International Journal of Computer and Information Sciences, Vol. 9, No. 2, 1980

Querying Databases with a Domain-Oriented Natural Language Understanding System[1]

Marco Bernorio,[2] Marco Bertoni,[2] Arnaldo Dabbene,[2] and Marco Somalvico[3,4]

Received October 1978; revised April 1979 and December 1979

This paper presents the DONAU (Domain Oriented NAtural language Understanding) system. The system can extract, from a sentence expressed in natural language, the useful information that is necessary in order to provide either an appropriate command for a robot or an acceptable query to a database system. The DONAU system, being adapted for such different versions, is intended to provide a contribution of quite general significance in the field of natural language understanding and within the general area of artificial intelligence. In fact, while a first version of DONAU, which has been developed and successfully tested on UNIVAC 1108 computer, is devoted to the semantic domain of robotics, a second DONAU version for querying databases has been constructed. Thus, the DONAU architecture has been conceived and developed in order to provide an experimental and formalizable result that is of general value, and that therefore can be applied to semantic domains of a different type as well.

KEY WORDS: Artificial intelligence; natural language understanding; syntax; semantics.

[1] A preliminary version of this paper was presented at the IIASA Workshop on Natural Language for Interaction with Data Bases, Laxenburg Schloss, Austria, January 1977.

[2] Milan Polytechnic Artificial Intelligence Project, Istituto di Elettrotecnica ed Elettronica, Politecnico di Milano, Milan, Italy.

[3] Stanford Artificial Intelligence Laboratory, Computer Science Department, Stanford University, Stanford, California.

[4] Present address: Milan Polytechnic Artificial Intelligence Project, Istituto di Elettrotecnica ed Elettronica, Politecnico di Milano, Milan, Italy.

0091-7036/80/0400-0141$03.00/0

1. INTRODUCTION

Natural language (NL) understanding has been for many years one of the central research goals of artificial intelligence.[6,16] Different approaches to the problem have been followed, and different experimental results have been obtained. The content of this paper is devoted to providing a conceptual and experimental contribution to this field, based on the invention of a particular modular architecture for an NL understanding system, and centered on the realization and experimentation of the DONAU (Domain Oriented NAtural Understanding) system, in a version intended to provide NL programming language of a robot.[3,14] The purpose of this paper is the study of the problem of NL understanding with the goal of obtaining a highly modular system of general validity.[5,12,13] Such a system is intended to be able to extract from an NL input sentence (IS) information useful for constructing an order to be executed by a robot or a query for a database.[1,9,15]

A preliminary version of a DONAU system was developed and tested on the UNIVAC 1108 of the Milan Polytechnic Artificial Intelligence Project (MP-AI Project).[2,7] At the present time, a first version, devoted to the semantic domain (SD) of robotics, is working; a second version, devoted to the SD of the International Institute for Applied System Analysis (IIASA) database, has been also successfully tested. In fact the DONAU system is based on general criteria, on which research has already been done, and which provide the possibility of developing different DONAU versions devoted to different SD's. On the other hand, the already available DONAU version on robotics cannot be considered as having reached a final state of development. Study and evolutional changes are still being done in order to increase and improve the modularization of the system, and in order to make more practicable and simple any change of SD of application.

The reason the first DONAU version was devoted to robotics is related to the existence, within the MP-AI Project's Robotics Laboratory, of the SUPERSIGMA robot devoted to the assembly of complex mechanical systems. We think that in the near future it will be possible to interact with SUPERSIGMA in Italian. Moreover, the possibility of programming a robot in NL represents a typical example of the research of artificial intelligence and the technical exigencies of robotics finding an interesting connection point. Also, the SD of robotics presents, in relation to NL understanding, a practical and realistic environment (namely industrial robotics) that enables one to clarify some controversial points about how to face such research problem.

After the SD of robotics was chosen for the development of a first DONAU version, a simple assembly robot was simulated on the UNIVAC 1108 computer. This simulated robot corresponds in many ways to the

SUPERSIGMA robot, so that, in the future, a direct connection between DONAU and SUPERSIGMA will be easily established. The simulator was developed by using both the LISP and MICROPLANNER languages.[10,11,17] The commands, which can be executed by the robot, are expressed by means of patterns of MICROPLANNER theorems. Therefore, the control language of the robot is, at the present time, MICROPLANNER.[4,8,18] It is interesting to observe that it will be possible to substitute, in the future, this simulator with a module of command SUPERSIGMA. In order to develop a new DONAU version on a database, we think that such a robot-oriented module will be replaced by another module that will deal with the query of a database.

After the construction of the robot's simulator, we selected a significative set of sentences, related to man–machine interaction, that we have called interaction protocol (IP). The various modules dedicated to performing the semantic analysis were constructed with the guidance of the IP. Such goal-oriented construction of the semantic analyzer constitutes one of the innovative characteristics of our system.

Moreover, a clever examination of each IS of the IP enables us to reach a deeper understanding of the NL subset which is related to the actual SD considered by the DONAU system. All the peculiarities, ambiguities, and problems related to NL understanding are considered in a more precise and realistic way, by having under consideration such IP.

The analysis of an IS is done according to the conceptual order of (1) syntactic analysis, (2) semantic analysis, (3) operative information extraction, (4) legality control, and (5) interaction execution. The syntactic analysis makes use of the PIAF system, developed at the University of Grenoble, France, and working also for Italian language. To an IS as input of the PIAF system correspond as output a few syntactic structures (SS's). The semantic analysis is performed by a semantic descrimination network (DN) which is based on a hierarchy of model lists (ML's); such a DN is an important part of the modular architecture of the DONAU system. A uniform technique has been developed that enables the construction of a different DN when a different SD and, hence, a different IP (i.e., a different DONAU version) is selected.

The operative information extraction is performed by another module, which operates on the results of the matching between the different SS's and the DN. Such results, contained in information lists (IL's), are processed automatically, within another module, by means of elimination rules which are based on domain-oriented information sets (IS's), which specify the partial ordering of information content assigned to various words of the lexicon.

IS's can be changed as well, within a general technique, when a new SD, and hence a new DONAU version, is developed. The legality control is made in order to check that the extracted operative information (OI) is consistent with respect to the interaction world, namely with the state of the world. Thus, it can eliminate operational ambiguities, and it produces, at its output, the executable interaction (EI). The module for such control is general; its output is provided to the robot's simulator which then executes the commands.

In Sec. 2 we illustrate the role of syntactic and semantic analysis in NL understanding in a restricted SD. In Sec. 3 a detailed example of understanding of an IS, taken from a first DONAU version on robotics, is presented. In Sec. 4 the design criteria of a second DONAU version for querying databases are illustrated. In Sec. 5 concluding remarks and directions for future research are presented.

2. NATURAL LANGUAGE UNDERSTANDING IN A RESTRICTED SEMANTIC DOMAIN

In this section we will describe, in general way, how NL understanding is structured and analyze an NL input sentence (IS) performed by the DONAU system. The NL understanding process can be divided, as a first approximation, into the following phases:

1. syntactic analysis;
2. semantic anslysis;
3. operative information extraction;
4. legality control;
5. interaction execution.

The syntactic analysis is made by the PIAF system, which was developed at the University of Grenoble[5,8]; this system works currently for French but can be adapted to Italian as well. It is evident that, in this way, we have organized the syntactic and semantic analyses as two independent activities performed by two autonomous modules. Therefore we have not considered the solution of a close interaction between the two analyses, as it has been suggested by Winograd and other authors.[16,17] We have adopted this criterion of separation, which, by the way, has been followed by other authors as well,[6] in spite of the consequent decrease of efficiency, which probably arises because of the lack of cooperation between the two analyses.

Such a decision has been considered as a convenient one because of the possibility of having already available a module for performing the syntactic analysis, thus avoiding, for the present time, the need of developing such a module by ourselves. But the more important reason lies in the fact that the characteristics of the PIAF system are such that it can provide, in correspondence to one IS, only a very limited number of output SS's. Thus, in most cases, it is not cumbersome at all to examine all of them. At any rate such a design criterion enables one to proceed on a system organized within a modular structure. It is therefore possible to proceed to modifications and improvements of each block, the same time preventing a change of a module from influencing another module. The idea of considering the DONAU system as a software system, including all the most advantageous technological results of software engineering research, such as modularization, is one of the main characteristics contained in our research effort.

An inconvenience of having the syntactic analysis independent from the semantic analysis could lie in the fact that an IS, illegal on the syntactic level, could be eliminated during the syntactic analysis also when it might actually contain all the information necessary for providing a command to the robot. This inconvenience would be rather unacceptable, because it would oblige the user to face a heavy task in formulating the IS's. This difficulty is largely overcome by the possibility, provided by the PIAF system, of also analyzing IS's that are illegal, within some limits, thus constructing, also in these cases, the corresponding SS's. It is then the responsibility of the semantic analysis of processing such SS's, in order to carry on the successive understanding process. This advantageous PIAF characteristic, which has not yet been completely exploited, will provide, in the future, some interesting improvements.

As we have already mentioned, the syntactic analysis is followed by the semantic analysis. The width and the complexity of the Italian language make evident the impossibility of considering its machine understanding on a completely general level. We think that NL understanding, within the framework of artificial intelligence, should be always founded on the identification of a well-defined (and, hence, necessarily formalized) NL subset and of a well-specified SD of application. It is therefore necessary to select such well-specified SD with the identification of a specific task to be accomplished, where the NL interaction plays a precise role.

As a specific SD on which to develop the first version of our DONAU system, we have selected the SD of robotics. This choice was motivated by the fact that in our Laboratory of Robotics we have working the SUPERSIGMA robot, devoted to performing the assembly of complex mechanical systems. Thus, a DONAU version on robotics should provide, in the near future, a direct NL interaction between the user and SUPERSIGMA. The role of the NL interaction is the programming of SUPERSIGMA in processing and executing and assembly algorithm. The mechanisms that are actuated during the semantic analysis will be illustrated, in greater detail, in the Sec. 3. However, we will here outline some general criteria that have been utilized.

The first criterion is the adoption of a nondeterministic language for processing the various SS's from PIAF, in order to discriminate which of them should be considered as the correct one, hence eliminating syntactic ambiguities. Thus, we have selected MICROPLANNER as the language in which the semantic discrimination network (DN), devoted to the task of eliminating ambiguities, has been constructed; hence the DONAU system results are programmed both in LISP and in MICROPLANNER.[10]

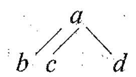
Moreover, we have designed the DN under the guidance of a very simple subset of the understandable NL called interaction protocol (IP). In fact the DN is made up of a certain number of model lists (ML's), hierarchically arranged, whose selection and definition are based on and guided by the IP. The IP has been defined with the idea of exemplifying some typical IS's from which an unambigous meaning should be obtained and an executable operative information should be extracted. Obviously, the IP selection should not be limited to the extension of the Italian subset which can be utilized in the NL interaction. This is, in fact, our case, since the IP is simply a subset of the understandable NL. Still we think that our DONAU system is a modular system, where each module is designed with some general technique and with some SD orientation. The IP plays the role of orienting the construction of the DN toward the selected SD.

The operative information extraction constitutes a successive phase of the understanding process. It is executed by a separate module which takes as its input the information constructed by the matching between the various SS's, of an IS, and the DN. Such constructed information, contained in information lists (IL's), is therefore processed, in order to eliminate semantic ambiguities, by this new module, which is based on the operation of particular elimination rules. The module, like the preceding two modules devoted to the syntax and semantic analyses, is based on an orientation toward the selected SD. This guidance is provided by the information sets (IS's), dependent on the SD, which specify the application of the elimination rules. By applying the elimination rules on the IL's, we obtain the output of this module, called operative information (OI). Each OI corresponds to an interaction with the real world (in our case the command to a robot, but, in another case, the query to a database).

The task of the successive phase, called the legality control, is performed by another module of the DONAU system, which processes the OI and proceeds to eliminate the operative illegalities related to IS's that are correct both on the syntactic and the semantic level, but that are inconsistent with the actual status and configuration of the interaction world (the robot or the database). Evidently, this module is also heavily related to the selected SD. The output, called executable interaction (EI), is processed by the last module of the DONAU system, which, in our DONAU version on robotics, is now constituted by the robot simulator and later, will be constituted by the interaction execution module connected with the SUPERSIGMA robot.

As we have already said, the syntactic analysis is performed by the PIAF system, while the successive phases (semantic analysis, etc.) are carried out by the DONAU system. Therefore, while DONAU activity will be discussed in greater detail in the following sections, it is worthwhile to illustrate the output of the PIAF system (i.e., of the syntactic analysis), which constitutes the input of the DONAU system in its present status. The PIAF system, after having received as its input the IS to be analyzed, provides as its output a small number of SS's which are alternative candidates of the syntactic interpretation of the IS.

It is the task of the successive DONAU's semantic analysis to perform the choice of that one candidate that is correct, not only syntactically, as is any one of the SS's, but semantically as well, as is the selected SS. While an SS, actually outputted by the PIAF system, is a tree, the input to the DONAU system is a parenthesized description of the same tree. For instance, to the tree

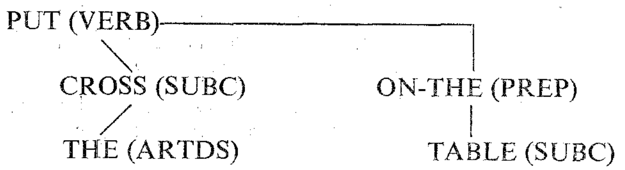


will correspond the list (a b c*d), i.e., a list defined recursively in this way: the first element is the root of the tree, and the asterisk separates the left subtrees from the right ones. The PIAF system constructs an SS in such a way that the root of a tree corresponds to an NL word which, in the IS, is preceded by the words contained in the left subtrees and is followed by the words contained in the right subtrees. (In the sequel to this paper we will present the English translation of any given IS considered. The corresponding IS in the Italian language, which is the actual input to the DONAU system, will be presented in square brackets after the English translation; this will be done only the first time that such an IS is considered.)

In order to illustrate in a better way the output of the PIAF system, let us examine the following example of IS:

PUT THE CROSS ON-THE TABLE
[METTI LA CROCE SUL TAVOLO]

The PIAF system provides, as its output, the two following SS's:



((PUT VERB)*((CROSS SUBC) (THE ARTDS)*((ON-THE PREP)* (TABLE SUBC)))

```
PUT (VERB)
    \
     CROSS (SUBC)
    /         \
THE (ARTDS)    ON-THE (PREP)
                  \
                   TABLE (SUBC)
```

((PUT VERB)*((CROSS SUBC) (THE ARTDS)*((ON-THE PREP)* (TABLE SUBC))))

In this example of IS, we can distinguish three parts: the verb (PUT [METTI]), the object (THE CROSS [LA CROCE]), and the place (ON-THE TABLE [SUL TAVOLO]). It is easy to observe that in only the first one of these two SS's can we clearly distinguish these three parts. These are the criteria that will guide the algorithms devoted to the execution of the semantic analysis, as it will be illustrated in the next section.

3. THE ANALYSIS OF NL INPUT SENTENCES

In this section we will discuss how it is possible to solve the problem of the extraction of the OI from an IS. At the DONAU's input we have some SS's provided by PIAF, which are represented, for our convenience, as LISP lists. We have to appropriately elaborate these lists, in the semantic context in which we operate, in such a way that we can recover the useful elements for the construction of a MICROPLANNER command for the robot.

Before examining in detail the mechanism that guides the semantic analysis, we want to point out how in an IS, which is apparently simple and easily understandable for human intelligence, there can be quite often some ambiguities and some lack of information. This is because of the fact that man often expresses himself either with sentences containing redundant information or with sentences requiring some deduction activity. On the other hand the computer, in order to operate in a correct way, needs the provision of precisely described information, which should be neither scarce nor redundant. Therefore it is necessary to insert, in the modules that deal with the analysis of the SS's the complete knowledge of the world in which the robot operates. This will allow the computer to understand the IS, i.e., to make the necessary deductions that are based on the information contained in the IS.

We will now show how the analysis of the SS's has been realized. This procedure can be considered as divided into three phases: in the first phase we discriminate, from the set of SS's provided by PIAF, a certain number of parenthesized structures; in the second phase we extract from such structures the information contained in the IS, which is sufficient in order to build the command to the robot; in the third phase we control the consistency of the extracted information with the physical reality of the robot's world. We will now examine, through a simple example, this process. Let us consider the following IS to be analyzed:

TAKE A TI FROM-THE TABLE
[PRENDI UNA TI DAL TAVOLO]

The first word that the Filter will take into consideration is the verb of the IS; in our case TAKE. Associated with this verb there are some ML's that correspond to the possible structures of a command beginning with this verb. In our example we find, in correspondence with the verb TAKE, the two following ML's:

(OBJ)
(OBJ PLACE)

The ML's indicate, as elementary semantic models, that the verb TAKE must necessarily be associated with the information of the object to be taken and, optionally, with the information of the place where the object to be taken is located. It is evident that, in our example, we find ourselves in the second case and that, in the IS, the two groups of words corresponding to OBJ and to PLACE are well identified. One SS, which is provided by PIAF in correspondence with the IS of our example, is the following one:

(TAKE*(TI A*) (FROM-THE*TABLE))

The elements OBJ and PLACE, belonging to the considered ML, are called meaningful elements (ME's). Each one of these ME's is associated with a set of ML's that distinguishes the various ways in which they can be further developed. In the example of the considered IS, we have the following correspondences:

OBJ → (SUBS ARTDS)
PLACE → (PREP SUBS)

Moreover, the elements SUBS, ARTDS, and PREP cannot any longer be developed; therefore, they are called basic meaningful elements (BME's). At this point it is clear how we can make a first choice of the SS's given by PIAF. Each SS is examined and, at each level of the syntactic tree, it is matched with the ML's of length equal to the number of syntactic subtrees whose roots belong to that level. This matching procedure is recursively repeated at all levels of the SS and, if the matching succeeds, then the SS examined is considered as a correct one. The set of ML's has been defined on the basis of the considered IP, but it has a more general validity because it enables the understanding of a large set of IS's which includes the IP as a small subset of its own. Furthermore, both the insertion of new ML's and the modification of the already existing ML's do not present excessive difficulty.

We will now discuss how the discrimination process, performed on the SS's by the Filter, produces the construction of the IL's that constitute the basis for the successive activity of extraction of the OI. During the semantic analysis, each BME which is encountered is utilized for further processing. In fact each BME that is obtained guarantees that the IS is correct, up to that point. Thus the BME contains an information that is very useful in order to build a command to the robot. Then the question arises of identifying which BME's are really of interest, and how to process the useful information that is associated with them. In order to solve these problems we have associated with each BME, that is encountered a property that is defined by the two following characteristics:

1. the position of the BME, namely the ME from which the BME has been derived;
2. the group of Italian language words to which the BME belongs (e.g., proper noun, common noun, couple of coordinate, etc.).

In correspondence with such properties, particular lists, namely the IL's, are associated. Each BME, which is associated with a particular property, is inserted in the corresponding IL. Intuitively, we can identify two categories of BME's. The first category provides information related to the part of the SS that is being considered. Such BME's (e.g., prepositions) give some information on which ML's have to be examined in the sequel of the discrimination process in order to control whether the considered SS is a correct one. The second category contains the information that is actually necessary in order to formalize a command to the robot. Only those BME's that belong to this second category are inserted in the IL's.

We can examine a simple example in order to illustrate the content of the IL's. Let us consider the following IS to be analyzed:

TAKE THE MODULE TI FROM-THE TABLE
[PRENDI IL MODULO TI DAL TAVOLO]

The correct SS, given by the PIAF system, is the following one:

(TAKE*(MODULE THE*TI) (FROM-THE*TABLE))

The correctness of this sentence is controlled by using the ML's as it has previously been illustrated. The processing of the semantic analysis will eventually provide a certain number of BME's that we will briefly discuss. MODULE and TI are BME's that both satisfy the same property of deriving from the same ME, OBJ, and of belonging to the group of substantives in the lexicon of the system. On the basis of this latter property, P_1, these two BME's are inserted in a corresponding IL, IL_1, which contains all the information concerning the object of the action. THE is a BME that satisfies a different property, P_2, since it derives from the same ME, OBJ, but it belongs to the group of articles in the lexicon of the system; therefore it is inserted in a different IL, IL_2.

When the BME TABLE is encountered, we can observe that it again satisfies the property of belonging to the group of substantives, like MODULE and TI. However, it has been derived from the ME PLACE, and therefore it is inserted in a new IL, IL_3, which contains all the information on the place of the action. The ME FROM-THE is a preposition and as has been previously specified, it is utilized only in order to more easily identify the ML's that are associated to the considered SS; it is inserted in a new IL, IL_4.

When we have completed the examination of the correctness of the SS, we have constructed the following IL's:

$$IL_1 = (\text{MODULE TI})$$
$$IL_2 = (\text{THE})$$
$$IL_3 = (\text{TABLE})$$
$$IL_4 = (\text{FROM-THE})$$

The two IL's, IL_1 and IL_2, are considered of particular interest for the construction of the command to the robot. As we can see, even from this simple example, the obtained information is still in a rather complex form, mainly because two BME's belong to the same IL_1.

In order to construct the command it is necessary to distinguish only one object of the action performed by the robot. Therefore, the system should be able to recognize the fact that TI is a particular module and, hence, that it contains all the information that is required for providing the command to the robot. This activity, which is executed by the Operative Information Extractor, is performed by utilizing a set of IS's; each IS is associated to each BME which is a substantive, and it distinguishes the content of information which is associated to such BME. In our example, to the two BME's TI and MODULE, we have associated the two following IS's:

$$\text{TI} \rightarrow \{\text{TI, MODULE, PIECE}\}$$

i.e., each TI is a module and is a piece;

$$\text{MODULE} \rightarrow \{\text{MODULE, PIECE}\}$$

i.e., each module is a piece.

By executing simple set inclusion operations on the IS's, it is possible to extract the OI on the basis of the IL's. More precisely, a binary relation of dominance $\prec$, defined on the set of BME's, is introduced. Given two BME's, x_1 and x_2, we say that $x_1 \prec x_2$ when the two corresponding IS's, IS_1 and IS_2, satisfy the relation $IS_1 \subset IS_2$. In this case we say that x_1 is dominated by x_2 and this assertion signifies that x_2 has a greater content of information than x_1. This dominance relation is iteratively applied to the BME's that belong to the same IL, by eliminating each time a BME that is dominated by another one. Thus, in our example, since MODULE is dominated by TI, we obtain the following new IL_1:

$$IL_1 = (\text{TI})$$

In order to verify the correctness of the OI that has been extracted from the IS, and in order to construct the order for the robot, the system must know the world in which the robot operates and its instantaneous situation. Therefore it is clear that the Legality Controller, which performs the check of the correctness of the OI, must be strictly dependent on the particular SD that has been chosen. In order to decide whether the OI is correct, we again start the processing activity by considering the verb contained in the IS. Each verb is associated with some ML's which characterize the order to the robot that is related to the use of that verb. In our example TAKE is related with the following ML:

(OBJ PLACE)

We try to construct this type of order with the use of the OI. If this can be done, this means that the OI is sufficient in order to formulate the command; in the opposite case, more detailed information will be requested. The control of the sufficiency of the OI does not necessarily complete the activity of the Legality Controller; in fact, the consistency of the OI with the instantaneous situation of the world also has to be checked. For instance, the following IS:

TAKE THE TI FROM-THE TABLE
[PRENDI LA TI DAL TAVOLO]

leads to the construction of the following OI:

(TAKE TI TABLE)

We also have to verify, because of the article THE, that on the table there is only one TI, in order that the order can be actually executed. In our case, then, the OI is legal and it becomes the EI ready for execution. Therefore, whenever the OI is controlled and is found to be sufficient and consistent, it becomes the EI that is communicated to the Formalizer which, by using the Robot Language, provides to the Executer, i.e., to the robot, the command for execution.

Let us consider, in conclusion, what happens when more than one SS is accepted by the Filter. In this case we can distinguish two cases:

1. All the accepted SS's lead to the construction of the same command.
2. Different SS's correspond to different commands.

In the first case we have an ambiguity only at the syntactic level, which is eliminated by successive semantic analysis. On the other hand, in the second case the ambiguity is also at the semantic level, i.e., the IS actually has more than one meaning. In this latter case the system asks for more information by providing an error message of semantic ambiguity. In a similar way when the OI examined by the Legality Controller either is not sufficient or is not consistent, an error message of operative illegality is provided.

4. TOWARD A DONAU VERSION OF DATABASE QUERY

The first version of DONAU that was developed was devoted to the programming of a robot for the manipulation of objects.[1,2] Subsequently, after the completion of the first DONAU version on robotics, the decision was made to develop a second DONAU version, oriented on a different semantic domain. In this way it was possible to control whether the architecture of DONAU, arranged in a modular way, was adapted, as desired when it was conceived, for its specialization on various semantic domains. Therefore, a new DONAU version, devoted to the semantic domain of querying databases, has been conceived and developed. More precisely, the query system has been constructed for the query in Italian, of the IIASA database, which contains information on the different types of energy resources in the world.

We will now discuss how it is possible to build a DONAU version dedicated to the query, in Italian, of a database. First, we must recognize that the problem of querying a database appears to be very similar to that of programming a robot. More precisely, the various phases of the NL understanding process are easily applicable to this new and very important application of NL understanding. Moreover, as has been illustrated in Sec. 2, the architecture of the DONAU system (see Fig. 1) is highly modular.

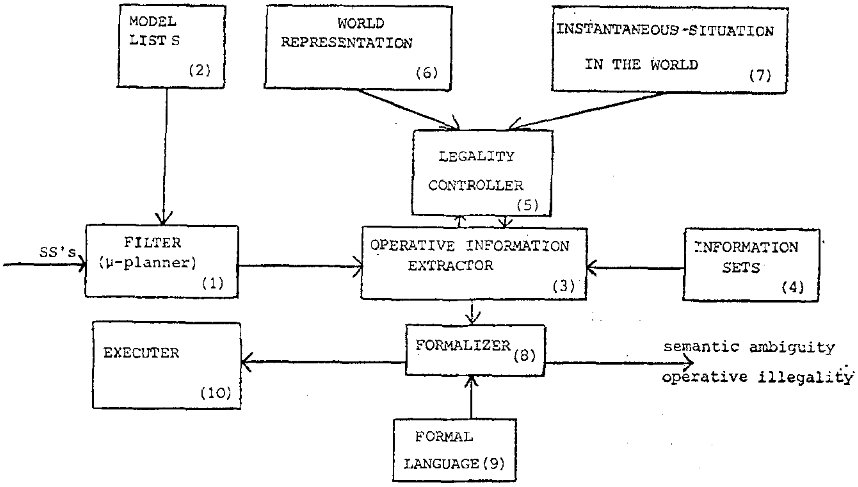


Fig. 1. Program architecture of the DONAU system.

We have clearly divided the blocks that are SD independent (blocks (1), (3), (5), and (8)) from the blocks that are SD dependent (blocks (2), (4), (6), (7), (9), and (10)). These six latter blocks need to be changed and the new ones, oriented toward the new SD of databases, have to be developed and inserted in their places. In this way a new DONAU version can be built by utilizing a conspicuous part of an old DONAU version. Both in the case of robotics and in the case of databases it is convenient to distinguish the IS's of an interaction in the following three classes: orders, questions, and descriptions.

Orders relate to the specification of the activity that the artificial system (either a robot or a database) has to perform (either an assembly operation or a data input/output). Questions refer to the request to the artificial system for knowledge about its characteristics and information about its structure and status, which are needed by the user for a better knowledge of its artificial behavior and available operating procedure. Descriptions indicate the specifications and modifications which the user applies to the structure and configuration of the artificial system.

As we illustrated in Sec. 2, the first step to be accomplished in order to develop the SD dependent blocks is the identification of a good IP. For this it is important to choose a set of typical IS's which are examples of common interaction. In our case these IS's should be selected from among the typical database queries that the new DONAU version should be able to deal with.

We will now give an example of IP oriented to database. This example is reduced to a minimal size, for providing understanding of just its main characteristics.

GIVE-ME THE PRODUCTION OF OIL OF-THE KUWAIT OF-THE 1975
[DAMMI LA PRODUZIONE DI PETROLIO DEL KUWAIT DEL 1975]

GIVE-ME THE PRODUCTION OF OIL OF-THE 1975
[DAMMI LA PRODUZIONE DI PETROLIO DEL 1975]

TELL-ME WHICH IS THE GREATEST PRODUCER OF OIL
[DIMMI QUALE E' IL MAGGIORE PRODUTTORE DI PETROLIO]

GIVE-ME THE PRODUCTION OF OIL OF-THE ARAB COUNTRIES
[DAMMI LA PRODUZIONE DI PETROLIO DEI PAESI ARABI]

TELL-ME HOW-MANY ARE THE PRODUCERS OF OIL
[DIMMI QUANTI SONO I PRODUTTORI DI PETROLIO]

The careful study of such IP, together with the analysis of the SS's that are provided by PIAF in correspondence with each IS of the IP, constitute the basis for the design of block (2), i.e., for the selection of the ML's that provide the semantic representation of the SD. The ME's that appear in the ML's related to the new SD of databases are almost the same ones that have been adopted in the ML's related to the old SD of robotics. On the other hand, the BME's of these new ML's are almost completely different from the ones considered in the old ML's.

We will now examine two IS's and we will illustrate how the DONAU version on databases will perform, on these two examples, the NL understanding process. Let us consider as examples of database queries, the following two IS's to be analyzed:

GIVE-ME THE PRODUCTION OF OIL OF-THE KUWAIT OF-THE 1975
[DAMMI LA PRODUZIONE DI PETROLIO DEL KUWAIT DEL 1975]

GIVE-ME THE PRODUCTION OF OIL OF-THE 1975
[DAMMI LA PRODUZIONE DI PETROLIO DEL 1975]

While the first query requests only the oil production of Kuwait in 1975, the second query asks for the oil production of the whole world in 1975. The correst SS, given by the PIAF system for the first IS, is the following one:

(GIVE-ME*(PRODUCTION THE*(OF*OIL) (OF-THE*(KUWAIT*(OF-THE*1975)))))

The first ML that is associated to GIVE-ME[DAMMI] is the following one:

(OBJ)

This ML indicates that the verb GIVE-ME must necessarily be associated with the information to be given. In fact, in the considered SS, the root of the tree (i.e., GIVE-ME) is followed by only one subtree, namely a right subtree, which corresponds to the only ME, OBJ, of the matched ML. Successively, the DONAU system selects the following ML's when it tries to match the considered SS with the hierarchical arrangement of the ML's. In a similar way the other ML's that are matched, in hierarchical way, to the considered SS are the following ones:

OBJ → (SUBS ARTDS SPEC PLACE-SPEC)
SUBS → (PRODUCTION)
ARTDS → (THE)
SPEC → (PREP SUBS)
PREP → (OF)
SUBS → (OIL)
PLACE-SPEC → (PREP PREP-SPEC)
PREP → (OF-THE)
PREP-SPEC → (SUBS TIME-SPEC)
SUBS → (KUWAIT)
TIME-SPEC → (PREP SUBS)
PREP → (OF-THE)
SUBS → (1975)

It is easy to control that the topology of the discrimination tree (i.e., the hierarchical organization of the previously indicated ML's) completely matches the topology of the SS (i.e., the corresponding syntactic tree, described in a parenthesized way, as illustrated in Sec. 2). In a similar way the correct SS given by the PIAF system for the second IS is the following one:

(GIVE-ME*(PRODUCTION-THE*(OF*OIL) (OF-THE*1975)))

Also in this case, as in the previous one, the first ML which is associated with GIVE-ME is the following one:

(OBJ)

Successively, the DONAU system selects the following ML's which completely match the SS when they are hierarchically arranged:

OBJ → (SUBS ARTDS SPEC TIME-SPEC)
SUBS → (PRODUCTION)
ARTDS → (THE)
SPEC → (PREP SUBS)
PREP → (OF)
SUBS → (OIL)
TIME-SPEC → (PREP SUBS)
PREP → (OF-THE)
SUBS → (1975)

At the end of the now-illustrated semantic analysis, the DONAU system constructs, as we described in Sec. 3, the IL's. In correspondence with the first IS, we have the following IL's:

IL_1 = (PRODUCTION OIL)
IL_2 = (KUWAIT)
IL_3 = (1975)

In correspondence with the second IS, we have the following IL's:

IL_1 = (PRODUCTION OIL)
IL_2 = (NIL)
IL_3 = (1975)

In the IL_1 it is clear that the BME OIL also contains the information expressed by PRODUCTION. Thus, the Operative Information Extractor, by applying (see Sec. 4) the following dominance relation:

$$\text{PRODUCTION} \prec \text{OIL}$$

will reduce the IL_1 to the following one:

IL_1 = (OIL)

Thus, in correspondence to the two considered IS's, the OI and, after the activity of the Legality Controller, the EI are provided to the Formalizer as the following two patterns of MICROPLANNER theorems:

(OIL KUWAIT 1975)
(OIL NIL 1975)

In this case we have made the assumption that the internal formal query language of the database is the MICROPLANNER language. Of course, if a different formal language is utilized, then the Formalizer will present the EI in its corresponding way.

5. CONCLUSIONS

We have illustrated the DONAU (Domain Oriented NAtural language Understanding) system which, in its two versions, has been devoted to robotics and to database querying. This is not a limitation, since the system was designed with a modular architecture that allows, in quite a simple way, the development of other DONAU versions dedicated to different applications.

In the first part of the paper we examined the robotics version that currently runs on the UNIVAC 1108 of the MP-AI Project. In the future we will connect the system with the SUPERSIGMA robot, which is now working in our Laboratory of Robotics. In the second part of the paper we discussed the development of a second DONAU version dedicated to the query of databases; an example of this type was illustrated in detail. We plan to use this new version in the near future and to test it on the IIASA database.

REFERENCES

1. M. Bernorio, M. Bertoni, A. Dabbene, and M. Somalvico, "Quasi Natural Language Understanding in the Semantic Domain of Robotics," Applied Robotics 77, Plsen, Czechoslovakia, October 1975.
2. M. Bernorio, M. Bertoni, A. Dabbene, and M. Somalvico, "Interazione Uomo Macchina in Linguaggio Quasi Naturale," *Proceedings Annual AICA Conference*, Milan, Italy, October 1976.
3. C. L. Bullwinkle, "Picnics, kittens and wigs: Using scenarios for the sentence completion task," 4th IJCAI, Tbilisi, USSR, August 1975.
4. J. Courtin, "Organization d'un Dictionnaire pour l'Analyse Morphologique," Seminaire de Theorie des Automates et Traitment Automatique des Langues, IRMA, Grenoble, France, 1973.
5. J. Courtin, "Un Système d'Analyse des Languages Naturelles: Application a la Correction Interactive de Textes," Thèse, Laboratoire de Informatique, Université de Grenoble, Grenoble, France, 1973.
6. A. Erchov, I. Mel'chuk, and A. Nariniany, "RITA. An experimental man-computer system on a natural language basis," 4th IJCAI, Tbilisi, USSR, August 1975.
7. G. Gini, M. Gini, and M. Somalvico, "Emergency recovery in intelligent industrial robots," 5th ISIR, Chicago, Illinois, September 1975.
8. E. Grandjean, "System PIAF Detecteur de Fautes d'Ortographie," Thèse, Université de Grenoble, Grenoble, France, 1974.
9. D. G. Hays, "Dependency Theory: A Formalism and Some Observation," Memo PM-4087-Ph, The Rand Corporation (1974).
10. C. Hewitt, "Procedural Embedding of Knowledge in Planner," Memo, AI Laboratory, MIT, Cambridge, Massachusetts (September 1971).
11. C. Hewitt, "How to use what you know," 4th IJCAI, Tbilisi, USSR, August 1975.
12. P. L. Miller, "An adaptive and natural language system that listens, asks, and learns," 4th IJCAI, Tbilisi, USSR, August 1975.
13. C. Schwind, "Generating hierarchical semantic networks from natural language discourse," 4th IJCAI, Tbilisi, USSR, August 1975.
14. G. W. Scragg, "Answering process questions." 4th IJCAI, Tbilisi, USSR, August 1975.
15. G. Veillon, "Modeles et Algorithmes pour la Traduction Automatique," These d'État, Université de Grenoble, Grenoble, France (1970).

16. T. Winograd, "Procedures as Representation of Knowledge in a Computer Program for Understanding Natural Language," MAC-TR-84, Project MAC, MIT, Cambridge, Massachusetts (1971).
17. A. Woods, "Transition network grammars for natural language analysis," *Comm. ACM* **13**:591–606 (1970).
18. S. Yoshida, "On the system of concepts relations and outline of the natural language systems," 4th IJCAI, Tbilisi, USSR, August 1975.

Printed in Belgium