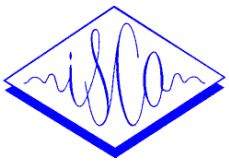

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

Static and Dynamic Predictions : A Method to Improve Speech Understanding in Cooperative Dialogues

François Andry

CAP GEMINI INNOVATION
118 rue de Tocqueville
75017 Paris - France

LIPN - Université Paris Nord
Avenue J.B. Clément
93430 Villetaneuse - France

ABSTRACT

We describe a method based on two independent and complementary *predictions mechanisms*, which is used to improve the recognition and the understanding performances of the SUNDIAL[1] speech and dialogue system. We exploit information produced by the dialogue manager component of the SUNDIAL system which consists of intentional contents (list of dialogue acts) and propositional contents (task types associated to a list of domain related semantic types). The *static predictions* mechanism corresponds to the determination of characteristic states of the dialogue related to the resolution of the task. The dialogue state reference is then transmitted to the recognition module, which activates a specific sub-lexicon and word-pair grammar. The *dynamic predictions* mechanism is based on two trials (semantic and dialogic), which are applied to the candidate strings produced by the parser. Tests show a significant definite improvement of the sentence understanding rate of the dialogue system with both kinds of mechanism.

I. INTRODUCTION

A way to improve the performance of speech understanding in dialogue systems is to predict the content of the next utterance of the user from the dialogue context. These predictions are usually restricted to cooperative task driven dialogues, within a limited domain, in which the user takes only few initiatives. High level knowledge provided by the dialogue level are used to constrain the recognition or the parsing phases [7] [8] [11] [12] [13].

In the SUNDIAL project, we deal with cooperative dialogues, using flight reservation as the application domain for the French prototype.

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

Let consider the dialogue context corresponding to the production of following utterance by the system (S) :

S : *April the 1st, at what time do you want to leave ?*

There is a variety of possible answers depending on the intention of the user and the content of the proposition associated with this intention. For example :

U_1 : *at 5:30.*
U_2 : *at 5:30 from Toulouse.*
U_3 : *Not the 1st, but the 3rd.*
U_4 : *Could you repeat the question please ?*

In U_1 and U_2 the user intention is to inform the system about the time and/or the departure place. In U_3 this is a correction on the date, and in U_4 the user asks for a repetition of the question.

The SUNDIAL Dialogue Manager (DM) has been developed in common by all partners of the project [10]. It is able to manage the dialogue steps [5], the task resolution, a linguistic history, and to modelise the belief states of both the system and the user [6]. After each production of an utterance of the system, the DM produces a set of predictions associated to possible moves of the user. This set is sent to the lower components of the SUNDIAL system (recognition and linguistic parser). Each set contains a list a possible dialogue acts (posi : positive answer to a Y/N question, nega : negative answer, inform : informative answer to an WH question, correction, echo ...) spanning on a specific propositional content (a task attribute and its value) if present.

From the intentional content and propositional content of the possible user utterances provided by the DM, we have designed two kinds of complementary mechanisms which are used to improve the recognition and the understanding rates of the SUNDIAL system.

II. PRINCIPLES

Using high level knowledge during the recognition level helps to reduce drastically the search space at later stages

of the speech understanding process. In SUNDIAL, the DM predictions are used at the recognition level to activate a dialogue contextual language model (word-pair grammar). This improves the quality of the word lattice produced by the recognition component [9]. Since the word-pair grammars related of the dialog context are built offline (from a corpus), this mechanism is called *static predictions mechanism.*

Although this first mechanism greatly improves the performances of the recognition phase, it still does not ensure the obtention of the best interpretation of the speech signal. The word-pair grammars overgenerate with regard to the grammar of the parser. Moreover, the discrimination of different contexts with short utterances is sometimes very poor since word-pair grammars are only local linguistic constraints. Therefore, we have designed a second mechanism which is applied at a latter stage in order to refine the results of the first one. This mechanism selects the most relevant utterance according the DM predictions among the set of best interpretations built by the linguistic parser. Since the linguistic constraints (semantic and dialogic) are built online from the DM predictions, this mechanism is called *dynamic predictions mechanism.*

III. STATIC PREDICTIONS

The principle of the static predictions mechanism is to spot a specific combinations of intentional and propositional contents inside the DM predictions in order to select a langage model (word-pair grammar) characteristic of the dialogue context. The main problem is to identify all possible combinations and to desing a method to build the word-pair grammars from all these combinations.

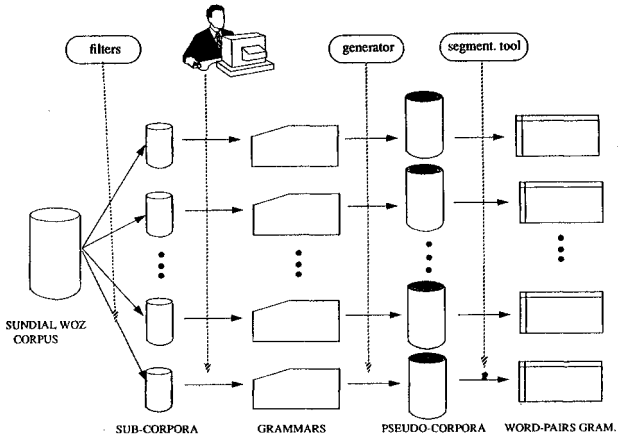

Fig.1 - Building-up word-pair grammars

Each word-pair grammar has been built offline (Fig.1) from a corpus obtained by a Wizard of OZ simulation [1]. From this corpus including more than 300 dialogues, we have automatically extracted a set a 16 different subcorpora, each of them corresponding to a different context usually related to the acquisition of a task parameter from the system (departure date, departure city, schedule, class, confirmation of the flight ...). Then, each sub-corpus has been encoded by hand in the form of regular grammars, homogenized and extended. A generator has been used to produce all possible utterances encoded by these grammars. Starting from a set of 3736 user utterances in the simulation corpus, the generator produces more than 370000 different user utterances. A segmentation tool is then used to build up the word-pair grammar. The task paramaters values such as places, dates, hours ... etc, which have been replaced in the grammar by generic symbols, are expanded at this step, multiplying the number of word-pairs. Measurements show that the word-pairs have a good coverage and are discriminant from one language model to an other one [3].

During the recognition phase, all word-pair subgrammars are loaded in memory. A system of bit masks is used to activate the sub-grammar corresponding to the dialogue context.

IV. DYNAMIC PREDICTIONS

The *dynamic predictions* mechanism is based on two trials (Fig.2) which are applied to the candidate strings produced by the acceptance phase of the parser (a later phase builds up the meaning of the best interpretation) [2]. The hypothesis kept corresponds to the string with the highest score from the trials. The acoustic score is taken into account when several strings present identical trial results.

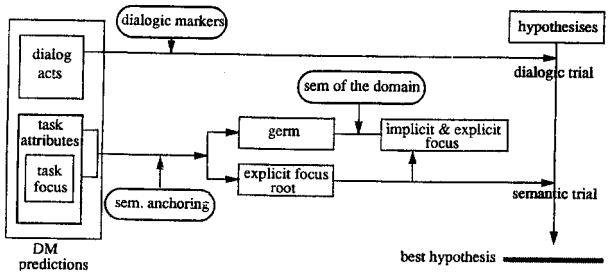

Fig.2 - Dynamic predictions mechanism in SUNDIAL

4.1 Semantic trial

The *semantic trial* is a matching between the semantic interpretation of the string, and the concepts that form the focus corresponding to the dialogue context.

The first step of the process determines dynamically the list corresponding to the focus. In order to be able to take into account shifts in the focus, we use the notion of *explicit* and *implicit* focus as in [4]. Elements of the explicit focus are concepts already present in the linguistic history of the speakers. It can be directly determined from the dialogue context provided by the DM. On the other hand, implicit elements of the focus can be obtained by propagation through the semantic network we used to represent the application domain knowledge.

From the propositional content provided by the DM, we first use inference rules which explain how a task type is expressed at the linguistic level. For example, the following rule explains that to express the time of departure (*sourcetime*) of a flight reservation, a user will employ an *hour_point* semantic type related to a departure event :

dbreserve:sourcetime===thedeparture:depart,hour_point

This operation called *semantic anchoring* provides two things : the *germ* from which all the elements of the focus will be determined, and the *root* of a subtree which is used to mark the element of the explicit focus.

Then, starting from the *germ*, we look for the concepts which could be the result of a focus shift. For this operation called extension, we use inheritance links and roles attached to the concepts of our semantic network. Pruning rules avoids both the extension of fine grain concepts and cycles during the process. The explicit focus elements are marked during the extension phase. Fig.3 shows an example of the extension process inside the semantic network.

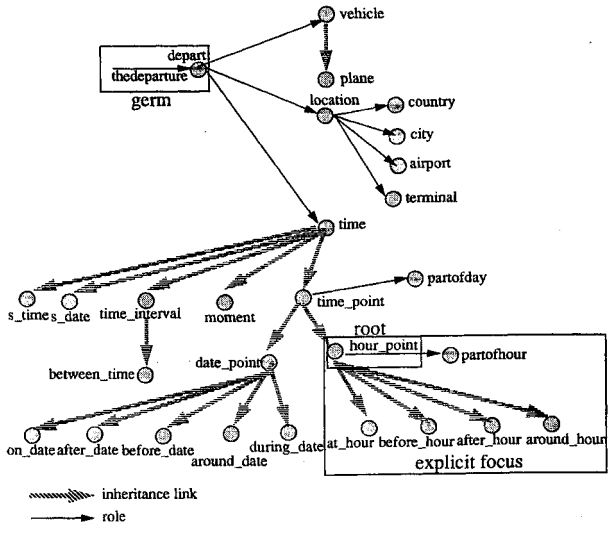

Fig.3 - Determination of focus elements
(extension phase)

The necessity of a semantic trial is determined by the nature of the predicted dialogue acts. For example, an "inform" dialogue act will trigger a semantic trial, whereas a "posi" dialogue act will not, since it does not span over a propositional content.

4.1 Dialogic trial

The *dialogic trial* corresponds to the spotting of characteristic dialogic markers at the surface form level. For example lexical items and frozen expressions like "*oui, d'accord, tout à fait, parfait, correct, ok*" are common ways to express a positive answer to a yes-no question in French. These dialogue markers are determined from the dialogue acts predicted.

V. EXPERIMENTS

A set of 72 different sentences including various linguistic phenomena have been recorded for 5 speakers (3 males + 2 females) in PSTN quality and for one speaker (1 male) in PABX quality. These utterances correspond to different dialogue contexts. They have been tested as input of the recognizer and parsing modules supplemented by either or both predictions mechanisms. Various rates are used to measure the contribution of the predictions. In the figure 4 and 5, the word recognition (WR), word accuracy (WA), sentence recognition (SR), and sentence understanding (SU) (acceptation of paraphrases) are presented for these two test sets.

		WR	WA	SR	SU
Stat.	Dyn.				
-	-	30	5	7	15
+	-	63	58	33	44
-	+	36	9	13	22
+	+	66	61	34	50

Fig.4 - Results for the test set in PSTN quality.

		WR	WA	SR	SU
Stat.	Dyn.				
-	-	54	29	29	29
+	-	84	85	60	69
-	+	54	25	30	32
+	+	85	82	61	73

Fig.5 - Results for the test set in PABX quality.

These figures show the contribution of predictions mechanisms for the recognition and understanding in dialogue contexte. In PSTN and PABX qualities, the static and dynamic predictions provide a gain of nearly 30 points on the word and sentence recognition (WR and SR), at least 35 points on the sentence understanding (SU), and more than 50 points on the word accuracy (WA).

The experiments indicate that the static predictions mechanism alone contributes the most, but the interpretation must be moderated by the fact that a language model itself independently of a dialog context, provides already an important gain.

A precise study on these test sets [3] shows that the contribution of dynamic predictions is higher on short utterances (from a 1 to 4 lexemes) often nominal groups and elliptical structures, than on long utterances. Two reasons for this result : the impact of word-pair contraints is more important for long utterances, and the probability for a short utterance to contain a focus element is smaller. The best results on long utterances are obtained when dialogic markers are used to discriminate the hypothesises.

The contribution of dynamic predictions can be increased if we take into acccount that the dynamic predictions mechanism favorised over-informative utterances. These utterances, which do not enter into the accounts of

the sentence understanding rate (SU), can be processed by the DM without affecting seriously the progress of the dialogue.

VI. DISCUSSION

To sum up, we have presented two mechanisms that improve the performances of the recognition and understanding of our oral dialogue system which deals with task driven cooperative dialogues. These mechanisms are complementary, and can be tested independently.

The *static predictions mechanism* helps to reduce the search space at an early stage of the understanding process. The exploitation of a simulation corpus to generate the static contexts could be generalized to similar applications.

The *dynamic predictions mechanism* based on a semantic preference process is relatively simple since it does not imply syntactic treatments. The mechanisms can be reused for other applications after the modification of the content of the semantic knowledge base. The nature of the links of the semantic network plays a very important role in the search of the focus elements.

Some improvements can be made to increase further the performances. We are studying the possibility of using the same grammar for the acceptance phase and the language modelling we use for the recognition. This will reduce the operations on the linguistic knowledge bases and increase the coherence of the linguistic processing. The exploitation of global predictions from the DM (list of all the attribute of the task) could help to reduce the size of the lexicon (values of attributes already instanciated such as city and airport names). The gain is estimated between 5 to 10 points. In addition, the last system surface form could be used to identify certain dialogue acts more precisely. This is the case for *posi, nega* and *correction* which can be realized by the reuse of the last system utterance.

ACKNOWLEDGEMENT

The author would like to thank its colleagues from the SUNDIAL project who have contributed to this research especially S. Thornton, F. Charpentier and F. Gavignet. Thank you also to D. Kayser, C. Fouqueré and J. Siroux for their comments concerning this subject.

References

[1] Andry F., Bilange E., Charpentier F., Choukri K., Ponamale M., Soudoplatoff S., "Computerised Simulation Tools for the Design of an Oral Dialogue System", ESPRIT Technical Week, Bruxelles, 12-15 Nov. 90.

[2] Andry F., Thornton S., "A parser for speech lattices using a UCG grammar", 2nd European Conference on Speech Communication and Technology, pp. 219-222, Genova, 24-26 Sept. 1991.

[3] Andry F., "Mise en oeuvre de prédictions linguistiques dans un système de dialogue oral Homme-machine coopératif", *Thèse de l'Université Paris Nord*, Mai 1992.

[4] Grosz B. J., *The Representation and Use of Focus in Understanding Dialogs*, in *Readings in Natural Language Processing*, Grosz, Jones and Webber (eds), Morgan Kaufmann Publishers, 1986.

[5] Bilange E., "A Task Independent Oral Dialogue Model", in *Proceedings of ACL*, Berlin, pp. 83-88, April 1991.

[6] Heisterkamp P., McGlashan S., Youd N., "Dialogue Semantics for an Oral Dialogue System", in *proceedings of the ICSLP 92 Conf. (this issue)*, Banff, Canada, Oct. 1992.

[7] Guyomard M., Siroux J., Cozannet A., "The role of dialogue in speech recognition : the case of the Yellow Pages Sytem", 2nd European Conference on Speech Communication and Technology, pp. 1051-1054, Genova, 24-26 Sept. 1991.

[8] Matrouf A.K., Gauvain J.L., Neel F., Mariani J., "Adapting Probability-Transitions in DP Matching Process for an Oral Task-Oriented Dialogue", in *ICASSP 90*, pp. 569-572, Albuquerque, 1990

[9] Niedermair G., "Linguistic Modelling in the Context of Oral Dialogue", in *proceedings of the ICSLP 92 Conf. (this issue)*, Banff, Canada, Oct. 1992.

[10] Peckham J., "Speech Understanding and Dialogue over the telephone : an overview of the ESPRIT SUNDIAL project", *Acoustic Bulletin*, 1990.

[11] Tomabechi H., Tomita M., "The Integration of Unification-based SyntaxSemantics and Memory-based Pragmatics for Real-Time Understanding of Noisy Continuous Speech Input", in *7th National Conference on Artificial Intelligence*, Vol 2, pp. 724-728, Saint Paul, Minnesota, August 1988.

[12] Yamaoka T., Iida H., "Dialogue Interpretation Model and Its Application to Next Utterance Prediction for Spoken Language Processing", 2nd European Conference on Speech Communication and Technology, pp. 849-852, Genova, 24-26 Sept. 1991.

[13] Young S., "The MINDS System : using context and dialogue to enhance speech recognition", pp. 131-136, Conf. DARPA 1989.