TALP-QA System for Spanish at CLEF-2004

Alicia Ageno, Daniel Ferrés, Edgar González, Samir Kanaan, Horacio Rodríguez, Mihai Surdeanu, and Jordi Turmo

TALP Research Center
Universitat Politècnica de Catalunya
{*ageno,dferres,egonzalez,skanaan,horacio,surdeanu,turmo*}*@lsi.upc.es*

Abstract

This paper describes TALP-QA, a multilingual open-domain Question Answering (QA) system that we have been developing during the last two years. The paper analyzes as well our participation in the CLEF-2004 Spanish monolingual QA task.

1 Introduction

This paper describes TALP-QA, a multilingual open-domain Question Answering (QA) system that we have been developing during the last two years. The paper analyzes as well our participation in the CLEF-2004 contest. Our aim in developing TALP-QA has been building a system as much as possible language independent, where language dependent modules could be substituted for allowing the system to be applied to different languages. A first preliminary version of TALP-QA for English was used for participating in TREC-2003 QA track (see [Massot et al, 2003]). From this initial version, current one, for Spanish, was built and has been used in CLEF-2004. An improved version, again for English, is planned to be used for TREC-2004.

In this paper we present the overall architecture of TALP-QA, we describe briefly its main parts, focusing on the components that have been changed more in depth from our initial prototype, and on the components involving processing of the Spanish language. We present as well a preliminary evaluation of the system presented in the CLEF-2004 evaluation for both factoid and definition questions.

2 System Description

2.1 Overview

The system architecture follows the most commonly used schema, splitting the process into three phases that are performed in turn. Several iterations can be carried out in order to achieve their goals but once one phase is finished there is no possibility to return to previous ones. There are three main subsystems (as shown in Figure 1), one corresponding to each phase:

1. Question processing (QP)
2. Passage retrieval (PR)
3. Answer extraction (AE)

These subsystems are described below. Previously we will describe some pre-processing tasks that have been carried out on the document collection (the EFE corpus in this case). As pointed out above, our aim is to get a highly language independent system. Language dependent components are included only within the Question Pre-processing and Passage Pre-processing components.

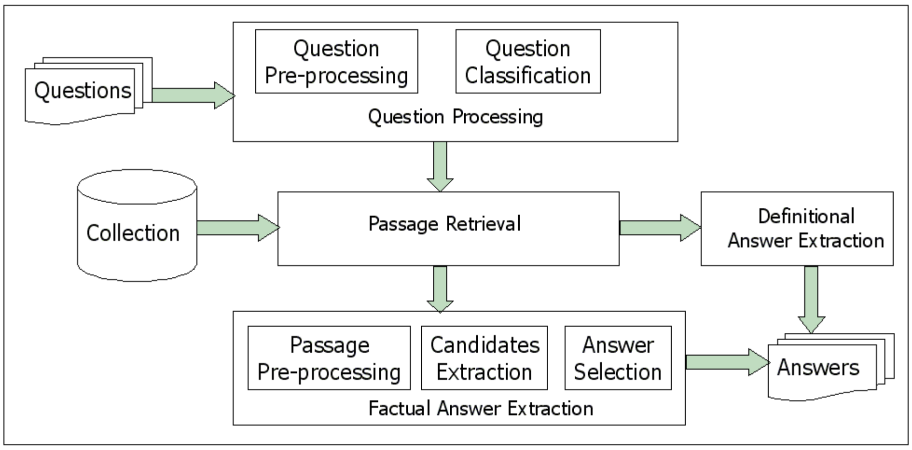

Figure 1: Architecture of TALP-QA system.

2.2 Collection Pre-processing

In order to perform the PR task we have used as Information Retrieval (IR) engine *Lucene*[1]. Before CLEF 2004 we indexed the whole EFE collections EFE 1994 and EFE 1995 (i.e. 454,045 documents). We pre-processed the whole collections with linguistic tools (described in next subsection) to obtain the part-of-speech (POS), lemmas and Named Entities (NE) of the text. This information has been used to built an index with the two aligned following parts:

- Lemmatized and NE recognized text: this part is built using the lemmas of the words and the results of the NE recognition module. This part is indexed and used in the PR module.
- Original text with NE recognition: the original text that is retrieved when a query succeeds to the lemmatized text.

As an additional knowledge source that will be used in the AE task, a tf*idf weight (at document level on the whole collection) is computed for all the words occurring in the collection.

2.3 Question Processing

The main goal of this subsystem is to classify the question regarding the kind of expected answer and to attach to it the information needed for the following subsystems. For PR the information needed is basically lexical (POS and lemmas) and syntactic, and for AE lexical, syntactic and semantic. We have tried to represent all these kinds of information using a language independent formalism. In particular we use the same semantic primitives and relations for the two languages (English and Spanish) involved in our system.

For CLEF 2004 (for Spanish) we have used a set of general purpose tools of NLP group of the UPC (see [Carreras et al, 2004] and [Atserias et al, 1998]). The same tools are used for the linguistic processing of both the questions and the passages. The question is analyzed with a pipe including the following processors:

- **FreeLing**, that performs tokenization, morphological analysis (including identification of quantities, dates, multiword terms, etc.), POS tagging and lemmatization.
- **Tacat**, a partial parser that obtains shallow nominal, prepositional and verbal phrases.
- **ABIONET**, a Named Entity Recognizer and Classifier that identifies and classifies NEs in basic categories (person, place, organization and other). See [Carreras et al, 2002].

[1] http://jakarta.apache.org/lucene

- **EuroWordNet**, we obtain and attach semantic information using *EWN*[2]: a list of synsets (with no attempt to Word Sense Disambiguation), a list of hyperonyms of each synset (up to the top of each hyperonymy chain), the EWN's Top Concept Ontology (TCO) class [Rodríguez et al, 1998], and the Magnini's Domain Codes (DC) [Magnini, Cavagliá, 2000].

- Gazetteers, we use three gazetteers with three types of information: acronyms, obtained using a Decision Tree approach [Ferrés et al, 2004], relations location-gentile (España-español, Spain-Spanish) and relations actor-action (escribir-escritor, write-writer).

The application of these linguistic resources and tools, obviously language dependent, to the text of the question is represented in two structures (an example is presented in Figure 2):

- **Sent**, that provides us with information for each lexical unit: the word form, the lemma, the POS (an Eagles compliant rich tagset was used), the semantic class of NE, the list of EWN synsets and, finally, whenever possible the verbs associated to the actor and the relations between locations and their gentile.

- **Sint**, composed by two lists, one recording the syntactic constituent structure of the question (basically nominal, prepositional and verbal phrases) and the other collecting the information of dependencies and other relations between these components.

QUESTION: *¿En qué país está Essen?*
(In which country is Essen located?)

QUESTION TYPE *where_location*

```
sent(...,
[...
(word("país"),lema("país"),p
os("NCMS000"),nil,nil,[n0520
9013,...]),
...
(word("Essen"),lema("Essen")
,pos("NP00G00"),location,nil
,[n05512526]),
...]
).
```

SENT

```
sint(...,[
  tk(1),
  ch(2,4,grup_sp,4,[
    ch(2,2,prep,2,[tk(2)]),
    ch(3,4,sn,4,[ch(3,3,espec_ms,3,[tk(3)]),
    ch(4,4,grup_nom_ms,4,[tk(4)])])]),
  ch(5,5,grup_verb,5,[tk(5)]),
  ch(6,6,sn,6,[ch(6,6,grup_nom_fp,6,[tk(6)])]),
  tk(7)],

[r(obj,3,4),r(subj,3,2)]).
```

SINT

```
env(... ,[
  state(5),
  i_en_proper_place(6),
  theme_of_event(5,6),
  participant_in_event(5,4),
  prep(4,2),entity(4),det(4,3),qu(3)]).
```

ENVIRONMENT

```
MC: [theme_of_event(5,6),state(5),
      entity(4),i_en_proper_place(6)]

OC: [type_of_location(5,5,i_en_state),
     participant_in_event(5,4),
     prep(4,2)]
```

SEMANTIC CONSTRAINTS

Figure 2: Results of the pre-process of a question.

Once this information is obtained we can get the information relevant to the following tasks:

- **Question type**. The most important information we need to extract from the question text is the Question Type (QT), because all the work the system has to perform for searching the answer is based on this issue. A failure on identifying QT practically disables the correct extraction of the answer. Currently we are working with about 25 QT (we have reduced the initial number of categories used in TREC-2003 for improving the accuracy of our classifier). The QT tries to focus the type of the expected answer providing as well additional constraints on it. For instance, when the expected type of the answer is a person, two types of questions are considered, *Who_action*, which indicates that we are

[2]EuroWordNet. http://www.hum.uva.nl/~ewn

looking for a person who performs a certain action and *Who_person_quality*, that indicates that we are looking for a person having the desired quality. The action and the quality are the parameter of the corresponding QT. The following are examples of questions correctly classified respectively as *Who_person_quality* and *Who_action* type:

- *¿Quién fue jefe del XII Gobierno de Israel? (Who was the head of the XII Israel government?)*
- *¿Quién ganó el Premio Nobel de Literatura en 1994? (Who won the Nobel Prize in Literature in 1994?)*

In order to determine the QT our system uses an Inductive Logic Programming (ILP) learner that learns, from a set of positive and negative examples, a set of weighted rules. We have used as learner the FOIL system [Quinlan, 1993]. A binary classifier (i.e. a set of rules) has been learned for each QT. As training set we have used the set of questions of TREC 8 and 9 (~900 questions) manually tagged and as test set the 500 questions of TREC 11. All these questions were previously manually translated into Spanish. For each classifier we have used as negative examples the questions belonging to the other classes. As features for classifying the following have been used: Word form, Word position in the question, Lemma, POS, Semantic class of NE (without sub-classing), Synsets together with all their hyperonyms, TCO, DC and subject and object relations.

The set of rules for each class has been manually revised and completed with a set of manually built rules (with lower weights) in order to assure a greater coverage. See below a couple of such rules:

- A learned rule:

```
regla(non_human_actor_of_action,A,1) :-
  first_position(A,B),
  next_position(B,C),
  is_tco(cObject,C),
  is_domain(dTransport,C).
```

- The same rule after transformation (performed for the sake of efficiency):

```
regla(non_human_actor_of_action,A,1,[],TT) :-
  sent(A,_,TT), TT=[_,W2|_],
  has_tco(W2,cObject),
  has_domain(W2,dTransport).
```

- A manual rule:

```
regla(non_human_actor_of_action,A,994,[T1,T3],T) :-
  sent(A,_,[T1|T]),
  the_lema(T1,lema("qué")),
  has_chunk_with_hyperonym(_,T,[T2|TT],
    [sArtifact,sObject,sAnimal],T3),
  the_pos(T2,pos("SP")),
  not(has_term_with_pos(TT,pos("AQ"),_)).
```

- **Environment.** The semantic process starts with the extraction of the semantic relations that hold between the different components identified in the question text. These relations are organized into an ontology of about 100 semantic classes and 25 relations (mostly binary) between them. Both classes and relations are related by taxonomic links. The ontology tries to reflect what is needed for an appropriate representation of the semantic environment of the question (and the expected answer). For instance, *Action* is a class and *Human_action* is another class related to *Action* by means of an *is_a* relation. In the same way, *Human* is a subclass of *Entity*. *Actor_of_action* is a binary relation (between a *Human_action* and a *Human*). When a question is classified as *Who_action* an instance of the class *Human_action* has to be located in the question text and its referent is stored. Later, in the AE phase, an instance of *Human_action* co-referring with the one previously stored has to be located in the selected passages and an instance of *Human* related to it by means of the *Actor_of_action* relation must be extracted as a candidate to be the answer.

The environment of the question is obtained from the *sint* and the semantic information included in *sent*. For performing this task a set of about 150 rules has been built. The environment extracted from a question is presented in Figure 2.

- **Semantic Constraints.** The environment tries to represent the whole semantic content of the question. Not all the items belonging to the environment are useful, however, for extracting the answer. So, depending on the QT, a subset of the environment has to be extracted. Sometimes additional relations, not present in the environment, are used and sometimes the relations extracted from the environment are extended, refined or modified. We define in this way the set of relations (the semantic constraints) that are supposed to be found in the answer. These relations are classified as mandatory, Mandatory Constraints (MC), (i.e. they have to be satisfied in the passage) or optional, Optional Constraints (OC), (if satisfied the score of the answer is higher). In order to build the semantic constraints for each question a set of rules (typically 1 or 2 for each type of question) has been manually built. A fragment of the rule applied in the example is presented in figure 3. The rule can be paraphrased as follows: If the relation *state(C)* holds in the environment, get recursively all the predicates related to C, then filter the appropriate ones to be included in MC and OC and finally extend these sets for the sake of completeness. The application of the rule results in the constraints shown in figure 2.

```
mandatory(Q,ManIni,CRC,ManFi2,OptFi3,where_location,1) :-
  …
  state(C,Q,Env),
  …
  get_related_tokens_in_env_rec(C,Env,SS,[qu]),
  filter_tuple_tokens(SS,Manx1,_,Opt1,
          [theme_of_event,time_of_event,location_of_event,which_entity],
          []),
  …
  filter_related_tokens(SS,
          [
                    [human_participant_in_event(C,_X)],
                    [participant_in_event(C,_X), i_en_proper_person(_X)],
                    [participant_in_event(C,_X), i_en_proper_organization(_X)],
                    [participant_in_event(C,_X), i_en_proper_named_entity(_X)]
          ],
          Manrel),
  …
  extend_mandatory(ManFi,Opt,V4,V2,ManFi1,OptFi,Q,Env,Env),
  extend_mandatory_1(ManFi1,OptFi,V1,V2,ManFi2,OptFi2,V11,V22,Q,SS1,Env),
  …
```

Figure 3: Semantic constraints of a question.

2.4 Passage Retrieval

This subsystem creates and retrieves dynamic passages using an iterative algorithm. The input of this algorithm is the information obtained in the question processing subsystem: *sint*, information of *sent* (part-of-speech and lemmas) and question classification. The goal of this algorithm is to extract a set of passages, where at least one passage contains the answer for the input question. Intuitively the algorithm relaxes the query context (defined by the set of keywords to be used and a keyword proximity value) if too few passages are found, and uses a stricter context (e.g. more keywords and/or closer proximity) if too many passages are extracted. From each query a priority is assigned to each non-stop question keyword, the best-scored keywords are selected to compose a boolean query to the IR system. We retrieve the top scored documents from Lucene, and we extract a set of passages per query in factual questions.

2.5 Factual Answer Extraction

After PR, for factual AE, two tasks are performed in sequence: Candidates Extraction (CE) and Answer Selection (AS). In the former, all the candidate answers are extracted from the best scored sentences of the selected passages. In the latter the best answer is chosen.

- **Candidates Extraction**. This process is carried out on the set of passages obtained from the previous subsystem. These passages are segmented into sentences. Each sentence is then scored according to its semantic content regarding the question using the tf*idf weighting of the terms from the question and their taxonomic neighbours occurring in the sentence, we will name semantic score this figure. See [Massot et al, 2003] for details.

  The linguistic process of extraction is similar to the process carried out on questions and leads to the construction of the environment of each candidate sentence. The rest is a mapping between the semantic relations contained in this environment and the Semantic Constraints extracted from the question. The mandatory restrictions must be satisfied for the sentence to be taken into consideration; the satisfaction of the optional constraints simply increases the score of the candidate. The final extraction process is carried out on the sentences satisfying this filter.

  The Knowledge Source used for this process is a set of extraction rules owning a credibility score. Each QT has its own subset of extraction rules that leads to the selection of the answer. An example of extraction rule is presented in figure 4. The rule can be paraphrased as follows: Look in MC for predicates *state(C)* and *location(X)* satisfied in the environment. Then look in the environment for the predicates, related to C, *location_of_event* and *location*. Make sure that the two locations are different and adjust the corresponding score.

  The process of application of the rules follows an iterative approach. In the first iteration all the semantic constraints have to be satisfied by at least one of the candidate sentences. If no sentence has satisfied the constraints, the set of semantic constraints is relaxed by means of structural or semantic relaxation rules, using the semantic ontology. Two kinds of relaxation are considered: i) moving some constraint from MC to OC and ii) relaxing some constraint in MC substituting it for another more general in the taxonomy. If no candidate sentence occurs when all possible relaxations have been performed the question is assumed to have no answer.

```
extract_contextual_answer_from_tokens(DS,SS,_,_,Env, where_location,1, MT,A1,Sc2,_):-
  satisfy_MT_esp_obl([state(C),location(X)],MT,_),Sc=10,
  satisfy_strict([location_of_event(C,A,DS,Env),location(A,DS,Env)]),
  X\==A,
  nth(A,SS,A1),
  nth(X,SS,A2),
  smooth_scr(SS,X,A,Sc,Sc1),
if(
satisfy_MT_esp_obl([type_of_location(_,_,TL)],MT,_),
(check_type_of_location(A1,TL,A2,Sc3),Sc3 > 0.4, Sc2 is (Sc1 + Sc3 * 10)/ 2),
Sc2 is Sc1).
```

Figure 4: One of the extraction rules applicable to the example.

- **Answer selection**. In order to select the answer from the set of candidates, the following scores are computed for each candidate sentence:
  - The rule score, which takes into account factors such as the confidence of the rule used, the relevance of the optional restrictions satisfied in the matching, and the similarity between NEs occurring in the candidate sentence and the question.

- The passage score, that uses the relevance of the passage containing the candidate.
- The semantic score, defined previously.
- The relaxation score, which takes into account the level of rule relaxation in which the candidate has been extracted.

Taking into account that the answer to a question can occur in different sentences/documents, the values for these scores are accumulated for all the sentences in which the same candidate occur. The resulting values are finally normalized and accumulated in a global score. The answer to the question is the candidate with the best global score.

2.6 Definitional Answer Extraction

The objective of definitional questions at CLEF 2004 is to obtain a fragment of text from the corpus explaining who is some person or what is some organization. These kind of definitions of persons, organizations or its acronyms are likely to appear as appositions. Therefore, definitions are extracted from words immediately before or after an occurrence of the question target.

Another important clue to obtain definitions is the apposition of a text in brackets next to the question target, usually associated with an explanation of the term or an expansion of an acronym, as in "OPEP (Organización de Paises Exportadores de Petróleo)". Also, words that occur frequently near the question target are likely to be part of its definition. For example, the word "presidente" in "El presidente Bill Clinton ha llegado esta tarde a Madrid.". Similar techniques have been applied, for example, in [Xu et al., 2003].

The approach taken to extract definitions can ve viewed as a three-step process:

1. **Question analysis and target extraction**. The question is analyzed with the same module as factoid questions. This module produces as output the target of the question and its type (human/organization). The type of the target allows to apply more specific heuristics to each question.

2. **Relative word significance computation**. We will call "relative significance" of a word stem to a measure of how related is a word stem to the question target; this relative significance is computed as follows. For each occurrence of the target in the corpus, a window with its 15 previous words and the 15 words after the occurrence is extracted. From each window extracted, adjectives and nouns (proper and common nouns) are selected and stemmed (stemming is important here in order to reduce the high morphological variability in Spanish). This +/-15 words window is expected to capture the context of the target, and observations recommend this number of words as an adequate distance, at least for Spanish.

   The number of occurrences of each stem in the context windows is computed, and then multiplied by the idf of the stem as computed from the whole corpus, in order to obtain its relative significance to the target. Moreover, there are two lists of stems (one for persons and one for organizations) that contain stems likely to appear in definitions of either persons (as professions, prizes, etc.) or organizations (words like "partido", "organización"). The significance of stems appearing on the corresponding list (depending on the question target type) is multiplied by a factor determined experimentally (3.2) in order to boost its importance.

3. **Selection of the most informative fragment**. The definition has to be selected from the corpus. Definitions are usually found in fragments that follow some high-level patterns, as "<def> (<target>)" or "<target> , <def>". To obtain the definition, for each occurrence of one of these patterns in the text, what we call its information density is calculated, that is, the sum of the relative significance of its words divided by the number of nouns and adjectives it contains. The definition is expected to contain between 4 and 15 non-stop words,

so the length of each definition is the one that maximizes its information density. The text fragment produced as final output is the definition with highest information density.

3 Results

This section describes some tables related with the results and the evaluation of our system in CLEF-2004. We evaluated the three main components of our system and the global results:

- **Question Processing**. This subsystem has been manually evaluated for factual questions (see Table 1) in the following parts: basic NLP tools (POS, NER and NEC), semantic pre-processing (Environment, MC and OC construction) and finally, question classification module.

Subsystem	Total units	Correct	Incorrect	Accuracy	Error
POS-tagging	1667	1629	38	97.72%	2.28%
NE Recognition	183	175	8	95.63%	4.37%
NE Classification	183	137	46	74.86%	25.14%
Environment	180	81	99	45.00%	55.00%
MC	180	77	103	42.78%	57.22%
OC	180	131	49	72.78%	27.22%
Q. Classification	180	105	75	58.33%	41.67%

Table 1: Results of Question Processing evaluation.

- **Passage Retrieval**. The evaluation of this subsystem has been done using the set of correct answers given by the CLEF organization (see Table 2). We participated in CLEF-2004 submitting two runs. In both runs we retrieved only the 1000 top documents (no passages) for definitional questions. These runs differ only in the parameters of the passage retrieval module for factual questions:
  - Windows proximity: in run1 the proximity of the different windows that can compose a passage was lower than run2's (from 60 lemmas to 80).
  - Threshold of minimum passages: the PR algorithm needs to relax the context or the keywords to obtain more passages if the number of extracted passages is lower than this threshold. These threshold's values are: 4 passages (run1) and 1 passage (run 2).
  - Number of top documents retrieved: we have chosen a maximum of 500 documents in run1 and a maximum of 1000 documents in run2.
  - Number of passages retrieved: in run 1 a maximum of 3000 passages, and in run2 a maximum of 50 passages.

Question type	Measure	run1	run2
FACTUAL	Accuracy (*answer*)	60.0% (96/160)	56.87% (91/160)
	Accuracy (*answer+docID*)	36.25% (58/160)	33.12% (53/160)
DEFINITIONAL	Accuracy (*answer*)	85.00% (17/20)	85.00% (17/20)
	Accuracy (*answer+docID*)	55.00% (11/20)	55.00% (11/20)

Table 2: Passage Retrieval results.

In this part we computed two measures: the first one (called *answer*) is the accuracy taking into account the questions that have a correct answer in its set of passages. The second one (called *answer+docID*) is the accuracy taking into account the questions that have a minimum of one passage with a correct answer and a correct document identifier in its set of passages.

- **Answer Extraction**. The evaluation of this subsystem for factual questions has been done in three parts: evaluation of the Candidates Extraction (CE) module, evaluation of the Answer Selection (AS) module and finally we have done an evaluation of the AE subsystem's global accuracy for factual questions in which the answer appears in our selected passages.

Subsystem	Measure	run1	run2
Candidates Selection	Accuracy (*answer*)	35.41% (34/96)	37.36% (34/91)
Answer Selection	Accuracy (*answer*)	70.58% (24/34)	79.41% (27/34)
Answer Extraction	Accuracy (*answer*)	25.00% (24/96)	29.67% (27/91)

Table 3: Factual Answer Extraction results.

- **Global Results**. our first participation in CLEF gives us these results (see Table 4).

Measure	run1	run2
Total Num. Answers	200	200
Right/Wrong	48/150	52/143
IneXact/Unsupported	1/1	3/2
Overall accuracy	24.00% (48/200)	26.00% (52/200)
Accuracy over Factoid	18.89% (34/180)	21.11% (38/180)
Accuracy over Definition	70.00% (14/20)	70.00% (14/20)
Answer-string "NIL" returned correcty	19.23% (10/52)	20.37% (11/54)
Confidence-weighted Score	0.08780 (17.560/200)	0.10287 (20.574/200)

Table 4: Results of TALP-QA system at CLEF-2004.

4 Evaluation and Conclusions

We provided answer to all 200 questions. From them we gave the exact answer to 48 and 52 questions in the run1 and run2 respectively. So the global accuracy of our system was 24% and 26% for run1 and run2 respectively. The discussion of the subsystems for the two kind of questions is presented next:

- **Factual questions**. The accuracy over factoid questions is 18.89% (run1) and 21.11% (run2). Although no direct comparison can be done with other evaluation in other language, we think that we have improved substantially these results in factual questions with respect of the results of the TREC-2003 (5.3%) in English.
  - **Question Processing**. The Question Classification subsystem presented an accuracy of 58%, a similar accuracy of the *environment*, MC and OC constraints. These values are influenced by the previous errors in POS, NER and NEC subsystems.
  - **Passage Retrieval**. In the PR we evaluated that 60% (run1) and 56.87% (run2) of questions have a correct answer in its passages. The evaluation taking into account the document identifiers shows that 36.25% (run1) and 33.12% (run2) of the questions are definitively supported.
  - **Answer Extraction**. The accuracy of the AE module for factual questions for which the answer occurred in our selected passages was of 25% and 29.67% for run1 and run2 respectively. It means that we achieved a significant improvement of our AE module since the results of this part in TREC-2003 were 8.9%. We expect to improve these results by reducing the error rate in the construction of the *environment*, MC and OC.
- **Definitional questions**. The definitional answer extraction module has obtained satisfactory results, 14 right definitions out of 20 proposed (70%). The errors occurred when the target appeared in a small number of fragments (<3), as the system could not be able to correctly determine the right set of significant words.

5 Acknowledgments

This work has been partially supported by the European Comission (CHIL, IST-2004-506909) and the Spanish Research dept. (ALIADO, TIC2002-04447-C02). Our research group, TALP Research Center, is recognized as a Quality Research Group (2001 SGR 00254) by DURSI, the Research Department of the Catalan Government.

6 References

[Atseries et al, 1998] J. Atseries, J. Carmona, I. Castellón, S. Cervell, M. Civit, L. Márquez, M.A. Martí, L. Padró, R. Placer, H. Rodríguez, M. Taulé, J. Turmo.
Morphosyntactic Analisys and Parsing of Unrestricted Spanish Text.
Proceedings of LREC-98. Granada, Spain.

[Carreras et al, 2002] Carreras, X., Márquez, L. and Padró L.
Named Entity Extraction Using Adaboost.
Proceedings of the CoNLL-2002. Shared Task Contribution. Taipei, Taiwan. September 2002.

[Carreras et al, 2004] Carreras, X., I. Chao, L. Padró and M. Padró
FreeLing: An Open-Source Suite of Language Analyzers
Proceedings of LREC-2004. Lisbon, Portugal. 2004.

[Ferrés et al, 2004] Daniel Ferrés, Marc Massot, Muntsa Padró, Horacio Rodríguez, Jordi Turmo.
Automatic Building Gazetteers of Co-referring Named Entities.
Proceedings of LREC-2004. Lisboa, Portugal, 2004.

[Magnini, Cavagliá, 2000] B. Magnini, G. Cavagliá
Integrating Subject Field Codes into WordNet.
Proceedings LREC-2000. Athens, Greece, 2000.

[Massot et al, 2003] Marc Massot, Daniel Ferrés, Horacio Rodríguez.
QA UdG-UPC System at TREC-12.
Proceedings of the TREC-2003. Gaithersburg, Maryland, United States, 2003.

[Quinlan, 1993] Quinlan, J. R.
FOIL: A midterm report.
Proc. of the sixth European Conf. on Machine Learning. Springer-Verlag, 1993.

[Rodríguez et al, 1998] H. Rodríguez, S. Climent, P. Vossen, L. Bloksma, W. Peters, A. Alonge, F. Bertanga, A. Roventini.
The Top-Down Strategy for Bulding EuroWordNet: Vocabulary Coverage, Base Conceps and Top Ontology.
Computer and Humanities 32. 1998, Kluwer Academic Publishers.

[Xu et al, 2003] Jinxi Xu, Ana Licuanan, Ralph Weischedel.
TREC2003 QA at BBN: Answering Definitional Questions.
Proceedings of the TREC-2003. Gaithersburg, Maryland, United States, 2003.