A ROBUST APPROACH FOR HANDLING ORAL DIALOGUES

Eric Bilange & Jean-Yves Magadur

CAP GEMINI INNOVATION
Paris Research Centre
118, rue de Tocqueville
75017 Paris, France
Phone : +33 (1) 40 54 66 66
e-mail: bilange@capsogeti.fr & magadur@capsogeti.fr

ABSTRACT

Present limits of speech recognition and understanding in the context of free spoken language (although with a limited vocabulary) have perverse effects on the flow of the dialogue with a system. Typically a non robust dialogue manager will fail to face with these limits and conversations will often be a failure. This paper presents some possibilities of a structural approach for handling communication failures in task-oriented oral dialogues. Several types of communication failures are presented and explained. They must be dealt with by the dialogue manager if we strike to have a robust system. The exposed strategies for handling these failures are based on a structural approach of the conversation and are implemented in the SUNDIAL system. We first recall some aspects of the model and then describe the strategies for preventing and repairing communication failure in oral conversations with a system.

1 INTRODUCTION

Despite the complexity of human-machine dialogue, the present limits of speech recognition and understanding techniques add further complexity. The troublesome aspect of these limits is clearly that when a dialogue manager fails to face properly misunderstandings, and failures in general, the conversation is tedious and often of no use. Thus one main aspect of a robust oral dialogue system is to be able to cope with misunderstandings of any type and to respect a minimum of ergonomy in use.

This paper exposes some techniques developed in the SUNDIAL[1] system (Speech UNDerstanding and DIALogue), a multi-user oral dialogue system over the telephone for database access. In this system the main objective is that the dialogue with the user must be efficient and robust such that computational problems are not apparent to the user, i.e., she must have the feeling to talk with an intelligent and normal partner.

The exposed techniques are obviously dependent of the conversation model we used. We recall so far the model in section 2 and 3, but the reader is referred to [Bilange 91a, 91b] for further details. The rest of the paper exposes several strategies to tackle with different types of communication failure.

First, let us introduce the problematic in a more concrete way and the typology of communication failures.

The dialogue manager receives as input a semantic representation of what the speech understanding layer has recognized from the user's utterance. One functionality is to interpret the current user's goal(s) in the context of the conversation and to appropriately react so that the task progresses on the right track. However, inputs may be corrupted in several ways:

- there may be confusions for some words,
- only a part of the utterance is understood,
- what is understood is not what the user said,
- nothing is understood.

So whenever the dialogue manager receives an input, it is aware of these problems. It must then check with the user whether they share the same information. In other words, the system

† SUNDIAL 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, SARITEL, SIEMENS, and SURREY Univerity.

has to be careful of providing enough *feedback* of its understanding in order to prevent failures. As we said earlier, this must be done in such a way that *the dialogue remains efficient and natural*.
The techniques used for providing feedback and tackling with communication failure are entirely based on a structural model of the conversation in which we formalized several dialogue strategies. Before describing our solutions, we first give a brief overview of the underlying theory.

2 THE DIALOGUE MODEL

The model is structural and functional [Moeschler 89, Bilange 91a, Bilange 91b]. The conversation is structured into four levels: *transactions, exchanges, interventions,* and *dialogue acts*. Each of these levels has functional relationships with the dominating and adjacent one. Fig. 1 presents the BNF syntax of the model.

A Dialogue is made of the following components: Transactions, Exchanges, Interventions, and Dialogue Acts (Da). The syntax of the model is as follows:

$$D \rightarrow \{T\}^+$$
$$T \rightarrow \{E\}^+$$
$$E \rightarrow \{I\}^+\{E\}^*\{I\}^* \mid \{E\}^+$$
$$I \rightarrow \{Da\}^+$$

Fig. 1: the BNF form of the model

• **Transactions:** are the outmost level. Analyses of corpora revealed that task-oriented dialogues are a collection of phases [Amalberti et al. 84, Ponamalé et al. 90], so-called transactions. In our domain, we can identify four main transactions: *dialogue opening, problem formulation, problem resolution,* and *dialogue closing*. The second and the third ones form a sequence that can be iterated or/and that can recursively occur during a conversation. One can think of transactions as discourse segments [Grosz and Sidner 86] that denote plan transitions' points at the task management level. During a transaction, the dialogue participants try to achieve a generic goal (open the dialogue, formulate the problem...). It has been also proved that participant roles depend on the type of the transaction and therefore participants' dialogue strategies vary according to the current transaction [Bilange 91b].

• **Exchanges:** are made up of interventions or of exchanges (sub-exchanges). An exchange carries a specific goal that may contribute to the transaction (the one it belongs to) or a goal dedicated to a communication clarification. An exchange has also three possible statuses: open, close or postponed. Once an exchange is closed, it is impossible to reenter it (e.g., if one wishes to discuss again about the same goal then a new exchange is opened). A postponed exchange is temporarily closed and may be re-opened later.

• **Interventions**: are the basic components of exchanges, and they are made up of dialogue acts. Three canonical illocutionary functions are attached to interventions: *initiative, reaction* and *evaluation*. Basically, initiatives[2] open exchanges: they introduce the goal of the exchange. Reactions react to initiatives (they may or may not be present) and evaluations evaluate the exchange (e.g., the status of the goal achievement: positive, negative, satisfying...).

Things can be a bit more complex since initiatives, reactions and evaluations may not be a mere collection of dialogue acts but rather a collaborative process. This is why exchanges may also have these illocutionary functions attached to them.
In oral human-machine interactions, it is crucial that *evaluations* can be performed by both user and system. Evaluating an exchange means verifying its completion, i.e., whether the underlying intention (goal) is reached or not. Therefore evaluations are of prime importance since the main side effect is that whenever the two dialoguees agree implicitly (a simple evaluation) or explicitly (an evaluative exchange) then the evaluated exchange can be closed (and thus all information exchanged in it can be certified as shared by both dialoguees).

• **Dialogue acts** are the basic components of interventions. Dialogue acts (as interventions) are monological units: they are performed by one participant as the result of an autonomous process. In one intervention (say move) one can perform more than one dialogue acts. At least one expresses the illocutionary function of the intervention, it is called the main act. Dialogue acts are actions with preconditions and effects. We describe them in the next section.

From this hierarchical description, one can build a structure that dynamically represents the current state of the conversation. It is called the dialogue structure. This object is continuously updated as the dialogue goes on. The dialogue structure may be thought as a tree where leaves

1 similar to *moves* in the litterature.

2 a shortcut to say "interventions that have an initiative illocutionary function".

dialogue act label	: req-for-spelling
dialogue act owner	: system
structural preconditions:	S = [... [E, i(s), r(u), [E_1^{Ev}, i(s), contest(u)]] ...]
	& E_1^{Ev} is a currently open exchange
structural effects	: S = [... [E,i(s),r(u), [E_1^{Ev}, i(s), contest(u), [E_2^{Ev}, req-for-spelling(s)]]] ...]

Fig.2 : a system dialogue act definition

are dialogue acts uttered by both speakers, and nodes are interventions, exchanges and transactions (see figure 4 for the representation of a dialogue excerpt).

3 DIALOGUE ACTS AS RULES

Dialogue acts come from the well-known theory of speech acts. We agree with Bunt, however, that a dialogue act must be defined with respect to the modifications on the context [Bunt 89]. An act is then uttered when the context fits the conditions associated to it. In turn, the production of an act modifies the context. Therefore, a dialogue act is a function which transforms a context into a new one. For Bunt, the context is the description of both dialogue participants' mental states: their knowledge, suspicions and beliefs.
However, some contextual aspects are difficult to encapsulate in a pure attitude model representation, especially those coming from structural indicators, denoted in the dialogue structure. Therefore, our notion of the context is Bunt's plus the dialogue structure [Bilange 91a]. The advantage of this approach is that some dialogue acts can be triggered if and only if certain patterns are present in the dialogue structure. This naturally captures the fact that performing dialogue acts must respect structural constraints and dialogue norms.
So, dialogue acts are triggered when certain conditions are met. These conditions are of two kinds: structural and/or non-structural. A structural precondition enforces the presence (or absence) of a certain pattern in the dialogue structure for the act, if triggered. A non-structural precondition is tied to the mental states of dialogue participants: task goals to achieve, mutual believes...

Figure 2 presents an example of a system dialogue act which has only structural preconditons and effects.
S denotes the dialogue structure. E denotes an exchange, made up here of one initiative (i(s)), one reaction (r(s)) and one evaluative exchange E_1^{Ev}. This evaluative exchange is in turn made up of one initiative and one reaction that is composed of one dialogue act: *contest*. ("s" and "u" denote the system and the user resp.).
From there, the system dialogue act req-for-spelling is triggered when an evaluation has been uttered by the system and that evaluation is contested by the user (the system's evaluation opens an evaluative exchange (E_2^{Ev}), thus it is an initiative and the user's contest is the reaction to that initiative). It is of course possible to define other dialogue acts based on the same idea: another act may be triggered when there are two, or three embedded evaluative exchanges instead of only one. It should be noticed that for req-for-spelling it is not necessary to have non structural preconditions and effects. Typically, only structural evidences are sufficient to trigger this act. This is what characterizes dialogue control acts.

4 HANDLING COMMUNICATION FAILURES

4.2 Failure prevention with feedback

As said earlier, evaluation purpose is mainly to close an exchange in providing a feedback on the outcome of the exchange intention. Evaluations are optional in essence, however a safe strategy for the system consists in using this opportunity to make clear what it understood.
One can perform an evaluation in several ways: either in explicitly checking one's understanding with a request for an acknowledgment or implicitly with a mere echo. The first solution blocks the conversation on a clarification whereas the second allows both dialogue participants to continue the conversation in moving to another topic. Obviously, the first behaviour is less risky for the system since the user may contest the evaluation anyway. However, the second behaviour is more natural and fluid. Therefore, we endowed the system with the capacity of using both behaviours with a preference for the second. Basically, in the oral context, the system makes its choices on the basis of

acoustic scores[3]. Three different behaviours are defined, based on the distribution of scores among three categories: *high*, *average*, and *low*.

- Low scores: only the evaluation is performed;
- Average scores: the evaluation is performed and the opening of a new exchange is allowed. An example is given with S_2 (see the dialogue below), where the evaluation concerns the destination and date (parameters obtained in the first exchange of the dialogue) and a new exchange is opened (the one concerning the solution).
- High scores: similar to the average score case. However, the system can generate two acts that can be merged in the same sentence which can, in some circumstances, strengthen the naturalness of the system's output. If the score were high in our dialogue, S_2 would have been: "there is a flight to *Rome* which takes off at 10.30 on *Tuesday*, is that ok?"

S_1 Flight reservation system. Formulate your request.
U_1 I'd like to go to Bonn on next Tuesday morning.
S_2 *Rome next Tuesday*, there is a flight which takes off at 10.30, is that ok?
U_2 No I want to go to BONN
S_3 ok, Bonn. There is a flight...

Whatever the scores are the system, while performing an evaluation, systematically predicts a possible user contest. These predictions are precise since the system knows exactly where and on what information contests can occur. U_2, for example, has been predicted and the prediction says that if the user contests the arrival city it is out of question to recognize the same value (i.e., Rome). Moreover, the system knows that if the user accepts an evaluation then the evaluated information is certified as shared (implicitly). This is the case of the departure date in our example (one can notice that the acceptance is not explicit, this is discussed in the next section). This is why in S_3, the system has to confirm only the departure city.
Scores are not the only information used by the system to plan its behaviour. The system can evaluate the degree of risk when merging evaluations with other acts. In our example, S_2 is considered as risky since there are simultaneously a twofold evaluation (city and date), a topic shift and a transaction shift (from problem formulation to problem resolution).

3 More precisely on a combination of acoustic scores and the perplexity.

The transaction shift is risky since the system closes a transaction where some parameters are not yet confirmed and the evaluation is also risky since correctly recognizing a contest means recognizing on what the contest is about between two possibilities. However, S_3 is less risky.

The evaluation principles enumerated above with the notion of risk in one utterance provide a good help for preventing failures in a very ergonomic fashion. This technique is well perceived by the majority of tested users.

4.2 Structural detection of some failures

In this section we illustrate how the structure of the conversation helps the system to detect failure situations.

Figure 3 (next page) shows two dialogues that differ only because of a failure in the second one. The structures of these dialogues are presented on the right hand side.
The system has tentatively opened a new exchange about the departure date, but before doing so, it has uttered an evaluation to close the previous exchange by echoing what it believes the departure and arrival cities are. At this stage, two possible continuations are:

- the user answers the question about the date; this means that she implicitly agrees one the system's evaluation. E_1 can then be closed;
- the user utters a disagreement about the evaluation.

These possible continuations are shown in figure 3. In the case of the user's contest (continuation II), the system's evaluation becomes the initiative of an evaluative exchange (E_3^{Ev}) and E_2 is postponed. The shape of the dialogue structure is typical in such a case of failure. This leads to the idea that one could examine the dialogue structure to detect communication failures and to trigger appropriate repair strategies.
Precisely, through dialogue act definitions, the system detects failures in triggering dialogue control acts. From now, implementing a repair strategy is easily done in adding new acts.
Actually, in the Sundial system a situation like the one in continuation II is treated as follow: the system triggers a req-for-spelling (see the preconditions in figure 2). This *strong* dialogue control act permits the system to send very accurate predictions, telling the speech un-

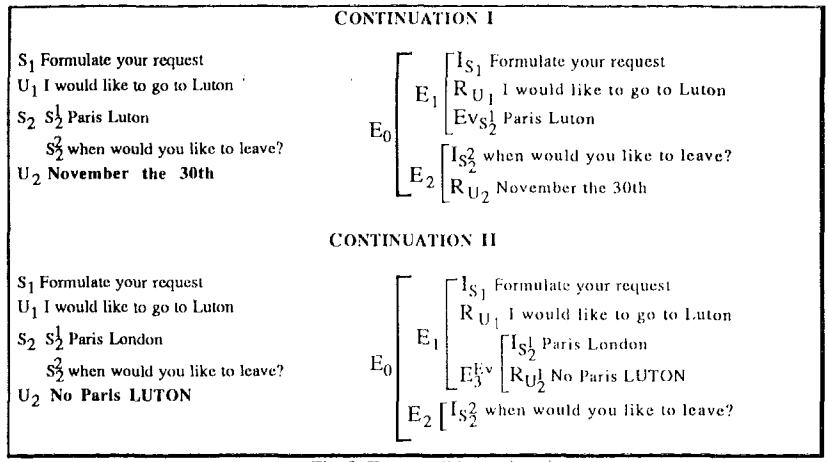

Fig. 3: Two possible continuations

derstanding layer that the user is about to spell the name of a city (with forbidden cities: the one misunderstood before). It has been observed that oral spelling is robust iff the speech understanding layer tries to recognize the input with a specific grammar which is done in the Sundial speech understanding component. So this strategy offers a very robust approach that can be easily generalized to proper names.

U_1	I would like to go to Luton
S_2	Paris London when would like to leave?
U_2	No Paris Luton
S_3	Could you spell the arrival city please?
U_3	L, U, T, O, N
S_4	So Luton. *And about the date?*

Fig.4: A repair sequence

Figure 4 shows one actual dialogue obtained with Sundial with this technique. In this dialogue, one can observe that the system has temporarily focused on the communication problem and once it is solved, it reintroduces the departure date topic.

We have examined here some possibilities of preventing and repairing failures in intensively taking into account the dialogue structure, specially at the level of dialogue act definitions. Dialogue control acts are of prime importance for obtaining accurate predictions that indeed help the speech understanding layer.

5 ACCEPTABILITY OF THE SEMANTIC INPUT

We examine now some cases in which a pragmatically doubtful input can be rejected by the dialogue manager with the help of structural evidences. Let us first examine the following scenario:

> *During its turn, the system has evaluated an exchange E_1 and opened a new one E_2. It is now the turn of the user (this situation is similar to continuation I in figure 3 before U_2).*

A system initiative may be of two types:

(i) an initiative such that the user can implicitly react,

(ii) an initiative such that the user can only perform an explicit reaction.

Figure 5 provides an example where the system's initiative in E_2 is of type (ii).

The last user's input seems quite surprising.

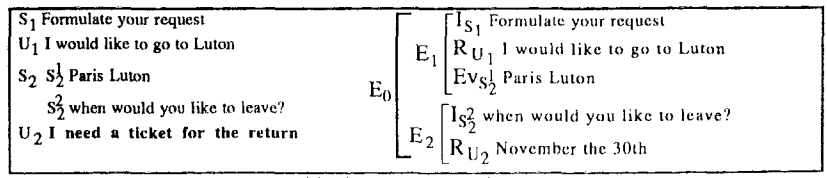

Fig. 5: A wrong user's input

dialogue act label	: *initiative*
dialogue act owner	: user
structural preconditions	: there is no currently open exchange of the form [E, i(s)] such that i(s) is of type (ii) and *other preconditions*
structural effects	: an exchange of the form [E, i(u)] is added in the current transaction and *other effects*

Fig.6: a skeleton for user's initiative definitions

The user is actually supposed to react to the system's initiative in E_2 or to the evaluation of E_1, or both. U_2 is a complete topic shift (not related to the problem formulation of the one-way -the current transaction-). From the system point of view, the most natural interpretation is that a recognition failure most probably occurred. This structural aspect (which captures normative evidences of the conversation) allows the inhibition of some user's initiatives interpretation when there exists, somewhere in the dialogue structure, a non-answered system's initiative of type (ii). This leads to the general definition of user's initiatives presented in figure 6.

Conversely, if the system initiative is of type (i), like in figure 7, the user's input can be accepted.
So, once an input has been rejected, the system must enter a repair process. In the case given in figure 5, the real user's utterance could be an answer on the departure date or a contest to the evaluation as well. The strategy is then to ask the user to repeat without changing the situation (except that the system remembers that there was a failure. If the failure continues then other strong control can be perform such as connect to a human operator).

S_1	At what time would you like to leave?
U_1	9 p.m.
S_2	9 p.m., *there is flight BA 123 ... is that ok?* *(initiative of type i)*
U_2	I need a ticket for the return *(acceptable user initiative)*

Fig. 7: a user initiative after a system initiative of type (i)

6 CONCLUSION

The structural model of the conversation used in the Sundial system offers great capacities to deal with speech pitfalls.
We have presented here some techniques to both prevent and repair misunderstandings. The benefit of the structure of the communication is to enrich, in a practicable way, the notion of context usually based on mental attitudes. This benefit allows to enrich dialogue act preconditions with structural patterns, as well as effects, which allows us to capture normative and natural aspects of task-oriented dialogues.
We have studied and exploited these capacities for at least a small set of possibilities. This approach has to be enriched to cover more situations. Clearly, this approach is complementary to the ones based on pure attitude models of both dialogue participants and this is where our system should be enlarged too.
The optimistic conclusion is to say that with the SUNDIAL system, dialogues never totally fail: failure/repair sequences often occur but at least the conversation always ends to the result envisaged by the user. This is what we observed after having tested a range of 20 naive users who were generally (around 90%) satisfied of the dialogues thay had with the system.

REFERENCES

Amalberti, R., Carbonell, N., Falzon, P. (1984) "Stratégies de contrôle du dialogue en situation d'interrogation téléphonique", in *Communication Orale homme-machine*, GALF-GRECO.

Bilange, E. (1991a) "A Task Independent Oral Dialogue Model", in Proceedings of the European ACL, Berlin, pp. 83-88, April.

Bilange, E. (1991b) "Modélisation du dialogue oral personne-machine par une approche structurelle", PhD Thesis, Rennes I University, December.

Bunt, H.C. (1989) "Information dialogue as communicative action in relation to partner modelling and information processing", in *The Structure of Multimodal Dialogues including voice*, D. Bouwhuis, M. Taylor, F. Néel (eds), pp. 1-19, North-Holland.

Grosz, B.J, Sidner, C.L. (1986) "Attention, Intentions, and the structure of discourse", in *Computational Linguistics*, **12**-3, pp. 175-204, July-September.

Moeschler, J. (1989) *Modélisation du dialogue, représentation de l'inférence argumentative*. Hermes, Paris.

Ponamalé, M., Bilange, E., Choukri, K., Soudaplatoff, S. (1990) "A computer-aided approach to the design of an oral dialogue system", in Proceedings of the Eastern Multiconference, Nashville, Tennessee, April.