

# A TASK INDEPENDENT ORAL DIALOGUE MODEL

Eric Bilange  
CAP GEMINI INNOVATION  
118, rue de Tocqueville, 75017 Paris. France  
and IRISA Lannion  
e-mail: bilange@crp.capsogeti.fr

## ABSTRACT

This paper presents a human-machine dialogue model in the field of task-oriented dialogues. The originality of this model resides in the clear separation of dialogue knowledge from task knowledge in order to facilitate for the modeling of dialogue strategies and the maintenance of dialogue coherence. These two aspects are crucial in the field of oral dialogues with a machine considering the current state of the art in speech recognition and understanding techniques. One important theoretical innovation is that our dialogue model is based on a recent linguistic theory of dialogue modeling. The dialogue model considers real-life situations, as our work was based on a real man-machine corpus of dialogues. In this paper we describe the model and the designed formalisms used in the implementation of a dialogue manager module inside an oral dialogue system. An important outcome and proof of our model is that it is able to dialogue on three different applications.

## 1 Introduction

The work presented here is a dialogue model for oral task oriented dialogues. This model is used and under development in the SUNDIAL ESPRIT project<sup>1</sup> whose aim is to develop an oral cooperative dialogue system.

Many researchers have observed that oral dialogue is not merely organized as a cascade of adjacency pairs as Schlegoff and Sacks (1973) suggested. Task oriented dialogues have been analyzed from different point of view: discourse segmentation (Grosz & Sidner, 1986), exchange segmentation with a triplet organization (Moeschler, 1989), initiative in dialogue (Walker & Whittaker, 1990), etc.

From a computational point of view, in task ori-

<sup>1</sup>This project is partially funded by the Commission for the European Communities ESPRIT programme, as project 2218. The partners in this project are CAP GEMINI INNOVATION, CNET, CSELT, DAIMLER-BENZ, ERLANGEN University, INFOVOX, IRISA, LOGICA, POLITECNICO DI TORINO, SARIN, SIEMENS, SURREY University

ented dialogues planning techniques have received a fair amount of attention (Allen *et al*, 1982; Litman & Allen, 1984).

In the latter approach there is no means to describe and deal with pure discursive phenomena (meta-communication) such as oral misunderstanding, initiative keeping, initiative giving etc. Whilst in the first approaches there is no attempt to develop a full dialogue system, except in Grosz's and Sidner's (1986) model that unfortunately does not cover all oral dialogue phenomena (Bilange *et al*, 1990b).

In oral conversation, meta-communication represents a large proportion of all possible phenomena and is not simple to deal with, especially if we strive to obtain natural dialogues. Therefore, we developed a computational model able to have clear views on happenings at the task level and at the level of the communication itself. This model is not based on pure intuition but has been validated in a semi-automatic human-machine dialogue simulation (Ponamale *et al*, 1990). The aim is to obtain a dialogue manager capable of natural behaviour during a conversation allowing the user to express himself and without being forced to respect the system behaviour. Thus we endow the system with the capabilities of a fully interactive dialogue.

Moreover, as a strategic choice, we decided to have a predictive system, as it has been shown crucial for oral dialogue system (Guyomard *et al*, 1990; Young, 1989), to guide the speech understanding mechanisms whenever possible. These predictions result from an analysis of our corpus and generalized by endowing the system with the capacity to judge the degree of dialogue openness. As a result, predicting the user's possible interventions doesn't mean that the system will predict all possibilities - only relevant ones. This presupposes cooperative users.

## 2 Overview of the Dialogue manager

The architecture of the SUNDIAL Dialogue Manager is presented in Fig. 1. It is a kind of dis-

tributed architecture where sub-modules are independent agents.

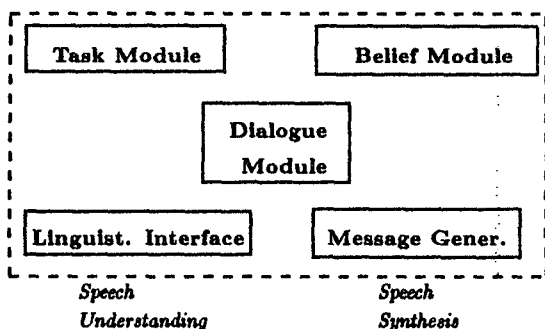


Figure 1. Architecture of the Dialogue Manager

Let us briefly present how the dialogue manager works as a whole. At each turn in the dialogue, the dialogue module constructs *dialogue allowances* on the basis of the current dialogue structure. Depending on whose turn it is to speak, these dialogue allowances provide either dialogic descriptions of the next possible system utterance or dialogic predictions about the next possible user utterance(s). When it is the system's turn, *messages from the task module*, such as requests for missing task parameters, *messages from the linguistic interface* module such as requests for the repetition of missing words, and *messages from the belief module* arising, for example, from referential failure, are ordered and merged with the dialogue allowances by the dialogue module to produce the next relevant system *dialogue act(s)*<sup>2</sup>. The resulting acts are then sent to message generator.

When it is the user's turn to talk, task and belief goals are ordered and merged with the dialogue allowances to form predictions. They are sent, via the linguistic interface module, to the linguistic processor. When the user speaks, a representation of the user's utterance is passed from the linguistic processor to the linguistic interface module and then on to the belief module. The belief module assigns it a context-dependent referential interpretation suitable for the task module to make a task interpretation and for the dialogue module to make a dialogic interpretation (e.g. assign the correct dialogue act(s) and propagate the effects on the dialogue history). This results in the construction of new dialogue allowances. The cycle is then repeated, to generate the next system turn.

This is necessarily a simplified overview of the processing which takes place inside the Dialogue Manager. A detailed description of the dialogue manager can be found in (Bilange *et al*, 1990a). The purpose of this paper is to describe some fun-

<sup>2</sup>This terminology is defined later.

damental aspects of the dialogue module. It is however important to state that the task module should use planning techniques similar to Litman's (1984)

### 3 Basis of the dialogue model

Task oriented dialogues mainly consist of negotiations. These negotiations are organized in two possible patterns:

1. Negotiation opening + Reaction
2. Negotiation opening + Reaction + Evaluation

Moreover negotiations may be detailed which causes sub-negotiations. Also, in a full dialogue, conversational exchanges occur for clarifying communication problems, and for opening and closing the dialogue. This description is then recursive with different possible dialogic functions.

A dialogue model should take into account these phenomena keeping in mind the task that must be achieved. An oral dialogue system should also take into consideration acoustic problems due to the limitation of the speech understanding techniques (soft-as well as hardware) e.g. repairing techniques to avoid misleading situations due to misunderstandings should be provided. Finally, as a cooperative principle, the model must be habitable and thus not rigid so that the two locutors can take initiative whenever they want or need.

These bases lead us to define a model which consists of four decision layers:

- Rules of conversation,
- System dialogue act computation,
- User dialogue act interpretation,
- Strategic decision level.

Now let us describe each layer.

#### 3.1 Rules of conversation

The structural description of a dialogue consists of four levels similar to the linguistic model of Roulet and Moeschler (1989). In each level specific functional aspects are assigned:

- *Transaction level* : informative dialogues are a collection of transactions. In the domain of travel planning, transactions could be : book a one-way, a return, etc. The transaction level is then tied to the plan/sub-plan paradigm. A transaction can be viewed as a *discourse segment* (Grosz & Sidner, 1986).
- *Exchange level* : transactions are achieved through exchanges which may be considered

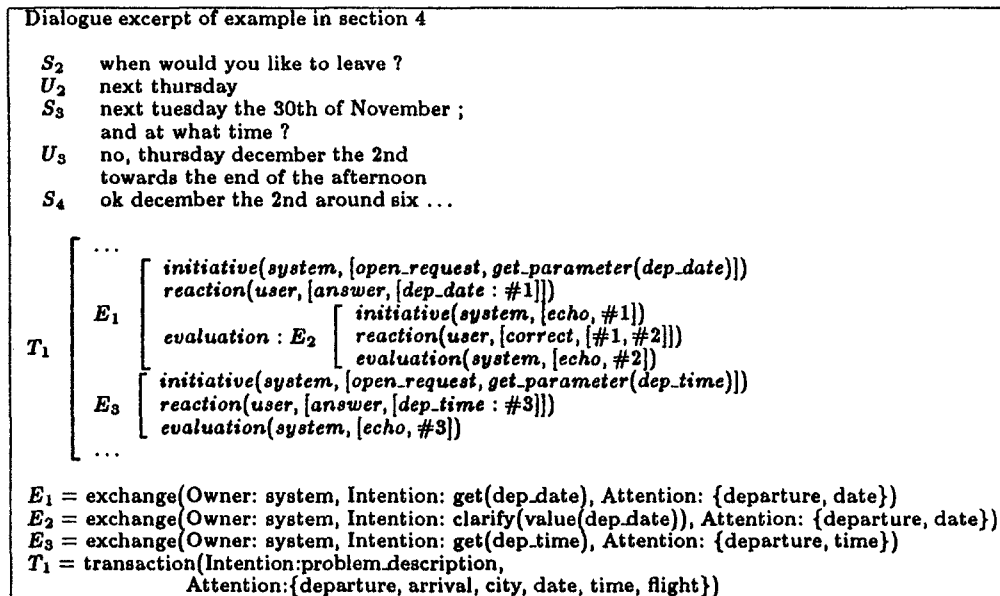


Figure 2. Dialogue history representation

as negotiations. Exchanges may be embedded (sub-exchanges). During an exchange, negotiations occur concerning task objects or the dialogue itself (meta-communication).

- *Intervention level* : An exchange is made up of interventions. Three possible illocutionary functions are attached to interventions: *initiative*, *reaction*, and *evaluation*.
- *Dialogue acts* : A dialogue act could be defined as a *speech act* (Searle, 1975) augmented with structural effects on the dialogue (thus on the dialogue history) (Bunt, 1989). There are one or more main dialogue acts in an intervention. Possible secondary dialogue acts denote the argumentation attached to the main ones.  
Dialogue acts represent the minimal entities of the conversation.

The rules of conversation use this dialogue decomposition and are organised as a dialogue grammar. Dialogue is then represented in a tree structure to reflect the hierarchical dialogue aspect augmented with dialogic functions. An example is given in Fig. 2. Now let us describe conversational rules through a detailed description of the functional aspects of the intervention level.

- *Initiatives* are often tied to task information requests, in task-oriented dialogues. Initiatives are the first intervention of an exchange but may be used to reintroduce a topic during an exchange. Intentional and attentional information is attached to initiatives and exchanges as in (Grosz & Sidner, 1986). When a locutor performs an initiative the exchange is attributed to him, and he

retains the initiative, since there is no need for discourse clarification, for the duration of the exchange. This is important as according to the analysis of our corpus the owner of an exchange is responsible for properly closing it and he has many possibilities to either keep the initiative or give it back.

The simplest initiative allowance rule *initiative-taking*, presented in Fig 3, means that the speaker *X* who has just evaluated the exchange *Sub-exchange* is allowed to open a new exchange such as it is a new sub-exchange of the exchange *Exchange* ( $\{ \cdot \}$  means any well-formed sequence according to the dialogue grammar). Moreover, the new exchange can be used to enter a new transaction. In this case the newly created exchange will not be linked as a sub-exchange (see section 3.2 below).

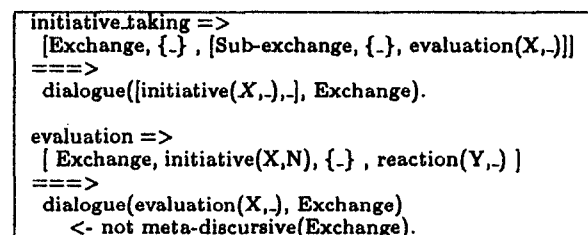


Figure 3. Two dialogue grammar rules

- *Reactions* obey the adjacency pair theory. Reactions always give relevant information to the initiative answered.

- *Evaluations*, both by the machine and the human, are crucial. To evaluate an exchange means evaluating whether or not the underlying intention is reached. In task-oriented dialogues evalu-

ations may serve task evaluations or comprehension evaluations in cases of speech degradations. An example of an evaluation dialogue rule is given in Fig 3. The rule *evaluation* permits when *X* has initiated an exchange and *Y* reacted that *X* evaluates this exchange. The evaluation cannot be made whilst there is no reaction taking place. This rule (as any other) is bidirectional: if *X* is instantiated by “*user*” then the generated *dialogue* ‘allowance’ is a prediction of what the user can utter. On the other hand, if *X* is instantiated by “*system*” then the rule is one of a “strategic generation”. Evaluations are very important in oral conversation and coupled with the principle of bidirectional rules, this allows to foresee possible user contentions and to handle them directly as clarifying subexchanges. The dialogue flavour is that the system implicitly offers initiative to the user if necessary, keeping a cooperative attitude, and thus avoids systematic confirmations which can be annoying (see example in section 4).

The structural effects of evaluations are not necessarily evident. When an evaluation is acknowledged (with cue expressions like “yes”, “ok” or echoing what has been said) the exchange can be closed in which case the exchange is *explicitly* closed. The acknowledgement may not have a concrete realization in which case the exchange is *implicitly* closed. In the latter case, closings are effective when the next initiative is accepted by the addressee. It is unlikely, according to our corpus of dialogues, that one speaker will contest an evaluation later in the dialogue. In the example in section 4, *S*<sub>2</sub> initiative is accepted because *U*<sub>2</sub> answers the question – the effect is then: *U*’s reaction implicitly accepts the initiative which implicitly accepts the *S*’s evaluation. Therefore, the exchange, concerning the destination and arrival cities, can be closed. We will describe later how such effects are modelled.

During one cycle, every possible dialogue allowance is generated even if some are conflicting. Conflicts are solved in the next two layers of the model.

### 3.2 Dialogue acts computation

Once the general perspective of the dialogue continuation has been hypothesised, dialogue acts are instantiated according to task and communication management needs. A dialogue act definition is described in Fig 4.

The premises state the list of messages the dialogue act copes with<sup>3</sup>. The conclusions are twofold: there is a description of the dialogic effect of the act and of its mental effect on the two

<sup>3</sup>We recall that these messages are received by the dialogue module internally (see section 3) or externally (see section 2)

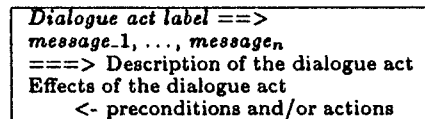


Figure 4. Dialogue act representation

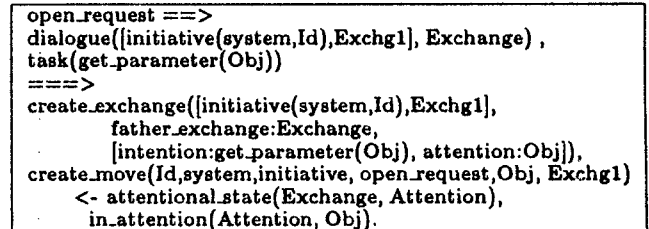


Figure 5. The open\_request dialogue act

speakers. We do not describe this last part as our model does no more than what exists in Allen *et al*’s work (82). Lastly, the preconditions are a list of tests concerning the current intentional and attentional states in order to respect the dialogue coherence and/or actions used for example to signal explicit topic shifts. Signaling this means introducing features in order that once the act is to be generated some rhetorical cues are included: “*Now let’s talk about the return when do you want to come back?*”, or simply: “*and at what time?*” when the discursive context states that the system has the initiative.

At this level all possible dialogue acts according to the dialogue allowances issued by the previous level are hypothesised. Discursive and meta-discursive acts are planned and the next layer will select the relevant acts according to the dialogue strategy.

In the next paragraphs, we describe the most important dialogue acts the system knows and classify them according to the function they achieve.

#### Combining task messages and dialogue allowances :

The dialogue model considers the task as an independent agent in a system. The task module sends relevant requests whenever it needs information, or information whenever asked by the dialogue module.

- *Initiatives and Parameter requests* : an initiative can be used to ask for one task parameter. The intention of the new created exchange is then tagged as “get\_parameter” whereas the attention is the requested object<sup>4</sup>. The act is presented in Fig. 5.

- The other identified possibilities are *initiative*

<sup>4</sup>This is a very simplified description. One can refer to (Sadek, 1990) to have a more precise view of what could be done.

and non topical information; initiative and task solution(s); transaction opening, initiative, and task plan opening; reaction and parameter value; transaction closing, evaluation and task plan closing in which case the act may not have a surface realization since exchanges in the transaction may have been evaluated which implicitly allows the transaction closing.

#### Dialogue progression control :

- *Confirmation handling*: Representations coming from the speech understanding module contain recognition scores<sup>5</sup>. According to the score rate, confirmations are generated with different intensity. The rules are :

- Low score : realize only the *evaluation* goal entering a clarifying exchange.
- Average score : a combination of *evaluation* and *initiative* is allowed, splitting them into two sentences as in "Paris Brest ; when would like to leave ?"
- High score : in that case, the *evaluation* can be merged with the next initiative as in "when would you like to leave for Bonn?"

- *Contradiction handling*. When the addressee utters a contradiction to an evaluation if any initiative has been uttered by the system, it is marked as "postponed". The exchange in which the contest occurs is then reentered and the evaluation part becomes a sub-exchange.

- *Communication management*. Requests for pauses or for repetition postpone every kind of dialogue goal. The adopted strategy is to achieve the phatic management and then reintroduce the goals in the next system utterance.

- *Reintroducing old goals*. As long as the current transaction is not closed the system tries to realize postponed goals if a dialogue opportunity (e.g. a dialogue allowance) arrives. When realizing the opportunity a marker is used to reintroduce the communicative goal if it has been postponed for a long time ("long time" refers to the length in the discourse structure from the postponement and the point where it is reintroduced). This involves the tactical generation of using a special case of rhetoric formulation.

- *Abandoning previous goals*. The concrete realization of dropping an exchange occurs when goals have been postponed and the transaction to which they belong is closed. The justification is simple : a transaction close is submitted to the addressee for evaluation. If he does not contest this closing then this implicitly allows the drop.

Only non crucial exchanges are dropped. If they

<sup>5</sup>Scores may be fuzzy. They only represent the confusion rate which occurs during the lexicalization of the acoustic signal.

were crucial to the transaction then they wouldn't have been dropped.

These communication management acts illustrate the interest of our dialogue model and offer new means to cope with dialogue failure comparing with recent techniques (Jullien & Marty, 1989).

### 3.3 Dialogue strategy modeling

In one running cycle, more than one dialogue act can be a candidate, this is due to the nondeterministic nature of the dialogue which is preserved until this step. For example, it is possible that the dialogue rules allow the system to take an initiative, evaluate an exchange, or react. Consequently a third layer of rules has been designed, in order to select the best candidate according to a general dialogue strategy. As our system is dedicated to oral dialogues the strategy is firstly oriented toward a systematic confirmation of system's understandings and secondly, as a general strategy, we decided to avoid too many embedded subexchanges. This avoids numerous topic shifts, especially implicit ones. The concrete realization of the latter is done by forcing the user to give explicit answers to problematic goals with utterances like "please answer yes or no".

## 4 A dialogue example

We present here a dialogue example obtained with our system in the field of flight reservations<sup>6</sup>. At present, there is no oral recognition system; user's utterances are entered manually in predefined format, including hypothesized acoustic scores and voluntary misrecognition.

S<sub>1</sub> flight booking service, how can I help you?  
 U<sub>1</sub> one ticket to Brest, please  
 S<sub>2</sub> Paris , Brest , when would you like to leave?  
 U<sub>2</sub> next thursday  
 S<sub>3</sub> next tuesday the 30th of November ; and at what time?  
 U<sub>3</sub> no, thursday december the 2nd towards the end of the afternoon  
 S<sub>4</sub> ok december the 2nd around six , just a moment please  
 U<sub>4</sub> thank you  
 S<sub>5</sub> there is one flight at six and another at nine thirty  
 U<sub>5</sub> nine thirty  
 S<sub>6</sub> ok nine thirty, Paris Brest, one-way or return?  
 U<sub>6</sub> no I'd prefer six o'clock

<sup>6</sup>Punctuation is here as oral transcription conventions. Therefore we suppressed some of them in user's utterances as we are not able, at present, to deal with them. Question marks mean that intonation rises and commas denote pauses.

*S*<sub>7</sub> ok, I'll book you on the six o'clock flight  
*U*<sub>7</sub> fine  
*S*<sub>8</sub> one-way or return ?  
*U*<sub>8</sub> no, just one-way thanks

Normally, the dialogue continues with the acquisition of the passenger name and address but now this is not included in the task management.

## 5 Conclusion

The exposed model and system takes into account previous works done in the field of dialogue management augmenting them with a more subtle description of dialogues. This allows us to respect our aims which were to obtain a generic dialogue module adaptable to different applications, a model well suited to oral communication and lastly a model capable of handling dialogue failures without any ad-hoc procedures.

The system is currently under development in Quintus Prolog on a Sun Sparc Station. We now have a first integrated small prototype which runs in three languages (English, French and German) and for three different applications: flight reservation, flight enquiries, and train timetable enquiries. This emphasizes the task independent and language independent aspects of the model presented here. At present, we have about 20 dialogue rules, 35 dialogue acts and limited strategy modeling.

## 6 Acknowledgements

I would like to thank Jacques Siroux, Marc Guyomard, Norman Fraser, Nigel Gilbert, Paul Heisterkamp, Scott McGlashan, Jutta Unglaub, Robin Wooffitt and Nick Youd for their discussion, comments and improvements on this research.

## 7 References

- Allen, J.F., Frisch, A.M., Litman, D.J. (1982) "ARGOT: the Rochester dialogue system". In *Proceedings Nat'l. Conferences on Artificial Intelligence*, Pittsburgh, August.
- Bilange, E., Fraser, N., Gilbert, N., Guyomard, M., Heisterkamp, P., McGlashan, S., Siroux, J., Unglaub, J., Wooffitt, R., Youd, N. (1990a) "WP6: Dialogue Manager Functional Specification". ESPRIT SUNDIAL WP6 first deliverable, June.
- Bilange, E., Guyomard, M., Siroux, J. (1990b) "Separating dialogue knowledge from task knowledge for oral dialogue management", In *Proceedings of COGNITIVA90*, Madrid, November.
- Bunt, H. (1989) "Information dialogues as communicative action in relation to partner modelling and information processing." In M. M. Taylor, F. Néel, and D. G. Bouwhuis, editors, *The structure of multimodal dialogue*, pp. 47-71. North-Holland.
- Grosz, B.J. and C.L. Sidner (1986) "Attention, Intentions, and the structure of discourse". *Computational Linguistics*, Vol. 12, No 3, July-September, 1986, pp. 175-204.
- Guyomard, M., Siroux, J., Cozannet, A. (1990) "Le rôle du dialogue pour la reconnaissance de la parole. Le cas du système des Pages Jaunes." In *Proceedings of 18th JEP*, Montréal, May, pp. 322-326.
- Jullien, C., Marty, J.C. (1989) "Plan revision in Person-Machine Dialogue". In *Proceedings of the 4th European Chapter of ACL*, April.
- Litman, D., Allen, J.F. (1984) "A plan recognition model subdialogues in conversations". University of Rochester report TR 141, November.
- Moeschler, J. (1989) "Modélisation du dialogue, représentation de l'inférence argumentative". Hermès pub.
- Ponamalé, M., Bilange, E., Choukri, K., Soudoplatoff, S. (1990) "A computer-aided approach to the design of an oral dialogue system". In *Proceedings of Eastern Multiconference*, Nashville, Tennessee, April.
- Sadek, M.D. (1990) "Logical Task Modelling for Man-Machine Dialogue". In *Proceedings of AAAI*, August.
- Schlegoff, E. A. and H. Sacks (1973). "Opening up closings". *Semiotica*, 7(4):289-327.
- Searle, J.R. (1975) "Indirect speech acts". In: P. Cole and J.L. Morgan, Eds., *Syntax and Semantics, Vol. 3: Speech Acts* (Academic Press, New York, 1975).
- Walker, M., Whittaker, S. (1990) "Mixed initiative in dialogue: an investigation into discourse segmentation". In *Proceedings of the Association of Computational Linguistics ACL*.
- Young, S.R. (1989) "Use of dialogue, pragmatics and semantics to enhance speech recognition". In *Proceedings of Eurospeech*, Paris, September.