Users' representation of the machine in human-computer speech interaction

René Amalberti, Noëlle Carbonell, P. Falzon

INRIA

UNITE DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.(1)39 63 55 11

Rapports de Recherche

N° 1125

Programme 8
Communication Homme-Machine

USERS' REPRESENTATION OF THE MACHINE IN HUMAN-COMPUTER SPEECH INTERACTION

René AMALBERTI
Noëlle CARBONELL
Pierre FALZON

Décembre 1989

RR-1125

Users' representation of the machine in human-computer speech interaction

La représentation de la machine pour l'utilisateur dans l'interaction orale homme-machine

René Amalberti°, Noëlle Carbonell°° & Pierre Falzon°°°[1]

Novembre 1989

Programme 8

° Centre d'Etudes et de Recherche de Médecine Aérospatiale
26 bd Victor, 75996 Paris, France

°° Centre de Recherches en Informatique de Nancy - INRIA
B.P. 239, 54506 Vandœuvre-les-Nancy, France

°°° Institut National de Recherche en Informatique et Automatique
Rocquencourt, B.P. 105, 78153, Le Chesnay, France

1. This text has been submitted for publication. Authors are listed in alphabetical order.

PAPIER RÉCUPÉRÉ ET RECYCLÉ

Résumé

Une expérimentation est conduite pour étudier l'applicabilité des modèles de la communication orale homme-homme à la communication homme-machine en langage naturel oral. Deux groupes de six sujets doivent obtenir des renseignements aériens, au travers d'un dialogue avec une interlocuteur distant. Les sujets du groupe machine pensent parler à un calculateur, ceux du groupe opérateur à un interlocuteur humain. Les deux groupes parlent en fait au même expérimentateur. L'étude est focalisée sur les représentations, chez les sujets, des capacités et connaissances de l'interlocuteur, et diffère des études analogues sur les points suivants:

- la tâche est plus complexe, donnant lieu à des échanges en langage naturel structurés, au lieu de paires de question/réponse en langage restreint;
- une attention particulière a été portée au protocole expérimental, de façon à éviter les biais pressentis dans d'autres études (notamment, les conditions sont strictement identiques pour les deux groupes);
- l'évolution temporelle est prise en compte (3 sessions, à une semaine d'intervalle).

Certains résultats confirment ceux de la littérature: les sujets du groupe machine tendent à contrôler et simplifier leur expression plus que les sujets du groupe opérateur. Néanmoins, la plupart des observations sont originales ou contredisent les résultats précédemment obtenus:

- les sujets du groupe machine parlent plus; aucune différence significative ne peut être détectée entre les deux groupes en ce qui concerne la plupart des aspects de l'usage du langage;
- en ce qui concerne l'évolution temporelle: les deux groupes évoluent vers plus de concision; les stratégies de traitement des scénarios-problèmes se modifient pour le groupe opérateur, mais restent stables dans le groupe machine.

Ces différences de comportement des deux groupes sont attribuées à des représentations différentes des capacités des interlocuteurs.

Mots-clés : Dialogue homme-machine, modèle de la machine, communication orale

Summary

We report an experiment designed in order to study whether models of human-human voice dialogues could be applied successfully to human-computer communication using spoken natural language. Two groups of 6 subjects were asked to collect information about air travel, through dialoguing with a remote interlocutor. Subjects of the machine group were led to believe they were talking to a computer, subjects of the operator group were told they were interacting with a human operator. Both groups actually talked to the same human experimenter. The study focuses on subjects' representations of interlocutor's skills and knowledge, and differs from previous analogous experiments in several respects:

- the task is more complex, giving rise to structured exchanges in natural language instead of question/answers pairs in a restricted language;
- specific attention has been paid to the experimental protocol, in order to avoid the possible bias of other studies (in particular, conditions are strictly identical for both groups);
- evolution over time has been considered (3 sessions, with a one-week interval).

Some results confirm those of the literature, namely that subjects of the machine group tend to control and simplify their linguistic expression more than those in the operator group. However, most observations are either original or in contradiction with previous results:

- subjects of the machine group are more talkative; no significant difference can be detected between the two groups with respect to most aspects of linguistic practice;
- concerning evolution over time: both groups tend to get more concise; with respect to scenario processing (i.e. problem solving), the strategies of the operator group evolve significantly, whilst those of the machine group remain stable.

These differences in the behaviors of the two groups are ascribed to different representations of the interlocutor's abilities.

Key-words : Man-machine dialogue, speech communication, model of the machine

SOMMAIRE

1 Motivations

For the time being, speech input/output is considered as an auxiliary communication medium inside multi-media human-computer interfaces and, with the exception of applications for the handicapped, it is never used as the sole or major information exchange channel. This results partly from the fact that the capabilities of currently available speech recognition algorithms or devices are restricted to isolated or connected word multi-speaker recognition; therefore, recognition parameters must be tuned up to actual speakers' speech characteristics through a preliminary training stage.

However, recent progress in speaker-independent continuous speech recognition and language understanding makes it possible to significantly increase the number of computer applications that could, in the near future, benefit from speech processing facilities. In particular, reliable speaker-independent continuous speech input/output devices would greatly increase the user friendliness and transparency of information/advising dialogue software intended for the general public; such users would appreciate "natural dialogue" which implies:

- speaker-independent continuous speech recognition;
- understanding of unrestricted natural language, especially interpretation of anaphora, broken or unfinished statements and all the phenomena that differentiate spontaneous speech from controlled written expression;
- flexible user-centered dialogue strategies, structured dialogue management allowing shared initiative, instead of menu selection or system driven strict question-answering.

Nevertheless, natural voice dialogues are not yet implemented, chiefly for the two following reasons:

- Is speech suitable for human-computer interaction ? Some authors in the HCI world doubt it and, considering speech as a communication medium dedicated to intercourse between humans, restrict the scope of its application to multi-activity contexts and interfaces for the handicapped (Shneiderman, 1987). We shall not investigate this issue in the absence of working continuous speech input/output prototypes; simulations involving human experimenters being liable to yield biased results in this context.

- Speaker-independent continuous speech recognition performs relatively well in the case of reading ; but the understanding and management of "natural" dialogues remain still to be mastered, due to some characteristics of spontaneous speech production, i.e. fast enunciation rhythm together with the presence of numerous hesitations and broken/incorrect syntactic constructs.

A possible short-term solution to this second problem could be to enforce speech constraints and language restrictions on users' oral productions. But, this strategy seems questionable: will users be able and willing to comply with all the constraints required by present available algorithms in order to achieve accurate interpretation of speech? Moreover, the mere presence of such constraints may induce them to discard speech as a human-computer communication medium, since the chief appeal of speech interfaces to users consists in the fact that they allow spontaneous expression from users and provide "natural" dialogue facilities.

For the near future at least, designers of continuous speech interfaces who aim at developing successful commercial products will have to define sets of speech constraints and language restrictions that should match two conflicting criteria :

- on one hand, users should be able to apply the envisaged constraints without effort nor reluctance, i.e. "naturally";

- on the other hand, these constraints should significantly reduce the complexity of speech recognition and understanding, so that the number of misunderstandings between the user and the system be reduced and communication efficiency maintained within acceptable limits.

But the design of such a "natural" set of constraints implies a careful preliminary thought on the notion of "natural" expression and dialogue :

- What are the exact characteristics of "natural" human-computer intercourse ?
- Should human-computer communication using continuous speech emulate human-human conversations ?
- Even if systems emulate human behaviour, will they influence dialogue progress and users' behaviours in the same way as human interlocutors do in the context of human-human dialogue? More generally, can results, especially dialogue or user models obtained from the observation of human-human dialogues or from controlled experiments, be applied to human-computer communication? Are they relevant in this latter context ?

Both speech scientists and interface designers overlook this crucial issue and usually take for granted that speech interfaces should emulate human-human communication.

In order to clarify this basic issue, a controlled experiment has been designed and carried out, the results of which may help to decide:

- whether models derived from the analysis of human-human task-oriented dialogues adequately describe human-computer oral communication and whether the design of human-computer interfaces using continuous speech should involve such models or specific ones ;

- more precisely, whether or not users behave in the same manner, have the same expectations and mental representations when they talk to a computer as when they dialogue with a human interlocutor.

We shall first survey the literature relating to this issue and, more generally, to "natural" human-computer dialogue involving written or spoken natural language input/output devices. Then, we shall describe the objectives and the protocol of a controlled experiment we designed and carried out, and we will discuss its results. More specifically, we shall try to :

- determine whether or not users behave similarly when talking to a computer or a human being;
- in the case of significant differences, bring out the specificity of users' behaviours, especially as regards :
  . linguistic practice,
  . dialogue strategies and task management,
  . users' expectations with respect to system performances and their a priori mental representations of system skills and knowledge concerning the task domain as well as linguistic understanding and dialogue management.

2 Human-human vs human-machine experiments

2.1 Verbal behavior in human-machine interactions

Human-human dialogues have often been assumed to be the reference model for human-machine interactions. However, human linguistic and reasoning processes could differ as a function of the nature (machine or human) of the interlocutor. Thus several studies have been devoted to the comparison between human-human and human-machine dialogues.

2.1.1 The Wizard of Oz paradigm

Studies have used different variations of the same experimental paradigm, the Wizard of Oz paradigm. Subjects have to fulfil a (more or less complex) task of information query. Some subjects are told that their interlocutor is another human; others are told that it is a machine. All subjects interact in fact with the same (human) interlocutor (another experimenter). Variations between studies concern mainly: the knowledge (by the experimenter) of the role he/she is playing (human or machine), the communication mode (speech or teletype), the quality of the voice heard by the subject (more or less free from distortion), the degree of constraint on the experimenter's production and comprehension, the type of measures made on the subjects' verbal behavior.

2.1.2 Experimental results

The observations below pool the results of Chin (1984), Kennedy et al. (1988), Meunier and Morel (1987), Morel (1985), Richards and Underwood (1984a), and Spérandio and Létang-Figeac (1986). According to these studies, in the human-machine condition, human verbal production is modified in the following ways:

- *less dialogue control acts*: less readbacks, less acknowledgments, less expressions aiming at closing or maintaining the communication;

- *less dialogue structuration*: less linkage between the themes, decrease in the role of request markers (markers are less numerous, less diverse, and meaningless); however, there is an increase in explicit communication planning;

- *attempts to "normativize" the language* (i.e. production of a language closer to written language): less indirect requests, less uncompleted sentences, more well-formed requests;

- *attempts to simplify the expression*: less contextual expressions (e.g. anaphora, ellipses, pronouns), repetitive use of the same syntactic expressions, reduction of the verbal group, smaller vocabulary (higher type-token ratio);

- *attempts to simplify the communication*: shorter conversations, decrease of the speech rate or volume, increase in the number of interventions, less chaining between the interventions (longer pauses, less interruptions, less overlaps), less digressions and comments, less expressions of politeness.

All these modifications tend to simplify verbal productions. This result has been considered very positively, since the use of a simpler language would alleviate the comprehension task for the machine.

2.2 Human adaptive behavior

2.2.1 A model of the interlocutor

According to several authors, the modifications reported above are caused by subjects' a priori model of the machine. This model would specify that machines have low abilities in language understanding and problem processing. This a priori model varies according to the level of knowledge in the computer field (Dumais and Landauer, 1982; Onorato and Schvaneveldt, 1986).

The adaptive behavior is not specific to human-machine dialogues. Human-human oral communication is sensitive to the nature of the interlocutor. Humans are able to adapt their language to the supposed knowledge of the dialogue partner. Incorrect (too high or too low) estimations of the interlocutor's knowledge have negative effects on performance; communication is optimal when estimations are accurate (Harris et al., 1980). Regulation processes may affect the lexical level (Anglin, 1977; Moscovici, 1967), the volume and content of the conversation (Krauss and Glucksberg, 1977), and the dialogue strategies (Deutsch, 1974; Falzon et al., 1986).

Thus, whatever the interlocutor (human or machine), verbal behavior is influenced by a model of the interlocutor. This model is elaborated during dialogue, based on an a priori representation and on a representation elaborated during dialogue of the interlocutor's technical and linguistic competences (fig. 1).

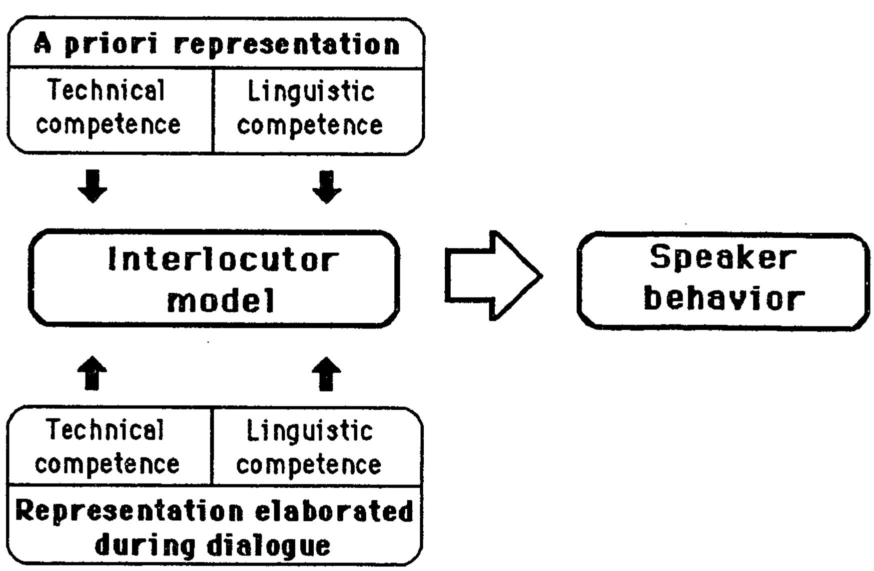

Figure 1

Speaker behavior as a consequence of the interlocutor model

2.2.2 Elaborating or inducing a model

The elaboration of the model during dialogue and the effects of the model on dialogue strategies have been shown very clearly in an experiment by Isaacs and Clark (1987). In the situation studied, couples of subjects had to collaborate in order to classify some objects. The task is to be completed several times (several trials). Subjects had various levels of competence concerning these objects. The authors indicate that efficiency in task completion is based on an evaluation, by each dialogue partner, of the knowledge of the interlocutor. In particular, this evaluation allows to optimize the vocabulary. Evaluations take place very early during dialogue: in the first quarter of the first trial, and even, most probably, in the first lines. Similarly, Falzon (1989; submitted for publication) notices the importance of the evaluation of the interlocutor's domain competence in a diagnosis task. Again, these evaluations occur very early during dialogue. Their effects bear both on dialogue management and problem processing.

Comparable language modifications may occur when the interlocutor happens to be a machine. The machine may be thought of as a special type of interlocutor: in the same way as humans have a model of what a child is able to understand, they have a model of what a machine is able to process. This model may affect their verbal productions. For some authors, the question has thus become: how could the machine influence users in such a way as to provoke an appropriate verbal behavior ("appropriate" meaning here: understandable by the machine) ?

Two experiments illustrate quite well this perspective. Richards and Underwood (1984b) have studied users' reactions to an information system (dealing with train timetables) which had various initial behaviors. Two factors were studied: the explicitness of the initial message (the system could state explicitly the information units to be provided, or could simply indicate it was ready to listen), and its politeness (presence or absence of polite expressions). Results indicate that polite initial messages provoke more polite requests. An identical result has been reported by Clark & Schunk (1980), in human-human communication.

Similarly, Zoltan-Ford (1984) has studied the possibility to obtain spontaneous restrictions in users' language by constraining the machine language. Results indicate that subjects exposed to computer outputs restricted in syntax and vocabulary use shorter commands and a smaller vocabulary

2.3 The a priori model of the machine

However, the results obtained contradict other observations, in particular the fact that subjects tend to overestimate the performances of the machines in comprehension and production. An illustration of these expectations is that, in Wizard of Oz experiments, subjects accept very easily the extraordinary (and miles ahead of the state of the art) abilities of the simulated machines (cf. Forrest et al., 1986). Inversely, subjects are very much disappointed when they face existing machines. Talbot (1985) notices that unrealistic expectations on the performance levels of existing machines lead to a subsequent lowering of confidence in the technology.

Thus, if the a priori model of the machine (of the layman) overestimates machine abilities, few differences should be observed when the interlocutor happens to be a machine.

2.4 A critical appraisal of the experimental conditions

2.4.1 Experimental bugs ?

However, results do indicate differences between human-machine and human-human differences. The origin of these differences could lie in some characteristics of the experimental conditions.

1. Experiments concern in some cases typed interaction, and in other cases oral interaction. This may explain the variability of some results (typed interaction increases the tendency to normativization of the language).

2. The characteristics of the voice heard by the subjects (in the case of an oral dialogue) varied: in some experiments (e.g. Richards & Underwood, 1984a; Spérandio & Létang-Figeac, 1986), the voice heard in the machine condition was artificially distorted (for reasons of plausibility). The acoustic characteristics of the speech signal might have influenced the results obtained in this condition (Waterworth, 1985).

3. In several experiments (e.g. Morel, 1985; Spérandio & Létang-Figeac, 1986), experimenters know the role they play, i.e. they know whether they are supposed to be a machine or a human interlocutor. This knowledge may have influenced their behavior, so that it is impossible to know if subjects' attitude is caused by their a priori model of the machine or by the behavior of the experimenter.

4. Experimental protocols have given various levels (and types) of constraints on the experimenter. For instance, in the experiment studied by Morel (1985) and Spérandio and Létang (1986), the behaviour of the machine was restricted in production (condition 1) or in production and comprehension (condition 2). Again, these restrictions have been introduced

in order to simulate a plausible machine (from the experimenters' point of view), taking into account the state of the art in human-computer speech communication.

In summary, in these researches, human-machine and human-human conditions differ not only in the supposed identity of the interlocutor (for the subject) but also in several other ways. This has two consequences:

- constraints on the verbal behaviour of the machine are such that they may interfere with the subjects' representation of its domain capabilities: thus, it is not possible to distinguish the effects of the representation of the machine linguistic abilities from the effects of the representation of the processing abilities of the machine (cf. Fig. 1);

- it is not possible to know whether subjects modify their verbal behavior because of their a priori model of machine abilities, or as a consequence of the constrained behavior of the machine.

2.4.2 Results obtained in other experiments

As a matter of fact, the results when the experimental conditions are kept similar limit or eliminate behavioral differences. For instance, Hauptman and Rudnicky (1987) have studied the effect of the interlocutor on the (oral) interaction with an electronic mail system (a "typed input" condition was also studied, but only for the machine interlocutor). Their main result is that the "typed input to machine" condition differs from the two others. About the two "oral interaction" conditions, the authors conclude that dialogues do not differ according to the number of utterances, the time needed to complete the task, the number of emitted words, the mean number of emitted words per utterance, the vocabulary, the use of pronouns, the well-formedness of the utterances, the type of utterances (questions, commands, assertions). This is all the more striking since the machine had been presented to the subjects as having low abilities, which should have resulted in restrictions in language use.[1]

To a large extent, the results reported by Kennedy et al. (1988) concur with these findings. The authors have conducted three "Wizard of Oz" experiments, using typed interaction and a blind experimenter (i.e. the experimenter is unaware of their role). The results do indicate some language modifications according to the condition, but, as the authors point out, there are globally very few differences between the two groups.

1. However, this may have been counterbalanced by the characteristics of the task, which led to very limited interactions: conversation was mostly limited to commands.

These studies do not provide a definite answer to the question of the degree of appropriateness of the human-human communication model to human-machine interaction. An experiment in which the sole factor of variation is the interlocutor's identity for the subject is reported below. Of course, we have attempted to avoid the bugs described above. Moreover, previous experiments have tended to neglect the time factor, i.e. the modifications that may occur due to the learning of the situation.

3 Methodology

The experimental situation attempts to simulate a quasi-natural situation of telephone inquiry. The experimental paradigm relies on the comparison of two groups of subjects equal in all respects save for the supposed identity of their interlocutor. For one group, instructions indicate that the interlocutor is another human while, for the other group, they indicate it is a talking machine. Of course the interlocutor is the same in both conditions and she is a woman.

The experimental protocol has to be controlled in several ways:

- the information provided about the identity of the interlocutor must be verisimilar, whatever the condition, and this without creating differences in the dialogue between these two conditions;

- the inquiry situation has to be plausible and sufficiently rich to elicit dialogues: scenarios have to be designed so that comparable dialogues can be observed in both conditions;

- since one of our interests is to evaluate the variations of verbal behavior over time, subjects will have to come back several times.

- subjects must be comparable: variables such as sex, age, education have to be controlled

- the experimenter must keep a constant attitude throughout the experiment, whatever her interlocutor. She must not be aware of the group to which the subject belongs. The experimenter must keep constant behavior over time: she must then be trained before the experiment.

3.1 Credibility of the experimental situation for both conditions

3.1.1 Voice quality

The problem is to define a telephone voice which can be judged as produced by a machine by subjects of the machine group, and as produced by another human by subjects of the experimenter group.

Standard telephone dialogue preserves most of the prosody of the human voice. For the machine group, such a voice quality will not be thought as verisimilar as produced by a machine. This solution has thus been rejected.

The use of a "vocoder" or of another similar machine poses a problem for the experimenter group: subjects of this group may find the voice too "machine-like". This solution has thus been rejected.

The solution we have chosen is to degrade minimally the sound signal, through an equalizer, which outputs a voice judged acceptable both from a machine and from another human. Several output voices ave been recorded and proposed to judges chosen at random. We have kept the voice quality that produced most hesitations in those judges as regards the identity of the interlocutor (55% of "human" judgements vs 45% of "machine" judgments).

3.1.2 Influence of the instructions

The credibility of the experimental situation has been strengthened by the instructions. Instructions pretended that the goal of the experiment was to test the acceptability of a new channel of transmission (through optic fibers), which tended to degrade the voice heard. The identity of the interlocutor was mentioned only as a subsidiary element.

3.2 The domain of inquiry and the scenarios

The domain of inquiries concern air timetables and fares. This domain is rich enough to elicit long dialogues, if the problems set to the subjects are not too simple. Two solutions were possible as concerns the scenarios provided to the subjects.

The first solution consisted in repeating a limited number of scenarios on all subjects. This solution was satisfying from the point of view of group comparability. However, we have rejected it, since the experimenter would not have been able to keep a constant attitude after a few repetitions of the same scenarios.

The second solution was to build as many scenarios as needed, each scenario being "played" only once. This solution could be applied only if a taxonomy of the different scenarios was devised, in order to obtain classes of scenarios and to propose in each session an equivalent set of scenarios.

Our final decision is close to that second solution. All scenarios are different, but are played twice, one time in each condition. In order to avoid the risk that the experimenter remembers a scenario, a delay has been introduced between the use of the two replications of the same scenario (e.g. if scenario w appeared for the machine group in session 1, it only appeared in session 3 for the other group). The experimenter in charge of the dialogue sessions did not participate in the elaboration of the scenarios or in their classification. During the dialogues, she was not aware of the scenario being played; she only perceived what the subject said about the problem.

In all, 200 scenarios have been elaborated, grouped in 9 classes:

- three classes of simple problems (timetables and fares of a trip of graded complexity);
- embedded problems, in which a trip is embedded in another one;
- optimization problems, in which the subject is requested to attempt to maximize some constraints;
- unfamiliar problems: the scenario mentions an unusual aspect (e.g. you have a sailboard in your luggage);
- conflicting problems: the scenario is impossible to fulfil without violating one of the constraints;
- embedded and optimization problems: these scenarios combine the two classes above;
- embedded and conflicting problems: these scenarios combine the two classes above.

At each session, each subject was given a booklet including: the objective of the experiment, the instructions he had to follow, a very simple training scenario, ten scenarios (4 simple problems, and one per other class). Each scenario was presented on a new page; at the bottom of the page, a question was asked about the quality of the communication. The last page of the booklet included a questionnaire about the quality of the communication. The wording of the questionnaire varied slightly according to the group the subject belonged to.

Scenarios appear in the booklet in a random order. Booklets were given to the subjects at random, save from the fact that the same booklet could be only used twice (one time in each condition) and that it could not be given twice in the same week.

3.3 Subjects

Subjects are all male, aged between 18 and 33, and have at minimum a high school education. Subjects participated in three sessions, with a minimum interval of one week between sessions. An interview was conducted after the subjects' last session: it concerned the opinions on the system, the difficulties they had met, the solutions they had chosen.

3.4 The experimenter

The experimenter had to be an expert in the domain (consultation about air travel), constant in her verbal productions, capable of controlling as much as possible her prosody. As previously mentioned, she should not be aware of the subject's group nor of the scenarios.

The domain expertise is composed of two dimensions: expertise in air travel and expertise in consultation dialogues. Training has concerned these two dimensions. The acquisition of air travel knowledge by the experimenter has been done by learning a (large) set of domain rules and by learning the use of a (paper) data base. The acquisition of consultative knowledge has been done through observations of real consultations and through a two-day training period in the experimental conditions (subjects who participated in this training sessions were not used later).

An additional training concerned communication, and was conducted in parallel. Simple rules were defined:

- no constraint on comprehension;
- no constraint on production except: avoid slang;
- attempt to control prosody;
- no constraint on dialogue management, as long as it remained stable from one session to the other; the experimenter was requested to be "natural and spontaneous";
- avoid personal topics, and more generally, avoid digressions: behave neutrally and professionally.

Assisting the experimenter soon appeared necessary. The fact that the experimenter did not dispose of the scenario meant that she had to find a solution to the problem while controlling her communication and that she might spend much time looking for the answer, which could hinder her credibility as a machine.

Thus, the experimenter was helped by an assistant. The assistant knew in advance the problems that were going to be set, and disposed of the elements of information that were necessary, and that she could prepare before the call. She listened to the dialogue, and waited for the subject to request information, which she then passed to the experimenter. This procedure allowed to speed up the dialogue without telling the experimenter the problem that was to come.

3.5 Physical layout

The equipments are set in two different rooms. The subject is given a telephone, in one room, which receives the distorted voice of the experimenter. The experimenter and her assistant stands in the other room, with a telephone, a tape-recorder (on which the telephone is connected) and all the necessary information.

3.6 Situation evaluation

The experiment had attempted to define an experimenter's behavior that would sound verisimilar both for an human and for a machine. This unavoidable constraint could provoke different bias: the machine behavior could sound unrealistic for a machine, too human; inversely, the experimenter's behavior could be judged too artificial, too machine-like; finally the experimenter's behavior could be judged as neither human nor machine-like, but simply strange.

This last bias is particularly dangerous. As a matter of fact, the goal of the experiment is to evaluate the relevance of the human-human model for man-machine interaction. Since human-human behavior is the reference point, its naturalness has to be tested

Thus an additional control experiment was conducted, in order to evaluate the naturalness of the gathered dialogues. A random sample of the dialogues has been selected and presented at two groups of judges (linguists and laymen). The judge had to fill a questionnaire about the dialogues. The questionnaire was based on the semantic differential technique (Osgood et al., 1957).

3.7 Data exploitation

The collected data consist of:
- over four hundred conversations by twelve subjects;
- post-dialogues and sessions opinion forms;
- post-experiment interviews with subjects.

All conversations have been transcribed (orthographic transcription). The analysis has focused on the dialogue corpus. The interviews have been used for suggesting explanations and discussions of the corpus results; the opinion forms have been used as complementary results.

Two categories of results are presented: results concerning the first session (synchronic point of view, 6 subjects per group) and results concerning the three successive sessions (dyachronic point of view, 3 subjects per group). The dialogs have been examined from three main viewpoints: the linguistic practice, the dialogue strategies and task management and the users'expectations with respect to system performances.

The different cues we collected cover three categories. They may concern dialogue, linguistics or scenario processing.

• Concerning dialogue

(1) average length of dialogues
(2) average length of utterances (an utterance has to be understood as any sequence of statements from the same speaker).

• Concerning linguistics

* Lexicon

(3) TTR
(4) vocabulary size
(5) average length of words in the vocabulary
(6) nominalization
(7) domain-dependent vocabulary versus general purpose vocabulary

* Syntax

(8) hesitations
(9) broken statements or uncompleted statements

(10) connection words between phrases (intra) or between utterances (extra)
(11) interruptions by the other speaker

* References
(12) nominal anaphora
(13) pronominal anaphora

* Stereotypes politeness phrases
(14) stereotypes politeness phrases

• Concerning scenario processing

(15) number of relevant data in the subject first request compared to the number of relevant informations given in the script. Relevance means related to the task (i.e. ex........)
(16) number of relevant data referred to by the subject in the whole dialogue compared to the number of relevant informations given in the script.
(17, 18) The same for total information (including relevant and irrelevant informations).

These cues have been selected because they enable to verify current hypothesis in the literature and our own hypothesis concerning human computer dialogue; all the mentioned hypotheses concern the machine group:

- Subjects control their linguistic expression and more generally their behavior (cues 5, 6, 8, 12).
- Subjects try to be more explicit, more precise (cues 1, 2, 3, 4, 5, 6, 7 , 12, 13)
- Subjects tend to assimilate the computer to human persons (humanizing) (cue 14)
- Subjects tend to control the dialog (cues 10, 11, 15, 16, 17, 18)
- Subjects tend to decompose more the problem, because they doubt the problem solving competence of the machine.

4 Results

Results will be presented in three sections : (i) results concerning the control of the experiment, (ii) results concerning the first session (with 6 subjects per group), (iii) results concerning the evolution of the two groups between the first and last sessions (with 3 subjects per group).

4.1 Pre-results

• Naturalness of the dialogues : we had to make sure that the constraints forced on the experimenter had not biased the experiment; in particular, dialogues in the operator condition had to be natural, not artificial. Two groups of judges were consulted : ten linguists and fifty people chosen at random. The judges had to listen to a few dialogues from the corpus collected during the experiment. Then, they were asked to characterize these dialogues and the interlocutors' behavior by filling in a form consisting of a set of semantic differentiators (with a five degrees scale). The linguists were interviewed afterwards and asked to comment on their judgements.

All judgements agree as regards the experimenter's behavior : she speaks slowly and is very calm. She is natural enough rather impersonal and laking in warmth; she is competent and efficient; she shows some clumsiness at the dialog level.
Thus, the judges consider the situation to be natural.

• Analysis of answers to the booklet questionnaires : The most interesting result concerns the discomfort provoked by voice degradation : subjects of the operator group feel far less at ease than subjects of the machine group (figure...). Since conditions were strictly identical, this result indicate that tolerance to voice degradation is higher when the interlocutor is thought to be a machine. In addition, this result also indicates that subjects strongly adhere to the instructions : subjects of the machine group did believe to be interacting with a machine and the reverse for subjects of the operator group.

4.2 Results concerning the first session

t tests (unpaired, two-tailed) have been computed using all dialogues of the twelves subjects in the first session (in all 131 dialogues). In some cases however the t test can only be computed on average values per subject (e.g. vocabulary size and TTR)

4.2.1 Vocabulary

• Vocabulary size

Specific of the machine group	216	Total machine: 563
Common to the two groups	347	
Specific of the operator group	158	Total operator: 505

• Mean vocabulary size in each group

Machine	252.8	+/-44.2
Operator	231.5	+/-38.6

The difference of mean vocabulary sizes is not significant (t test).

• Mean type-token ratio

Machine	.109
Operator	.120

The difference between TTRs is not significant (t test).

• Mean length of the words

number of syllables	words common to both groups		words specific to the machine group		words specific to the operator group	
1	127	(36.6%)	24	(11.1%)	18	(11.4%)
2	112	(32.3%)	85	(39.3%)	71	(44.9%)
3	72	(20.7%)	69	(31.9%)	50	(31.6%)
4	19	(5.4%)	32	(14.2%)	13	(8.2%)
5	3	(0.8%)	6	(2.8%)	5	(3.1%)
6			1	(0.6%)		
			mean: 2.587		mean: 2.487	

Table 1
Mean length of words common to both groups or specific to one group in the first session

The comparison of the two groups as concerns the length of the words is not significant (t test).

• Nature of the vocabulary

	Application-specific	Scenario-specific
Machine group	82	79
Operator group	71	53

Comparisons of the two groups as regards their respective use of application-specific and scenario-specific vocabulary are not statistically significant (test e for comparing two percentages. Application: 1.367, NS. Scenario: 0.6122, NS).

4.2.2 Other linguistic results

Table 2 presents additional results concerning the number of words and interventions per dialogue, word length, requests, unfinished sentences, connectors, fill-up words, calls for attention, and anaphora.

	Machine group	Operator group	t	Signif.	df
Number of words per dialogue	210.39	170.86	2.234	.02	129
Number of interventions per dialogue	26.39	22.34		NS	
Mean number of words per interventions	8.916	8.332		NS	
Mean word length (syllables)	2.588	2.497		NS	
Requests for repetition and for confirmation	0.108	0.081	1.9	.06	129
Unfinished sentences (+"heu", breaks, structure switches, ...)	0.144	0.205	2.304	.02	129
Connectors between interventions	0.085	0.125	2.628	.01	129
Connectors within interventions	0.066	0,046		NS	
Fill-up words	0.388	0.454	1.809	.07	129
Calls for attention and politeness	0.088	0.131		NS	
Anaphora (total)	3.667	3.441		NS	
Noun anaphora	2.175	0.882	4.485	.001	129
Pronoun anaphora	1.469	2.632	2.939	.004	129

Table 2 : Linguistic productions of the two groups in the first session

In summary

- The machine group speaks more (number of words per dialogue).
- The machine group control and pays more attention to their expression (fill-up words, unfinished sentences).
- The machine group monitors the functioning of the communication (requests for confirmation and repetition)
- The machine group does not make explicit the switches between the topics and themes (connectors between interventions)

A number of results are not significantly different in the two groups: TTR, vocabulary, politeness, connectors within interventions, interventions, word length, number of words per intervention.

4.2.3 Scenario processing

	Machine group	Operator group	t	Signif.	df
Content of the 1st request (contextual + pb relevant info)	O.38	O.46	1. 69	.O9	129
Content of the 1st request (pb relevant info only)	O.5O8	O.565		NS	
Content of the whole dialogue (contextual + pb relevant info)	O.942	0.958		NS	

Table 3
Information provided at the first request and in the whole dialogue in the first session

The machine group tends to provide more information in the first request (table 3).

4.2.4 Discussion of the first session results

Comparison with previous results

The tendency of the subjects of the machine group to control and simplify their expression is confirmed by the results of this experiment; several cues converge to indicate it.

However, two differences are worth noticing:

- subjects of the machine group speak more than subjects of the operator group (although the number of interventions per dialogue and the number of words per intervention do not differ significantly in the two groups);

- groups do not differ in vocabulary size, word length and TTR.

Most probably, these differences between our results and those of other authors are caused by the use of different experimental conditions, by a the better control of these conditions in this experiment, and by the higher level of complexity of the task in this experiment. The politeness phrases, although not pertaining directly to the linguistic level, are not less frequent in the machine group.

Most of the analyses made in other studies focus on the linguistic behavior. Other characteristics may be considered, such as the way subjects process the problems. The following section considers this aspect.

The model of the machine

The subjects' verbal behavior is the consequence of a model of the competences of their interlocutor: linguistic abilities, communication abilities, problem processing abilities for instance. Consequently, the observations of the behavior allow to access the subjects' model of the machine. We propose the following interpretation: the machine is seen as an interlocutor with low linguistic abilities, unreliable from the standpoint of communication and dialogue, not competent from the standpoint of problem processing (or may be lacking domain knowledge).

<u>Low linguistic abilities</u>: This appears in the tendency to normativize and simplify language (less unfinished sentences, more noun anaphora).

<u>Low abilities in communication and dialogue</u>: This appears in the attempts to monitor the functioning of the dialogue (requests for confirmation and repetition).

<u>Low abilities in problem processing (or lack of domain knowledge)</u>: This appears in several facts:

- first, subjects do not bother to provide the machine with indications allowing to follow their line of reasoning (less connectors between interventions; up to a certain extent, the fact that subjects of the machine group do not provide more within-intervention connectors although these subjects are more verbose indicate the same phenomenon). Subjects of the machine group do not sem to believe that dialogue may benefit from a better comprehension by their interlocutor of the succession of their topics; or, in other words, theses subjects believe that the machine is not able to grasp and follow these topics.

- second, subjects of the machine group provide less information in the first request. This fragmentation of the information indicates that subjects' estimation of the machine abilities to process information is limited. This is confirmed by some of the remarks of the subjects in the post-experimental interviews: subjects believe that the machine needs to have the problem decomposed into simpler problems.

- third, subjects of the machine group are more verbose (cf. number of words per dialogue). This result can be compared to those reported by other authors (e.g. Krauss & Glucksberg, 1977) in human-human communication, who observe that the volume of emitted information increases when the interlocutor is judged of low domain competence

4.3 The evolution across sessions

t tests (unpaired, two-tailed) have been computed using all dialogues of the three subjects per group that have participated in the first and third session (in all 128 dialogues).[1]

4.3.1 Vocabulary

Common vocabulary
Session 1 : 204
Session 3 : 176
Vocabulary size: x2= 2.396, 1df, NS

1. In some cases, the t test can only be computed on average values per subject (e.g. vocabulary size and TTR), instead on all dialogues.

Application-specific vocabulary

	Machine group	Operator group
Session 1	60/148	46 /87
test Σ	1.448 NS	0.239 NS
Session 3	69/141	31/61

Table 4
Evolution across sessions:
Use of application-specific vocabulary

Percentages are compared using a sigma test. No comparison is significant (table 4).

Scenario-specific vocabulary

	Machine group	Operator group
Session 1	62 /148	45/87
test Σ	1.674 NS	1.384 NS
Session 3	71/ 141	20/61

Table 5
Evolution across sessions:
Use of scenario-specific vocabulary

Percentages are compared using a sigma test. No comparison is significant (table 5).

4.3.2 Other linguistic results

Tables 6 to 17 present additional results concerning the number of words and interventions per dialogue, requests, unfinished sentences, connectors, fill-up words, calls for attention, and anaphora.

It is to be noted that, in some cases, the result observed on three subjects (in session 1) differs from that observed on 6 subjects (in the same session). The difference is here non significant, while it was significant with 6 subjects. This is the case for requests for repetition or confirmation and for fill-up words.

Number of words per dialogue	Machine group	t test	Operator group
Session 1	213.2	2.12 S.O3 62ddl	164.14
t test	NS		1.852 S.06 64 ddl
Session 3	176,2	2,553 TS.O1 62ddl	128.12

Table 6
Evolution across sessions:
Number of words per dialogue

Number of interventions per dialogue	Machine group	t test	Operator group
Session 1	24.48	NS	23.79
Test	NS		NS
Session 3	22.68	NS	19.34

Table 7
Evolution across sessions:
Number of interventions per dialogue

Number of words per intervention	Machine group	test	Operator group
Session 1	8.98	NS	8.63
Test	1,94, S.O5 ddl 6Oddl		1.97, S .O5 ddl 64
Session 3	7.89	NS	7.2

Table 8
Evolution across sessions:
Number of words per intervention

Requests for confirmation or repetition	Machine group	test	Operator group
Session 1	O.1O2	NS	0.1
Test	NS		NS
Session 3	O.O7	NS	0.081

Table 9
Evolution across sessions:
Requests for confirmation or repetition

Unfinished sentences	Machine group	test	Operator group
Session 1	0.129	2.02 S.04	0.232
Test	NS		NS
Session 3	0.123	NS	0.162

Table 10
Evolution across sessions:
Unfinished sentences

Fill-up words	Machine group	test	Operator group
Session 1	0.418	NS	0.452
Test	NS		2.189 TS .03
Session 3	0.408	NS	0.361

Table 11
Evolution across sessions:
Fill-up words

Connectors between interventions	Machine group	test	Operator group
Session 1	0.083	3.193, S .002	0.163
Test	NS		NS
Session 3	0.081	3.385, TS .001	0.155

Table 12
Evolution across sessions:
Inter-intervention connectors

Connectors within interventions	Machine group	test	Operator group
Session 1	0.066	NS	0.046
Test	NS		NS
Session 3	0.089	NS	0.068

Table 13
Evolution across sessions:
Within-intervention connectors

Calls for attention	Machine group	test	Operator group
Session 1	0.113	NS	0.09
Test	NS		NS
Session 3	0.107	NS	0.086

Table 14
Evolution across sessions:
Calls for attention

Anaphora (total)	Machine group	test	Operator group
Session 1	3.094	NS	3.258
Test	NS		NS
Session 3	2.594	NS	2.129

Table 15
Evolution across sessions:
Use of anaphora

Nominal anaphora	Machine group	test	Operator group
Session 1	2.065	3.09 TS.OO3	0.829
Test	2.773 TS.OO7		NS
Session 3	0.844	NS	0.548

Table 16
Evolution across sessions:
Use of nominal anaphora

Pronominal anaphora	Machine group	test	Operator group
Session 1	0.871	2.722 S.008	2.324
Test	2.171 S .03		NS
Session 3	1.75	NS	1.581

Table 17
Evolution across sessions:
Use of pronominal anaphora

The two groups speak less and less (number of words per intervention or per dialogue), but the trend is significantly less strong for the machine group. Fill-up words decrease significantly in the operator group, but not in the machine group.

4.3.3 Evolution in problem processing

Concerning scenario processing (tables 18, 19, 20):

- for the operator group, there is a very significant decrease in the number of information provided in the first request and in the whole dialogue, and this for problem relevant information as well as for all information.

- inversely, there is very little evolution for the machine group (non-significant variations). The only significant result is in an opposite direction: subjects of the machine group increase the number of information given in the first request.

1st request (context info + pb relevant info)	Machine group	t test	Operator group
Session 1	0.405	NS	0.449
t test	NS		3.036, S 0.004
Session 3	0.405	1.852, S.06	0.309

Table 18
Evolution across sessions:
Relevant or contextual information mentioned in the first request

1st request (pb relevant info only)	Machine group	t test	Operator group
Session 1	0.553	1.718, S.09	0.668
t test	1.746 S.08		2.604, S 0.01
Session 3	0.668	2.63 , S .001	0.506

Table 19
Evolution across sessions:
Problem-relevant information mentioned in the first request

All dialogue (pb relevant info only)	Machine group	t test	Operator group
Session 1	0.758	NS	0.79
t test	NS		1.89, S .06
Session 3	0.827	2.7 S .008	0.698

Table 20
Evolution across sessions:
Problem-relevant information mentioned in the whole dialogue

4.3.4 Interpretation of the evolution

The two groups progress, but non on the same dimensions. The machine group mainly evolve in the use of anaphora, increasing its use of pronominal anaphora to the expense of noun anaphora. This result can be explained in the following way: the machine group realizes that its effort towards explicitness is useless, and switches to using more pronominal anaphora, thus tending to behave as the operator group in that respect. The evolution of the machine group concerns the model of the machine linguistic abilities.

The operator group progresses mainly with respect to hesitations and scenario processing, whilst the machine group is relatively stable on these aspects. Three interpretations may be thought of:

1. The machine group could evolve if more practice time was available. In this hypothesis, only the evolution speed would distinguish the two groups. It could be assumed that the machine group has more to learn than the operator group: it must first assess the machine linguistic competences. However, there is no sign of the beginning of an evolution concerning problem processing.

2. The machine group is stable because the subjects' model of the machine would specify that machines are not capable of learning through practice. Some subjects have indeed mentioned this opinion in the post-experimental interviews.

3. The observed phenomenon corresponds to a very general one in the context of human-computer interaction: the tendency of the subjects to stick to the first acceptable functioning, sufficient to achieve the goal they seek (cf. Norman, 1983). Subjects are not seeking optimality, but effectiveness (Maïs, 1988). In this perspective, and knowing that they interact with a

machine and not with a human operator, subjects would stick to the first strategy yielding a satisfactory result.

None of these interpretations can be supported directly by our data. The third interpretation, however, appears as the most likely one, considering our knowledge of human-computer interaction.

5 Conclusion

5.1 Problem processing as the key factor

As a result of the analysis above, it appears that the machine group speaks more, is more careful as regards its expression, and doubts the machine linguistic and processing abilities. This group seems to progress more as concerns its model of the machine linguistic competence than as concerns the machine processing competence. It monitors alone problem processing, does not bother too much about facilitating the understanding (by its machine interlocutor) of problem processing. It behaves more as the user of a tool than as the participant to a collective task. In that sense, it seems there is no collective processing of the problems, which may explain the absence of evolution of the machine group.

This seems to constitute the main element of response to the question set in the introduction of this study: can human-computer communication be analyzed using the model of human-human communication ? The results tend to provide a negative answer, since the collective dimension of the task is absent in the human-machine situation.

5.2 Implications for the design of human computer interactive systems

The situation we studied (complex task, no constraints in production or comprehension for the simulated machine) does not give rise to the modifications of behavior observed in other experiments: vocabulary size, word lengths, number of interventions for instance are equivalent in both groups. What is more, some of the results we obtain contradict the results of other experiments: in particular, subjects of the machine group are more verbose than subjects of the operator group.

We do not believe that these differences are caused by an experimental artifact. On the contrary, we have paid special attention in making sure that the two experimental conditions were strictly identical (save for the supposed identity of the interlocutor), which was not he case in most other available experimental studies.[1]

Globally, the linguistic behavior of the subjects of the machine group is at least as complex as the behavior of subjects of the operator group. Thus, what provokes, in other studies, the modifications of the subjects' linguistic behavior is not the fact of speaking to a machine, but the behavior of the simulated machine.

Moreover, the better experimental control and the repetition of the sessions make apparent two facts:

- from a linguistic standpoint, the evolution goes towards complexification (more numerous pronominal anaphora);

- from the standpoint of problem processing, results indicate differences concerning the problem solving strategy, and particularly the absence of collective resolution.

Consequently, human-computer interface designers have the choice between two directions:

- one possibility is to increase the quality of speech recognition and understanding (rather than to develop sophisticated dialogue strategies or systems of assistance for problem solving);

- another possibility is, in order to simplify the task of the machine, to constrain the user's linguistic production.

Concerning this second possibility, it has been seen above than previous experiments have concluded to various alterations of verbal behavior as a response to a variously constrained machine behavior (various restrictions in comprehension or production). The trick is then to devise a machine behavior that gives rise to a machine-understandable user behavior. The general scheme is presented in Fig.2.

---

1. The counterpart is that experimental conditions of this study may appear as unrealistic considering the state of the art in the domain of unrestricted speech recognition and comprehension.

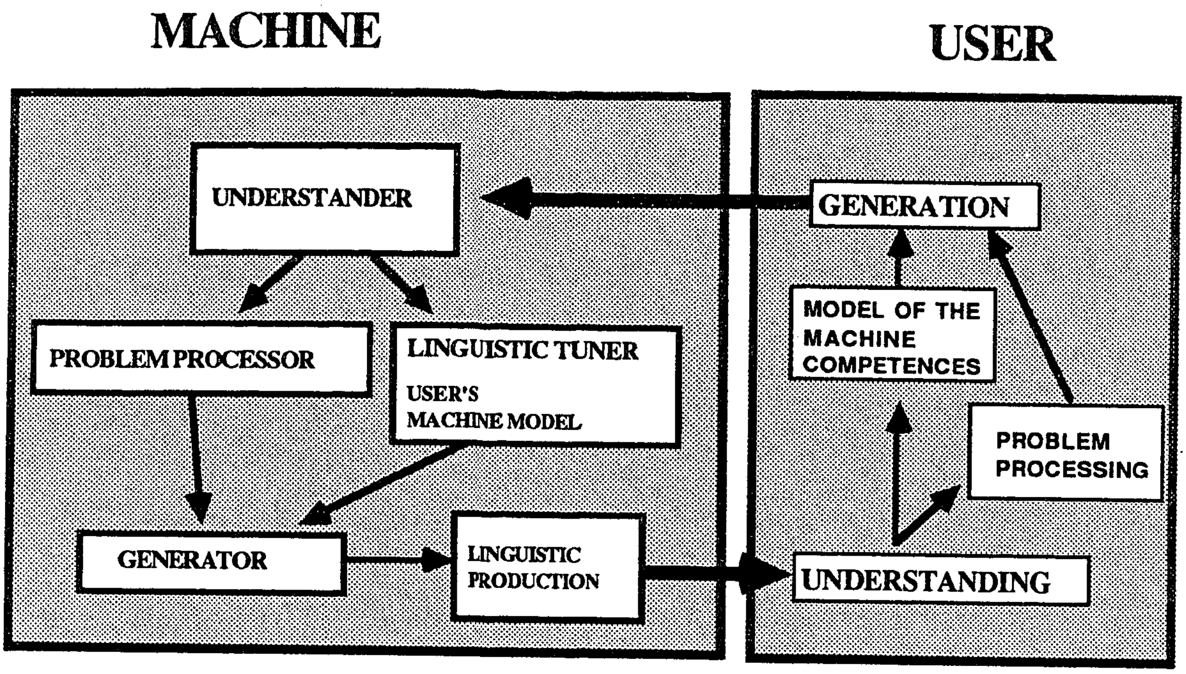

Figure 2
A model of user-computer interaction processes

The important component of this figure is the linguistic tuner. Its task is to assess whether the user's linguistic input is acceptable, considering the machine processing abilities, and, if not, to alter appropriately the generation characteristics of the machine outputs. Following our results and those of previous studies, it is assumed that, as a consequence of these alterations, the user will modify their outputs. The problem is, of course, to generate in this way appropriate modifications in the user's outputs, that is a machine-understandable language. The resulting user language should be acceptable to the machine but also to the user: it should not be felt as unbearably constraining.

REFERENCES

Anglin, J.M. (1977) Word, object, and conceptual development. New York: Norton.

Chin, D. (1984) An analysis of scripts generated in writing between users and computer consultants. AFIPS Conference Proceedings, 53, 637-642.

Clark, H.H. & Schunk, D.H. (1980) Polite responses to polite requests. Cognition, 8, 111-143.

Deutsch, B.G. (1974) The structure of task oriented dialogs. In L.D. Erman (Ed.), Proceedings IEEE Symposium on speech recognition. New York: IEEE.

Dumais, S. & Landauer, T. (1982) Psychological investigations of natural terminology for command and query languages. In A. Badre & B. Shneiderman (Eds), Directions in human-computer interaction. Norwood, NJ: Ablex.

Falzon, P. (1989) Diagnosis dialogues: Modelling the interlocutor's competence (submitted for publication).

Falzon, P., Amalberti, R. & Carbonell, N. (1986) Dialogue control strategies in oral communication. In K. Hopper & I.A. Newman (Eds) Foundation for human-computer communication. Amsterdam: North Holland.

Forrest, M.-A., Reilly, R. & Sheehy, N. (1986) Preliminary pilot experiments. In Dialogue studies: Pilot phase (Deliverable 3). Report of ESPRIT Project 527 (CFID) Communication Failure in Dialogue: Techniques for Detection and Repair.

Harris, G., Begg, I. & Upfold, D. (1980) On the role of the speaker's expectations in interpersonal communication. Journal of Verbal Learning and Verbal Behavior, 19, 597-607.

Hauptman, A.G. & Rudnicky, A.I. (1987) Talking to computers: an empirical investigation (Technical Report CMU-CS-87-87-186). Pittsburgh: Department of Computer Science, Carnegie Mellon University.

Isaacs, E.A. & Clark, H.H. (1987) References in conversation between experts and novices. Journal of Experimental Psychology: General, 116 (1), 26-37.

Kennedy, A., Wilks, A., Elder, L. & Murray, W.S. (1988) Dialogues with machines. Cognition, 30 (1), Oct. 1988.

Krauss, R.M. & Glucksberg, S. (1977) Social and nonsocial speech. Scientific American, 236 (2),100-105.

Maïs, Ch. (1988) Pour des systèmes d'aide à la réalisation de procédures sous-optimales. In AFCET (Ed.) Proceedings of the Ergo & IA Conference, Biarritz, October 1-6, 1988.

Meunier, A. & Morel M.-A. (1987) Les marqueurs de la demande d'information dans un corpus de dialogue homme-machine. Cahiers de Linguistique Française, 8.

Morel, M.-A. (1985) Analyse linguistique d'un corpus d'oral finalisé (Centre de renseignements SNCF à Paris). Report to the Greco n°39 Communication Parlée.

Moscovici, S. (1967) Communication processes and the properties of language. In Berkowitz (Ed.) Advances in experimental social psychology (Vol. 3). London: Academic Press.

Norman, D. (1983) Some observations on mental models. In Stevens & Gentner (Eds.) Mental models. Hillsdale, N.J.: Lawrence Erlbaum.

Onorato, L.A. & Schvaneveldt, R.W. (1986) Programmer/non-programmer differences in specifying procedures to people and computers. In E. Soloway & S. Iyengar (Eds), Empirical studies of programmers. Norwood, NJ: Ablex.

Richards, M.A. & Underwood, K. (1984a) Talking to machines: how are people naturally inclined to speak? In E.D. Megaw (Ed.), Contemporary Ergonomics 1984. London: Taylor & Francis.

Richards, M.A. & Underwood, K. (1984b) How should people and computers speak to each other. In B.Shackel (Ed.) Proceedings of INTERACT'84, 1st IFIP Conference Human-Computer Interaction, Londres, 4-7 Septembre 1984.

Shneiderman, B. (1987) Designing the user interface: strategies for effective human-computer interaction. Reading, Mass.: Addison-Wesley.

Spérandio, J.-C. & Létang-Figeac, C. (1986) Simulation expérimentale de la synthèse vocale en dialogues oraux de communication homme-machine. Etude ergonomique. Report to the GRECO n°39 Communication Parlée. Paris: IRAP.

Talbot, M. (1985) Speech technology: is it working? In P. Johnson & S. Cook (Eds.) People and computers: Designing the interface. Cambridge: Cambridge University Press.

Waterworth, J.A. (1985) The cognitive effort of processing synthetic speech. Proceedings of the 11th International Symposium on Human Factors in Telecommunications Conference, Cesson Sévigné, France, 9-13 September 1985. Boston, Mass: Information Gatekeepers Inc.

Zoltan-Ford, E. (1984) Reducing variability in natural language interactions with computers. In M.J. Alluisi, S. de Groot & E.A. Alluisi (Eds.), Proceedings of the Human Factors Society 28th Annual Meeting, San Antonio, Texas, Oct.20-26.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique