The LIMSI ARISE System*

L. Lamel, S. Rosset, J.L. Gauvain, S. Bennacef††,

M. Garnier-Rizet†, *B. Prouts*†

Spoken Language Processing Group, LIMSI-CNRS, 91403 Orsay, France

†VECSYS, 3 r. de la Terre de Feu - Les Ulis, 91952 Courtabœuf, France

†† Now with the Vecsys Company

January 26, 2003

Keywords: spoken language systems, speech recognition, speech understanding, natural language understanding, information retrieval dialog

Number of pages: 25

1 table

8 figures

Published in *Speech Communication*, vol 31 (4):339-354, August 2000.

*This work was partially financed by the LE-3 project 4223 ARISE.

List of Tables

List of Figures

Abstract

The LIMSI ARISE system provides vocal access by telephone to rail travel information for main French intercity connections, including timetables, simulated fares and reservations, reductions and services. Our goal is to obtain high dialog success rates with a very open interaction, where the user is free to ask any question or to provide any information at any point in time. In order to improve performance with such an open dialog strategy, we make use of implicit confirmation using the callers wording (when possible), and change to a more constrained dialog level when the dialog is not going well.

Le système ARISE du LIMSI est un serveur téléphonique d'informations sur les trains grandes lignes en France. Il fournit les horaires, les prix et réservations simulés, les réductions ainsi que les services. Notre but est d'obtenir des taux de succès de dialogue élevés avec une interaction très ouverte, offrant à l'utilisateur la possibilité de poser des questions ou de donner des informations à n'importe quel moment. Afin d'améliorer les performances avec une telle stratégie, nous avons utilisé la confirmation implicite en fonction des formulations des utilisateurs ainsi qu'un dialogue plus contraint lorsque l'interaction se passe mal.

Das vom LIMSI entwickelte ARISE System erlaubt den vokalen Zugriff auf Zugreiseinformationen. Neben Fahrplanauskünften zwischen den wichtigsten französischen Intercity-Verbindungen simulierte das System Zugfahrpreise, Ermäβsigungen und anderen Serviceleistungen. Unser Ziel ist es, hohe Erfolgsraten im Mensch-Maschine Dialog zu erhalten. Erreicht werden soll dieses Ziel aber unter Benutzung einer sehr offenen Dialogstrategie, bei der der Benutzer zu jedem Zeitpunkt dem System Fragen stellen oder Informationen liefern kann. Um die Erfolgsraten mit solch einer offenen Dialogstrategie zu verbessern, benutzen wir die Technik oler impliziten Konfirmation, bei der (wenn möglich) die Formulierungen des Benutzers wiederverwendet werden. Sollte der Dialog schlecht verlaufen, so wechseln wir zu einem begrenzteren Dialog-Niveau.

1 Introduction

The LIMSI ARISE (Automatic Railway Information Systems for Europe) system provides vocal access to rail travel information. The demonstration system, development of which has been partially financed by the EC through the LE-3 ARISE project, provides timetables, simulated fares and reservations, and information on reductions and services for the main French intercity connections.

The system providing information for the main intercity connections is largely based on the spoken language system developed for the LE-MLAP RAILTEL project (Bennacef et al, 1996; Lamel et al, 1996; Lamel et al, 1997). Callers are able to obtain information taken from the French National Railways (SNCF) static timetables and additional information about services offered on the trains, fares and fare-related restrictions and reductions. The system is based on a modular architecture and contains six modules: a medium vocabulary, real-time, speaker-independent, continuous speech recognizer which transforms the acoustic signal into the most probable word sequence; a natural language understanding component which first carries out a literal understanding of the text string output by the speech recognizer, and then reinterprets the query in the context of the ongoing dialog producing a semantic frame; a mixed-initiative dialog manager who, with its goal of providing information to the user, ensures communication between the user and the DBMS; a database retrieval component which accesses the RIHO database provided by SNCF to extract the requested information using a pseudo-SQL request generated from the semantic frame; a generation component which outputs a natural language response based on the dialog state, the caller's query, and the information returned from database access; and a synthesizer which concatenates speech units corresponding to the generated text string, producing very natural sounding synthetic speech.

Compared to our RAILTEL system, the main advances in ARISE are in dialog management, the use of confidence measures, dialog state-dependent models, an optional spell mode for city/station names, a time-out facility to aid the caller, and in the inclusion of a barge-in capability which allows a more natural interaction between the user and the machine. In the next section we provide an overview of the system. Section 3 focuses on the dialog strategy, highlighting recent advancements. In Section 4 we report on user trials carried out with recruited subjects as well as punctual evaluations carried out by the SNCF, a partner in the ARISE project.

2 System Overview

The main components of the spoken language system are shown in Figure 1. The system runs on a Unix workstation with a telephone interface. The phone server manages the low-level interface with the phone hardware (pickup/hangup); controls the data organization, creating a directory for each call to store the signal and log files; and ensures the synchronization of the speech recognizer, the dialog

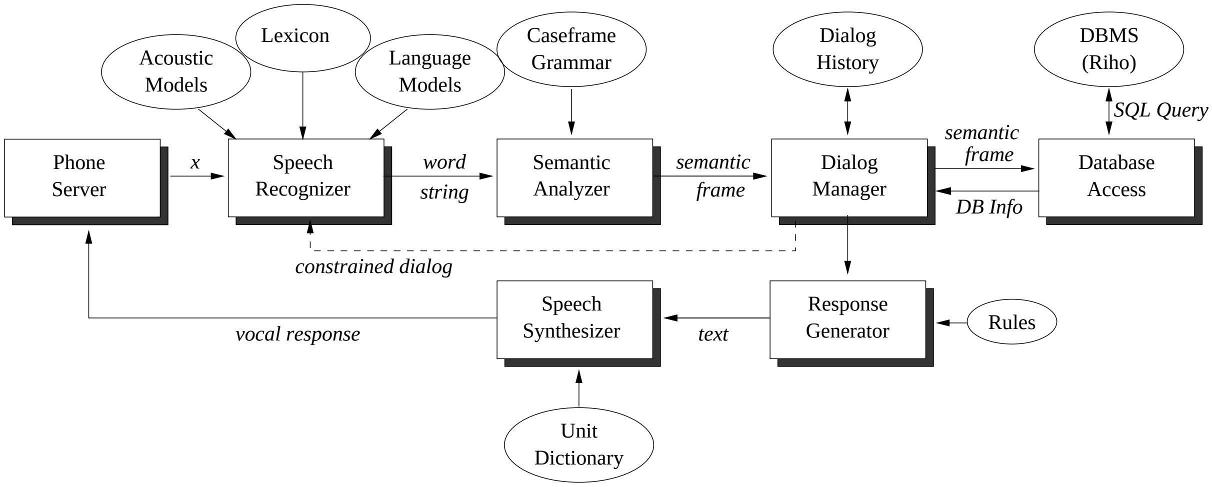


Figure 1: Overview of the ARISE system for spoken language information retrieval. **x** is the input speech signal.

manager and the speech synthesis module. Multiple telephone lines can be connected to the telephone interface thus allowing multiple systems to run on the same machine.

The speaker independent, continuous speech recognizer (Gauvain et al, 1994) transforms the acoustic signal into the most probable word sequence. The recognition vocabulary contains about 2000 words, including over 500 station names. Speaker independence is achieved by using acoustic models which have been trained on speech data from a large number of representative speakers, covering a wide variety of accents and voice qualities. The feature vector contains 12 MFCC cepstral coefficients computed on the 0.3-4kHz telephone band and their first and second order derivatives. The recognizer uses continuous density HMM with Gaussian mixture for acoustic modeling and *n-gram* backoff language models (Katz, 1987). Context-dependent phone models can be used to account for allophonic variation observed in different contextual environments. The acoustic models have been trained on about 36 hours (8500 calls) of data. The *n*-gram statistics are estimated on the transcriptions of spoken queries. Since the amount of language model training data is small, some grammatical classes (such as cities, days, months, etc) are used to provide more robust estimates of the *n*-gram probabilities.

Real-time synchronous decoding required implementation of a new approach to cepstral normalization and energy normalization (Gauvain, Gangolf and Lamel, 1996). Other techniques used to obtain real-time performance include lexicon trees, distributed LM weights, Gaussian shortlists and

Conches	kOS kOSx(.)
Paris-Est	pari{.}Est
les	le(.C) lez(V)
mais	mE me mEz(V)
dix-neuf	diznXf diznXv(V)

Figure 2: Some example lexical entries. Phones in { } are optional and phones in ( ) specify the context in which the pronunciation can occur, where V is a vowel, C a consonant, and "." silence.

gender-dependent acoustic models. The implementation used for the experiments reported in this gender-independent models and carries out only a single decoding pass with a bigram class language model.

The recognition lexicon is represented phonemically using a set of 37 units, including silence, breath noise and filler word specific phonemes. The lexicon contains multiple names for some cities, which are generated by rule: *Bordeaux-Saint-Jean / Bordeaux*, *Paris-Gare-de-Lyon / Gare-de-Lyon;* and for different syntactic forms: *Le-Mans*, *au-Mans*. Each lexical entry is described phonemically, using a set of 37 symbols (including phones corresponding to silence, breath noise and hesitations). Frequent pronunciation variants are explicitly described, and multiple pronunciations to handle the liaison phenomenon. The lexicon also contains frequent pseudo and filler words (*bon*, *ben*) and hesitations (*euh, ah*). Some example lexical entries are shown in Figure 2.

In order to reduce the number of understanding errors due to speech recognition, a confidence score is associated with each hypothesized word (Wessel, Macherey and Schlüter 1998; Bouwman, Sturm and Boves, 1999). If the score is below an empirically determined threshold, the hypothesized word is marked as uncertain. The natural language understanding component can choose to ignore these uncertain words or the dialog manager can decide to start a clarification subdialog. On average, rejection tends to lead to a longer dialog, since some correct words are ignored, however the overall dialog success rate is improved. An optional spell mode is authorized for station names, to improve recognition performance with a large number of cities since this is critical for the task. In our current implementation the output of the speech recognizer is the best word sequence with a confidence score, however, the recognizer is also able to provide a word lattice. The confidence score estimator is based on a logistic regression with an intercept term and 4 predictors: 1) the word duration; 2) the language model score (the average LM log-likelihoods of the current and following words); 3) the word acoustic score minus the best possible acoustic score over all the HMM states; and 4) the average number of active HMM states over the duration of the word.

The text string output by the recognizer is passed to the natural language understanding component. The first step is to preprocess the string in order to normalize the lexical forms and to use local syntax rules to identify and label some unambiguous concepts. Three types of rewrite are used: general processing rules which remove filler words, convert numbers to digits, reduce inflected verb forms, and gender and plural forms; task-dependent processing rules, which group synonyms and application specific words into classes; and language-dependent processing rules to process expressions for negation, negotiation and politeness.

Speech understanding consists of literal understanding followed by a reinterpretation in the context of the ongoing dialog. In literal understanding, the semantic analyzer carries out a caseframe analysis (Bruce 1975, Fillmore 1968) to determine the meaning of the query, and builds an appropriate semantic frame representation. Compared with (Bennacef et al, 1994) the caseframe grammar has been simplified and rendered language-independent. The slot-value pairs are ordered according to their order in the query, and it is possible to have several slots of the same type. Keywords are used to select an appropriate case structure for the sentence without attempting to carry out a complete syntactic analysis. The concepts needed to carry out the main ARISE ticketing task concern train times, connections, fares and reservations (including reductions and other constraints). Other concepts are used to handle general information available about reductions and services. The concepts have been determined by analysis of queries in the training corpora. The literal understanding module takes the minimum of decisions so as to avoid misinterpretations in the case of uncertainity. For example, if the sentence contains several cities (this happens frequently), it is better to provide a semantic frame with multiple city slots. The ambiguity must then be resolved by the dialog manager taking into account the dialog context.

Contextual understanding consists of interpreting the utterance in the context of the ongoing dialog, taking into account common sense and task domain knowledge. The semantic frames resulting from literal understanding are reinterpreted using default value rules. At the same time qualitative values are transformed into quantitative ones. Semantic frames corresponding to the current utter-

Q1) *Je veux aller demain matin de Paris à Marseille.* (I would like to go from Paris to Marseille tomorrow morning)	
`<train-time>`	*After interpretive rules*
`from: paris`	
`to: marseille`	
`relative-day: demain` *(tomorrow)*	Oct 1, 1998
`morning-afternoon: matin` *(morning)*	07:00-12:00

Figure 3: Example query and result of semantic analysis after literal and contextual understanding

ance are then completed using the *dialog history* in order to take into account all the information previously given by the user, as well as the questions posed by the system.

The dialog manager then either prompts the user to fill in missing information or uses the semantic frame to generate an SQL-like request to the database management system. The caller is required to supply four key items: the departure and arrival stations, the date and approximate time of travel. The day and time can be specified exactly (March 14th) or in a relative manner, such as *next Monday, early morning, late tomorrow afternoon*. Interpretative and history management rules are applied prior to generation of the DBMS request. These rules are used to determine if the query contains new information, and if so, if this information is contradictory with what the system has previously understood. If a contradiction is detected, the dialog manager may choose to keep the original information, replace it with the new information, or enter into a confirmation or clarification subdialog.

The database retrieval component uses a copy of the static SNCF train information (RIHO database). Post-processing rules are used to interpret the returned information prior to presentation to the user, taking into account the dialog history and the content of the most recent query.

The dialog manager generates a dialog act in the form of a frame. The natural language generation component converts this generation semantic frame into a natural language response, which is played to the user. The form of the natural language response depends on the dialog context, and whether or not the same information was already presented to the user. Our aim is to give a direct response to the caller, highlighting new information. Careful attention has been paid to construction of sentences that contain the appropriate information and the generation of natural-sounding utterances (Bennacef et al, 1996). Messages are synthesized by concatenation of variable-sized speech units stored in the form of a dictionary (Lamel et al, 1993). The resulting synthetic speech is rated as very natural by users.

A barge-in capability is often considered to be very important for usability. The ability to interrupt the system required modifications to several modules. Primarily, recording and speech recognition must be active at all times, especially when the system is synthesizing a response. Software-based echo cancellation, applied to the recorded signal is used to remove the echo of the known synthetic speech in order to be able to detect if the caller is speaking. When speech is detected, synthesis is stopped. In order to be able to assess the use of this capability, the time of the partially played response and location in the text are noted. Although not yet taken into account, the dialog manager could make use of this partial response (i.e. knowing what was actually heard by the caller). There are also dialog situations in which barge-in is disabled to ensure that the caller hears the entire message, such as the introductory message and during a system directed dialog. A new feature in the system is the incorporation of a time-out to reprompt the caller if no speech has been detected during a predetermined period. In the current system the time-out is 5s.

3 Dialog Strategy and Management

The dialog management aspect of the system has become more and more important as we have gained experience with spoken language dialog systems. The dialog manager is the central controller of the entire system as it manages contextual understanding, the dialog history, information retrieval and response generation, and ensures a smooth interaction between the user and the computer. The information retrieval dialog is divided into three phases: main information exchanges preceded and closed by formalities (Bennacef, Neel and Bonneau-Maynard, 1995; Bennacef et al, 1996). The dialog manager maintains the dialog history which is used to complete missing information in the semantic frame. The dialog context is used to provide default values for required slots and to define the *dialog-state* which is used to determine what dialog strategy to apply. By dialog strategy, we refer to the decisions taken by the dialog manager at a given point in the dialog, which are determined by a consideration of the task functionality and ergonomic design.

In order to be closer to a real dialog situation, representatives from LIMSI and VECSYS visited the Douai SNCF Information Service to observe how the human-human telephone dialogs are performed and what strategies are used by the human operators. The following main objectives were the basis for the new dialog strategy:

- *To never let the user get lost.* This is a very important rule in dialog systems. The user must always be informed of what the system has understood. Considering that most users will be unfamiliar with talking to a machine, it is important that s/he is kept aware of what the system has understood. For example, when the system has recognized a new item, implicit confirmation is used by formulating a message that starts with the new item.

  ***Le matin**, vous avez un train... (**In the morning**, there is a train ...)*

  This allows the caller to rapidly check and correct the system if necessary.

- *To use direct responses to user questions.* After analyzing dialogs with our system it was apparent that the responses of the system needed to be more accurate and should answer directly to the user's request. For example, if a caller asks for the arrival time of the next train, only this information is returned, so the user does not have to search for the desired information among other information.

  ***Le train suivant arrive** à Brest à 15h25. (**The next train arrives** in Brest at 15:25.)*

- *To explain unexpected system responses.* Sometimes the system cannot find a solution satisfying the user's request, and needs to relax the specified constraints. In this case the provided solution may surprise the caller if an appropriate explanation is not provided. For example, in order to arrive in Marseille in the morning, a traveler must leave Paris the evening before.

- *To give to the user the opportunity, at each step, to correct the system.* This capability is needed to be able to correct for recognition errors, but also the user may correct what s/he said or change their mind. Even though users are able to correct the system at any moment, we have observed that they do not always do so (see the next point).

- *To avoid misunderstanding.* Since our experience suggests that we cannot expect users to correct the system, it is important to minimize recognition errors. This was the motivation for rejecting unreliable hypothesized words.

  In case of misrecognition or misunderstanding, the system lets the caller know what's happening by generating the following message before asking again for a city or a date.

Je suis désolée, je n'arrive pas à comprendre. Veuillez-indiquer votre ville de départ. (I'm sorry, I don't understand. Please give your departure city.)

Although not an objective, the SNCF indicated that the overall interaction was too long, and requested that the dialog be shortened. As a result, when possible short system responses are provided, so as to keep the caller in tighter contact with the system, and to make for a more natural dialog.

In addition to these objectives, our objective is to develop a dialog system that is easy to use, flexible, allowing the user as much liberty as possible. This implies that a mixed-initiative dialog strategy is desired, with the system only taking the initiative to guide the user or to clarify an error or resolve a conflict. The dialog strategy thus also supports navigation and negotiation. The caller can switch at will among the system functionalities (schedule information, fares, reductions, reservations, services) with no specific navigation commands.

The system supports negotiation for the primary information task as well as for other tasks, such as requests for information about fares and reductions, etc. Negotiation by our definition is when the caller refuses a solution proposed by the system, and can modify, enlarge or refine the constraints. For example, the system may propose an itinerary based on the time requested by the caller that requires changing trains. The caller can specify that s/he wants a direct train, and the system will propose only direct trains (if there are any) or respond that there are none.

Since negotiation is available, the system proposes only one train,[1] the one corresponding most closely to the caller's request. If a time range (early morning, 2-4 pm) has been given, the train closest to the middle of the specified range is proposed. The user is given the opportunity to ask for another train (the preceding/following one, an earlier/later train, the first/last train).

Constraint relaxation is used in retrieving timetable information in order to provide a more co-operative dialogue and response. For example, the system relaxes constraints on the departure time when no train corresponds to the user's request, allowing the system to return the closest train after or before the specified time. If the user has specified a time range (e.g., early morning) only one train is proposed to the user, that closest to the middle of the specified time. In this case an explanation is provided for why the system response does not correspond exactly to what had been asked for. Some examples are:

[1] In the RAILTEL system, information was given for up to 3 trains.

*Le train **le plus proche** correspondant à votre demande part à... (The **closest** train to your request leaves at ...)*

Il n'y a pas de TGV correspondant à votre demande. (There is no TGV corresponding to your request.)

Le premier train part à 5h30. (The first train leaves at 5:30.) (If the user asked for a train before 5 o'clock.)

Depending upon the dialog state, different confirmation strategies can be invoked. A study of user behavior with different types of confirmation strategies is described in (Lavelle, de Calmès and Pérennou, 1998). In our system, by default implicit confirmation is used. Implicit confirmation plays an important role in keeping the user aware of what the system has understood. This is handled by including the new information (supplied by the caller) in the system response. This direct feedback is illustrated by the dialog in Figure 4. When explicit confirmation is required, the system prompts the caller with a yes/no question. The caller is allowed to accept or reject a system suggestion or to modify it. For example, expressions such as *je ne veux pas partir de Lille (no, I don't want to leave from Lille),* or *je ne pars pas à 8 heures mais à 20 heures (I'm not leaving at 8 o'clock but at 20 o'clock)* are handled by the system.

A two-level dialog strategy has been implemented, in which a mixed-initiative dialog where the user can provide any information at any point in time is combined with a system-directed dialog in case a problem is detected in obtaining departure and arrival station names and the date of travel. When the second level, or constrained dialog is active, the speech recognizer makes use of a dialog-state dependent language model and more accurate acoustic models. (These models are not used elsewhere for decoding speed reasons.) In this mode, barge-in is disabled. An example of a dialog which is largely user-initiated is shown in Figure 5. The final part of the dialog (concerning the reservation) is guided by the system.

A constrained dialog can be initiated by the system if the user does not respond to the system prompt for one of the four basic items (departure city, arrival city, date, time), or in cases where the information received by the system is contradictory with what was previously understood. Such constrained dialogs apply only to the departure and arrival cities, and the travel date. For example, if the system understands a change in the departure or arrival city, one of the following strategies is

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: Je voudrais un train pour Paris. (I want a train to Paris.)

S: *Alors, **vous allez à Paris**, de quelle ville partez-vous? (So, **you want to go to Paris**, where are you leaving from?)*

U: de Bordeaux (from Bordeaux)

S: *Alors, **de Bordeaux-St-Jean** à Paris, quel jour partez vous ? (**From Bordeaux-St-Jean** to Paris, what day are you traveling?)*

U: demain matin (tomorrow morning)

S: ***Le matin**, dimanche 21 septembre, vous avez un train direct de Bordeaux-St-Jean à Paris à 10 heures 22 arrivant à 13 heures 40. Qu'en pensez-vous? (**In the morning, Sunday, September 21st**, there is a direct train from Bordeaux-St-Jean to Paris at 10:22 arriving at 13:40. Is this ok?)*

U: Et le train suivant? (and the next train?)

S: ***Le train suivant** part à 10 heures 42 et arrive à 13 heures 45. (**The next train** leaves 10:42 and arrives at 13:45.)*

U: Quel est le prix du billet en première classe? (what is the price of a first class ticket?)

S: ***Le prix du billet** Bordeaux-St-Jean Paris est de 549 francs **en première classe.** (**The price of a first class ticket** from Bordeaux-St-Jean to Paris is 549 francs.)*

U: D'accord, je réserve sur ce train. (Ok, I'd like to reserve on this train.)

S: *Voulez-vous une place fumeur ou non fumeur? (Smoking or non-smoking?)*

U: non fumeur s'il vous plaît (non smoking, please)

S: *Vous avez réservé en première classe dans le train direct de Bordeaux-St-Jean à Paris partant à 10 heures 42, arrivant à 13 heures 45 le dimanche 21 septembre. Au revoir, la SNCF vous souhaite un bon voyage. (You have a reservation in first class on the 10:42 train direct from Bordeaux-St-Jean to Paris, arriving at 13:45, Sunday Septembre 21st. Goodbye, the SNCF wishes you a good trip.)*

Figure 4: Example dialog illustrating direct feedback of what has been understood by the system. The relevant part of the system response is shown in bold.

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: Je voudrais l'horaire d'un train direct de Paris à Grenoble vendredi prochain dans la matiné (I'd like to know the time of a direct train from Paris to Grenoble next Friday morning)

S: *Veuillez patienter... (Please wait ....)*

S: Le matin, vendredi 26 septembre vous avez un train direct de Paris-Gare-de-Lyon à Grenoble à 9h48 arrivant à 12h54. Qu'en pensez-vous? Cet horaire vous convient-il? *(In the morning, on September 26th, there is a direct train from Paris to Grenoble at 9:48 am arriving at 12:54 pm. Is this ok?)*

U: et un peu plus tard? (and a little later?)

S: *Le train suivant part à 14h36 et arrive à 17h41. Qu'en pensez-vous? Cet horaire vous convient-il? (The next train leaves Paris at 14:36 and arrives at 17:41. Is it ok?)*

U: Est-ce qu'il y a un bar dans ce train? (Is there a snack bar on this train?)

S: *Oui, vous avez un bar dans ce train. Souhaitez-vous une autre information? (Yes, there is a snack bar on this train. Would you like any other information?)*

U: Bon, je réserve. (Good. I'd like to reserve.)

S: *En quelle classe voulez-vous voyager ?(What class are you traveling in?)*

U: en seconde classe (in second class)

S: *Zone fumeur ou non fumeur ? (Smoking or non-smoking?)*

U: fumeur, si possible (smoking, if possible)

S: *Vous avez réservé en seconde classe, en zone fumeur, dans le train direct partant de Paris à 14 heures 36, arrivant à Grenoble à 17 heures 41 le vendredi 26 septembre. (You have a reservation in second class, smoking zone, Friday September 26, on the train leaving Paris at 14:36, arriving in Grenoble at 17:41.)*

Figure 5: Example dialog illustrating highly user-initiated dialog.

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: des horaires de train (timetable information)

S: *Oui, de quelle ville partez-vous? (Ok, where are you leaving from?)*

U: Paris

S: *Bon, vous partez de Paris, quelle est votre destination? (You are leaving from Paris, what is your destination?)*

U: non (recognition error: user said Dinan)

S: *Je suis d'esolée, je n'arrive pas à comprendre, indiquez votre destination, par exemple Lille. (I'm*
n
sorry, I did not understand. Indicate your destination, for example Lille)

U: Di- (user hesitation results in word fragment)

Constrained dialog

S: *Donnez votre ville d'arrivée et épelez la si vous le souhaitez. Par exemple Lille L I 2L E. (Give your arrival city and spell it if you want, for example Lille L I L L E)*

U: Dinan D I N A N

S: *Alors, de Paris à Dinan quel jour partez vous? (From Paris to Dinan, what day are you leaving?)*

...

Figure 6: Excerpt illustrating a constrained dialog for the arrival city.

used depending upon the state of the dialog: the system may choose to ignore the information; it can ask for an explicit confirmation of the new city; or it can ask the user to repeat the information. If the caller changes one of these items during the confirmation request, implicit confirmation is used in the following system prompt. The directiveness of the prompt increases if the user does not supply the requested information. An example of a constrainted dialog for the arrival city name is given in Figure 6. In this example, when the dialog manager did not understand the destination, the caller was first provided with an example, and then given the option of saying and spelling the city name.

The results of an informal assessment of the modified dialog strategy indicate that the overall dialog time is reduced, and the caller assessment (via questionnaires) is positive.

Closely related to the changes in dialog managment, the generation component has been modified to produce texts which directly integrate information from the user's request. These responses aim to provide only the information that is new and/or of high relevance. In previous versions, the responses, while assessed favorably by subjects, tended to recapitulate too much information and were thus too long and repetitive. The synthesis dictionary unit was extended to the larger vocabulary and completed with additional formulations. New recordings were made in accordance with the changes to the generation component. Questionnaires completed by callers indicate that the new generation strategy is well accepted.

4 User Trials to Assess Progress and Usability

Evaluation plays an integral role in system development (Madcow 1992; Eagles 1998), and can be considered as an ongoing activity. Different types of evaluation can be used, each with their particular strengths and costs. It is advantagous when the evaluation can be carried out automatically, although this still requires labeling of the test data. Such types of evaluation are usually applied to individual system components, particularly the speech recognizer and the semantic understanding component. A multilevel error analysis can be used to distinguish between errors due to a particular component and those propagating from preceding stages (Lamel et al, 1997). This error analysis requires human intervention. However, when experimenting with new user interfaces and dialog strategies, it may be impossible to reuse previously recorded data. Therefore, it is often useful to carry out an informal assessment of system performance and capabilities and how these are perceived by users.

ARISE	*#Calls*	*#Queries*	*#Words*	*#Distinct*
Aug97	2787	36.4k	179.7k	2529
Dec97	6130	84.5k	412.3k	3677
Mar98	6545	90.4k	437.9k	3764
Oct98	9789	144.9k	654.1k	4512

Table 1: Summary of data collection for ARISE. Word fragments are not counted.

An important concern is obtaining realistic user trials (RailTel 1995a, 1995b; Blasband .. 1998; Baggia, Castagneri, and Danieli, 1998; Sanderman et al, 1998). These are obviously needed to properly evaluate the prototype or potential service, but can be risky if done too prematurely. Evidently the user trials to be as realistic as possible, however, particularly in the early stages of system development it may not be possible to have access to the final user. As a consequence, we recruit subjects on ongoing basis to provide data for system development and evaluation.

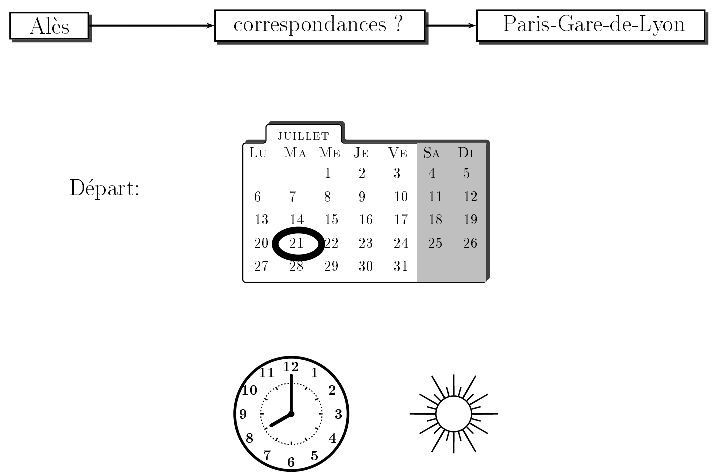
In addition to these user trials, the prototype system underwent periodic evaluations carried by the SNCF (a partner in the ARISE project) in order to assess usability and performance (Gitton et al, 1998). LIMSI subjects are recruited via advertisements in local newspapers. Each subject receives a short instruction set (1 page), and calls the system 10 times. In 4 of the calls the user is given a predefined scenario[2] to solve, and in the remaining 6 calls the subject is asked to obtain whatever information s/he would like to have. Some example scenarios are given in Figure 7. Subjects are also given a randomly selected subset of the stations known to the system. At the start of each call, the subject enters a DTMF code to allow for easy tracking and processing. After each call, the caller is required to complete an evaluation questionnaire and return it LIMSI. In addition to a qualitative assement, the caller provides the information obtained from the system, and is free to give any additional comments. The call questionnaire is shown in Figure 8. In 1998 we recorded over 3700 dialogs from 580 subjects. Table 1 summarizes the cummulative characteristics of the data collected with different versions of the ARISE system. The number of distinct words found in the corpora is relatively small compared to the total number of words.

[2] Different types of formats have been explored and are used to define the scenarios. These range from short descriptions, situations, and both graphic and textual presentations.

1- Situation

Le 30 mai 1998 vous devez vous rendre de Paris à Ancenis pour une rencontre sportive. Vous devez absolument arriver avant 8h30 afin de vous préparer pour l'entrainement. Renseignez-vous sur le prix de votre billet et les réductions possibles. (You need to go from Paris to Ancenis for a sporting event on May 30, 1998. You need to arrive before 8:30 am to warmup. Find out the price of your ticket and about any possible reductions.)

2- Pictural Scenario



Renseignez-vous sur les prestations et choisissez un train avec bar.
(Ask about onboard services and choose a train with a snack bar.)

3- Open Scenario

Le système ARISE dispose d'informations sur : (the Arise system can provide information about)

- les horaires (timetables)
- les tarifs (prices)
- les réductions (discounts)
- les prestations à bord des trains (onboard services): wagon-lit, train couchette, bar, restaurant

Le système ARISE permet de : (the Arise system enables you to)

- demander des informations (ask for information)
- faire des réservations (make a reservation)

Le système ARISE connaît les trajets entre 400 villes. (Arise knows about connections between 400 cities).

Figure 7: Example scenario types used in data collection.

QUESTIONNAIRE

Date:

Indicate the information requested:

Indicate the information obtained:

Respond yes or no to the following questions.

Question	**YES**	**NO**
ARISE is easy to use.		
I got the information I wanted.		
The systeme is fast.		
I got the impression the system understood me.		
I understood what the system said.		
I found the responses too long.		

Comments on this call:

Figure 8: Caller questionnaire used for data collection. (Translated from French)

The periodic tests carried out by the SNCF (with the exception of the final assessment) involved subjects recruited by a hostess at a Parisian train station. The subjects were asked to test a new, but experimental service, and were given a gift certificate for their participation. Each subject called the system three times, carrying out an open scenario that s/he wrote down prior to each call. Subjects completed a short questionnaire after each call and a longer one after the final call. Despite the differences in recruitment, the general characteristics of these calls (in terms of dialog success, overall call duration, the number of exchanges, vocabulary, types of requests and typical problems) are essentially the same as those made by the subjects we recruit.

The results reported here concern a total of 271 dialogs, 108 calls recorded in the spring of 1998, and 163 calls recorded in November 1998. A set of 50 dialogs recorded during a two-day period (May 27-28) were used to measure performance just prior to the last SNCF user tests held in June'98.[3] There were an average of 19 exchanges per call, with an average total duration of 229 seconds. The dialog error rate in obtaining timetable information was 20% . A dialog error is considered to have occurred if the system returned the wrong information given the constraints supplied by the caller. In 26% of the dialogs some difficulty was experienced in obtaining supplementary information.

On a set of 58 calls from the June'98 SNCF user trials[4] the dialog error in obtaining timetable information was 16%. The reservation failure rate was 11% (reservations require specifying the class of travel, seating preference and reduction). A higher error rate (30%) was obtained for diverse questions, due in part to functionality limitations. 12% of the dialogs ended without a closing formality (ie. the caller hung up) without either the system or the caller saying goodbye. The internal test set of 50 calls are seen to have been a relatively accurate indicator of performance during the SNCF trials.

These results are substantially better than the results of the November'97 SNCF evaluation before most of the modifications presented in this paper were implemented. On 80 calls the timetable information failure rate was 47.5% and the reservation failure rate was 35.7%. In November'97, 52% of the calls were terminated without a closing formality. The June'98 calls are longer, averaging 15 exchanges (167 seconds), compared to 10 exchanges in November (114 seconds). Although more performant, the two-level dialog has increased the length of the dialog. The assessment by SNCF

[3] The 50 calls consist of all the calls recorded during this two-day period, removing only calls with no queries.

[4] The calls from June 3rd are not used here, due to experimental problems, such as the subject speaking with the experimentor, or interference due to simultaneous recording. Results are given for calls from June 4 and 5, 1998.

subjects has improved from a satisfaction level of 5.9 to 12.7 (out of 20).

We have carried out an analysis of the use of the barge-in capability on the 58 calls of the June'98 SNCF test set. The callers were aware that they could interrupt the system if they so desired. Users interrupted the system in 72% (42) of the calls, speaking during 122 of 958 system responses (13%). When barge-in was observed during a call, it was used to interrupt, on average, 3 responses. Barge-in was observed in a variety of contexts, most frequently used to respond to questions before they were finished. For example, if the system is uncertain about a station name, the caller is prompted to say and optionally spell the city name. Almost 40% of the interruptions occurred on this type of prompt. In almost 25% of the instances, the system responded after a long pause (the endpoint was assumed) or after a time-out, but the subject seemed to be engrossed in their thoughts, and unaware of what the system was saying. In contrast to our expectations, barge-in was only rarely used (6% of the 122 cases) to correct the system, and usually to change the date of travel.

In November'98 the SNCF carried out the final performance assessment within the context of the ARISE project (Gitton et al, 1998). Subjects were recruited via a marketing survey company. The overall dialog error rate in obtaining timetable information on the 163 calls was 21.5%. If calls where the user made no attempt to correct the system when the system gave incorrect information are eliminated, the dialog error rate on the remaining 151 calls was 15.3%.

5 Discussion and Conclusions

Enabling efficient, yet user-friendly interaction for access to stored information by telephone is quite difficult. Most existing services are quite directive, restricting the caller to use limited formulations and to provide only the requested information. Some laboratory prototypes allow a more open, user-initiated dialog, but performance is generally lower than what can be obtained with more restricted dialog stuctures.

Our goal is to obtain high dialog success rates with a very open dialog, where the user is free to ask any question or to provide any information at any point in time. The dialog management aspect of the system has become more important as we have gained experience with spoken language dialog systems. We have adpoted a two-level dialog strategy which combines a mixed-initiative dialog with a system-directed dialog in case a problem is detected in obtaining departure and arrival station names or the travel date. A system directed dialog can also be invoked if the system recognizes information which is contradictory to earlier information. We make use of implicit confirmation using the callers wording or a rephrased version (when possible), and explicit confirmation in the context of constrained subdialogs. We have also incorporated confidence measures in a simplistic manner where uncertain words can be ignored in later processing steps.

Our system supports negotiation for the primary information task as well as for other tasks, such as requests for information about fares and reductions, etc. Negotiation by our definition is when the caller refuses a solution proposed by the system, and can modify, enlarge or refine the constraints. For example, the system may propose an itinerary based on the time requested by the caller that requires changing trains. The user can then specify that s/he wants a direct train, and the system will propose only direct trains (if there are any) or respond that there are none. The user is free to move from one subtask to another. For example, while making a reservation the system asks the user if s/he is eligible for a reduced fare. At this point the user may ask the system what types of reductions are possible.

In the ARISE project our work was organized into three main development cycles, where at the end of each development cycle an evaluation of the current system was carried out by the SNCF in close coordination with us. The test conditions were defined conjointly and the subjects selected by the SNCF. Both LIMSI and the SNCF analyzed the calls on an objective basis. The SNCF also analyzed the subjective assessment of the system as expressed by the subjects' in response to a detailed questionnaire. User trials carried out in June and November'98 had an overall timetable information dialog error rate of about 15% compared with almost 50% in November'97.

When interpreting the results of the usability trials it is important to remember that recruited subjects are more or less representative of the general population. We noticed that some callers did not attempt to correct the system when wrong information was returned. This is in accordance with our preliminary observations on the use of barge-in. Barge-in which was not frequently used, perhaps due to a lack of familiarity with the system, and was also used in the manner we had anticipated (to correct misrecognized items). This may be partially due to the experimental conditions, as callers do not really need the information they are asking for, and therefore may not notice (or care about) the errors. We expect that real users will at least attempt to correct the system.

Another important issue that was highlighted during the SNCF user trials is that users do not distinguish the functionalities of the service from the system responses. Although the system was able to detect some out-of-functionality requests, and responded that it was unable to handle these, such responses are not satisfactory for users. For example, a caller that wanted to reserve 3 seats was informed that he could only reserve one seat at a time. From the spoken language system developer's point of view, this response is correct, although the user may not agree. In such cases we may have a successful dialog, but an unhappy caller. It is therefore crucial that representative users be involved early on in system development to ensure that the resulting service fulfils their needs.

ACKNOWLEDGMENT

We would like to thank the SNCF for providing the information database RIHO for use in the ARISE project and for their contribution to assessment by carrying out the user trials.

References

[1] Madcow (1992), "Multi-site Data Collection for a Spoken Language Corpus," *Proc. DARPA Speech and Natural Language Workshop*, Harriman, NY, pp. 7-14, February.

[2] RAILTEL (1995a), "Definition of the evaluation methodology for the Field Trials," RAILTEL/MAIS *Project deliverable D4*, Saritel, June.

[3] RAILTEL (1995b), "Results of Field Trials," RAILTEL *Project deliverable D8*, November.

[4] EAGLES (1998), *Handbook of Standards and Resources for Spoken Language Systems, Vol 3: Spoken Language System Assessment* Editor: Mouton de Gruyte.

[5] G. Adda, M. Adda-Decker, J.L. Gauvain, L. Lamel (1997), "Le système de dicté vocale du LIMSI pour l'évaluation AUPELF'97," *Proc. Journées Scientifiques et Techniques du Réseau Francophone d'Ingénierie de la Langue de l'AUPELF-UREF*, Avignon, France, pp. 35-40, April.

[6] P. Baggia, G. Castagneri, M. Danieli (1998), "Field Trials of the Italian ARISE Train Timetable System," *Proc. IEEE IVTTA'98*, Torino, Italy, pp. 97-102, September.

[7] S.K. Bennacef, H. Bonneau-Maynard, J.L. Gauvain, L.F. Lamel, W. Minker (1994), " A Spoken Language System For Information Retrieval," *Proc. ICSLP'94*, Yokohama, Japan, **3**, pp. 1271-1274, September.

[8] S.K. Bennacef, L. Devillers, S. Rosset, L.F. Lamel (1996), "Dialog in the RAILTEL Telephone-Based System," *Proc. ICSLP'96*, Philadelphia, PA, pp. 550-553, October.

[9] S.K. Bennacef, F. Neel, H. Bonneau-Maynard (1995), "An Oral Dialogue Model based on Speech Acts Categorization," *Proc. ESCA Workshop on Spoken Dialog Systems*, Vigsø, Denmark, pp. 237-240, Spring.

[10] M. Blasband (1998), "Speech Recognition in Practice: The ARISE Project (Automatic Railway Information System for Europe)," La lettre de l'IA numéros 134-135-136, *Proc. NIMES'98*, pp. 207-210, Nîmes.

[11] G. Bouwman, J. Hulstijn (1998), "Dialog Strategy Redesign with Reliabilty Measures," *Proc. 1st Int. Conf. on Language Resources and Evaluation*, Granada, Spain, June, pp. 191-198.

[12] G. Bouwman, J. Sturm, L. Boves (1999), "Incorporating Confidence Measures in the Dutch Train Timetable Information System Developed in the Arise Project,", *Proc. IEEE ICASSP-99*, Phoenix, AZ, **1**, pp. 493-496, March.

[13] B. Bruce (1975), "Case Systems for Natural Language," *Artificial Intelligence*, **6** pp. 327-360.

[14] Ch.J. Fillmore (1968), "The case for case," in *Universals in Linguistic Theory*, Emmon Bach & Robert T. Harms (eds.), Holt, Rinehart and Winston, Inc.

[15] J.L. Gauvain, J.J. Gangolf, L. Lamel (1996), "Speech Recognition for an Information Kiosk," *Proc. ICSLP'96*, Philadelphia, PA, pp. 1672-1675, October.

[16] J.L. Gauvain, L. Lamel (1996), "Large Vocabulary Continuous Speech Recognition: from Laboratory Systems towards Real-World Applications," *Institute of Electronics, Information and Communication Engineers* , J79-D-II:2005–2021, December.

[17] J.L. Gauvain, L.F. Lamel, G. Adda, M. Adda-Decker (1994), "Speaker-Independent Continuous Speech Dictation," *Speech Communication*, **15**, pp. 21-37, September.

[18] S.M. Katz (1987), "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer," *IEEE Trans. ASSP*, **35**(3), pp. 400-401, March.

[19] L. Lamel, S.K. Bennacef, S. Rosset, L. Devillers, S. Foukia, J.J. Gangolf, J.L. Gauvain (1997), "The LIMSI RailTel System: Field trials of a Telephone Service for Rail Travel Information," *Speech Communication* **23**, pp. 67-82, October.

[20] L.F. Lamel, J.L. Gauvain, B. Prouts, C. Bouhier, R. Boesch (1993), "Generation and Synthesis of Broadcast Messages," *Proc. ESCA-NATO Workshop on Applications of Speech Technology*, Lautrach, Germany, pp. 207-210, September.

[21] L. Lamel, J.L. Gauvain, S.K. Bennacef, L. Devillers, S. Foukia, J.J. Gangolf, S. Rosset (1996), " Field Trials of a Telephone Service for Rail Travel Information," *Proc. IEEE IVTTA-96*, Basking Ridge, NJ, pp. 111-116, October.

[22] L. Lamel, S. Rosset, J.L. Gauvain, S.K. Bennacef, M. Garnier-Rizet, B. Prouts (1998), "The LIMSI ARISE System," *Proc. IEEE IVTTA'98*, Torino, Italy, pp. 209-214, September.

[23] L. Lamel, S. Rosset, J.L. Gauvain, S.K. Bennacef (1999), "The LIMSI Arise System for Train Travel Information" *Proc. IEEE ICASSP-99*, Phoenix, AZ, **1**, pp. 501-504, March.

[24] C.A. Lavelle, M. de Calmès, G. Pérennou (1998), "A Study of User's Behaviors in Different States of a Spontaneous Oral Dialog with an Automatic Inquiry System," *Proc. IEEE IVTTA'98*, Torino, Italy, pp. 118-123, September.

[25] C. Popovici, P. Baggia (1997), "Specialized Language Models using Dialogue Predictions", *Proc. IEEE ICASSP-97*, Munich, Germany, pp. 815-818, April.

[26] A. Gitton et al (1998), "Spoken Language Understanding and Dialog Aspects: Final evaluation," *Arise Deliverable D124*, December.

[27] A. Sanderman, J. Sturm, E. den Os, L. Boves, A. Cremers (1998), "Evaluation of the Dutch Train Timetable Information System Developed in the Arise Project," *Proc. IEEE IVTTA'98*, Torino, Italy, pp. 91-96, September.

[28] F. Wessel, K. Macherey, R. Schlüter (1998), "Using Word Probabilities as Confidence Measures," *Proc. IEEE ICASSP-98*, Seattle, May, pp. 225-228.

List of footnotes.

Footnote 0: This work was partially financed by the LE -3 project 4229 ARISE.

Footnote 1: In the RAILTEL system, information was given for up to 3 trains.

Footnote 2: The 50 calls consist of all the calls recorded during this two-day period, removing only calls with no queries.

Footnote 3: The calls from June 3rd are not used here, due to experimental problems, such as the subject speaking with the experimentor, or interference due to simultaneous recording. Results are given for calls from June 4 and 5, 1998.

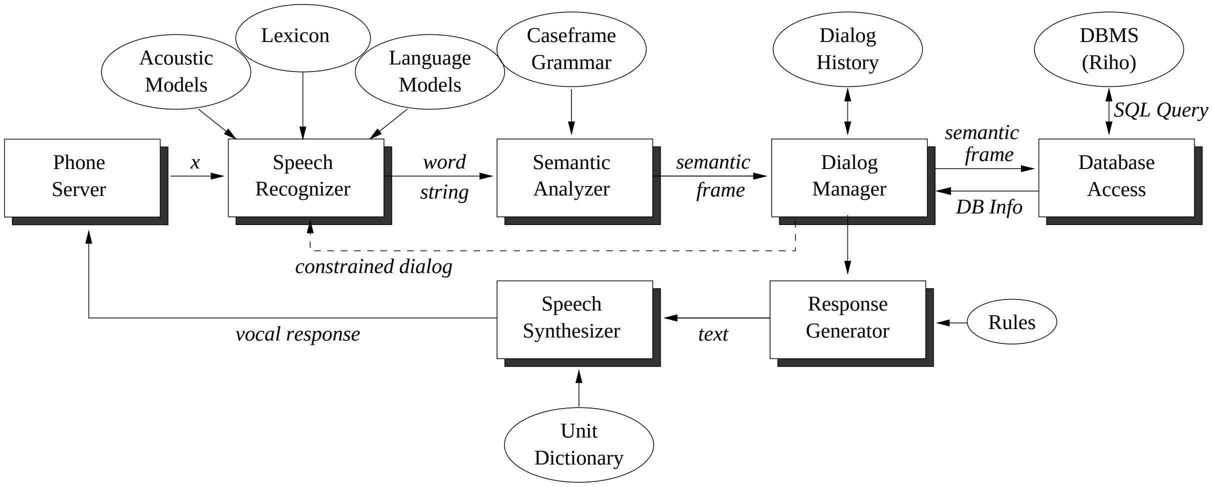


Figure 1. Overview of the ARISE system for spoken language information retrieval. **x** is the input speech signal.

Conches	kOS kOSx(.)
Paris-Est	pari{.}Est
les	le(.C) lez(V)
mais	mE me mEz(V)
dix-neuf	diznXf diznXv(V)

Figure 2. Some example lexical entries. Phones in { } are optional and phones in ( ) specify the context in which the pronunciation can occur, where V is a vowel, C a consonant, and "." silence.

Q1) *Je veux aller demain matin de Paris à Marseille.*

(I would like to go from Paris to Marseille tomorrow morning)

`<train-time>`	*After interpretive rules*
`from: paris`	
`to: marseille`	
`relative-day: demain` *(tomorrow)*	Oct 1, 1998
`morning-afternoon: matin` *(morning)*	07:00-12:00

Figure 3. Example query and result of semantic analysis after literal and contextual understanding

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: Je voudrais un train pour Paris. (I want a train to Paris.)

S: *Alors, **vous allez à Paris**, de quelle ville partez-vous? (So, **you want to go to Paris**, where are you leaving from?)*

U: de Bordeaux (from Bordeaux)

S: *Alors, **de Bordeaux-St-Jean** à Paris, quel jour partez vous ? (**From Bordeaux-St-Jean** to Paris, what day are you traveling?)*

U: demain matin (tomorrow morning)

S: ***Le matin**, dimanche 21 septembre, vous avez un train direct de Bordeaux-St-Jean à Paris à 10 heures 22 arrivant à 13 heures 40. Qu'en pensez-vous? (**In the morning, Sunday, September 21st**, there is a direct train from Bordeaux-St-Jean to Paris at 10:22 arriving at 13:40. Is this ok?)*

U: Et le train suivant? (and the next train?)

S: ***Le train suivant** part à 10 heures 42 et arrive à 13 heures 45. (**The next train** leaves 10:42 and arrives at 13:45.)*

U: Quel est le prix du billet en première classe? (what is the price of a first class ticket?)

S: ***Le prix du billet** Bordeaux-St-Jean Paris est de 549 francs **en première classe.** (**The price of a first class ticket** from Bordeaux-St-Jean to Paris is 549 francs.)*

U: D'accord, je réserve sur ce train. (Ok, I'd like to reserve on this train.)

S: *Voulez-vous une place fumeur ou non fumeur? (Smoking or non-smoking?)*

U: non fumeur s'il vous plaît (non smoking, please)

S: *Vous avez réservé en première classe dans le train direct de Bordeaux-St-Jean à Paris partant à 10 heures 42, arrivant à 13 heures 45 le dimanche 21 septembre. Au revoir, la SNCF vous souhaite un bon voyage. (You have a reservation in first class on the 10:42 train direct from Bordeaux-St-Jean to Paris, arriving at 13:45, Sunday Septembre 21st. Goodbye, the SNCF wishes you a good trip.)*

Figure 4. Example dialog illustrating direct feedback of what has been understood by the system. The relevant part of the system response is shown in bold.

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: Je voudrais l'horaire d'un train direct de Paris à Grenoble vendredi prochain dans la matiné (I'd like to know the time of a direct train from Paris to Grenoble next Friday morning)

S: *Veuillez patienter... (Please wait ....)*

S: Le matin, vendredi 26 septembre vous avez un train direct de Paris-Gare-de-Lyon à Grenoble à 9h48 arrivant à 12h54. Qu'en pensez-vous? Cet horaire vous convient-il? *(In the morning, on September 26th, there is a direct train from Paris to Grenoble at 9:48 am arriving at 12:54 pm. Is this ok?)*

U: et un peu plus tard? (and a little later?)

S: *Le train suivant part à 14h36 et arrive à 17h41. Qu'en pensez-vous? Cet horaire vous convient-il? (The next train leaves Paris at 14:36 and arrives at 17:41. Is it ok?)*

U: Est-ce qu'il y a un bar dans ce train? (Is there a snack bar on this train?)

S: *Oui, vous avez un bar dans ce train. Souhaitez-vous une autre information? (Yes, there is a snack bar on this train. Would you like any other information?)*

U: Bon, je réserve. (Good. I'd like to reserve.)

S: *En quelle classe voulez-vous voyager ?(What class are you traveling in?)*

U: en seconde classe (in second class)

S: *Zone fumeur ou non fumeur ? (Smoking or non-smoking?)*

U: fumeur, si possible (smoking, if possible)

S: *Vous avez réservé en seconde classe, en zone fumeur, dans le train direct partant de Paris à 14 heures 36, arrivant à Grenoble à 17 heures 41 le vendredi 26 septembre. (You have a reservation in second class, smoking zone, Friday September 26, on the train leaving Paris at 14:36, arriving in Grenoble at 17:41.)*

Figure 5. Example dialog illustrating highly user-initiated dialog.

S: *Quels renseignements désirez-vous? (What information do you want?)*

U: des horaires de train (timetable information)

S: *Oui, de quelle ville partez-vous? (Ok, where are you leaving from?)*

U: Paris

S: *Bon, vous partez de Paris, quelle est votre destination? (You are leaving from Paris, what is your destination?)*

U: non (recognition error: user said Dinan)

S: *Je suis d'esolée, je n'arrive pas à comprendre, indiquez votre destination, par exemple Lille. (I'm sorry, I did not understand. Indicate your destination, for example Lille)*

n

U: Di- (user hesitation results in word fragment)

Constrained dialog

S: *Donnez votre ville d'arrivée et épelez la si vous le souhaitez. Par exemple Lille L I 2L E. (Give your arrival city and spell it if you want, for example Lille L I L L E)*

U: Dinan D I N A N

S: *Alors, de Paris à Dinan quel jour partez vous? (From Paris to Dinan, what day are you leaving?)*

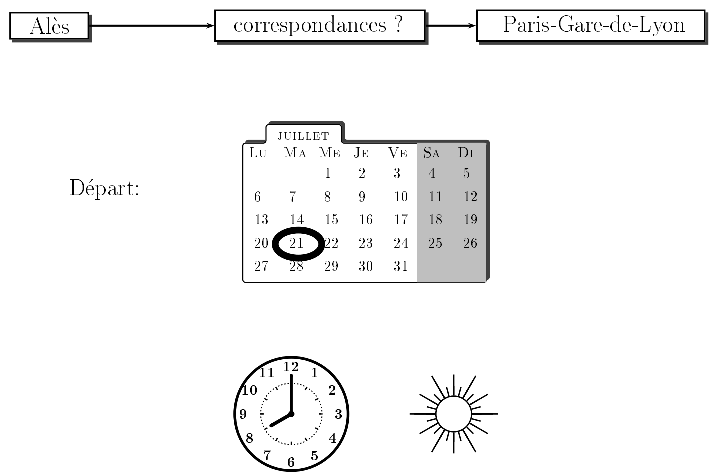
...

Figure 6. Excerpt illustrating a constrained dialog for the arrival city.

1- Situation

Le 30 mai 1998 vous devez vous rendre de Paris à Ancenis pour une rencontre sportive. Vous devez absolument arriver avant 8h30 afin de vous préparer pour l'entrainement. Renseignez-vous sur le prix de votre billet et les réductions possibles. (You need to go from Paris to Ancenis for a sporting event on May 30, 1998. You need to arrive before 8:30 am to warmup. Find out the price of your ticket and about any possible reductions.)

2- Pictural Scenario



Renseignez-vous sur les prestations et choisissez un train avec bar.
(Ask about onboard services and choose a train with a snack bar.)

3- Open Scenario

Le système ARISE dispose d'informations sur : (the Arise system can provide information about)

- les horaires (timetables)
- les tarifs (prices)
- les réductions (discounts)
- les prestations à bord des trains (onboard services): wagon-lit, train couchette, bar, restaurant

Le système ARISE permet de : (the Arise system enables you to)

- demander des informations (ask for information)
- faire des réservations (make a reservation)

Le système ARISE connaît les trajets entre 400 villes. (Arise knows about connections between 400 cities).

Figure 8. Example scenario types used in data collection.

QUESTIONNAIRE

Date:

Indicate the information requested:

Indicate the information obtained:

Respond yes or no to the following questions.

Question	**YES**	**NO**
ARISE is easy to use.		
I got the information I wanted.		
The systeme is fast.		
I got the impression the system understood me.		
I understood what the system said.		
I found the responses too long.		

Comments on this call:

Figure 8. Caller questionnaire used for data collection. (Translated from French)

ARISE	*#Calls*	*#Queries*	*#Words*	*#Distinct*
Aug97	2787	36.4k	179.7k	2529
Dec97	6130	84.5k	412.3k	3677
Mar98	6545	90.4k	437.9k	3764
Oct98	9789	144.9k	654.1k	4512

Table 1. Summary of data collection for ARISE. Word fragments are not counted.