Linguistic and Logical Tools for an Advanced Interactive Speech System in Spanish

Jordi Álvarez, Victoria Arranz, Núria Castell, and Montserrat Civit

TALP Research Centre, Universitat Politècnica de Catalunya,
Barcelona, Spain
{jalvarez,varranz,castell,civit}@talp.upc.es

Abstract. This paper focuses on the increasing need for a more natural and sophisticated human-machine interaction (HMI). The research here presented shows work on the development of a restricted-domain spontaneous speech dialogue system in Spanish. This human-machine interface is oriented towards a semantically restricted domain: Spanish railway information. The paper focuses on the description of the *understanding module*, which performs the language processing once the dialogue moves have been recognised and transcribed into text. Following the morphological, syntactic and semantic analysis, the module generates a structured representation with the content of the user's intervention. This representation is passed on to the *dialogue manager*, which generates the system's answer. The dialogue manager keeps the dialogue history and decides what the reaction of the system should be, expressed by a new structured representation. This is sent to the natural language generator, which then builds the sentence to be synthesised.

1 Introduction

The increasing need for telephone information systems allowing a communication that is as natural as possible for the user has boosted the interest and research on the topic. Work has already been done on the development of both application modules (cf. *TIDAISL* project [1]), or full applications for a specific functionality (cf. *IDAS* project [2]). At a higher level of elaboration and user-friendliness there are other systems that offer more interaction during the information exchange. Some examples of the latter systems, and also related to our domain of interest, are the following: *ARISE* (*Automatic Railway Information Systems for Europe*) [3], and *TRAINS* [4], both about railway information, and *ATIS* (*Air Travel Information System*) [5], about flights. Furthermore, the *TRINDI* (*Task-Oriented Instructional Dialogue*) [6] project should also be mentioned, which focuses on a more generic technology, i.e., multi-application and multi-language, for the creation of a dialogue movement engine.

The work described in this paper is part of the project "*Development of a Spontaneous Speech Dialogue System in a Semantically Restricted Domain*", which involves six research groups from several Spanish universities[1]. The main objective of

[1] Combining both Speech and NLP groups (cf. http://gps-tsc.upc.es/veu/basurde). This project is partially financed by the CICYT (TIC98-0423-C06) and by the CIRIT (1999SGR150).

L. Monostori, J. Váncza, and M. Ali (Eds.): IEA/AIE 2001, LNAI 2070, pp. 519-528, 2001.

this project is of a twofold nature: a) it aims at constructing a human-machine oral interface that allows the user obtain the necessary information about a train trip (and not only regarding timetables, as it happens in *ARISE*), while b) providing a relatively user-friendly communication exchange. The user should be capable of asking for a wide range of information, concerning both the data stored in the database (DB) or that within the dialogue history that has been created throughout the dialogue. Moreover, he should be able to correct any system's errors or demand for clarifications, always expressing himself by means of spontaneous natural language.

The paper has been structured as follows: section 2 describes the construction of the corpora to be used. Section 3 provides an overview of the whole system architecture, while sections 4 and 5 focus on the understanding module and dialogue manager, respectively.

2 Corpora Construction

The collection of dialogues in spontaneous speech from real users is an essential step towards the building of a spoken language dialogue system for real use [7]. When building this type of system, it is of major importance to study real user speech and language usage. Therefore, the data collected to build the corpus must comprise a set of examples which is representative enough of the type of queries to take place, as well as large enough to contain a rich variety of dialogue moves, turns, from both speakers. Further, the corpus must be built upon relatively simple sentences.

Given that there is no public corpus available of such characteristics for Spanish, its building has become an objective of the project in itself. In fact, the building of the corpus has consisted in the development of two different *corpora*: an initial *human-human* corpus [8], based on real conversations between users and personnel from a telephone railway information system; and a *human-machine* corpus [9] that has been obtained by means of the *Wizard of Oz* technique [10]. The former has been created as a reference for the initial study of the language involved in a dialogue system of this kind, and also as a guide to elaborate the scenarios upon which to base the latter. The latter is the corpus used for the main task of the research. Several instances have been created for every scenario type (based on a number of objectives and variants already defined). This has allowed us to establish a set of about 150 different situations. In addition, there is an open scenario that the informant specifies, thus obtaining 227 dialogues. Finally, this human-machine corpus is the one currently used as test bed for the understanding module and for the design of the dialogue manager.

3 System Architecture

In order to illustrate the system architecture, figure 1 shows its main components. As it can be observed, this architecture follows the standards of other such systems [7]. The initial stage is that of speech recognition, which translates the spoken utterance into a word sequence. In the present work, this is carried out by a speech recogniser that belongs to one of the project partners. However, despite its working rather well,

the recogniser needs to incorporate the treatment of spontaneous-speech extra-linguistic phenomena, such as pauses, hesitations, coughs, background noise, etc.

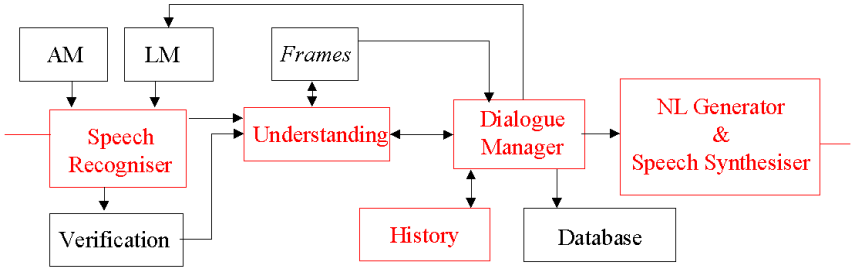

Fig. 1. System Architecture

Once the spoken utterance has been recognised and transcribed into written text, this is sent to the understanding module. This module is in charge of the linguistic processing of the text, its understanding and the generating of a formal representation (*frame*) with the extracted meaning that is to become the input of the dialogue manager module. The dialogue manager controls all the exchanges taking place within the dialogue, establishing, when necessary, the interactions with the dialogue history and the DB. Whatever information is extracted and has to be communicated to the user as the system's answer is passed on to the next module, the Natural Language (NL) generator, to be converted into sentences that the following and final module, the text-to-speech synthesiser, will produce as a spoken response for the user.

4 Understanding Module

This module performs the NLP tasks required to obtain the representation of the user's utterance. Figure 2 shows its architecture and below follows a detailed description.

4.1 Input to the Module

Ideally, the recogniser should provide the correct orthographic transcription for the user's utterances. Nevertheless, erroneous transcriptions also take place, which is something expected when working with spontaneous speech. The recogniser's errors the understanding module has to face can be basically classified into three types:

- Excess of information in the recognition, providing words that do not represent any part of the user's utterance. For example:
  - *user*: "sábado treinta de octubre" (Saturday 30 October).
  - *recogniser*: "**un tren que o** sábado treinta de octubre" (**a train that or**...).
- Erroneous recognition: recognised words do not match those uttered by the user, such as:
  - *user*: "gracias" (thank you).
  - *recogniser*: "sí pero ellos" (yes but they).

- Grammar errors in a broad sense, such as:
  - Lack of preposition+determiner contractions: "*de el*", instead of "*del*".
  - Misuse of the indefinite article instead of the cardinal: "***un*** *de octubre*", while it should be "***uno*** *de octubre*" (1 October).
  - Erroneous orthographic transcriptions causing changes in grammatical categories: "*qué/que*" (what/that), "*a/ha*" (to/has), "*e/he*" (and/have), etc.

Further to the recogniser's errors, there are also transcription problems that are due to the use of spontaneous speech, such as the following:

- Syntactic disfluencies: for instance, the syntactically incorrect sentence caused by the repetition of information:
  - *user*: "a ver los horarios de los trenes que van de Teruel a Barcelona **el este** próximo viernes y que vayan de Barcelona a Teruel **el próximo que vuelvan de Barcelona a Teruel el próximo** domingo".
- Other disfluencies: lexical variations, pauses, noises, etc.

As it can be expected, all these problems become added difficulties for the understanding module. The solutions adopted to deal with them, even if only partially, are mainly the following three:

- Adapting the recogniser to the task domain.
- Adapting, within their limitations, the understanding module components (cf. section 4.3) so as to increase their robustness when facing such problems.
- The possibility of closing the entry channel when the recogniser believes it relevant is also being considered.

Despite all these difficulties, the understanding module will have to be capable of providing a representation of the user's query, even if not a complete one.

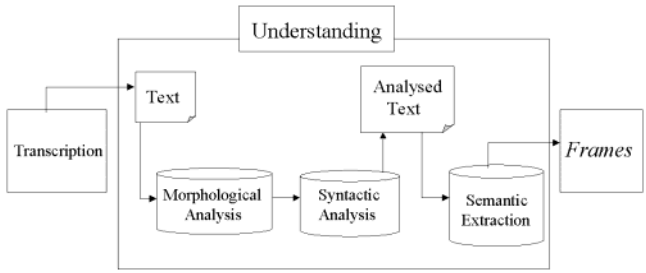

Fig. 2. Understanding Module

4.2 Morphological Processing

Once the recogniser's task is finished and thus having the utterance transcription, the morphological analysis of the text is performed. Then, its output undergoes a

syntactic analysis (cf. section 4.3), which precedes the generation of semantic *frames* (cf. section 4.4) to be sent to the dialogue manager (cf. section 5).

The morphological analysis is carried out by means of MACO+ (*Morphological Analyzer Corpus Oriented* [11]), which has been adapted for the task domain. The linguistic knowledge employed by this tool is organised into classes and inflection paradigms. Further, forms are considered from an orthographic point of view. Adapting this tool to the domain allows to reduce partially the ambiguity that would be generated by a general language analyser. For example, eliminating verb *salgar*[2] from the lexicon allows the analyser to select, without any ambiguity, verb *salir* when encountering verb form *salgo*. In addition, the fact of working with a smaller lexicon (and with a restricted grammar within the syntactic analyser) does also help to reduce the system execution time, which is of major importance for a dialogue system of these characteristics.

The result of this morphological analyser is a set of possible morphological labels per word with their corresponding lemma. Therefore, a disambiguation process is required, which is performed by RELAX (*Relaxation Labelling Based Tagger* [12]), selecting the appropriate label among all those provided. The set of labels employed is based on the *EAGLES* guidelines [13]. Figure 3 shows the morphological (already disambiguated) analysis of the sample sentence: *Me gustaría información sobre trenes de Guadalajara a Cáceres para la primera semana de agosto*[3]:

me yo PP1CSO00
gustaría gustar VMCP1S0
información información NCFS000
sobre sobre SPS00
trenes tren NCMP000
de de SPS00
Guadalajara guadalajara NP000C0
a a SPS00
Cáceres cáceres NP000C0
para para SPS00
la la TDFS0
primera primero MOFS00
semana semana NCFS000
de de SPS00
agosto agosto NCMS000
. . Fp

Fig. 3. Morphological Analysis and Disambiguation

This particular application has required some slight tag adaptations. This has been done so as to identify within the tag itself some domain-specific terms referring to both *city* and *station* names. The adaptations have consisted in using digit #6 in the morphological label (that of semantic case) to mark them as “C” and “E”[4].

4.3 Syntactic Analysis

Further on the linguistic processing, the morphologically analysed and disambiguated sentence is then syntactically processed by the shallow parser TACAT [14]. TACAT's output is a series of phrasal groupings with no internal dependency made explicit.

---

[2] *Salgar* means "give or put salt to/on" and it does not belong to the task domain, but it shares some basic conjugated verb forms with *salir* ("to leave/exit"), such as *salgo*.

[3] I would like some information about trains from Guadalajara to Cáceres for the first week of August.

[4] "C" stands for *ciudad* (city) and "E" for *estación* (station).

Generally, TACAT works with a context-free grammar for unrestricted domains. Here, though, it has been adapted so as to treat our domain. The adaptations applied have basically consisted in re-writing some grammar rules in order to incorporate lexical information. This avoids certain syntactic ambiguity and speeds up the process. The main rules affected are those referring to dates, timetables and proper names (covering both cities and stations). The rules for timetables are the following:

```
sn-h ==> j-fp, grup-nom-h.                                          %las doce horas
grup-nom-h ==> numer-fp, n-fp(horas).                               %doce horas
grup-nom-h ==> grup-nom-h, coord(y), grup-nom-m.                    %doce horas y diez minutos
grup-nom-m ==> numer-mp, n-mp(minutos).                             %diez minutos
grup-nom-m ==> numer-mp, coord(y), numer-mp, n-mp(minutos).%treinta y cinco minutos
```

By means of these rules, the referred information is propagated towards the high nodes in the analysis tree. This helps the semantic searches performed with PRE+ (cf. section 4.4) to be more immediate and direct. Nevertheless, some rules covering structures that do not occur in this domain have been removed for this particular application. Figure 4 shows the syntactic analysis for the sample sentence above.

```
[ { pos=>S }
[ { pos=>patons } [ { pos=>pp1cso00 , forma=>"Me" , lema=>"yo" } ] ]
[ { pos=>grup-verb }
                [ { pos=>vmcp3s0 , forma=>"gustaría" , lema=>"gustar" } ] ]
[ { pos=>sn }
                [ { pos=>ncfs000 , forma=>"información" , lema=>"información" } ] ]
[ { pos=>grup-sp }
                [ { pos=>sps00 , forma=>"sobre" , lema=>"sobre" } ]
                [ { pos=>sn }
                [ { pos=>ncmp000 , forma=>"trenes" , lema=>"tren" } ] ] ]
[ { pos=>grup-sp }
                [ { pos=>sps00 , forma=>"de" , lema=>"de" } ]
                [ { pos=>sn }
                [ { pos=>np000c0 , forma=> "Guadalajara" , lema=> " Guadalajara" } ] ] ]
[ { pos=>grup-sp }
                [ { pos=>sps00 , forma=>"a" , lema=>"a" } ]
                [ { pos=>sn }
                [ { pos=>np000c0 , forma=>"Cáceres" , lema=> " Cáceres" } ] ] ]
[ { pos=>grup-sp }
                [ { pos=>sps00 , forma=>"para" , lema=>"para" } ]
                [ { pos=>sn }
                [ { pos=>tdfs0 , forma=>"la" , lema=>"la" } ]
                [ { pos=>mofs00 , forma=>"primera" , lema=>"primero" } ]
                [ { pos=>ncfs000 , forma=>"semana" , lema=>"semana" } ] ] ]
[ { pos=>grup-sp }
                [ { pos=>sps00 , forma=>"de" , lema=>"de" } ]
                [ { pos=>sn }
                [ { pos=>ncms000 , forma=>"agosto" , lema=>"agosto" } ] ] ]
[ { pos=>punto }
                [ { pos=>Fp , forma=>"." , lema=>"punt" } ] ] ]
```

Fig. 4. Syntactic Analisis

4.4 Semantic Extraction

Previous to the semantic extraction stage, a thorough study has been carried out of a representative part of the dialogue corpus. The aim of this study is to define the type

of restricted information that should be considered. The semantic representation to be generated is based on the concept of *frame*, which can be easily translated as a query to the DB. The concept of *frame* has been previously used [15] and it functions as a summary of the dialogue turn. This implies that for every user turn that is sent to the understanding module a *frame* will be generated, i.e., a pattern holding all extracted semantic information. Moreover, a *frame* can have two different types of information:

1. *Concepts*: specific information the user is enquiring about.
2. *Cases*: restrictions applied to the concepts.

In order to clarify these classification concepts, figure 5 offers a specific *frame*. The sample sentence in section 4.2 becomes a concept (query) about a train departure time (*Hora-Salida*) and the restrictions applied are: 1) departure city (*case Ciudad-Origen*) is *Guadalajara*, 2) destination city (*case Ciudad-Destino*) is Cáceres, and 3) departure date interval (*case Intervalo-Fecha-Salida*) is the first week of August.

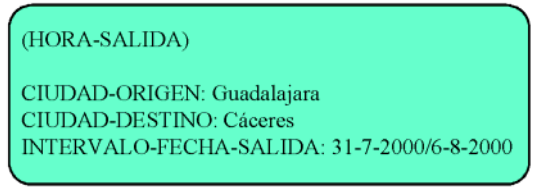

Fig. 5. Example of *Frame*

Once this has been established, the going from syntax to semantic *frames* is performed by means of a semantic extraction system implemented in PRE+ [16]. PRE+ is a production rule environment that runs on PERL and allows us to declare rules for information extraction. The formalisation of these rules requires the morpho-syntactic representation of the text, together with the study of the *markers* (different ways) used to express a query, confirmation, correction, etc. Figure 6 shows an example of the diversity of expressions used to ask for information.

Querría saber...	Me gustaría saber...
Querría pedir...	Podría decirme...
Quisiera saber...	Me gustaría que me dijera...
Quisiera obtener...	Desearía saber...
Quería información...	Deseo información sobre...
Me gustaría información...	Llamaba para saber...

Fig. 6. Query Markers

PRE+ establishes conditions and actions. The former contain the syntactic patterns and lexical items to search for in the word strings. The latter represent the extraction method for each pattern. The way a specific PRE+ rule functions could be paraphrased as follows: “if in a syntactic tree there is a prepositional phrase, *grup-sp*, with a daughter node whose terminal item has lemma *de* (of/from) or *desde* (from), and another daughter node containing the name of a city, *np000c0* (cf. section 4.2),

this proper name represents the *Ciudad-Origen case* (restriction) to be extracted for the query". Figure 7 shows the PRE+ rule used to extract such information.

```
(rule CiudadOrigen3
      ruleset CiudadOrigen
      priority 10
      score [0,_,1,0]
      control forever
      ending Postrule
      (InputSentence ^tree <+a>tree_matching(
            [{pos=>grup-sp}
                  [{lema=> de|desde}]
                  [{pos=> np000c0, forma=>?forma}]
            ]))
      ->
      (?_ := Print(CiudadOrigen,?forma))
      (?_ := REM(CiudadOrigen,X,+a)))
```

Fig. 7. Example of PRE+ Rule for *CiudadOrigen*

Further to the conditions and actions, this type of rules allow to establish explicitly the way in which they should be applied (in a static/dynamic manner, which priority and control should they have, etc.) and the location of the concept to be extracted within the domain hierarchy. Given both the speech recogniser's problems and those caused by spontaneous speech (cf. section 4.1), semantic information is approached as locally as possible. This implies a search for and extraction of information from the lowest nodes of the syntactic tree, by means of phrases and lexical items.

5 Dialogue Manager

This is one of the main components of the system since it directs the system behaviour. The interpretation and administration of the content in a user's utterance would not be possible without this module's supervision, and neither would the co-reference resolution. This module must decide which the system reaction is: inquiring about information in order to complete a query to the DB, asking for repetitions, offering information, etc.

The dialogue manager is implemented using YAYA [17], which is a terminological reasoning system. The dialogue manager strategy is expressed in a declarative manner by means of a set of axioms. The reasoning engine combines both a) the facts (*frames*) generated by the understanding module, and b) the facts that represent the dialogue history, with the axioms that represent the strategy, in order to generate the facts that represent the system reaction (this includes, if possible, access to the DB).

Below follows an example of a concept axiom, where the user has not specified the departure date for a trip and thus the system needs to enquire about this information:

```
(:and informacion-usuario (:not (:some fsalida :top)))  ≤
   (:some (:inv rdominio) (:and speech-act-usuario (:some rrespuesta pregunta-fecha-salida)))
```

The dialogue manager output is also based on the concept of *frame*. The NL generator receives a *frame* comprising all the information required to build the system answer. Then, the sentence is synthesised by the text-to-speech translator.

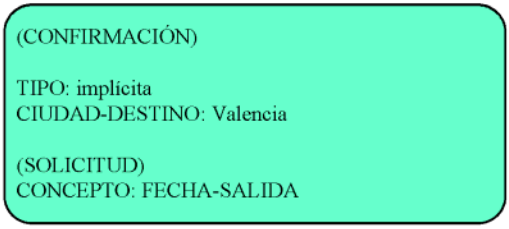

Fig. 8. Dialogue Manager Output *Frame*

For example, the NL generator can build the following system answers: "*¿Cuándo desea viajar a Valencia?*", *A Valencia, ¿cuándo desea viajar?*"[5] (cf. fig. 8), based on the *frames* received from the dialogue manager in a hypothetical situation where the date of travelling has not been specified by the user.

The dialogue manager will also inform the recogniser about the type of *speech act* that is expected in the next user turn. In order to generate such predictions, the dialogue manager has a speech act grammar at its disposal. This grammar has been designed by studying the available corpora and evaluating the possible types of speech act. The types of speech act already expected from the user are: *consulta* (with several subtypes depending on the aim of the query), *falta de comprensión, confirmación, afirmación, negación* and *cierre*[6].

6 Conclusions

This paper aims to present the work done on the development of a spontaneous-speech dialogue system for a semantically restricted domain, such as a railway information system for Spanish.

The present article focuses on the understanding and dialogue manager modules, although the remaining modules are also developed by now. However, no evaluation of the whole system can be provided for the time being since the system integration is currently taking place and its testing is planned within the next few months. In spite of this, and at a sub-module level, some conclusions have already been drawn: the morphosyntactic analysis tools (which already contain a lexicon and grammar adapted for the task) have already proved their efficiency during corpus analysis. Furthermore, the production rule environment PRE+ used for the semantic extraction has shown to be appropriate for the task and robust in an information extraction environment. In addition, the use of a declarative tool to implement the dialogue manager allows the development of the module prototype in a relatively short time and it also allows easy

[5] "When do you wish to travel to Valencia? ", "To Valencia, when do you wish to travel?"

[6] *Query, not understood, confirmation, affirmation, negation* and *closing*.

modifiability. Therefore, once the evaluation of the whole system has been done it will be decided whether its translation to a conventional tool is required.

Last but not least, the importance and cost of the corpora construction and analysis tasks should also be emphasised. In fact, one of the project objectives was the development of a Spanish speech corpus of such characteristics. Once tagged and standardised, this corpus can become a valuable resource for future work.

References

1. Ferreiros, J., Macías-Guarasa, J., Gallardo, A., Colás, J., Córdoba, R., Pardo, J.M., Villarrubia, L.: Recent Work on Preselection Module for Flexible Large Vocabulary Speech Recognition System in Telephone Environment. In *Proceedings of ICSLP'98*, Sidney (1998)
2. San-Segundo, R., Colás, J., Montero, J.M., Córdoba, R., Ferreiros, J., Macías-Guarasa, J., Gallardo, A., Gutiérrez, J.M., Pastor, J., Pardo, J.M.: Servidores Vocales Interactivos: Desarrollo de un Servicio de Páginas Blancas por Teléfono con Reconocimiento de Voz (Proyecto IDAS: Interactive Telephone-Based Directory Assistance Service). In *IX Jornadas Telecom I+D*, Barcelona-Madrid (1999)
3. Lamel, L., Rosset, S., Gauvain, J.L., Bennacef, S.: The Limsi Arise System for Train Travel Information. In *Proceedings of ICASSP'99* (1999)
4. Allen, J.F., Miller, B.W., Ringger, E.K., Sikorski, T.: A Robust System for Natural Spoken Dialogue. In *Proceedings of ACL'96* (1996)
5. Cohen, M., Rivlin, Z., Bratt, H.: Speech Recognition in the ATIS Domain Using Multiple Knowledge Sources. In *Proceedings of the ARPA Spoken Language Systems Technology Workshop*, Texas (1995)
6. *TRINDI* project: http://www.linglink.lu/le/projects/trindi.
7. Giachin, E., McGlashan, S.: Spoken Language Dialogue Systems. In: Young, S., Bloothooft, G. (eds.): *Corpus-Based Methods in Language and Speech Processing*. Kluwer Academic Publishers (1997)
8. Bonafonte, A., Mayol, N.: Documentación del corpus INFOTREN-PERSONA. Project Report BS14AV20, UPC, Barcelona (1999)
9. Sesma, A., Mariño, J.B., Esquerra, I., Padrell, J.: Estrategia del Mago de Oz. Project Report BS52AV22, UPC, Barcelona (1999)
10. Fraser, N.M., Gilbert, G.N.: Simulating Speech Systems. *Computer Speech and Language*, Vol. 5 (1) (1991)
11. Carmona, J., Cervell, S., Màrquez, L., Martí, M.A., Padró, L., Placer, R., Rodríguez, H., Taulé, M., Turmo, J.: An Environment for Morphosyntactic Processing of Unrestricted Spanish Text. In *Proceedings of LREC'98*, Granada (1998)
12. Padró, L.: *A Hybrid Environment for Syntax-Semantic Tagging.* PhD Thesis, UPC, Barcelona (1997)
13. *EAGLES* group: http://www.ilc.pi.cnr.it/EAGLES96/home.html.
14. Castellón, I., Civit, M., Atserias, J.: Syntactic Parsing of the Unrestricted Spanish Text. In *Proceedings of LREC'98*, Granada (1998)
15. Minker, W., Bennacef, S., Gauvain, J.L.: A Stochastic Case Frame for Natural Language Understanding. In *Proceedings of ICSLP'97* (1997)
16. Turmo, J.: PRE+: A Production Rule Environment. Working Report LSI-99-5-T, UPC, Barcelona (1999)
17. Alvarez, J.: The YAYA Description Logics System: Formal Definition and Implementation. Research Report LSI-00-40-R, UPC, Barcelona (2000)