Multimodal Communication between Software Designers and Software Users

Clarisse Sieckenius de Souza[1], Raquel Oliveira Prates, and Flávio Miguel Varejão[2]
Departamento de Informática, PUC-Rio
Rua Marquês de São Vicente 225
22453-900 Rio de Janeiro, RJ
email: {clarisse,raquel,fvarejao}@inf.puc-rio.br

Abstract

Multimodal interface design represents significant costs in terms of cross-referenced representation languages, abstraction tracking, and inter-modal coherence and consistency. This paper stresses a perspective of designer-to-user multimodal communication and reports the results of small-scale exploratory experiments meant to shed light on the nature and extent of some of the theoretical issues involved in Multimodal Interface design. Based on a discussion about such issues, it proposes some topics for a research agenda in the field.

Keywords: multimodal interface design, natural language descriptions and explanations, representation languages, computer semiotics

1. Introduction

Multimodal Interfaces have extended the interactive capacity of computer systems by allowing people to resort to a variety of communicative modes (e.g. natural and visual languages) and media (e.g. audio and video) in search of achieving an efficient and effective usage of software tools. Although standard views of HCI center on "**h**uman-**c**omputer" **i**nteraction alone (as the acronym suggests), we follow those that have broadened this scope and proposed that HCI is in fact part of "human-human" interaction mediated by computer systems (e. g. Kammersgaard, 1988; Nadin,1988; Andersen, 1990; de Souza, 1993). The shift implies that multimodality is not only a resource users may get hold of to communicate with computers, but also that text, graphics, video, audio, virtual reality, and whatever is there in years to come, are means of communication for designers to get their ideas and products across computer devices to people who are willing to use them and enjoy them.

---

[1]Clarisse Sieckenius de Souza has a research grant from CNPq.
[2]Flávio Miguel Varejão is an Assistant Professor at Espírito Santo Federal University, Computer Science Department.

This stand may sound like a truism in the field of Multimedia (MM) and Hypermedia (HM) applications, but we will not be focusing on these. Rather, we intend to explore the fact that any typical Graphic User Interface (GUI) design nowadays involves the same nature of choices faced by MM/HM designers. The difference lies in that the message sent from non-MM/HM designers to users is less frequently one of documentation and entertainment, and more often one of problem-solving, programming, and design itself. An increasing awareness of the importance of extended computer literacy, end user programming, and direct acquisition of knowledge from users in a very wide spectrum of applications arises from the fact that no designer in the world can possibly anticipate all of future users' needs and fancies (Nardi,1993). It must be the case, then, that users have to be empowered with the capacity to write their own programs and applications, by means of parameter configurations, macro recording, or programming in the small (Myers,1992).

Thus, the scenario we envisage for future software applications is one in which multimodal interfaces will be designed to achieve the goal of helping users understand what computer application designers mean when they write software. Such understanding, as any other, should open a new range of possibilities for creative behavior and foster significant progress. However, interface designers should provide users with as many different perspectives on software as possible, including intelligent explanation and help systems, direct manipulation and visualization graphics, which all turn designers into actual software writers (or composers), rather than computer programmers.

In the following we will describe a small-scale exploratory experiment carried out with a prototype layout system. Participants have used the prototype to build 2D layout models of typical university offices. The standard furniture included desks, chairs, bookshelves, and cabinets, as well as computer equipment. Subjects have accomplished different memory tests whose goal was to indicate to us, designers, how textual and visual interface modes supported differently the various cognitive tasks involved in sense-making. With such indications we expected to estimate the complexity of choices in optimal distribution of message contents across different modes of communication provided by computer interfaces, and costs in intermode consistency.

Our conclusion is that an integrated self-explanatory problem-solving environment represents major challenges for designers in terms of representation languages and inter-mode coherence and consistency. Because different cognitive tasks are better supported by different modes of communication, designers of extensible modularized applications will have to devote much of their efforts to analysing their own interpretations about computer tasks and deciding about how to encode such interpretations.

Section 2 reports the experiment we have run, and Section 3 presents results achieved with the testing. Section 4 states our conclusions in the light of some cognitive (Kirsh and Maglio,1994; Stenning and Oberlander,1995) and semiotic (Eco,1988; Nadin,1988; Andersen,1990; de Souza,1993) theories, and our proposal of additional items for the current research agenda in the field.

2. Layout Configurations: Understanding Descriptions and Retrieving History

Layout problems have been selected for the experiment because of several reasons. Firstly, they illustrate very clearly some of the effects of multimodal interfaces. Because programs can be geared by both textual (e.g. menus and command language) and direct manipulation graphic interfaces, the choice of using one and/or another can be tested as to which consequences arise. Secondly, they represent a class of design (i.e. synthesis) problems users are likely to encounter in the future if they ever build novel applications by combining modules of pre-programmed software. Thirdly, the rationale for some layout configuration does not necessarily emerge from the sheer observation of the design product in its final state. The history of design, with backtrackings and loops, reveals a lot about a designer's intent (Garcia and de Souza, forthcoming) and plays a major role in sense-making. Last, but not least, the correspondence between direct manipulation operations on visual objects and words used in natural language to refer to them shows evidence of numerous meaning adaptations and extensions which have a considerable impact on representation languages embedded in computer applications.

Our prototype system was a very simple layout program in which users could decide how to furnish a typical university office with desks, bookshelves, cabinets, and computers. The interface adopted a direct manipulation style, and users were able move and locate visual objects on a gridded canvas space. The canvas was a 2D top view of the office area, including the position of the door and window. Pieces of furniture were visually represented by quadrilateral shapes equivalent to their top surface. Computers were represented as squares; desks as larger squares (for sake of distinction from cabinets and bookshelves); cabinets and bookshelves were represented by equal rectangular surfaces and distinguished from each other by spectrally distant colors (dark green and bright yellow). The room surface was structured as a sequence of 9 numbered sub-spaces or regions with special markings for doors and windows (see Figure 1).

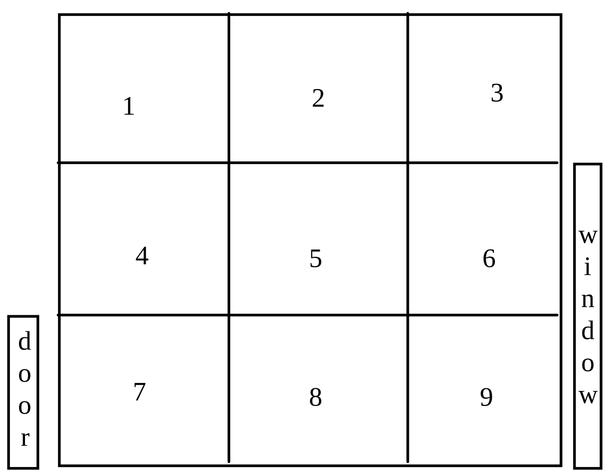

Figure 1: Spatial structuring of an office

Although trivial, the grid proved to be a convenient space representation to support direct manipulation at the interface level, since all the constraints about object locations could be easily stated in terms of such regions. For example, desks could be placed anywhere except in region 7; bookshelves could only be placed against walls (i.e. not in regions 5 and 6); and so on. Moreover, the attributes of each region allowed for inferences as to the location of furniture against perpendicular walls (e.g. the only available wall for a bookshelf in region 9 is that perpendicular to the window). Objects have been characterized by physical constraints relevant for the task. Thus, computers have been constrained to be placed only on top of desks, cabinets and bookshelves have never been allowed to be put against the same wall, nor have cabinets and desks. Bookshelves, however, have been allowed to be placed above desks, against walls.

All users' manipulations were logged in a file whose structure is shown in Figure 2. Once again, this structure is extremely simple but effective. Figure 2 illustrates the location of a desk and a computer in the same region. The fact that the computer is ON the desk is not explicit in the structure. The implementation of drag'n'drop actions forces computers to be on top of desks, and thus consistency is enforced by the program itself (being abstracted in the underlying representation)

STEP 1	MOVE	DESK-1	from(OUTSIDE) to(REGION1)
STEP 2	MOVE	COMPUTER-1	from(OUTSIDE) to(REGION1)
...	...	...	...

Figure 2: A portion of a logged layout section

This representation was selected because of its good coupling with the visual mode in which layout tasks are usually performed. However, we have added two other functionalities to the system. One was the generation of textual printouts with a natural language description of the office layout, and the other was a narrative text accounting for the layout process history. As mentioned above, these features play a major role in design documentation applications and clearly characterize the need for intelligent multimodal interfaces (Maybury,1993). We will not stress the differences between the two modes in retrieving layout configurations and histories, because it is intuitive that a visual representation is required for adequate understanding of spatial configuration descriptions, and that a textual narrative provides abstractions and powerful focusing mechanisms for understanding processes over time. We will, however, report tests and results performed with this prototype involving memorization of layout states and history from both textual and visual inputs in an attempt to highlight the underlying requirements in terms of representation languages, inter-modal consistency and coherence, and the importance of arbitrary encodings of our own interpretations about real-world situations.

In our prototype, users could playback graphically all of their steps in solving a layout configuration problem. The system must only read the log file, and the subsequent frame animation could be generated to reproduce every operation users performed. This feature alone certainly supports the design rationale documentation process, but it also imposes some cognitive loads on users. To mention the most obvious one, this movie-like strategy conceals some abstract concepts a natural language report of the process could directly reveal. For instance, the use of the word RETURN to signal that a certain piece of furniture is being moved *back* to a place where it stood *before* has no equivalent in frame animation. The same is true for such other concepts and words as EXCHANGE (or SWAP) and CENTER. Natural language is more powerful to express such global abstract patterns, and a user's perspective about the problem is automatically expanded to account for time periods that are longer than one frame (e.g. RETURN, after "n" frames), or space ranges that refer to more than one region (e.g. the CENTER of the room).

We have started our experiment by trying to use exactly the same log file as input to both an animation program and a text generation program. Text generation was a simple readout of the log file structure, with only a few syntactic refinements like coordination and pronominal anaphora. Since the generated text was very poor in capturing some evident perceptual abstractions, an immediate upgrade was made. The second text generator used a Brazilian Portuguese grammar and dictionary in which notions such as RETURN, SWAP or EXCHANGE and CENTER appeared in narrative text. Therefore, the path a certain object followed before "returning" to its original space could be left undetermined in narrative text, although this was not true of the "playback" mode in which all details were offered to visual observation. The semantics of these words was a collapsed sequence of steps looping back to the original state. In other words, a layer of abstraction was built on top of the flat structure.

Nevertheless, whereas in the visual interface mode the animated representation of subsequent state sequences was successfully rendered, in the narrative text a number of new problems emerged. Two of these have been selected to demonstrate the costs of having multimodal interfaces: the spatial reference problem and the higher-order abstract move pattern problem.

The spatial reference problem amounts to the low-quality text provided by a fixed-frame reference system which was computationally attractive to reproduce the layout history visually. This system only supported text as that in (a), whereas a much more desirable rendition of it would clearly be that shown in (b).

(a) "Pedro colocou a primeira mesa na região 1 e a segunda mesa na região 3. Em seguida, ele posicionou o primeiro armário na região 2." [Automatic Generation which translates into: "Peter placed the first desk in region 1 and the second desk in region 3. Next, he put the first cabinet in region 2."]

(b) "Pedro colocou a primeira mesa na região 1 e a segunda mesa em frente a ela, contra a parede. Em seguida, ele posicionou o primeiro armário entre as duas mesas." [Manual Generation which translates into: "Peter placed the first desk in region 1 and the second desk in front of it, across the room. Next, he put the first cabinet between the two desks."]

The major difference between (a) and (b) is the inter-object dynamic nature of the reference system used in (b) for the realization of adequate spatial expressions. For (b) to be automatically generated, the original fixed-frame system must be replaced by or translated into an object-oriented, indeterminacy-tolerant, hierarchical and dynamic system in which the possibility of inter-object spatial references may override office space fixed references (Ioerger,1994).

The higher-order abstract move pattern problem amounts to the low-quality of text provided by a flat representation of state sequences (as in (c) below) as compared to articulated representations of different state change patterns (as in (d) below). Such notions as RETURN (or GO BACK), for example, are the result of computations performed upon the whole history (or larger portions) of problem-solving steps, in which the existence of two statements saying that OBJECT X MOVED to(LOCATION Y) actually means that the object was returned to a previous position. By introducing this feature in the text generation system, we have been able to evaluate the considerable computational cost it adds to a simple readout of subsequent historical steps, i.e. that of a dedicated knowledge base for the textual interface mode.

(c) "Pedro moveu a segunda mesa da região 9 para a região 3.[...] A seguir, ele moveu a segunda mesa da região 3 para a região 9." [Automatic Generation which translates into: "Peter moved the second desk from region 9 to region 3. [...] Then, he moved the second desk from region 3 to region 9."]

(d) "Pedro moveu a segunda mesa da região 9 para a região 3.[...] A seguir, ele retornou a segunda mesa para a região 9." [Automatic Generation which translates into: "Peter moved the second desk from region 9 to region 3. [...] Then, he returned the second desk to region 9."]

In order to estimate the effects of a suboptimal fixed-frame referencing system in a multimodal environment, we have carried out tests exploring static and dynamic reconstitutions of layout configuration states. By comparing results obtained by subjects with visual and textual input, we tried to have indications about the extent to which inadequate underlying knowledge representation systems affect interface quality. We looked for answers to two related questions:

(I) Is any of the current prototype's output modes sufficient to account for both a layout description of configuration states and a report of the layout process and history?
Supposition: None is.
(II) Is any of the output modes not fit for one retrieval task (state/process information) or the other?
Supposition: The Visual Mode is not fit to retrieve history.

We first configured an arbitrary initial layout state, to which we applied subsequent changes until we eventually reached a final state. The whole process was recorded in a backlog file (see Figure 2). From this file we automatically generated a narrative text in Brazilian Portuguese (Prates and dos Santos,1994) and an animated playback sequence. A descriptive text in the same language was also generated to represent the final layout picture in the last frame of the animation.

Two groups of 4 computer-literate subjects have made a set of 6 tests each: (a) memorizing and reproducing a layout state described in NL text; (b) memorizing and reproducing the exact steps of a layout process reported in NL text; (c) reproducing a layout process described in NL text permanently accessible for reading; (d) reproducing and transforming a layout state into another, both depicted on a screen permanently accessible for viewing; (e) memorizing and reproducing a layout state represented on a screen; (f) memorizing and reproducing the exact steps of a layout process represented by an animated sequence of screens.

Since our main interest was to assess how difficult it was for subjects to memorize textual and visual representations of states and processes, tests (c) and (d) were dummy tests. They were introduced just as a buffer between the other two sets of tests, because the configurations in textual and visual representations were exactly the same.

Subjects were introduced to the interface, and were left free to experiment for a while until they thought they knew how to operate. Groups performed the tests in different order. Group TV performed first the **t**extual memorization tests and then the **v**isual ones, whereas group VT performed the tests in reversed order. Each memorization test was carried out by first allowing subjects to read or playback the material at their own pace; then, the material was taken away from them and the experiment began. Because dummy tests did not require memorization, the material for these was not taken away from subjects.

3. Interpretation of Results

Table I[3] shows the global results achieved by both groups. Briefly, we can summarize a large number of tabulations and analyses by saying that the TV Group showed a homogenous pattern of success across the four tests. However, the group's pattern of failure showed a pronounced difficulty in Narrative Tests, in both Textual and Visual Modes. Moreover, the best Success/Failure ratio in this group was found in the Visual Description Test; the second best was in the Textual Description. The VT Group, however, showed a more heterogeneous behavior. Best results were found in the Textual Narrative Test (29 successful retrievals), against the lowest score in the Visual Narrative Test (19 retrievals). Not surprisingly, in this group, the best and worst performances were those in the Narrative Tests, which are indeed much more difficult than the Descriptive ones. Best scores were obtained with the Textual Mode and the worst with the Visual Mode. Also, this group had much better success/fail ratios than the TV Group. With the exception of the Visual Narrative Test, where both groups had 50% of success in the test, the VT Group had a pronouncedly superior performance than the TV Group.

Table I: TV and VT Group's Global Performance Scores

Group	VisDesc OK	VisNarr OK	TxtDesc OK	TxtNarr OK
TV	19	19	16	17
VT	23	19	21	29
Group	VisDesc Fail	VisNarr Fail	TxtDesc OK	TxtNarr Fail
TV	10	19	12	19
VT	5	18	7	7

[3]Table I shows absolute values of right/wrong operations involved in the task. Discrepancies of one item in total scores are due a TV Group subject's placing of extra objects in the office space.

We have confirmed our intuitions that using a Visual Mode to represent layout process history causes cognitive difficulties for users, but found that reasoning made from a visual representation of the layout model seems to benefit subsequent reasoning of the same kind made from a textual representation. The reverse order testing, however, going from reasoning about textual representations to reasoning about visual representations, does not show evidence of the same benefits, although mental images may have played a role in the TV Group's performance with visual description memorization tests.

The TV Group had just about as many problems in the Visual Narrative Test as in the Textual Narrative Test. The VT Group, however, had almost no problems with the Textual Narrative Test. The only few occurrences of mistakes (7, against 19 in the other group) were 4 omissions of furniture and 3 wrong positionings. Such performance presented a very strong contrast with the group's performance in the Visual Narrative Test (18 failures, against 19 in the other group) with pronounced frequencies of mistakes in the order of actions and omissions.

Fine-grained observations showed that the barren fixed-frame structure model from which positions were realized in NL text apparently caused unexpected difficulties in the processing of narrative texts. Also, the order of events was easier to retrieve from textual representations than from visual ones, thus favoring Textual Modes instead of Visual Modes for process retrieval information (de Souza, Prates and Varejão, 1995).

We have asked ten other subjects, with the same background as those who participated in the experiment, to freely write Brazilian Portuguese texts describing and narrating the configuration states and sequences used in the tests. In absolutely all texts the reference system used was not fixed: As expected, all subjects used a dynamic inter-object reference system to describe and narrate how objects were placed in the office room. Other interesting features of such texts were a more extensive use of the door and window as landmarks in the space to be furnished and a sophisticated perspective of the observers, who often viewed themselves inside the office and reported what they could see from where they were. Additionally, in accordance with existing research results about the indeterminacy problem in spatial reasoning from NL inputs (Ioerger,1994), subjects proposed descriptions and narratives that did not translate into one and only semantic model. All texts showed evidences of loosely determined semantic models, if not of simultaneous distinct semantic models (since some subjects clearly changed referring strategies within the same span of text).

One last experiment was made, in which we manually generated NL texts with one controlled improved feature — a dynamic object-oriented reference system for object locations, instead of the fixed frame used in the first experiment. Thus, the narrative and descriptive texts included expressions like: "Peter placed the first desk in region 1 and a computer on top of it", or "he placed a bookshelf in region 2 and another against the opposing wall, across the room". In other words, we made more extensive use of spatial anaphora, but all within the same (though more abstract) semantic model. Some solutions, however, were still not present in our new text, which made use of the underlying grid and referred to positions in terms of the "regions".

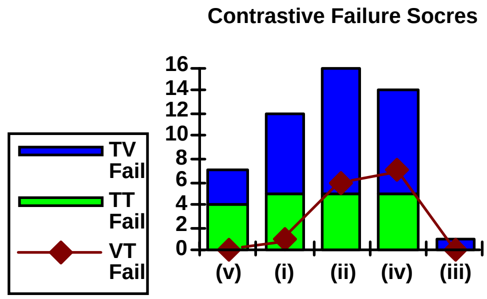

Chart I: Failure Scores in VT, TV and TT Groups' Tests with Textual Input only

We selected a third group of four subjects, not present in any of the previous experiments, and asked them to make only two tests: Textual Description and Textual Narrative. Contrastive results are illustrated in Chart I, which also classifies mistakes by types. (i) Users were able to remember an object, but they were mistaken about its position. (ii) Users were able to remember a position , but they placed the wrong object in it. (iii) Users introduced spurious objects in spurious positions. (iv) Users omitted an object from the layout. (v) Users were able to recall a correct step in the process, but they performed it in the wrong order.

As can be seen, this new Group, which we have coined the TT Group, has been virtually as successful as the TV Group regarding mistakes of type (v), and has been systematically more successful than the latter in all other cases. Also, the TT Group has equaled the VT Group performance with types (ii), (iv) and (iii), even without having benefited from previous memory tests from visual representations.

The indications from TT Group scores is that the improvement of text quality actually impacted performance in a positive way. Nevertheless, further improvements could still be introduced in the textual input, so that subjects had even less cognitive loads than they did in this last test. The obvious needed refinement in our text is abandoning the grid structure and regions, altogether, and adopting cognitively-salient landmarks of rooms (doors, windows, and walls), along with inter-object reference systems for spatial anaphora. This improvement, however, is clearly excessive for the Visual Mode Interface, which needs very little improvement to perform efficient and effective renditions of (sequenced) layout states.

4. Discussion and Conclusion

Knowledge representation issues in connection to Multimodal Interface Design have been previously singled out as critical (Arens, Hovy and Vossers,1993). We would like to resort to Stenning and Oberlander's Abstract Representation Systems (Stenning and Oberlander,1995) to highlight what we think is an important line in our experiment. The authors have sketched a cognitive theory of graphical and linguistic reasoning, and argued that graphical representations limit abstraction and aid processibility: diagrams, for instance, instantiate circumscribed interpretations of abstract concepts and help focus on important aspects of complex concepts, leaving details behind. They have distinguished three different kinds of representation systems. Minimal Abstraction

Representation Systems (MARS) allow for exactly one semantic model (i.e. leave virtually no room for abstractions). Limited Abstraction Representation Systems (LARS) allow for the construction of more than one semantic model for a fixed expressive system. And finally, Unlimited Abstraction Representation Systems (UARS) allow not only for more than one semantic model per expressive system, but actually support variations of the expressive system itself.

Our experiment with a multimodal prototype interface shows that graphics may indeed use MARS (fixed frame model), but text is best with LARS (indeterminacy-tolerant model). In the latter, higher-order abstractions are conveyed by selected lexical items in natural language. The direct manipulation graphic interface supports operations that are represented as uniquely determined semantic expressions, and this is why the design process may be completely played back in graphic mode. Although natural language sentences can be easily generated in correspondence with each and every semantic expression in the log file, the resulting text contributes very little to users' goals. One of the examples provided in the paper is the semantics of the word EXCHANGE (e.g. "The user exchanged the locations of the two desks"). The exact path the user chooses to follow in the sequential interface steps is irrelevant to communicate the user's goal, and is best left underdetermined.

Some challenging topics for our research agenda, in this respect, are: (a) graphics and natural language text may be mapped onto MARS, LARS, and UARS - how can designers control levels of abstraction in semantic models *across* different modes of communication with users? (b) more abstract levels are better suited to provide focal information in text and graphics - if different modes are supported by different representation systems, are inferences drawn in different modes *always consistent* with each other? (c) can graphics in layout configuration problems correspond to LARS, and account for psychological indeterminacies involved in spatial reasoning (Ioerger,1994)?

As a tentative approach to the above issues, our experiments suggest that inter-mode control of abstraction can only be achieved at the expense greater perspicuity in representation languages (which runs contrary to implementation efficiency criteria). Moreover, since history is better captured when expressed in words than graphics, arbitrary abstractions encoded in the representation language and generator lexicon by the system's designer will directly impact users' perception and inferencing. Text and graphics must indeed be used together to balance out biased interpretations due to media rhetoric.

Experimental research with games involving spatial reasoning (Kirsh and Maglio,1994) has shown that players of Tetris, for instance, do not always flip shapes in a direct attempt to fit them into the available slots at the bottom of the screen, but often do it to gain insight about object characteristics and problem-solving opportunities. Such actions are called epistemic actions (as opposed to goal-directed actions). In graphic playback mode, epistemic actions are not distinguished from goal-directed ones. Users are directly exposed to every single step along design, with no markings of relevance and intentionality. In text, however, goal-directed text plans (Hovy,1988) may clearly privilege finality in action over exploratory (lateral) behavior.

The above considerations lead us to other challenging issues: (d) can and should goal-directed behavior be distinguished from epistemic behavior in both text and graphics? (e) in search of design patterns (or in the so-called *data mining* or *knowledge discovery* activities), for example, should textual renditions be provided before or after graphic playback?

Multimodality, especially when modes have overlapping semantic models, underlines the arbitrariness of interpretations and encondings involved in computer programming, which is a legitimate semiotic problem. Being devoted to investigating the nature of sign production and interpretation, Semiotics may provide the necessary connections between computational, psychological, linguistic, artistic, and design issues involved in multimodal interfaces (Nadin,1988). Narrow local perspectives on isolated modes alone may cause damage to the global communicative system (as seen in the fixed-frame spatial reference model of our experiment, which was adequate to animation but inadequate for text generation). Every encoding excludes a wide variety of other possible encodings and reveals much about the mind which selected it as best. The same is true of interpretations. Thus, software programming (including data structuring, knowledge representation, and interface design) is very much like writing or composing (Andersen,1990), and computer programs tell much about psychological and culture values of those who wrote them (Eco,1988).

Further research issues related to semiotic aspects that arise in multimodality include: (f) if users in the near future are expected to modify and extend off-the-shelf programs, how can software designers provide them with a solid expressive set of MARS and LARS of their own mental models in generating the original system? (g) what kinds of modes can be complementary, supplementary, or alternative to each other in conveying software design rationale to end users?

All of the above cannot but hint at some research avenues in the future of multimodal designer-to-user communication via computer programs. Our modest experiment has provided us with an instant perspective on exciting possibilities involved in HCI, and given us a manageable platform on which to start testing the complexities of multimodal rhetoric and grammaticalization of interface languages. In the future, we shall be pursuing the connections between end user programming and direct acquisition of knowledge from users (de Souza and Edmonds,1997) to check the constraints imposed by designers when they first select representation language primitives for software, a problem that is not but a generalization of the problems encountered in the experiment reported in this paper.

Acknowledgements

We would like to thank students of the 1994 graduate class of Interactive Computational Linguistics (PUC-Rio, Departamento de Informática) for producing components of the layout system prototype, and all the subjects who have agreed to participate in the experiment, lending us their time and support.

References

Andersen, P.B. (1990) *A Theory of Computer Semiotics.* Cambridge. Cambridge University Press.

Arens, Y., Hovy, E.H. and Vossers, M. (1993). On the Knowledge Underlying Multimedia Presentations. In Maybury (Ed.) *Intelligent Multimedia Interfaces*. AAAI Press, Menlo Park, pp. 328-354.

de Souza, C.S. (1993). The semiotic engineering of user interface languages. *International Journal of Man-Machine Studies*, **39**, pp. 753-773.

de Souza, C.S. and Edmonds, E. (1997) *End User Programming meets Knowledge Acquisition*. Unpublished manuscript.

de Souza,C.S.; Prates,R.O. and Varejão,F.M. (1995) On Classes and Costs of Multimodal Interfaces. in Lucena,C.J.P. (Ed.) *Monografias em Ciência da Computação*. Departamento de Informática, PUC-Rio. **MCC 12/96**

Eco, U. (1988) On Truth: A Fiction. In Eco, Violi and Santambroggio (Eds.) *Meaning and Mental Representations*. Bloomington,In. Indiana University Press.

Garcia, A.C.B. and de Souza, C.S. (forthcomming) ADD+: Including Rhetorical Structures in Active Documents. To appear in *AIEDAM - Special Issue on Design Rationale*. London. Academic Press.

Hovy, E. (1988) Planning Coherent Multisentential Text. In *Proceedings of 26th ACL Annual Meeting*. New York State University at Buffalo. Buffalo,NY. ACL-NYSU. pp. 163-169

Ioerger, T.R. (1994). The Manipulation of Images to Handle Indeterminacy in Spatial Reasoning. *Cognitive Science*, **18**, pp. 551-593.

Kammersgaard,J. (1988) Four different perspectives on human-computer interaction. *International Journal of Man-Machine Studies*. **28**. pp.343-362

Kirsch, D. and Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, **18**, pp. 513-549.

Maybury, M. (Ed.). (1993). *Intelligent Multimedia Interfaces*. AAAI Press, Menlo Park, pp. 328-354.

Myers, B. (1992) *Languages for Developing User Interfaces*. Boston, Ma. Jones and Bartlett.

Nadin, M. (1988) Interface Design and Evaluation — Semiotic Implications. In Hartson and Hix (Eds.) *Advances in Human-Computer Interaction.* Vol.2. Norwood,NJ. Ablex.

Nardi, B. (1993) *A Small Matter of Programming*. Cambridge,Ma. MIT Press.

Prates, R.O. and dos Santos, A.L.S.C. (1994). Program developed for Final Project for Interactive Computational Linguistic Course. Departamento de Informática, PUC-Rio.

Stenning, K. and Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation. *Cognitive Science*, **19**, pp. 97-140.

Svendsen, G.B. (1991). The influence of interface style on problem solving. *International Journal of Man-Machine Studies*, **35**, 379-397.

For reference, please use:

de Souza, C.S.; Prates, R.O.; Varejão, F.M. (1997) Multimodal Communication between Software Designers and Software Users. In: III Workshop em Sistemas Multimídia e Hipermídia no Brasil (WOMH '97), 1997, São Carlos. Anais do III Workshop em Sistemas Multimídia e Hipermídia no Brasil, 1997. p. 93-105.