



A KNOWLEDGE-BASED APPROACH TO THE DESIGN OF A MAN-MACHINE DIALOG SYSTEM BY VOICE

Noëlle CARBONELL, Jean-Paul HATON, Jean-Marie PIERREL*

ABSTRACT

We are developing a knowledge-based system capable of understanding oral task-oriented dialogs and process pseudo-natural sublanguages (with few syntactic restrictions) with large vocabularies (several thousand words) in a multi-speaker environment. Information Centers provide a wide range of potential applications, in as much as they deal with the general public, i.e. with a large number of unfrequent and untrained speakers. In this paper, we define on the one hand, the various knowledge sources necessary for understanding and managing natural task-oriented dialogs, and on the other hand, the different types of dynamic information involved in the processing of such dialogs. We then present and comment the architecture of a knowledge-based system that could efficiently operate on multiple knowledge sources and data.

INTRODUCTION

Continuous speech understanding, i.e. the recognition and semantic interpretation of full sentences, plays a prominent part in the management of oral man-machine interfaces. But current understanding systems prove inadequate when the processing of natural dialogs is considered : the capabilities of an oral dialog processor cannot be restricted to the mere understanding of statements [1]. Such a system should also be able to :

- organize speaking turns and generate appropriate replies to the human user, more generally, control the overall dialog progress ;
- track the successive topics the user broaches upon in his speech and find out his goal(s), hence the actual task he requests the system to perform ;
- understand whole dialogs, which is more complex than understanding isolated sentences.

In this paper, we first define on the one hand, the various knowledge sources necessary for understanding and managing natural task-oriented dialogs, and on the other hand, the different types of dynamic information involved in the processing of such dialogs.

We then present and comment the architecture of a knowledge-based system that could efficiently operate on the multiple knowledge sources and data mentioned in the preceding paragraph. In these two parts, aspects of our long-term project will be analyzed as an illustration. The variety of knowledge sources involved in the speech communication process together with the great undeterminism about facts and data make it necessary to implement sophisticated search strategies and heuristics. The problem is relatively similar to the one encountered in the field of computer vision [2] and it constitutes a major challenge of Artificial Intelligence.

FUNCTIONALITIES AND ARCHITECTURE OF THE DIALOG SYSTEM [3]

This paragraph describes the architecture of the system we are presently implementing. We have chosen a multi knowledge-based approach, in order to achieve two major goals :

* Pattern Recognition and Artificial Intelligence Group
CRIN / INRIA-Lorraine - B.P. 239 - 54506 Vandoeuvre les Nancy cedex, France.

- **flexibility** : our aim being the development of a system that could accommodate a wide range of applications, the knowledge sources relating to a specific application (namely, the lexicon and the application universe) are defined as global parameters in the system,
- **modularity** : the general system architecture consists of five main processors, each of these operating on a particular type of data and using one or several specific knowledge source(s). The concept of separation between knowledge and processes allows for permanent interaction between designer and system, therefore promoting incremental development.

In this respect we depart from existing architectures in-as-much as processors in our system interact and cooperate instead of merely processing sentences in hierarchical order. Moreover, in order to be efficient the analysis strategy should be adapted to the nature and volume of information available at a given moment.

The systems's components

Our system is made up of five autonomous processors, each corresponding to a main knowledge source used in the understanding of speech. These processors range from the level closest to the signal to the level furthest from the signal. We describe them briefly below :

- Acoustic-Phonetic (**APHON**) :
segmentation of the signal and phonetic labeling of the different segments ; more generally, construction of a representation of a user's statement in the form of a phonetic lattice,
- Prosodic (**PROSO**) :
detection in the signal of prosodic marks indicating :
 - . word or syntagm boundaries,
 - . the nature of a statement (assertion, question, contestation),
- Lexical (**LEX**) :
construction of a lexical representation of a statement in the form of a word lattice, using the acoustic-phonetic lattice generated by APHON,
- Syntactic-Semantic (**SYN-SEM**) :
elaboration from the lexical lattice associated with a segment of one or more syntactic-semantic representations of the same statement (use of case grammars),
- Dialog (**DIAL**) :
for each interaction with the user :
 - . interpretation of the syntactic-semantic representations of the user's statement according to the current dialog history, active script(s), application universe and task goal,
 - . the generation of an appropriate reply.

More generally, this procedure takes charge of the dialog management and the task fulfilment (i.e. the satisfaction of the user's request), which leads to complex reasoning and dialog history updating.

Exchanges of information between processors

Each processor has at its disposal specific static knowledge sources ; for example, a dictionary is associated with the lexical processor, providing for each lexical entry, phonological and syntactic-semantic information besides phonetic descriptions of words.

Moreover, at each step of the recognition process, every component can exchange dynamic information about the current statement with each of the others. Dynamic information exchanges between two processors may be characterized as follows :

(a) from a processor **P** towards a processor **Q**
(from a higher level towards a lower level)

- predictions from **P** about a part of the statement that has not yet been processed by **Q**, within the framework of a top-down analysis ; these predictions enable **Q** to reduce the number of concurrent hypotheses to be examined when processing this part of the statement (analysis focusing),
- hypotheses from **P**, that **P** requests **Q** to confirm before adopting them definitively.

(b) from a processor **Q** towards a processor **P**
(from a lower level towards a higher level)

- on the one hand, adoption or rejection of a prediction provided by **P** (according to strict selection criteria), on the other hand, validation or invalidation of a hypothesis that **P** wishes to confirm (according to compatibility criteria less strict than the former).
- results from the processing by **Q** of a statement, within the framework of a bottom-up analysis from left to right or from the middle outwards ; for example, if **Q** designates the lexical processor and **P** the syntactic-semantic processor, then these results are constituted by the lexical lattice generated by **Q** for a given part of the statement.

The system's functioning and strategies

While defining the functions of the various processors we tried to ensure the maximum autonomy for each one, which allows for some parallelism at a global level, thus reducing the system response time without affecting accuracy.

At the understanding level, the present global strategy of the system may be described as follows :

- SYN-SEM supplies DIAL with a set of syntactic-semantic representations classified according to recognition criteria computed from phonetic, lexical, syntactic-semantic scores,
- DIAL applies a "best-first" strategy based on criteria that differ from those used by the other components, in order to choose from amongst the different representations of a statement the one that it will interpret first : amongst representations whose recognition score is superior to a given threshold and which are compatible with the dialog history, DIAL will select the one that will enable it to come nearest to the current goal(s).

In the area of continuous speech understanding, the object-oriented architecture described here is somewhat original in as much as :

- it consists of autonomous processors, each associated with a specific knowledge base and capable, in the course of the understanding process, of exchanging information about the current statement with the others ; which may significantly improve understanding results,
- it allows for better integration of the dialog component into the understanding process,
- it accomodates flexible analysis methods,
- it overcomes the drawbacks of current structures, namely : fixed strategies (as in architectures controlled by supervisors) as well as the inefficiency linked to "blackboard" type structures. For, in our system, control strategies are essentially dynamic and take into account the nature of available dynamic information ; moreover the activation of the different processors is not determined by incoming data but triggered by events,
- it enables some degree of parallelism, which can improve the system response time if implemented on a multi-processor,

- finally, this architecture is open to modification and evolution. It allows incremental development of the system thanks to its modularity and to the fact that the representations of the various information in the system has been specified right from the start.

All these features enable our system to approximate the mental activity of human understanding, at least what we understand about it, thanks to psycho-linguistic researches and the recent progress in artificial intelligence. Our system functions as a cooperative expert society.

CONCLUSION

The design and implementation of man-machine dialog systems using continuous speech is amongst the most difficult challenges of Artificial Intelligence. The number and the variety of the knowledge sources involved in the speech communication process, as well as the large indeterminism of both data and knowledge make it necessary to use sophisticated search strategies and reasoning techniques. That makes the problem more difficult than written language understanding and draws similarities with the field of computer vision.

Most speech understanding systems developed so far are basically sentence interpreters with very restrictive dialog capabilities. We have presented in this paper the principles and present state of our project of a task oriented speech dialog system. The basic idea is to fully integrate the dialog level amongst the other knowledge sources and processes in order to ensure a better understanding. The use of an object-oriented representation of knowledge and of interactions between knowledge bases results in an architecture of expert society that has been presented and discussed.

ACKNOWLEDGEMENTS

This work has been partly supported by CNRS GRECO "Communication Parlée". The authors gratefully acknowledge the active participation in this research of several members of the group, Bernard Mangeol, Philippe Morin, Pierre Mousel and Azim Roussanaly.

REFERENCES

- [1] J.M. Pierrel, "Aspects of Man-machine Dialog", in "Fundamentals in Computer Understanding", J.P. Haton (ed.), Cambridge University Press, 1987.
- [2] J.P. Haton, "Intelligence artificielle en compréhension automatique de la parole : état des recherches et comparaison avec la vision par ordinateur", T.S.I., vol. 4, n° 3, pp. 265-287, 1985.
- [3] N. Carbonell and J.M. Pierrel, "Architecture and Knowledge Sources of a Human Computer Oral Dialogue System", Proceedings of OTAN Workshop "Structure of Multimodal Dialogues including Voice", Corsica, Sept. 1986.