Issues in the Design of a Voice Man Machine Dialogue System Generating Written Medical Reports

P. Nugues, P.O. ElGuedj, F. Cazenave, and B. de Ferrière

Institut des Sciences de la Matière et du Rayonnement
Caen, France

***Abstract*— In this paper, we describe the syntactic and semantic issues of a voice man machine dialogue system. This system, featuring a voice recognition circuit, allows the dictation of medical analysis reports and automatically generates a written document. An object oriented kernel is designed to represent the linguistic and medical entities. The syntactic parser is adapted to voice input and enables the physician to dictate using natural language. A semantics analyzer manages the medical entities and the agenda of the report. This analyzer allows to detect certain errors or inconsistencies, and to warn the user, using synthesized audio messages. The audio interface has been merged into a graphical environment to allow keyboard and pointing device controls, as well as graphical alerts.**

I. Introduction

Voice man machine dictation systems can be used to automate the reporting of medical analyses. These systems involve several techniques of speech and language processing: voice recognition, syntactic parsing, semantics analysis, and man-machine dialogue [1, 2]. They allow the physician's report to be directly captured by a computer and to be immediately available for printing, modification, etc.

The dictation system we describe uses an external word recognition circuit from the industry and is based on a modular architecture composed of three Knowledge Sources (KSs). These Knowledge Sources carry out the dialogue, syntactic parsing, and semantics analysis (Fig. 1).

Earlier research has produced the knowledge sources architecture and a first prototype [3]. Present efforts have focused primarily on the design of an object oriented kernel, on the improvement of the syntax and semantics KSs, and on the building of a new prototype.

The object oriented kernel encompasses all the linguistic and medical entities handled by the system and allows their representation into classes of objects. These classes are shared, in a consistent way, using the same object oriented language, by the three Knowledge Sources.

The syntactic parser uses contextual rewriting rules and is adapted to oral natural language. It receives a stream of words from the physician, and creates structured linguistic objects concurrently with the dictation. Once the syntactic structure of the sentence and the category of each word has been determined, the semantics analyzer establishes relationships between these words and interprets them.

The dialogue audio interface processes text edition controls such as corrections, and controls the synthesis of warning or error messages. It is supplemented by a graphical environment that allows keyboard and pointing device controls, as well as graphical alerts.

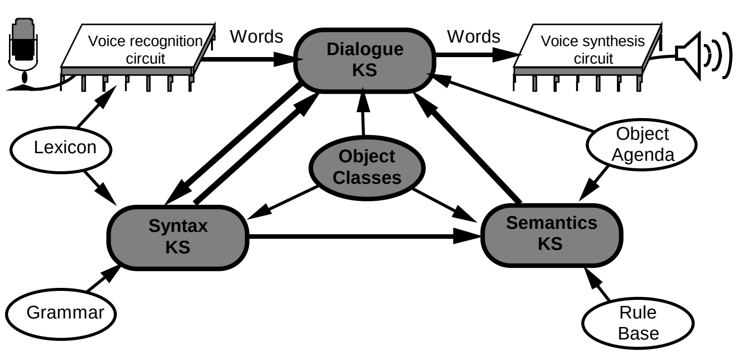

Fig. 1. The Knowledge Sources Architecture

II. The Object Oriented Representation

The object oriented kernel is implemented on top of Prolog. It has been inspired by [4] but we have significantly extended the proposed implementation. This kernel allows the creation of classes with variable and method slots, and of objects which are instances of classes. Classes and objects pertain to a hierarchy and inherit slots from their superclasses. Methods are activated using a message passing mechanism.

Syntactic classes correspond to word and phrase categories (parts of speech). For each sentence of the report, an instance of the "Sentence" class is created. The "Sentence" class notably contains the slots: subject noun phrase, verb phrase, object noun phrase, and prepositional phrases. These slots are assigned with instances of category classes which are created according to the result of the sentence parsing. For example, the slot "subject noun phrase" will contain an object from the class "Noun Phrase." The slots of this latter class include notably: determiner, list of adjectives, noun, number, and gender. They are assigned with the corresponding words of the parsed sentence.

Semantic classes represent the medical entities and the structure of an analysis report, which is for the moment limited to reports of Holter ambulatory monitoring. These classes are organized as: "Themes," "Subjects," and "Attributes."

"Themes" map the main divisions of a report. A theme would correspond, on a written document, to a paragraph. During the dictation, the transitions from a theme to another are monitored using an automaton, supplemented by rules that determine the legality of these transitions. Theme slots include notably: error and help messages, and the agenda of the next possible themes.

Within a particular Theme, "Subjects" represent the possible concepts that can be mentioned. For example, "Heart Rhythm," is a "Subject" pertaining to the "Heart Waves" theme. Certain subjects must be mentioned, others are optional. The physician may tackle the different subjects of the agenda of a theme in any order, but the compulsory ones (which are dynamically defined) must all be tackled before proceeding to another theme.

A Subject is described using "Attribute" objects corresponding to groups of adjectives associated to this subject. For example, "Speed" is an attribute of "Heart Rhythm." Attribute slots include the value given through the dictation and properties such as the domain of possible values, etc. In addition, subjects and attributes include a slot defining a list of constraints that the presence of a subject or the specific value of an attribute may impose to other subjects or attributes.

III. A Syntax Compatible with Incomplete Sentences and Word Recognition

The syntax KS parses the sentences and checks their acceptability. The syntactic parser uses extended Definite Clause Grammar (DCG) rules [5]. Linguistic objects are created concurrently with the parsing process and possibly retracted in case of backtracking. The structure of a sentence, for example, is expressed by the following simplified rule:

```
sentence(Obj) -->                  (1)
  nounP(Nobj), verbP(Vobj),        (2)
  {createObject(sentence, Obj),    (3)
  assign(Obj, nounPhrase, Nobj),   (4)
  assign(Obj, verbPhrase, Vobj)}.  (5)
```

which signifies that a sentence is a noun phrase followed by a verb phrase (1), (2). For each sentence, a `sentence` object: `Obj` is created (3), whose `nounPhrase` slot is assigned (4) with the noun phrase object: `Nobj`, returned from (2). `verbPhrase` (5) is processed in a similar way.

The grammar includes context sensitive rules whose basic syntactic units are: Subject Noun Phrase, Verb Phrase, and "Complements." Complements include phrases such as adjectival complement, object noun phrase, and several types of prepositional phrases. Context sensitivity enables to restrain the number of possible structures and notably to check, at the syntactic level, the presence (resp. absence) of compulsory (resp. prohibited) complements that certain verbs may request.

A standard DCG parsing is not practicable for voice input because it only succeeds with a correct and complete sentence. To accept a continuous sequence of words from the recognition circuit, we have extended the DCG rules by augmenting them with a "`freeze`" predicate.

This predicate allows the parsing of an incomplete sentence, represented as a list of words. When the end of the list is reached, it freezes the parsing until it receives a new word. On reception of a word, `freeze` allows to process it and to append it to the sentence in progress or possibly to reject it if not syntactically correct. The sentence is ended by uttering the word "period." The `freeze` predicate is compatible with DCG formalism and must be inserted before each terminal symbol (word) in the right–hand side of the rules.

Besides, since the recognition circuit, which operates at the phonetic level, can't make any difference between homonymous words, the lexicon includes the phonetic spelling. Parsing is performed considering words together with their phonetics.

IV. Semantics Verification

From the linguistic objects, the semantics KS creates the medical objects representing the concepts the physician describes. In addition, the semantics KS allows to check whether a sentence makes sense, in itself, and with respect to previously created objects.

Once the category of each word has been identified, it is necessary to bind words together to create the medical objects. The bonds are determined using the "`getAssociations`" predicate which retrieves, for each noun of a sentence object, the associated adjectives. It considers the trivial case where the noun and the adjectives are in the same phrase. It also examines verb adjectival complements and, recursively, relative propositions. `getAssociations` allows thus to detect all the subjects of a sentence, to create the corresponding objects, and considering the adjectives, to create the attribute objects.

Once the medical objects are created, verifications are performed, notably on the consistency between different attribute values. These verifications are implemented through the `constraintsList` slot of the attribute objects which include constraints of the form: (value 1, constraint type, destination attribute, value 2). The constraint type can be: compulsory, optional, or prohibited, and signifies, for example, that if the "value 1" is given, it prohibits the "value 2" of the destination attribute. Once an attribute value is given, the corresponding constraints are propagated using a Truth Maintenance System [6]. If the value is removed, the corresponding contraints are retracted. Constraints can also be propagated to subjects.

A conflict is detected, for example, when an attribute object is created with a prohibited value or when a prohibited value is transmitted to an existing subject object. In this case a message is issued; nevertheless, the physician may override this warning and the corresponding object will be created.

V. Merging Graphical and Voice Interfaces

The messages emitted by the system combine speech synthesis and a graphical interface. These messages correspond to alerts and questions, and also to dictation help for a physician novice to the system.

The graphical interface (Fig. 2) provides a help through the list of subjects and themes which can be tackled. Notably, within a current theme, compulsory and optional subjects, together with the authorized values of their attributes, are specified in two different windows ("Sujets nécessaires" and "Sujets facultatifs"). These windows are updated concurrently with the creation of new objects and the propagation of new constraints. Once all the compulsory subjects are described, the next possible themes appear in the "Thèmes" window.

While the graphical interface provides a quite detailed help, the speech synthesis interface enables the physician to concentrate on the analysis without looking at the computer. Voice messages are then used to only warn or to question the physician when an error or a conflict occurs. These error messages try to manage the following trade-off: to be as explanatory as possible, but also as short as possible, because sound messages tend to be tedious if overused.

To notify an error we use different sound short cuts. The sound interface signals also, for instance in case of a syntactical error, where the error occurred by repeating the current sentence deprived of the faulty word. This type of error is mostly due to a confusion made by the voice recognition circuit. The graphical interface may also provide an alert box as well as an optional explanation using a text window.

VI. Implementation and Results

A prototype was first developed on a 386SX computer then ported to a Macintosh II SI to ease the graphical interface development.

The voice recognition is performed by an Introvoice module [7]. This module is driven by a resident software which operates concurrently with the dialogue and language processing program. Recognition is operated on isolated words which are sent to the keyboard buffer of the computer. Text-to-speech synthesis uses the Apple Sound Chip. The software was developed using the AAIS Prolog language and totals more than 7,500 lines (400 predicates).

The prototype was tested on Holter analyses. To implement it, we developed a syntax covering a subset of the French language, with a total number of 180 words, and the associated semantics. This prototype was presented to the French National College of Cardiology where it met a very positive welcome. It was also positively assessed and granted by the French Agency for Technology Transfer.

VII. Conclusion

We have presented a prototype of a voice dictation machine dedicated to medical use. We have detailed certain issues concerning notably the syntactic and semantic features.

The merging of graphical and audio interfaces provides different message modalities from the computer. The graphical interface displays a rather detailed help while the audio interface frees the physician from continuously watching the screen.

Efforts are in progress to make the system provide vocal help and answer questions such as for example: "And now what?" or "What are the possible values?"

We hope that these efforts will provide a useful aid for the physicians in their reporting tasks.

References

[1] R. Kurzweil, *The Age of Intelligent Machines*, MIT Press, 1990.

[2] J.M. Pierrel, *Dialogue oral homme-machine*, Hermès, 1987. (In French).

[3] P. Nugues, E. Bellegarde, H. Coci, R. Pontier, and P. Trognon, A voice man machine dialogue system generating written medical reports, *Proc. of the Annual Int. Conf. IEEE/EMBS*, Vol. 13, 1991.

[4] Object Oriented Programming in Prolog, E.P. Stabler, *AI Expert*, Oct. 1986.

[5] F.C. Pereira and D.H. Warren, Definite Clause Grammars for Natural Language Analysis, *Artificial Intelligence*, Vol. 13, 1980.

[6] J. Doyle, A Truth Maintenance System, *Artificial Intelligence*, vol. 12, 1979.

[7] *Introvoice IV user's guide*, The Voice Connection, Irvine, 1989.

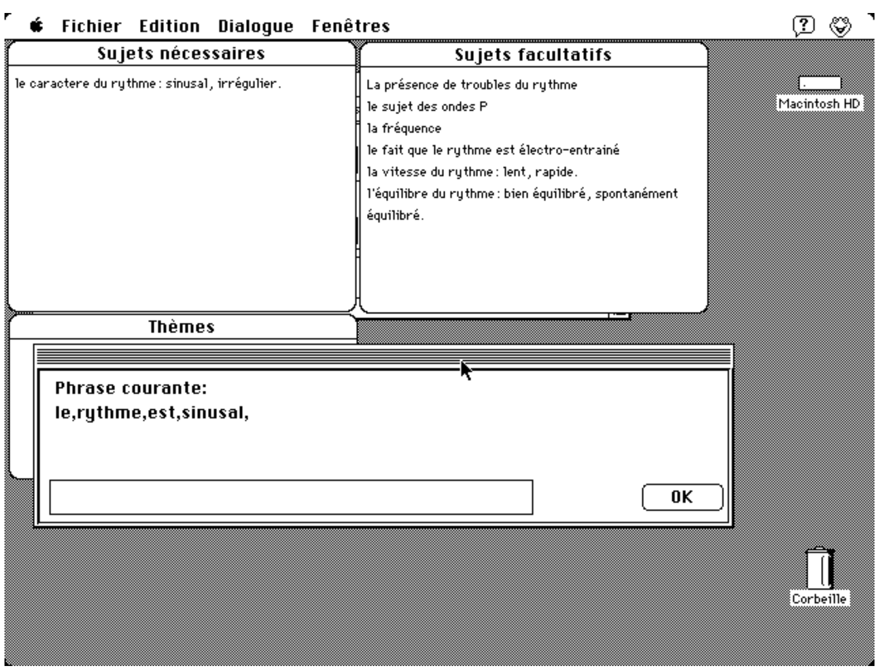

Fig. 2. The System Graphical Interface

Pierre NUGUES
Institut des Sciences de la Matière et du Rayonnement
6, boulevard du Maréchal Juin,
F-14050 CAEN CEDEX, FRANCE
Phone: 33-31-45-27-22
E-mail: pnugues@ismra.ismra.fr