



CONFIRMATION STRATEGIES TO IMPROVE CORRECTION RATES IN A TELEPHONIC INQUIRY DIALOGUE SYSTEM

C. Alexia Lavelle, Martine de Calmès, Guy Pérennou

IRIT, Université Paul Sabatier,
118, route de Narbonne,
31062 Toulouse, France
{lavelle,decalmes,perennou}@irit.fr

ABSTRACT

In the framework of ARISE project (Automatic Railways Inquiry Systems for Europe) we developed DEMON, a real time spontaneous speech dialogue system allowing mixed initiative. This system gives information for train connection.

On this system three different confirmation strategies were developed and compared with regard to erroneous confirmation rates.

The first strategy was implicit confirmation question oriented. Those questions were used as widely as possible. Refutation for those questions appeared to be quite low. The reordering of parameters in implicit questions to match the order of mental operations required to answer the question and then the introduction of semi-implicit questions to replace implicit ones by increasing refutation and direct correction rates have helped to increase dialogue success rate without lengthening dialogues.

we observed on correction rates depending on the confirmation question type. Actually, for 95% of explicit confirmation questions holding badly understood parameters users tried to correct, whereas they tried for only 27% of implicit confirmation questions.

We then assumed this had to be considered as the result of an inappropriate dialogue strategy or question design. As suppressing implicit confirmation would have slow down the dialogue [6], which is not tolerable for telephonic applications we have been working on next versions of DEMON (DEMON_1 and DEMON_2) on redesigning implicit questions and application cases for both question types.

We will firstly present the three confirmation strategies used in the three versions of the system and the performed experiments. We will then present results for understandability and efficiency of questions and evolution on success rate and length of dialogues.

1. CONFIRMATION STRATEGIES

A. *DEMON_0: Explicit and implicit single sentence questions*

In this first version of the system implicit confirmations were used as largely as possible. They were used each time there was at least one parameter to confirm and they included every parameter to confirm at the stage they were uttered. They were made of one single sentence following rules of well-formed French questions.

For example, if Monday the 5th of January 1998 as to be confirmed as departure day and departure time as not been provided yet, the question was:

“A quel heure désirez vous partir le lundi 5 janvier 1998?”

“At what time do you which to leave on Monday the first of January?”

Explicit confirmation questions were used only if there was no parameter left to fulfil. They were straightforward well-formed questions.

The same confirmation as above would lead to a question as:

“Désirez-vous partir lundi 5 janvier?”

“Do you which to leave on Monday the 5th of January?”

INTRODUCTION

In the framework of ARISE LE3-4229 project (Automatic Railways Inquiry Systems for Europe) [1], we developed at IRIT a real time spontaneous speech dialogue system allowing mixed initiative called DEMON [2,3]. This system has been developed on a PHILIPS platform and gives over the phone actual information from SNCF'S (French Railway National Service) train timetable database for the connections between the 600 most frequently asked for train stations.

Tests of the first version of DEMON (DEMON_0) in November 1997 showed that about 29% of users got an information about a connection which did not match the one they asked for. This means that 29% of did accept a badly understood value for one parameter of the connection they were asking for. After [4] and [5] we will call those situations *erroneous confirmations*.

The most obvious explication for this phenomenon was that users involved in our experiments were not actual users with real needs. This can explain a certain amount of non-correction but this can not explain the difference

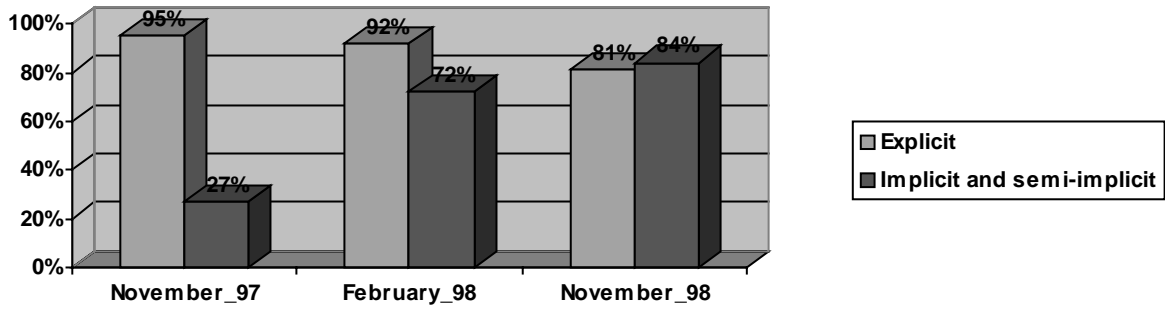


Figure 1 – Refutation Rates

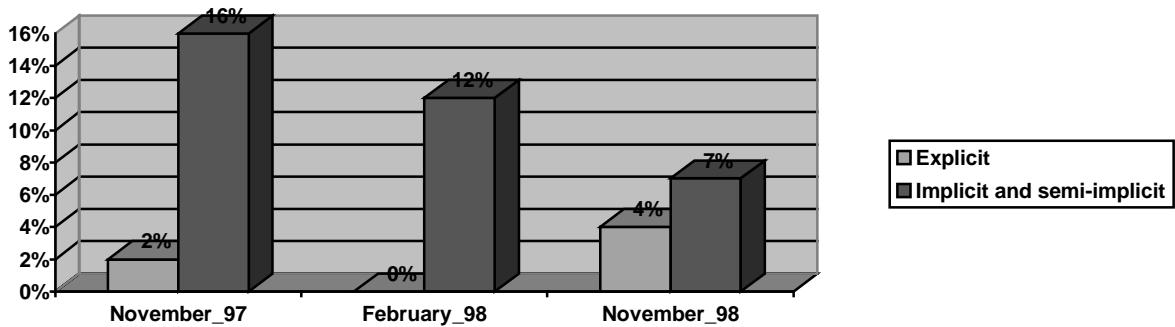


Figure 2 - No answer rates

B. DEMON_1: Explicit and implicit two sentence questions

In this second version, implicit questions were redesigned to match the order of mental operation necessary to answer the question. They were therefore divided in to sentences: first a statement of parameters to confirm, then a question on parameters still to fulfil.

The question for the situation described above was then :
 “Vous partez lundi 5 janvier 1998. A quelle heure?”
 “You are leaving on Monday the 5th of January 1998. At what time?”

Application cases for both explicit and implicit questions remained unchanged.

C. DEMON_2: Explicit and semi-implicit questions

In this last version we introduced semi-implicit questions to replace implicit ones. Those carry the same information as implicit questions : a question on new information and a request for confirmation but, as explicit questions, they clearly state they can be denied. In the situation described above the confirmation question would now be:

“Vous partez lundi 5 janvier 1998. En cas d’erreur corrigez moi, sinon , indiquez l’heure de votre départ.”
 “You are leaving on Monday the 5th of January 1998. In error case, correct me, otherwise, indicate your departure time.”

Application cases for explicit and semi-implicit questions were also redefined. Semi-implicit questions only include confirmation request for concept of the same semantic type. In this applications we consider two semantic types: place for departure and arrival train station and time for departure or arrival day and time.

2. EXPERIMENTS

Observations shown here are based on transcription of recorded calls. We study here 3 corpora of transcribed calls. See table 1. They have been obtained in the framework of ARISE project in collaboration with SNCF (French Railways National Service).

| Corpus | Number of call | Callers |
|-----------------------|----------------|---------------------------|
| November_97 (DEMON_0) | 41 | French railways employees |
| February_98 (DEMON_1) | 88 | French railways users |
| November_98 (DEMON_2) | 200 | French railways users |

Table1 – Summary of experiments

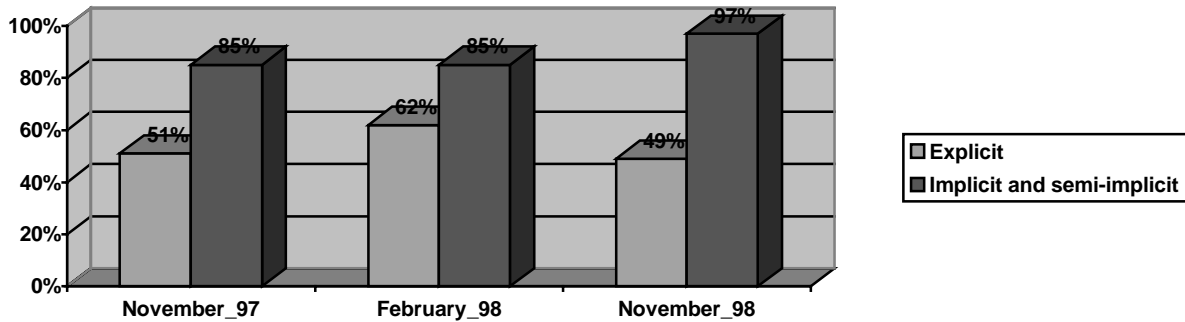


Figure 3 – Direct Correction Rates

The first experiment was performed in November 1997 with SNCF's employees and led to corpus November_97. For this test, participants were asked to call at most three times. The second and third involved SNCF's customers respectively calling from Paris Saint-Lazare train station in February 1998 and from home in November 1998. For those two last experiments participants respectively performed three and two calls.

3. UNDERSTANDABILITY AND EFFICIENCY OF QUESTIONS

Explicit, implicit and semi-implicit confirmation questions were compared in two different ways. Their understandability or answerability and their efficiency in the dialogue.

Two measures are used to evaluate understandability of questions: the *Refutation Rate* and the *No Answer Rate*.

Figure 1 displays the *Refutation Rate* for all three corpora. This measure represents the percentage of confirmations denied over the number of confirmations to be denied (i.e. including badly understood parameters).

The *No Answer Rate*, displayed in Figure 2, represents the percentage of confirmation questions that were not answered in the first place and therefore had to be repeated at least once over the total number of confirmation questions asked. When we look at both those figures we observe that both rates stayed rather stable for explicit questions. This is no surprise since the design of those questions did not change between the three experiments and the application cases change only in the last experiment. The increase of the amount of explicit questions holding more than one parameter induced by this change in application cases might explain the slight decrease of correction rate in corpus November_98. If we consider the implicit and semi-implicit questions the differences are a lot more significant. In DEMON_1 the redesign of questions parameters to match the order of mental operation led to an increase of 45% in *refutation rate*. Then again when we added to those questions a statement that they are actually questions the correction rate increased by 12% and very closely equals the rate obtained for explicit questions. We can also notice that no answer rate

drops from 14% and 12% for DEMON_0 and DEMON_1 to 7% for DEMON_2.

Nevertheless to make it clear for users that they can correct a system statement is not enough. Actually, to deny a confirmation two answer types have been observed. Users either answer "no" or some synonymous or state the correct information with or without adding a clear denial statement. Clearly, if the user answers immediately with the correct information we save an extra dialogue turn asking for the correct information. A question leading to such an answer is considered to be more efficient in the dialogue. The *Direct correction rate*, percentage of direct correction over the total number of denied confirmations, shown in Figure 3, is so our measure for efficiency.

With this regard for all three strategies, implicit and semi-implicit confirmation questions are more efficient than explicit ones. Looking separately at explicit and implicit questions, considering the small amount of data we have, we will be very careful concluding anything. We might suppose that the increase in *direct correction rate* observed for DEMON_2 is partly due to the clear request for correction in semi-implicit confirmations.

4. SUCCESS RATES AND DIALOGUE LENGTH

The overall aim of improving our confirmation strategy was of course to improve the success rate of the system. This means increasing the number of dialogues providing the user with the information he or she requested and reducing or at least not increasing the dialogue length.

Figure 4 shows the success and failure rate for the three corpora. We identify three types of dialogue result:

- *Success* if the user has obtained the information he or she asked for.
- *Failure* if the user or the system has ended the call before all parameters about the intended journey had been given and therefore before the system was able to give any information for relevant travel.

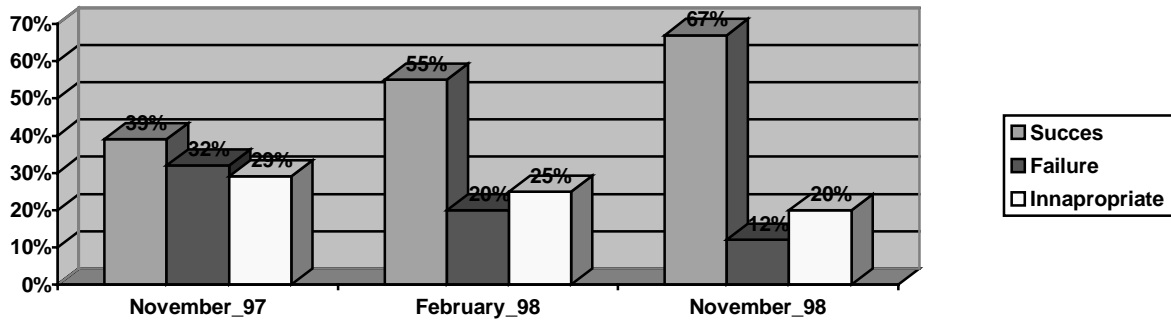


Figure 4 - Dialogue Success and Failure Rates

- *Irrelevant answer* if the system provided information for some travel which did not match exactly one the user asked for.

Improving refutation rate seems to lead to a real benefit for dialogue success rates. Inappropriate answer rate which generally correspond to people accepting badly understood parameters decrease by 4% between each experiment. The failure rate which mostly correspond to people hanging up because they have been misunderstood but do not know how to correct nearly decrease by 10%.

Table 2 shows the average number of dialogue turns per transaction, where *dialogue turn* stands for a system question and the user's answer.

| Corpus | Success | Failure | Inappropriate |
|-------------|---------|---------|---------------|
| November_97 | 12 | 7.5 | 12.3 |
| February_98 | 12.2 | 8.5 | 13.4 |
| November_98 | 9.6 | 8.4 | 12.3 |

Table 2 – Average number of turns per transaction

The limitation of semi-implicit question to hold only parameters of the same semantic type in DEMON_2 induced an increase in the number of question to be asked. It was likely that this would have increased the number of turns per dialogue. It proves not to be the case. Maybe this was counterbalance by the fact most implicit questions were answered and nearly all of them with direct corrections.

CONCLUSION

From those results it appears that the reordering of parameters in implicit confirmations to match the order of required mental operation and then the introduction of semi-implicit confirmations, by increasing refutation and direct correction rates, have helped to increase dialogue success rate without lengthening dialogues.

ACKNOWLEDGMENTS

The authors acknowledge the invaluable contribution of *Philips Research Labs Aachen* for providing the Dialogue workstation used during ARISE Project.

REFERENCES

- [1] *M. Blasband*, "Speech Recognition in Practice the ARISE Project (Automatic Railway Information Systems for Europe)", *La lettre de l'IA numéros 134-135-136*, proceedings NIMES'98, Nîmes, France, pp.207-210.
- [2] *G. Pérennou, M. de Calmès, C.A. Lavelle, R. Tronel*, "DEMON: Un système de dialogue oral spontané pour l'accès téléphonique aux informations d'horaire de train – Problèmes de robustesse", *La lettre de l'IA numéros 134-135-136*, proceedings NIMES'98, Nîmes, France.
- [3] *C.A. Lavelle, M. de Calmès, G. Pérennou*, "A Study of Users' Behaviors in Different States of a Spontaneous Oral Dialogue with an Automatic Inquiry System", *IVTTA'98*, Turin, Italy, September 1998.
- [4] *M. Danieli, E. Gerbino, L.M. Moisa*, "Dialogue Strategies for improving the usability of Telephone Human-Machine Communication", *Dialogue Systems, Bring Speech and NLP Together in Real Applications*, Madrid, Spain, 1997
- [5] *L. van Haren, M. Blasband, M. Gerristen, M. van Schijndel*, "Evaluating Quality of Spoken Dialogue Systems: Comparing a Technology-focused and a User-focused Approach", *First International Conference on Language Resources & Evaluation*, Granada, Spain, 1998, pp 655-660
- [6] "Modeling Dialogue Control Strategies to Relieve Speech Recognition Errors", *Y. Niimi, Y. Kobayashi*, *EUROSPEECH'95*, Madrid, Spain, September 1995, pp 1177-1180