Design issues in a semiotic description of user responses to three interfaces

(Short Title: Semiotics, users and design)

KECHENG LIU, GEOFF CRUM, KRISTIAN DINES

Appeared in Int. J. of Behaviour & Information Technology, Vol 17, No.3, pp175-184, 1998

Abstract

Semiotics has been used as a theoretical basis for systematising the increasingly interdisciplinary nature of human-computer interaction. This paper focuses on the descriptive qualities of semiotics which are useful for understanding the user perspective of computer systems. The end goal is to assess the application of such semiotic description in interface design. An experiment is presented in which volunteers were asked to use and describe three interfaces created with the Apple Macintosh product Hypercard. A semiotic interpretation is given to their responses based on six principles taken from the relevant literature. The evaluation is then used to lay down a set of general guidelines for interface design.

1. Introduction

Semiotics is the study of signs. John Locke first used the term at the end of the seventeenth century to cover one of the three branches of science but its main founder is the American philosopher Charles Sanders Peirce (1931/1958). Charles Morris (1938) summarised 'semiotics' as a study covering three disciplines of logic, mathematics and linguistics. The study of signs is the cornerstone of structuralist thought initiated by Ferdinand de Saussure in his linguistic philosophy and this has been reapplied by later structuralists to other disciplines; for example Claude Levi-Strauss to anthropology, Jaques Lacan to psychology, Roland Barthes to literature (relating myth and narrative), and Michel Foucault to, among other things, history, politics and archaeology (Hoy 1994). Other academics, not necessarily of the structuralist tradition, have used semiotics as a basis for describing particular communication systems such as psychology (Mahl & Schulze 1964), psychiatry (Oswald 1964), music (Doubravora 1988), architecture (Guerri 1988), the relation between literature and psychology (Greimas 1990), education and learning (Houser 1994 and Macfarlane 1990) and the social implications of linguistics (Halliday 1978).

Recently the subject of information systems has received similar attention. In particular, semiotics has been identified by a variety of writers as a useful way of understanding the computer interface because of the computer's nature as a very special kind of communication medium. It is both an artefact in the sense of holding a set of messages left by the designer and an interactive tool which the user can communicate with. The interactive capacity of the interface makes it far closer to spoken language than other methods of communication such as literature or art, yet the restricted set of responses laid down by the designer puts the computer into a special category of its own. For this reason it is an exciting area for semiotic study.

From a computing perspective a semiotic approach is useful because it provides a framework in which to reconcile the perspectives of the system held by both designer and user. Conventional analyses, derived from human-computer interaction (HCI) practice, see the interface in design or engineering terms, and user responses are subsumed into those conventions. However, so far semiotic contributions have tended to be quite theoretical, and where practical applications have been examined their treatment has tended to be quite cursory. This paper attempts to give greater attention to a more specific practical example - the Apple product Hypercard. Though Hypercard is not new to semiotic work (Andersen 1990b), the intention here is to concentrate on three specific interfaces created with the package. Various semiotic principles will be adopted to structure an understanding of user interpretations of the three systems in a short experiment. An assessment of the results will then be made in order to draw out some helpful guidelines for interface design.

2. Theoretical Background

The recent interest in semiotics and interface design has been influenced greatly by the work of Peter B. Andersen (1990a, 1990b). As a result, much attention will be given to his theory in this paper. Reference will also be made to a paper by de Souza (1993) which adds an important perspective to Andersen's framework. The discussion will be rounded off with reference to related articles by other writers which, though not directly about semiotics, address related issues.

2.1. *Andersen's Theory Of Computer Semiotics*

Andersen (1990a) takes as one area for his source material the linguistic school of Structuralism, in particular the Structuralist theory of Glossematics laid down by Louis Hjelmslev (1963). Structuralism asserts that what is fundamental in a system of signs is the meaning which those signs create. This meaning derives from the interrelation of the signs in a system. Meaning cannot be based on a sign in isolation - its meaning depends on other signs in the world. Glossematics uses a formalised method of analysing such structural systems. These are broken down into smaller and smaller `units'. The justification for treating a unit as such is based on its relative independence with regard to other potential units in the text and the relations it contracts with them. If part of the system seems to function as a unit then it should be treated independently and meaning relations can be examined by looking at the way it relates to others. The units continue to be broken down using the same principle until the final `elements' can no longer be further analysed. These units at the lowest level are known as taxemes (Andersen 1990a). They are not atomic particles but rather dictate the most simple of semantic oppositions such as `possible/not possible'. Andersen applies this method of analysis to computer interfaces. At the highest level is the task. The task is independent in its own right unlike more elementary components. A task is comprised of actions. Actions are dependent on one another to function. They therefore provide little room to manoeuvre for the user - for this reason Andersen posits that the larger a task (i.e. the more numerous its actions) then the less flexible a system is; the more elementary a task (the fewer its number of actions) - the more plastic/free its manipulation. Within actions are indicators. These show what can be done and what has been done. They are the smallest units in a sign system

from which the larger context of tasks can be interpreted. An example of an indicator is the current form of an icon depending on the process selected (perhaps it changes shape during a certain function). Below indicators are the elementary semiotic units - the taxemes. These are the basic semantic oppositions, binary values, from which the indicators are built.

The notions of ‘form’ and ‘substance’ are highly relevant to human communication. Andersen focuses, for instance, on the `form' of the unitary sets within a system. Form in this context means the `essential' meaning that distinguishes a unit from another unit. Substance can best be distinguished here as that part of the expression of a sign which does not contribute to the sign's essential meaning but merely acts as decoration or is incidental in its occurrence. Andersen uses as an example the flipping of a coin. The essential meaning is derived from the opposition of heads and tails. Incidental to this is the expression of whether the coin falls in the centre or at the edge of the table. This would be called the substance.

The form/substance paradigm is very important in interface design. For Andersen, what is fundamental to a system is that it expresses processes which it is designed to allow a user to recognise and manipulate. Interface design then, cannot be separated from functionality. Form represents the functions of different symbols on the screen - functionality being the essential meaning (purpose of a system function). Incidental to this are the decorative effects - the substance that gives the look and feel to an interface.

In addition to the form/substance distinction Andersen (1990b) examines the relationship between configuration and process. He likens this opposition to two types of `chain' - a concurrent chain, which represents signs available in a system at any single point in time, and a sequential chain, which shows the sequence of events that results from the transformation of an individual sign. Owing to Andersen's emphasis on function as the essential part of a system sequential chains are promoted as the most fundamental parts of an interface. The concurrent chains provide the environment in which these occur within the context of the Hypercard system (Andersen 1990b). Sequential syntagmata (patterns of sequential chains) refer to the paths through a stack. Concurrent syntagmata (patterns of concurrent chains) refer to signs that occur simultaneously on the same card. Andersen describes this subject in terms of a discussion of narrative structure as opposed to a simple description of the links between cards. The choice between concurrent and sequential syntagmata is derived from the nature of the subject matter that is going to be represented - that is its plot or `story line'. Some subjects will clearly have a stronger narrative slant and will be mainly sequential while subjects with less linear structure will have a network of interrelating cards, with fewer constraints on sequential movement through the system.

Andersen places the ideas given already into a framework of semiotic principles for the proper programming interfaces and then categorises different types of sign and genres of sign use. The first principle is perceptibility. This dictates that all functions of a system must be represented at the interface. The second is sign morphology which states that signs are either permanent features (such as the overall appearance of the Macintosh desktop which remains fairly static) or transient features (such as the visual effect of zoom in the Macintosh environment), or control features. The third principle is that actions belong to signs - signs perform actions but it is not possible to perform an action if it is not represented by a sign. Forthly sign classes should be defined by their combinatorics - they can be combined to form more

complex signs and can be broken down into more elementary signs. The fifth principle, which is concerned with form and substance, says that concepts are relational - the form of a sign is not represented by absolute values but is relative to the form of other signs around it.

2.2. *De Souza's Homomaterial/Heteromaterial Distinction*

De Souza (1993) gives a very comprehensive framework in which certain semiotic principles are applied to interface design. She adopts four parameters proposed by Eco in his Theory of Sign Production (TSP) (Eco 1977), and then maps these to four guidelines for interface design. The first guideline suggests that designers of user interface language should use the signs in the same way as they are codified in the intended application context. The second one proposes that in the user interface design one should make use of the indexical relationship between signs and their referents (for instance, smoke is an index of a fire). The third one is concerned with the heterogeneity of domain objects and computer-modelled solutions and the homogeneity of representation for direct manipulation. The last guideline asserts that users' culture and the social norms are the ultimate rules that will govern the use of signs created in the user interface, and therefore any user interface design have to follow these rules.

The third guideline about the link between domain object, computer-modelled solutions and representation has particular relevance to the user interface design in a Hypercard environment. The guideline is derived from Eco's principle of the importance of the medium selected for the type of meaning being communicated. This medium is either heteromaterial - the sign medium is of a different representation system to the object being identified, or homomaterial - the sign medium is of the same representation system as the object. In her guideline for interface design de Souza notes how functions which are enabled solely by the computer environment should be expressed as homomaterial signs. She gives as one example a selected text block in a word processor. The text can be cut and pasted as a single object, which is an extremely easy task in the computer environment and so is understood as a sign referring to that medium. Functions on a general level which mirror to some extent real world behaviour can be represented heteromaterially, as in the use of the desktop analogy.

Andersen highlights the need to view adherence to real world analogies in computer interfaces with caution (Andersen 1992). He claims that sometimes departure should be made from transitional forms, not only because those can not be as adequately represented by the computer but that the new potential of the computer can not be exploited. Areas which he identifies as being more wholly computer based, with the potential for innovation, include the interactive capability of computer systems, the combination of different forms of media, and special narrative techniques which can be implemented in the computer environment but no where else. He cites as an example of the latter a moveable viewing window or `spot'. Only through this can the contents of the screen be seen clearly and the spot can be moved with the mouse so as to reinforce the feeling of exploration - an effect which can be achieved in a computer system but not in a more traditional medium such as a book.

2.3. *Specialised Approaches*

Other writers have contributed to semiotics (either explicitly or implicitly) in specialised areas. A small selection of these is given here.

2.3.1. *Icons:* Familant and Detweiler (1993) attempt to give a precise definition of icons and the relation between an icon and the `thing' it represents. They first distinguish icons from other signs such as indices (signs left by other signs), and symbols whose expression only conforms to the signified object accidentally - the choice of symbol for an object being arbitrary in terms of logical entailment, it is fixed instead wholly by cultural convention. Social influence and convention does affect iconic meaning. An iconic relation between signal (the physical thing carrying the encoded sign) and denotative referent (the thing being referred to) is a mapping of the signal's `feature set' to the `features found in the referent's feature set'. The choice of features is socially determined.

The authors then proceed to give a taxonomy of different types of icon. The main distinction is between direct and indirect reference. Direct reference involves a single denotative referent. Indirect reference involves at least two referents - the sign referent and denotative referent. The sign relation occurs between signal and sign referent (what the signal directly represents). A further referent relation occurs between sign referent and denotative referent. An example of indirect reference is in a movie or a television programme that a gun shot ‘bang’ may be associated with a death of someone in the show. But does the reference stop there in the show?

Blakenberger and Hahn (1991) show experimental evidence of some of the effects of icon design. They reveal that screen position of icons is one of the most influential features of icon design affecting user response times.

2.3.2. *Inconsistency:* Reisner (1993) develops a framework in which to explain inconsistency in interface design which she calls Agent Partitioning Theory (APT). The focus is placed on the way the designer and user employ different assignment rules. Differences create inconsistency, which Reisner cites as the difference between peoples' views about which things are similar in a system.

2.4. *Summary Of The Semiotic Principles For Interface Design*

Six main semiotic principles associated with interface design can be summarised from the discussion so far: glossematic hierarchical analysis of structure, form/substance distinction, configuration/process distinction, homomaterial/heteromaterial distinction (the world of reference), taxonomy and feature mapping (iconic representation) and the APT explanation for inconsistency. Structure (identified by glossematic analysis and the configuration/process distinction by Andersen) and consistency can be grouped together because consistency establishes structure (Tero & Briggs 1994). The subject of structure and consistency has a direct effect on the navigation in a hyperspace, and is often cited as reinforcing user navigation strategies in a system (Payne & Howes 1992, Wright & Lickorish 1994). Determination of what is and what is not functional in a system and what simply adds aesthetic effect is explained by the form/substance distinction, which will be covered by the heading of aesthetics in later sections.

3. The Experiment

A research project was conducted in which the semiotic principles raised in the previous section were used to interpret user responses to three Hypercard stacks.

3.1. *Aims Of The Experiment*

The experiment aimed at testing the semiotic principles through the implementation and critical analysis of a system interface. Specifically it attempts to examine whether the principles provide greater precision in interpreting the rather vague language users often adopt in their responses, and to analyse how the results reflect some of the theoretical assumptions. Finally a summary of design issues and guidelines would be drawn from this evaluation of the results.

3.2. *Design Of The Experiment*

The experiment was conducted with ten volunteers who were all under-graduate, computing science students, and all familiar with the use of Hypercard. Each volunteer was asked to give his/her responses to the three Hypercard stacks by an assessor who followed the instructions and questions laid out in the questionnaire. The user could only use the mouse cursor to select buttons on a card, the keyboard was not used. The testing of the responses of each volunteer lasted approximately twenty minutes.

The computing resources used for the experiment were Apple Macintosh LC II computers with Hypercard 2.1. The Hypercard was chosen because of its popularity at the time and its capability of managing different data types such as text, graphics and sound. In the Hypercard, a stack is the largest unit in a program and is made up of one or more cards. Each card can be thought of as a single screen of information and consists of fields of text, graphics and buttons. Buttons provide the main functionality of Hypercard.

Each of the programmed interfaces in the experiment is a stack of 42 cards. Each gives a representation of a set of facts taken from the Encyclopaedia Britannica (Anon 1979). These cover the history of the Germanic peoples. The subject area has been chosen because it consists of several interrelated categories which can be used to take full advantage of Hypercard's button linkage facility. The categories represented are: language, religion, politics, economics, distribution and law. Within these are several subcategories so that a degree of hierarchical layering can be incorporated into the programs. The three stacks are called Germania, Germania2, and Germania3 respectively. Germania and Germania3 share the same structure of card links and are different only in terms of cosmetic effects - Germania employs a book metaphor so that each card looks like two facing pages of an open book. Germania2 shares the book metaphor employed by Germania but has a more complex structure to allow more complicated navigation. On the other hand, Germania3 employs a homomaterial approach - the appearance of the cards is more obviously part of a computer object because Germania3 does not use a real world analogy such as a book to represent its functionality.

The three stacks employed different visual effects associated with the links between the cards and users' comments about them seem to reveal an awareness of structure. In Germania the first card took the appearance of a closed book and could be `opened' with a button taking the user to the contents card. This function was accompanied by Hypercard's `barn door open' effect in which the screen parts form a

vertical at the centre to reveal the next card. A close button on every other card took the user back to the `closed book' card accompanied by the opposite effect. Every other button utilised a `wipe' effect to simulate turning the page of a book. Germania2 shared these features except that some of the wipes were replaced with `scroll' effects (where the whole screen showing the next page was `rolled' over the current one). Germania3 employed different open/close visual effects and the other buttons used zooms characteristic of the Macintosh desktop.

3.3. Contents Of The Experiment

The experiment was guided by structured questionnaires. Each volunteer was asked to describe what he or she thought of the three stacks from a questionnaire. This gave instructions to the volunteer about which cards to access and asked questions about what the volunteer thought of the effectiveness of their functionality and aesthetic layout. These questions were grouped under the four semiotic sections, following the semiotic principles summarised in the last chapter.

The first section, Navigation, covers basically two issues: structure and consistency. The purpose of the section was to test the users' responses to the structure and consistency of the stacks. The issue of structure is analysed into the components with Glossematics and is perceived by the user as a question concerning the combination of configuration of the stacks and the process of using the cards. The issue of consistency reinforces the question of structure. The second section is Iconic Representation, which concerns taxonomy and feature mapping. The questions in this section gauged volunteer responses to the effectiveness of icons represented on the cards. Questions about aesthetics are found in the third section which concerns the form/substance distinction. The questions aim to assess volunteer responses to visual features of the three stacks - these included 'visual effects' (animated sequences provided by Hypercard which can occur when a user moves from card to card), and screen design. The last section on the `world of reference' refers to the homomaterial/heteromaterial distinction drawn from de Souza's paper. The book analogy can be viewed as a heteromaterial reference to an object in the `real world' used as an analogy for the purpose of the Hypercard stacks.

3.4. *Results and Evaluation*

The results of the experiment are summarised in table 1 and are evaluated under the four semiotic categories.

3.4.1 *Navigation:* Volunteers rarely made explicit references to structure in each stack. One volunteer likened the stacks to `tree structures' of screens. He observed this fact when using a button commonly available on a lot of the cards which employs the icon of a back pointing hand. It takes the user back to the previous card. However, implicit reference to structure was made by volunteers who also recognised that the back pointing hand took them to the previous screen. There appeared to be an awareness of where cards were in relation to one another and the function of the back pointing hand seems to have reinforced this. Several volunteers commented on the need for extra buttons leading to specific screens to aid navigation. This was because they found difficulty in using the hand icons in terms of remembering where they had come from (the hand not giving any indication of this). Presumably these users also

wanted a freer means of navigation - showing a bias towards the constellation type of syntagma as opposed to the more constrained subordination and interdependence relationships raised by Andersen. Moving backwards and forwards one card at a time is a more constrained way of navigating through a stack than being able to skip several cards in its structure to get to a particular destination.

Difficulties of navigation revealed evidence of both structure and consistency. Extra functionality (in the form of additional buttons and links) was added in Germania2 and the new options available flouted some of the conventions in the other two stacks. This confused some users whose expectations of Germania2 were based on button positions set up in Germania. Here is a good example of Reisner's APT theory taking effect - the user's and designer's view of the functionality of particular signs differs so that errors are made. It also shows evidence of the importance of button position cited by Blankenberger and Hahn (1991).

One volunteer observed that the visual effects of zooming into or out of a button reinforced where the user was coming from and going to, so this is some evidence of visual effects contributing to a realisation of structure. Another volunteer observed how the open/close effects in the Germania stack were differentiated from the wipe effects in the same stack, thereby reinforcing the different types of screen - perhaps what she was indicating was the reinforcement of the exterior and interior of the book. One volunteer noted that the scroll effect appeared to move a second book on top of the first, but that functionally all the screens were actually part of the same book - this perhaps shows an awareness of the cards being part of the same ‘unit’ or stack.

Andersen's Glossematic structure of computer interfaces also appears to manifest itself in the results. Evidence of the effect of taxemes on users can be seen for instance in the difficulty certain users found in hitting a button with the mouse pointer. A taxeme opposition which might be the focus of this problem could be target area/nontarget area. This opposition shows the user where they can click to trigger the function of a button. They gradually build up a picture of where the target area is depending on a combination of other taxemes and indicators. This can be revealed in the comments of users - for instance, one said that bold text seemed to differentiate button labels from other text on the screen. Bold text of a particular size can thus be viewed as an indicator which helps locate the target area. The taxeme target area manifests itself when it `works', that is something happens when the mouse button is clicked and the pointer is in the right area. Actions and tasks were contained implicitly in the instructions given to users in the questionnaire.

3.4.2. *Iconic Representation:* What is looked for here is a mapping of the users' descriptions of the icons to the semiotic descriptions of icons in terms of taxonomy, feature set mapping and so on covered in the theory laid out by Familant and Detweiler. Evidence of feature mapping seemed to manifest itself at three levels: *appropriate mapping* - where the volunteer thought the icon had those features he felt were necessary to represent the subject; *deficient mapping* - where not enough features were included; and *wrong mapping* - where the whole icon was inappropriate to the subject. Mapping in all three categories can be found with the use of icons on the main contents card of the Germania stack. For an critical analysis of the (in)proper use of the icons and their ‘iconicity’ these icons will be presented in Figure 1, as they appeared on the card.

Appropriate feature mapping was not consistent for any icon - perhaps revealing the subjective nature of user interpretation - different people held different conceptualisations of what features made up what objects. For instance, nine out of ten volunteers thought the Language icon of the contents card was satisfactory yet one said he would find difficulty knowing what the icon conveyed if the button label had not also been present. Similar results were obtained for the Law icon. The few reservations that did occur took the same tone. The Law icon was described by one volunteer as being `a bit weak', another said that it was ambiguous and another that it could represent writing rather than law. These criticisms show two factors - a deficiency of the number of features used in the representation and a use of the wrong types of features. The different responses to the icons indicate a slide from `good' to `bad'. The language icon, for instance, would occur near the positive end of this scale. The features of its signal show a face and bubble containing lines. This is a conventional image (albeit stylised here) used in cartoons to represent a person speaking. The language icon is an example of indirect reference in which 'a person speaking' is the sign referent. The function of speech which this referent possesses is also shared by the denotative referent `language' and so there is a strong link between the sign and denotative referents. The convention of the cartoon speech bubble issuing from a face is also strong reinforcing the link between signal and sign referent.

The Law icon would occur closer to the opposite end of the scale. It shows a hand writing on paper. The sign referent is `writing' and the shared function with law is the fact that law is commonly written down. However, the function of writing is a denotative referent in itself and for some volunteers was stronger than `written law'. In other words for these volunteers the features of the law icon were the `wrong' ones to represent the subject of law. For one volunteer the icon could have been improved by replacing the piece of paper with a scroll or will, adding an extra feature shared by the feature set of law. This shows evidence that the volunteers' interpretation of the Law icon as a bad icon was based on deficient mapping rather than a wrong mapping of features.

One icon which all volunteers had difficulty with was the Distribution icon. They could not understand what the subject being represented was unless they then went to the Distribution card and read the introductory text there. The subject, according to that card, is: `the geographical distribution and migration of the different Germanic tribes across northern Europe'. The icon consists of two denotative referents - population distribution and migration. This seems to suggest that multiple reference appears to be harder to grasp than when a single dominant denotative referent is used.

3.4.3. *Aesthetics:* The distinction between form and substance is not clear cut, rather the two categories tend to `grade' into each other. Different volunteers vary on how useful, functional, they find a particular image or visual effect. Certain images, however, appear to be more universally accepted as contributing to the functionality of the program. In the light of this the form/substance distinction discussed by Andersen can be viewed as occurring on a scale. At one extreme, `pure form', occur examples such as button position (already shown to invariably be associated with function by volunteers). Below this are features of each stack whose usefulness depends on the interpreter. Almost all volunteers agreed that the book image was `effective', preferring it to no book used in Germania3 (Figure 2). Some claimed this was because the book helped identify the purpose of the program, that it was a

`knowledge base' conveying information. However, others said that though they preferred the use of the book image it provided no real function.

The visual effects of Germania and Germania2 helped to reinforce the book analogy for some users. It could be claimed that the effects indirectly helped to reinforce the overall function of the stack. The volunteer who observed the differentiation of the open/close effects from the turn page effect perhaps showed how opposition created meaning (or at least reinforced it) just as in Andersen's theory the recognition of function is dependent on the opposition of form at the interface. However, once more, others said that the visual effects had no real functional purpose.

At the lowest end of the scale (nearest to substance) are features which tend to be missed by users unless prompted to describe them. A good example of this is the difference between the hand icons and arrow icons (the latter replacing some of the hand icons in Germania2 but still retaining the same functionality). One volunteer identified one of the essential features of form of these icons - namely that in both cases shape was that of a `pointer' pointing in the direction to be taken through the screens. The shape of that pointer was fairly insubstantial - even those who said they preferred one or the other admitted this.

What appears to emerge are three levels of functional involvement. The highest level is the most functional, the lowest, the least functional. The level in between occurs when there is a reinforcement of function, but that if removed the function would still be represented by another sign. The divide between the middle and lowest levels is presumably dependent on the extent of the contribution of the sign to its function, whether other signs representing the same function are very strong, whether the sign itself is particularly eye catching and so on.

3.4.4. *World Of Reference:* The use of the book analogy is heteromaterial - it uses a real world object to represent functionality that occurs within the world of the computer. The way the stacks operate actually differs from a book and so, theoretically, the homomaterial representation of hypertext in Germania3 should be preferred by users. However, this did not appear to be the case. The book was seen as useful for reinforcing the purpose of the program. One volunteer even suggested using zooms with the book analogy - showing that at least for him, the complete realism of the book analogy was not necessary. His comments are interesting in the light of Andersen's own remarks about exploiting the full potential of the computer environment. The zooms reinforce the location of the button from which a card is accessed where as the wipes contribute no such extra functionality.

Only two volunteers showed any sign of disenchantment with the book analogy and this was restricted to visual effect - both thought that the visual effects were not realistic enough - one claiming that a book `does not open like that', the other that the wipe was not a realistic depiction of a page turning. This is not strong support for de Souza's homomaterial/heteromaterial distinction - but perhaps further testing on more restricted examples is needed. The problem with the Germania stacks is that they are quite complex - the book analogy enforces their overall purpose and this perhaps overrides more specific examples of functionality and form which do not adhere to the `behaviour' of a book in the real world.

4. Issues Arising From The Experiment

This section presents an assessment of the experimental procedure and gives some guidelines for designers which arise from the evaluation of the results.

4.1. *Assessment Of The Experimental Procedure*

The method was fairly structured and this probably influenced user responses where they might not have commented otherwise. However, the approach was taken to avoid volunteer reticence and inconsistency of navigation through the stacks between subjects. The choice of subjects, all computing students but with some variation of background, had a significant effect on results. The ability of subjects to articulate their views of the interfaces varied according to the extent of their exposure to computing. The participants are computing students in different years. Those who had more computing experience tended to describe aspects of the system in more explicit computing terms. Other backgrounds influenced responses in other ways, particularly in the way the subjects interpreted icons. The cultural variation obviously created problems for keeping the experiment as consistent as possible, but it did provide interesting results which highlighted the effects of social background on the interpretation of signs - a major concern for several of the theories introduced earlier.

4.2. *Guidelines For Designers*

These will be discussed under the four categories applied to the questions and results of the experiment.

Navigation: An implicit user awareness of structure was manifested in the results of the experiment. Both consistency and structure were shown to be closely interrelated and they established user understanding of the system's functionality. Designers should therefore make structure as consistent as possible so that functionality is reasonably predictable - otherwise the user's view of the system may deviate from that of the designer - leading to errors being made. The use of visual effects, imagery and button positioning should also be consistent, which will help to reinforce user awareness of structure.

Iconic Representation: Focus should be placed on the correct mapping of feature sets between signal and referent. Users' culture can influence the interpretation of signals used in an interface, because certain meanings may be assigned to the signals in a cultural context. The choice of signal and its features should take account of the user culture - a designer needs to pick features which are as universally acceptable to the target users as possible. One way of achieving this would be to investigate the users' culture and habit by interviewing a selection of users in order to discover what is and what is not an acceptable signal. The semiotic principles can assist in interpreting the answers taken in the experiment to understand the nature of the problem. For instance, a user who says that an icon is `weak' may actually mean that `the icon is a signal with an insufficient number of features to indicate the denotative referent'.

Difficulties surrounding unusual subjects to be represented can be avoided, where possible, by recategorisation of the subject matter - splitting up the subject being covered into smaller components to avoid more than one obvious denotative referent being represented.

Aesthetics: The scale of the effect of the form/substance distinction has been shown to be ill-defined. However, some awareness of the relative degrees of representation along it will aid the designer to utilise features of appearance appropriately where needed. Consistent and obvious button location, for instance, is a good way of expressing an important function because this seems to have the largest influence on the scale. At the opposite end, superficial features can be added to make a system more attractive to the eye. The bottom line is to identify those aspects of form which express functionality at the interface. It is these features which must be highlighted the most to aid a user's understanding of a product. Consistency of substantial aspects is aesthetically pleasing but is not so important for expressing functionality.

World Of Reference: The real world analogy of the book in Germania and Germania3 was appreciated on a general level by users. It reinforced the purpose of the stacks, that is, to represent data, just as a book is written to hold information. It is only where the specifics of such an analogy become unworkable in the computer environment that they should not be used - but this has been shown to be a very subjective matter. As with iconic representation, utilisation of analogy can be made more effective by interviewing potential users - gauging the extent to which realism matters in representation.

5. Conclusions

A presentation of the application of semiotic theory to computing by different writers has been given in this paper. A research project was conducted in which the principles raised there have been used to assess the interpretations of three Hypercard interfaces that were specially designed for the experiment. From the research, several guidelines have been produced for designers, showing how user expectations can be governed by semiotics and divided into different categories. However this taxonomy may not be precise enough as, for instance, structure and consistency being closely associated with one another and the identification of form in the aesthetics category being reinforced by their application. The overall aim has been to incorporate a self-consciousness in evaluating the results. It is hoped that this self-consciousness can be translated to interface design as has been attempted in the four guidelines. It may also equip designers with tools for understanding more precisely the language users adopt in describing a system - either in interviews during the initial design and implementation stages - or in later suggestions which are used to create an updated version.

The approach taken in this paper has been informal. The ideas raised from the research can be used as heuristic principles to assist designers' in 'decision-making' during the interface design. However, the principles should be viewed as helpful pointers rather than fixed rules. A more formal design method may eventually be developed but in such a subjective area the descriptive concerns taken here may still prove very useful.

References

ANDERSEN, P. B. 1990a, *A theory of computer semiotics* (Cambridge University Press, Cambridge).

ANDERSEN, P. B. 1990b, Towards an aesthetics of hypertext systems: a semiotic approach, in A. Rizk, N. Streitz and J. André (eds), *Hypertext: concepts, systems and applications. Proceedings of the European Conference on Hypertext, INRIA, France.* (Cambridge University Press, Cambridge).

ANDERSEN, P. B. 1992, Vector spaces as the basic component of interactive systems: towards a computer semiotics, *Hypermedia*, **4**, 53 - 76.

ANON 1979, *Encyclopaedia Britannica Macropaedia*, **8**, 15 - 44, (University of Chicago Press, Chicago).

BLAKENBERGER, S. and HAHN, K. 1991, Effects of icon design on human-computer interaction, *International Journal of Man-Machine Studies*, **35**, 363 - 377.

DE SOUZA, C. S. 1993, The semiotic engineering of user interface languages, *International Journal of Man-Machine Studies*, **39**, 753 - 773.

DOUBRAVORA, J. 1988, Denotation and connotation of fugue in music and visual arts in the twentieth century, in M. Herzfeld and L. Melazzo (eds), *Semiotic theory and practice, Proceedings of the Third International Congress of The IASS Palermo 1984* (Mouton de Gruyter, division of Walter de Gruyter & Co., Berlin).

ECO, U. 1977, *A Theory of Semiotics* (The MacMillan Press, London).

FAMILANT, M. E. and DETWEILER, M. C. 1993. Iconic reference: evolving perspectives and an organising framework, *International Journal of Man-Machine Studies*, **39**, 705 - 728.

GREIMAS, A. J. 1990, *Narrative semiotics and cognitive discourses* (Pinter Publishers, London).

GUERRI, C. F. 1988, Semiotic characteristics of architectural design based on the model by Charles Sanders Pierce, in M. Herzfeld and L. Melazzo (eds), *Semiotic theory and practice, Proceedings of the Third International Congress of the IASS Palermo 1984* (Mouton de Gruyter, division of Walter de Gruyter & Co., Berlin).

HALLIDAY, M. A. K. 1978, *Language as social semiotic: the social interpretation of language and meaning* (Edward Arnold, Arnold).

HJELMSLEV, L. 1963, *Prolegomena to a theory of language* (The University Winsconsin Press, Menasha, Winsconsin). Translated from `Omkring Sprogteoriens Grundlaeggelse'. University of Copenhagen 1943.

HOUSER, N. 1987, Toward a Peircean Semiotic Theory Of Learning, *The American Journal Of Semiotics*, **5**, 251 - 274.

HOY, D. C. 1994, Introduction, in D. C. Hoy (ed), *Foucault: A Critical Reader* (Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, UK).

MACFARLANE, A. G. J. 1990, Interactive Computing: a revolutionary medium for teaching and design, *Computing & Control Engineering Journal*, 149 - 158.

MAHL, G. F. and SCHULZE, G. 1964, Psychological research in the extralinguistic area, in T. A. Sebeok, A. S. Hayes and M. C. Bateson (eds), *Approaches to Semiotics* (Mouton & Co. Publishers, The Hague, The Netherlands).

MORRIS, C. 1938. *Foundations of a Theory of Signs* (University of Chicago Press, Chicago).

OSWALD, P. B. 1964, How the patient communicates about disease with the doctor, in: T. A. Sebeok, A. S. Hayes and M. C. Bateson (eds), *Approaches to Semiotics* (Mouton & Co. Publishers, The Hague, The Netherlands).

PAYNE, S. J. and HOWES, A. 1992, A task-action trace for exploratory learners, *Behaviour & Information Technology*, **11**, 63 - 70.

PEIRCE, C.S. 1931/1958, *Collected Papers* (Harvard University Press, Cambridge).

REISNER, P. 1993, APT: a description of user interface inconsistency. International *Journal of Man-Machine Studies*, **39**, pp. 215 - 236.

TERO, A. and BRIGGS, P. 1994, Consistency versus compatibility: a question of levels?, *International Journal Of Human-Computer Studies*, **40**, 879 - 894.

WRIGHT, P. and LICKORISH, A. 1994, Menus and memory load: navigation strategies in interactive search tasks, *International Journal of Human-Computer Studies*, **40**, 965 - 1008.

Table 1. The experiment results.

Category	Question	% of Users Responded
Navigation	Explicit awareness of structure	10
	Shape inconsistency, correct generalisation	100
	Location inconsistency, correct generalisation	80
	Reinforcement of visual effects on structure	20
	Reinforcement of analogy on structure	80
Iconic Representation	Language button of contents card as a good icon	90
	Law button of contents card as a good icon	50
	Distribution button of contents card as a good icon	0
	Correct identification of subject represented by Distribution icon	0
Aesthetics	Book analogy as reinforcement of function	30
	Book analogy as expression of function	0
	Visual effects as reinforcement of function	20
	Visual effects as expression of function	20
	Button shape as reinforcement of function	0
	Button location as reinforcement of function	20
World of reference	Positive response to book analogy	80
	Zooms as reinforcement of structure	10
	Book appearance as reinforcement of structure	10

Captions to figures

Figure 1. The main contents card of Germania. (The Germania and Germania2 Stacks use the book analogy. This figure shows the icons appeared on the card, and they are referred in the discussion.)

Figure 2. The main contents card of Germania3. (Following the homomaterial representation, this stack does not use the book analogy.)