1 INTRODUCTION

In this paper we present a general model of communication applied to the special case of dialogue. Our broad perspective aims to account for the many facets of human dialogue within a single theoretical framework. In particular, our project's aim of incorporating relevant non-verbal communicative acts from the person-machine interface makes it essential that the description of communication be sufficiently broad.

The model described here takes as its starting point the communicative utterance or act. It considers the higher-order structures into which communicative acts may be incorporated, but does not detail their internal composition. It is in this sense that the model provides a framework for the formal treatment of dialogue.

2 COMPONENTS OF THE MODEL

A full description of the adopted dialogue model has been given in Egan, Ferrari, Harper, et al. (1987). It relies on a double description of dialogue: a syntactic analysis of dialogue structure and a semantic-pragmatic description of the communication context. The basic units are:

- meaningful expression (ME): Any physical act carrying a non-contextual meaning;
- communicative act (CAct): An instance of ME issued by a specific "issuer" and received by a specific "receiver";
- communicative situation (CS): The CAct together with all the relevant facets.
- communicative situation structure (CSS): A larger aggregation of "CSs that provide a bridge into the intentional component of the dialogue model.

Each of these components is discussed in more detail below.

2.1 Communicative Acts and Dialogue Structure

The syntactic component of the dialogue model relies on the fact that, if we examine a dialogue or any other communicative exchange, it is possible to observe in the sequence of communicative acts, subsequences which follow regular patterns. These patterns can be catalogued in a form which expresses their significant regularities. This approach leads to a descriptive method very similar to the formal description of language in terms of a vocabulary of terminal symbols (the communicative acts), a vocabulary of auxiliary symbols (a collection of labels), and a set of productions (discourse patterns). Within the definition of a communicative act, provision is made for gestural information accompanying an utterance, such as a deictic gesture involving a mouse or some other pointing device (in the context of person-machine interaction).

The idea of treating discourse segments like phrases in a sentence is not new (cf. Burton, 1981). However, the nature of the entities involved is rarely fully clarified. In Christie, Egan, Ferrari, et al. (1983), a dialogue classification system was presented, based on the system of classification of Burton (1981). It consisted of a set of functional labels divided into the following five hierarchical levels, from lower to higher,

acts: {marker, summons, elicitation, reply, ...},
moves: {delineating, sketching, ...},
exchanges: {explicit, boundary, conversational, ...},
transactions: {exchange, ...},
interactions: {transaction, ...}

The labels at the act level are defined in terms of functional labels assigned to expressions, such as "starter", structurally realized by a statement, a question, or a command; "informative" structurally realized by a statement; "elicitation", structurally realized by a question. These, together with their functional definitions, represent a closed set of elements. The labels at a higher level are all defined in terms of patterns of labels of the immediately lower level. This set of rules may be regarded as the set of productions, which generates communications. In this way, a dialogue/communication is adequately described in terms of a formal generative grammar. An ATN-like grammar of dialogue in these terms has been described in Egan, Forrest, Gardiner et al. (1996) and Reilly (in press).

2.2 Communicative Situations

The semantic-pragmatic description relies on the notion of "communicative situation" (CS). A CS is a way of representing the communicative exchange together with its context. It consists of facets, which are aspects of the CS that occur with a certain regularity in all CSs of a given sort. Facets may be formally conceived of as "sorted regularities" in the scene where communication takes place, therefore a CS may be described as

CS = (f₁, f₂, ...)

where the subscripts identify the sort of the facet.

It is relatively easy to identify the sort of the more frequent regularities, such as who the issuer is (f₁), who the receiver is (f₂), etc., and to consider them as constituent elements of a CS, around which other facets become, from time to time, relevant.
Situation Semantics has been shown (cf. Egan, Ferrari, Harper, et al, 1987) to have some advantages for the representation and the treatment of a CS, provided that certain extensions and modifications to the original description of a discourse situation are carried out. In communication, since more than one, and often more than two, participants are involved, each with different attunements to the CS and different perceptions of what in Situation Semantics is called the speaker's connections, more than one classification of the same CS is possible. In the best case, where participants understand a CS in the same way, communication is successful, otherwise some failure occurs. In general we can assume, that participants in a communicative event are able first to classify, and then understand situations on the basis of the situation types they share. In the spirit of SS, we assume that these CS-types are the description of regularities observed in actual communications. An important consequence is that a new notion, relevance or relevant property, is established in terms of the more frequently observed regularities.

We can, then, describe the facets of the communicative situation in terms of properties of that situation, where the notion of relevance intervenes at two levels. At the first level, the set of properties is not defined a priori. Different properties are relevant to the interpretation of different utterances in different situations. Some of these are involved more frequently than others in the process of understanding, and may be considered more fundamental than others to a CS. These seem to be the roles of issuer, receiver, location, communication mode, illocution. The communication mode, i.e., whether communication happens face-to-face, by telephone, or in any other way, may affect both the form of the message and the referring expressions. By illocution, the traditional illocutionary force is meant, although a more fine-grained classification of such acts is intended (cf. Christie, Egan, Ferrari, et al, 1985). Also, other facets of a CS may occasionally become relevant to the understanding of an utterance.

At the second level, each property of a CS is taken to be a role participating in an intersecting set of regularities which qualify its sort. Thus, the property:

\[[x]\ll<1, \text{ saying, x, } \alpha\ll<1, \alpha>\ll<1]\]

describes some indeterminate x saying α, and participating in those situations where it is "regular" (nomie) that some x says α.

By further specification we can assume that

\[[\text{a-tourist}]\ll<1, \text{ saying, a-tourist, } \alpha\ll<1, \alpha>\ll<1]\]

participates in those situations in which x is of type a-tourist. In Harwise and Perry's (1985) notation this would be given as:

\[[x] \text{ in S: a-tourist, x, yes}\]

where S is the set of situation-types in which a tourist is involved.

Both properties and types classify real objects that become relevant to a discourse situation in accordance with the relations participants are attuned to. On the basis of this notion of relevance, it is possible to define a large set of types of properties which may or may not appear in one or the other CS. A receiver makes use of these classificatory devices to classify and understand any specific CS with which he or she is presented.

![Figure 1: Structural components of the model](image)

2.3 Communicative Situation Structures

The Communicative Situation Structure (CSS) is equivalent in level of analysis to the phrase segment of the Grosz and Sidner (1986) model. The three components of the CSS (see Figure 1) are the communicative act component (CAct), the communicative situation component (CS), and certain properties specific to the CSS itself. A CSS can consist of a number of CSs, and these in turn can consist of a number of CActs. The nature of CActs and CSs has already been discussed above.

A number of factors serve to distinguish one communicative situation from another. These can involve any change in the context of the dialogue for example, a change in location or a change of speaker. In the case of person-machine communication it is most likely to involve a change in the speaker or a change in some aspect of the computer's visual display.

A number of communicative situations go to make up a CSS. What distinguishes one CSS from another is a change in the purpose of the CSS. The CSS is also the repository of information about what entities in the dialogue are currently in focus. This information is used in the resolution of anaphora.

2.4 Structural Relationships

A CSS can be related to another CSS in a limited way. The relationship can only be hierarchical, and it represents a route through which information relating to the focus of attention can be transmitted. If the focus of attention is on one CSS, definite noun phrases and anaphora in general can be resolved either from entities in focus within the current CSS or from the focus space of a CSS that is connected to the current one.

Figure 2 represents a structured collection of CSSs. As can be seen, they consist of a number of tree fragments, rather than one large tree. Such a situation can occur if the purpose of a dialogue is to achieve a number of distinct goals, which cannot be integrated under a dominating CSS.

3 PRAGMATIC DIMENSIONS

3.1 Attentiveional State

The disembodied arrow in Figure 2 represents the current focus of attention. The focus of attention sets bounds on what are valid targets for anaphoric reference within a CSS. This focus shifts automatically as a new CSS is created. It can also be shifted by one or other of the dialogue participants explicitly requesting a shift of focus back to a previous topic in the dialogue. However, there is a constraint on this shift. When moving from one tree fragment to another, the focus of attention can only shift to the top-most node of the
target tree. From there, it may traverse the subordinate nodes of the tree to locate the appropriate CSS. This restriction reflects the fact that when a dialogue participant returns to a previously active topic in the dialogue, he or she tends to proceed from the general to the specific aspect of that topic. Traversal of the CSS tree from top to bottom represents such a transition.

The component of the model operated upon by the attentional mechanism is the focus space. This consists of a list of items that we call discourse objects. The entities on the list can either have properties in their own right, or can inherit them from higher up in a classification hierarchy. The reason for having highly structured objects in the focus space, is to allow for the resolution of anaphoric reference of the following type (after Sidner, 1979):

A: I saw John's Irish Wolfhound yesterday.

B: Yes. They're really big dogs.

In (B) the phrase they're does not refer back to any specific entity mentioned in (A), but rather to the class of dogs of which John's is a member. In order successfully to resolve this reference, knowledge needs to be available to the resolution process concerning the class of entities to which the specific Irish Wolfhound mentioned belongs. The way this is achieved in the model described here, is to allow the entities in the focus space to inherit properties via a classification hierarchy.

3.2 Intentional Structure

As has been pointed out in the description of the dialogue structures, the topmost element of the structural hierarchy (the CSS) contains a pointer into a structure representing the purpose of the CSS. Croz and Sidner refer the set of such CSS purposes as the intentional structure of the dialogue. In essence the CSS purposes are elements in the plan underlying the dialogue. In the case of a person-machine dialogue system, they are the actions that the user wishes the system to perform. There are two relationships that can hold between elements of the Intentional structure and these are dominance and satisfaction-precedence. These represent goal/subgoal and pre-condition relationships, respectively. The hierarchy of Intentional elements is more or less isomorphic to the dialogue structure, as can be seen in Figure 3. Here, the dialogue structure is represented by white boxes and the underlying intentional structure by shaded boxes. Also note that the intentional structure may be expanded by an inferential process, without there being a corresponding node in the dialogue structure.

The specific details of the intentional structure is dependent on the dialogue domain, unlike the dialogue structure. In the following example of an application of the model, the domain is that of database interaction with the user performing the specific task of tabulating data about students' ages and courses. Each intentional component represents an action of tabulation and the place that the action has in the intentional hierarchy is determined by the complexity of the table requested (or inferred).

4 A SAMPLE APPLICATION

The following dialogue (except $8$) was collected as part of a corpus of simulated person-machine dialogues collected for the studies described in Egan, Harris, Harper, and Holly (1986). $8$ is inserted to illustrate how an inferred intention can be used by the system to direct the dialogue.

U1: How many students, both male and female, under 16 or younger in the year degree course?

S2: There are no students of that age group in the College.

U3: Again in the 3 year degree course, how many male and female students in the following age groups: 19 20 21 22 23 25 or older.

S4: Here is the table.

U5: Total number of both male and female students in this course of study.

S6: 153 males and 559 females.

U7: Please supply a breakdown of both male and female students in the graduate course.

S8: Do you wish to see a complete sex by age by course breakdown?

Figure 4 illustrates the unfolding of both the dialogue and intentional structures (the numbers in the boxes correspond to utterances). The intentional structure underlying S8 is inferred on the basis that the user has asked for the same breakdown for two courses, therefore he or she may wish to have a three-way breakdown for all courses. This inference then gives rise to utterance S8, which is incorporated into the dialogue structure. The left of Figure 4 represents the state of the dialogue and intentional structures up to and including utterance
Figure 4: A Structural analysis of the sample Dialogue

U7. The right of the figure represents the structures after S8.

In U5, the reference to an unspecified course (underlined) requires that a referent be found. The bi-directional links in the discourse structure allow information from the focus spaces of the connected nodes to be accessed in the resolution process. Thus, the anaphoric reference in U5 can be resolved by accessing the focus space of utterances 3 and 4.

Note that the small disembodied arrows in Figure 4 indicate the current attentional state of the dialogue.

5 CONCLUSION

The dialogue model outlined above is under-specified in a number of important aspects. For example, no algorithmic description has been provided that can generate and utilise the data structures of the model.

The research programme, of which the work described here is a part, is still in the early stages of implementation. However, a number of implementation decisions have already been made which give some indication of what the final system will look like.

Both the dialogue and intentional structures are to be represented using a frame-based language. The frames will be connected in a network. The instantiation and interconnection of the frames will be the job of a general control algorithm, while the filling of many of the slots in the various frames will be demon driven. That is, associated with each slot will be a function that is activated when data is required for the slot, such as when the frame containing the slot is instantiated. Limited use will be made of the inheritance mechanism of the frame system. Inheritance will be mainly used for the inheritance of focus-space information. The feature of frames that will be most utilised is that of demon-driven slot filling.

REFERENCES

Barwise, J., & Perry, J. (1983). Situations and attitudes. Cambridge, MA: Bradford/MIT Press.

Burton, D. (1981). Analyzing spoken discourse. In M. Coulthard & M. Montgomery (Eds.), Studies in discourse analysis. London: Routledge & Kegan Paul.

Christie, R., Egan, O., Ferrari, G., Gardiner, M., Harper, J., Reilly, R., & Sheehy, N. (1987). Deliverable 2 of ESPRIT Project 5521: A dialogue classification system. Educational Research Centre, Dublin.

Egan, O., Ferrari, G., Harper, J., Prodano, I., Reilly, R., Sebastiani, F., & Sheehy, N. (1987). Deliverable 7 of ESPRIT Project 5521: An Integrated Formal Model of Dialogue. Educational Research Centre, Dublin.

Egan, O., Forrester, M.A., Gardiner, R., Reilly, R., & Sheehy, S. (1986). Deliverable 3 of ESPRIT Project 5521: Dialogue Studies - Pilot Phase. Educational Research Centre, Dublin.

Egan, O., Harper, J., Harris, J., & Reilly, R. (1986). Deliverable 4 of ESPRIT Project 5521: Dialogue Studies - Main Phase. Educational Research Centre, Dublin.

Grosz, B. (1977). The representation and use of focus in dialogue understanding. Unpublished PhD thesis, University of California, Berkeley.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. Computational Linguistics, 12, 175-204.

Reilly R. (in press). An ATN-based grammar for the structural analysis of dialogue. In M. E. Sharkey (Ed.), Modelling cognition: An annual review of cognitive science. Norwood, NJ: Ablex.

Sidner, C. L. (1981). Toward a computational theory of definite anaphora resolution in English discourse. Technical Report 547, MIT Artificial Intelligence Laboratory, Cambridge, MA.

This paper reports research carried out as part of the CFTD project, supported in part by ESPRIT programme of the CSQ (Ref. B8/84 AQ 5527).