their work and ours are different. Reiter and colleagues address the problem of new hypertext being dynamically generated as the user performs further requests by clicking the available buttons in the current text. The fact that, as the user moves further, new specific text is generated presupposes that all the relevant knowledge is coded in an internal format and can be expressed linguistically on request. The typical browsing modality is not really the goal for Reiter and colleagues. For instance the user is not supposed to move around the network with a bird's eye view just to have an idea about what information is available for certain areas. Instead, hypertext buttons are a way for guiding and limiting the local follow up requests that the user may wish to express (see also Moore and Swartout 1990). Our view is different: we maintain that natural language and hypertexts can be combined, yielding their specific contribution, both in input and in output. In the first place, hypertexts exist exactly for the kind of information that it is difficult to express in an internal format (for instance art

[1] Phil Cohen suggested this possibility.

critics' views). So they constitute *per se* a level of representation with a very limited number of entrypoints. In the second place we believe that the role of hypertext in a natural language centered interactive system is to consent to the user to explore the area surrounding the focus of her attention and to provide a noncommitted modality of interaction, characterized by a fast turnaround. Of course the generator starts from an internal representation also in our case but the result is a text linked to other preexisting text in a personalized way.

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. 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.

Many points are to be explored: apart from more particular aspects, 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 potentiality because cultural tourism is of 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.

On the other hand we think that as we understand more about how exploration (as in hypermedial interaction) and goal-oriented behaviour (as in natural language) can be coherently modelled and integrated, a new perspective may be opened for intelligent interfaces that goes beyond what we are used to and amplifies our cognitive capabilities.

Acknowledgement

We would like to thank J. Slack for many helpful discussions

References

André, E., Finkler, W., Graf, W., Rist, T., Schauder, A., Wahlster, W., "WIP: The Automatic Synthesis of Multimodal Presentations". *This volume*.

André, E., Rist, T., "The Design of Illustrated Documents as a Planning Task". *This volume*.

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.

Arens, Y., Hovy E.H., "How to Describe What? Towards a Theory of Modality Utilization". In *Proc. 12th Cognitive Science Conference*, Cambridge, MA, 1990.

Arens, Y., Hovy, E.H., Vossers, M., "The Knowledge Underlying Multimedia Presentations". *This volume*.

Bonarini, A., "Modeling Issues in Multimedial Car-Driver Interaction". *This volume*.

Brown, D.C., Kwasny, S.C., Chandrasekaran, B., Sondheimer, N.K., "An Experimental Graphics System with Natural Language Input". *Computer and Graphics*, 4, pp.13-22. 1979.

Burger, J., Marshall, R., "The Application of Natural Language Models to Intelligent Multimedia". *This volume*.

Carenini, G., Pianesi, F., Ponzi M., and Stock, O., "Natural Language Generation and Hypertext Access". To appear on *Applied Artificial Intelligence, An International Journal.*

Cohen, P.R., Dalrymple, M., Moran, D.B., Pereira, F.C.N., Sullivan, J.W., Gargan Jr, R.A., Schlossberg, J.L., Tyler, S.W., "Synergetic Use of Direct Manipulation and Natural Language", *Proc. CHI 89*, pp. 227-233, Austin, Texas, May 1989.

Conklin, J., “Hypertext: An Introduction and Survey”, *IEEE Computer*, 20(9):17-41, 1987.

Feiner, S.K., McKeown, K.R., "Automating the Generation of Coordinated Multimedia Explanations". *This volume*.

Feiner, S.K., Litman D.J., McKeown, K.R., Passonneau, R.J., "Towards Coordinated Temporal Multimedia Presentations". *This volume*.

Franconi, E., “The YAK Manual: Yet Another KRAPFEN”, *IRST Manual no. 9003-01*, Trento, Italy, 1990

Goodman, B.A., "Multimedia Explanations for Intelligent Training Systems". *This volume*.

Grosz, B.J., "The Representation and Use of Focus in Dialogue Understanding". *Tech. Note 151. SRI Project 5844*, SRI, July 1977.

Grosz, B.J., Sidner, C.L., "Attention, Intentions and the Structure of Discourse". *Computational Linguistics*, 12 (3) July-September 1986.

Hajicovà, E., "Focusing: a meeting point of Linguistics and Artificial Intelligence". In P. Jorrand and V. Sgurev (eds.), *Artificial Intelligence II "Methodology, Systems, Applications"*, Elsevier Science

Publishers (North Holland) 1987.

Halasz, F.G., “Reflections on NoteCards, Seven Issues for the Next Generation of Hypermedia Systems”, *Communications of the ACM*, 31(7):836-852, July 1988

Hobbs, J. and Shieber, S., "An algorithm for generating quantifier scopings", *Comp. Ling* Volume 13 1987

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*, 14 (3):5-22, September 1988

Kay, M. "Algorithm Schemata and Data Structures in Syntactic Processing", *Technical Report CSL-80*, Xerox Palo Alto Research Centers. Palo Alto, California, 1980.

Kobsa, A., Allagyer, J., Reddig, C., Reithinger, N., Schmauks, D., Harbusch, K., Wahlster, W., "Combining Deicitic Gestures and Natural Language for Referent Identification". *Proc. of the 11th Conf. on Computational Linguitics*, Bonn RFT, 1986.

Lavelli, A., Magnini, B., Strapparava, C., "An Approach to Multilevel Semantics for Applied Systems". In *Proc. 3rd Conference on Applied Natural Language Processing*, pp. 17-24, ACL, March 31-April 3, Trento, Italy, 1992.

Lavelli, A., Stock, O., "When Something is Missing: Ellipsis, Coordination and the Chart", *Proc. of the 13th International Conference on Computational Linguistics, COLING 90*, pp. 184-189, Helsinki 1990.

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

McKeown, K.R., "Text Generation", Cambridge University Press, 1985.

Maybury, M.T., "Planning Multimedia Explanations Using Communicative Acts". *This volume*.

Neal, J.G., Shapiro, S.C., "Intelligent Multimedia Interface Technology". *Proc. on the Workshop on Architectures for Intelligent Interfaces: Elements and Prototypes*. Monterey CA USA, 1988.

Nielsen, J., "The Art of Navigating through Hypertext", *Communications of the ACM*, 33(3), July 1990.

Oviatt, S.L., Cohen, P.R., “The Effects of Interaction on Spoken Discourse”, *Proc. of the 27th Meeting* *of the ACL*, pp. 126-134, University of British Columbia, Vancouver, Canada, June 26-29, 1989.

Paris, C.L., “Combining Discourse Strategies to Generate Descriptions to Users Along a Naive/Expert Spectrum”, *Proc. of IJCAI-87*, pp. 626-632, Milano, Italy, August 23-28, 1987.

Pianesi, F., "Head Driven Bottom Up Generation and Government and Binding: A Unified Perspective". In H. Horacek and M. Zock (eds), *New Concepts in Natural Language Generation,* Pinter, London, 1993, to be published.

Reiter, E., Mellish, C., Levine; J., "Automatic Generation of On-Line Documentation in the IDAS Project". In *Proc. 3rd Conference on Applied Natural Language Processing*, pp. 64-71, ACL, March 31-April 3, Trento, Italy, 1992.

Samek-Lodovici, V., Strapparava, C., "Identifying Noun Phrase References, the Topic Module of the AL Fresco System". In *Proc. ECAI 90*, pp. 573-578, Stockholm, Sweden, August 6-10, 1990.

Scha R., "Logic foundation for question answering", Philips Research Labs MS 12.331, 1983

Slack, J.M., Conati, C., "Modeling interest: exploration of an information space", 1991. To appear in *Acta Psychologica on Cognitive Ergonomics*

Stallard D. "A terminological simplification tranformation for NL question answering system". *Proceedings of the 24th Meeting of the ACL*, pp. 241-246, Columbia University, new York, June 10-13, 1986.

Stein, A., Thiel, U., Tißen, A., "Towards coherent hypermedia navigation by pragmatic dialogue modeling". Arbeitspapiere der GMD 580, GMD, Darmstadt, Germany, 1991

Stock O., "Parsing with Flexibility, Dinamic Strategies and Idioms in Mind". *Computational Linguistics* 15(1):1-18, MIT Press, Cambridge, Mass., 1989

Stock, O. “Natural Language and Exploration of an Information Space: The ALFresco Interactive System”, *Proceedings of IJCAI-91*, pp. 972-978, Sydney, Australia, August 24-30, 1991.

Stock, O., "A Third Modality of Natural Language?". In *Proc. ECAI-92, 10th European Conference on Artificial Intelligence*, pp. 853-862, Vienna, Austria, August 3-7, 1992

Strapparava, C., "From Scopings to Interpretation: The Semantic Interpretation within the ALFRESCO System". In E. Ardizzone, S. Gaglio, F. Sorbello (eds.), *Lecture Notes in Artificial Intelligence 549*, Springer-Verlag, 1991.

Trigg, R.H., Moran, T.P., Halasz, F.G., “Tailorability in NoteCards”. In *Proc. of Interact 87 2nd IFIP*

Conference on Human-Computer Interaction, Stuttgart, 1987.

Wahlster, W., "User and Discourse Models for Multimodal Communication". In J.W. Sullivan and S.W. Tyler, eds., *Architectures for Intelligent Interfaces: Elements and Prototypes*, Addison-Wesley, 1988

Westerstahl, D., "Quantifier in formal and natural language", *Report No. CSLI-86-55*, Palo Alto, 1986

Whittaker, S., Stenton, P., "User Studies and the Design of Natural Language Systems". *Proc. of the 4th Conference of the European Chapter of the ACL*, pp. 115-123, Manchester, UK, April 10-12, 1989

Woolf, B., Suthers, D., Cornell, M., "Using 'Live Information' in a Multimedia Framework". *This volume*.