"How much context do you need?" An experiment about the context size in Interactive Cross-language Question Answering.

Borja Navarro, Lorenza Moreno-Monteagudo, Elisa Noguera, Sonia Vázquez,
Fernando Llopis and Andrés Montoyo.
Grupo de Investigación en Procesamiento del Lenguaje y Sistemas de Información.
University of Alicante
(borja,loren,elisa,svazquez,llopis,montoyo)@dlsi.ua.es

Abstract

The main topic of this paper is the context size needed for an efficient Interactive Cross-language Question Answering system. We compare two approaches: the first one (baseline system) shows user whole passages (maximum context: 10 sentences). The second one (experimental system) shows only a clause (minimum context). As cross-language system, the main problem is that the language of the question (Spanish) and the language of the answer context (English) are different. The results show that large context is better. However, there are specific relations between the context size and the knowledge about the language of the answer: users with poor level of English prefer context with few words.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

General Terms

Experimentation

Keywords

Contextual Information, Interactive Question Answering, Question Answering

1 Introduction

In an Interactive Question Answering system, the decision about the correctness of the answer in factotum questions (or usefulness, satisfaction, or helpfulness in analytical questions) depends on the linguistic context in which the possible answer appears [6]. The user decides according to the context. Besides previous knowledge about the topic and the question itself, the context is the main source of information available for the user in order to decide about the correctness of the answer shown by the system. According to the context, he/she decides if it is necessary a refinement of the question or not.

However, there is a specific problem in Interactive Cross-language Questions Answering: the language in which the answer (and the context of the answer) appears is different from the language

of the user and, therefore, the language of the question. The user must deal with a language with null or passive knowledge about it.

There are two approaches to this problem: to translate the possible answer with its context to the language of the user, or to look for other alternative methods of interaction. As in previous years, due to the problems of Machine Translation, we are interested in alternative methods of interaction with the user, avoiding the use of Machine Translation systems [7] [8].

The specific question in this experiment is how much context the users need in order to achieve a satisfactory interaction with the system in a language different from the one of the query.

We have run two systems. The first one (baseline system) is an Information Retrieval System based on passages. This system shows a complete passage of 10 sentences: the maximum context shown to the user.

The interaction with the user has been improved with two elements:

1. A Name Entities Recognition system. The NE that appears in the passages and in the query, plus the NE of the possible answer, are marked with different colors.

2. Also, the set of synonyms of each (disambiguated) word of the question is shown to the user. If he/she thinks that it is necessary, he/she can re-run the IR system with the synonyms. That is, the user decides if it is better to use an extended query or not.

The second system (experimental system) is a preliminary version of a Question Answering system based on syntactic-semantic patterns. This system calculates the syntactic-semantic similarity between the question and the possible answers. Both are formally represented by means of syntactic-semantic patterns, based on the subcategorization frame of the verb. The system shows user only the clause in which appears the possible answer. A clause is a linguistic unit smaller than the sentence: it is the minimum context.

In both systems, the users can see the whole document, if it is necessary.

Together with this primary objective about the context size, we have two secondary objectives:

1. As questions are written in a natural language, it is necessary to disambiguate it. We have applied a Word Sense Disambiguation method based on Relevant Domains for the disambiguation of the question.

2. We are developing methods of syntactic-semantic similarity between the question and the possible answer in a bilingual/multilingual framework. As we said before, the experimental system is a QA system based on the syntactic semantic similarity between the verbal subcategorization frame of the question and the verbal subcategorization frame of the possible answers. In this experiment we have obtained preliminary evaluation results.

In the next section, the process of disambiguation, translation and expansion of the question is explained. The baseline system (IR-n system) is explained in section 3 and in the section 4 the QA system based on syntactic-semantic similarity. At the end of the paper, the results and some problems founded will be shown.

2 Question translation, disambiguation and expansion.

As standard situation, the mother tongue of users is Spanish. The questions are written in Spanish and the answers in English. The users have passive knowledge of English: they can understand some words/sentences in English, but they can not formulate a question in English correctly.

The words of the questions were disambiguated with a Word Sense Disambiguation system based on Relevant Domains.

2.1 WordNet Domains and Relevant Domains

WordNet Domains (WND)[5] is an extension of WordNet 1.6 where each synset is annotated with one or more domain labels selected from a set of about 250 hundred labels hierarchically organized.

WND allows to connect words belonging to different subhierarchies and to include into the same domain label several senses of the same word. Thus, a single domain label may group together more than one word sense, obtaining a reduction of the polysemy. Furthermore in WND the same domain label can be associated to synsets belonging to different syntactic categories. Therefore using domain labels we can establish relations between synsets that belong to different syntactic categories.

In this work, WND is used to collect examples of domains associations to the different meanings of the words. With this information we obtain a new resource named Relevant Domains (RD).

To obtain RD, WND glosses will be used to collect the more relevant and representative domain labels for each word. So the first step is using a POS-tagger to obtain all syntactic categories and lemmas of each gloss. We use Tree-tagger [13]. Once the results of the POS-tagger have been obtained, the second step is assigning the domain associated to the gloss analyzed for each word (each gloss has associated one or more domain labels). This process is done with all glosses in WND. Finally with all this information we can proceed to obtain the new resource RD.

Our purpose is to obtain a resource that will contain all words of WND glosses with all their possible domains organized in an ascending way because of their relevance in domains. In order to do so we first need to collect the most representative words of a domain. So we use the Mutual Information formula (1) as follows:

$$MI(w, D) = \log_2 \frac{Pr(w|D)}{Pr(w)} \tag{1}$$

$w : word.$
$D : domain.$

Intuitively, a representative word would appear in a domain context most frequently. But we are interested on the importance of words in a domain, that is, the most representative and common words in a domain. We can appreciate this importance with the Association Ratio (A.R.) formula:

$$AR(w, D) = Pr(w|D) \log_2 \frac{Pr(w|D)}{Pr(w)} \tag{2}$$

Formula (2) shows A.R. that is applied to all words with noun grammatical category obtained from WND glosses. Later, the same process is applied to verbs, adjectives and adverbs. A proposal in this way has been made in [12], but using Lexicography Codes of WordNet Files. Once the results are obtained, we sort (by means of A.R.) all the collected domains for each word.

An example of the domains sorted by A.R. for word "organ" is shown on table 2.1:

2.2 WSD method

Our WSD method is unsupervised and it is based on the hypothesis that words appeared into the same context have their senses quite related. In this case, as context we can take a sentence or a window of words that contains the ambiguous word.

For collecting context and the domains of each word sorted by A.R. we need a structure named context vector. Furthermore, each polysemic word in the context has different senses (with their corresponding glosses) and for each sense we need a structure that contains the most representative domains sorted equally by the Association Ratio formula. This structure is named sense vector. In order to obtain the correct word senses into the context, we must measure the proximity between context vector and sense vectors. This proximity is measured by using cosinus between both vectors, that is, the higher the cosinus is the more proximity between both vectors.

Association Ratio for organ	
Domain	**A.R.**
Surgery	0.189502
Radiology	0.109413
Sexuality	0.048288
Optics	0.048277
Anatomy	0.047832
Physiology	0.029388
Music	0.012913
Psychoanalysis	0.010830
Genetics	0.009776
Medicine	0.009503
Entomology	0.002788
...	...

Table 1: Association Ratio obtained for word "organ"

2.3 Application to interactive task

Part one. First of all we need to disambiguate initial questions. This task needs an automatic translation of questions from Spanish to English.

Step one. Obtaining the automatic translation of questions.
For obtaining the automatic translation of each question we use three different translators: We have use three machine translation (MT) systems available on the web: Systran Babelfish[1], Reverso Soft.[2], and Google[3]. Each one provides its own translation.

Step two. Selecting the appropriate translation.
Between all translations we select those words more frequent. If there isn't any word in common between the three translations we select all words obtained.

Step three. Obtaining the correct sense of words.
For this purpose we use our method Relevant Domains [15] to obtain the disambiguation of words selected. This method uses the words of questions as context to construct word sense vectors and select the appropriate sense of each word.

Part two. The next step is using the information provided by our Relevant Domains disambiguation system to expand each question.

Step one. Obtaining synonymous words.
Once we have obtained the correct sense of each word we intend to expand each question with a list of synonymous words. That is, we add more information selecting all synonymous words to each word disambiguated.

This task is possible thanks to the fact that words are disambiguated, so we have only one sense per word. Each sense has associated a synset in WordNet that contains one or more synonymous words. With this new information users have the possibility of selecting more words that can appear associated with the answer.

[1] http://babelfish.altavista.com/
[2] http://www.elmundo.es/traductor/
[3] http://www.google.com/language_tools

Our method obtains the English disambiguation of questions but there isn't any problem because we have a direct association of English words and Spanish words with the EuroWordNet [16]. So for each English word we have a Spanish word with its synonymous in Spanish. So this is the information that users will employ to the iCLEF task.

Step two. Calling the Passages Retrieval system IR-n.

With the words selected by users we have the information necessary to call the IR-n system for obtaining the possible paragraphs with the correct answer to each question. The expansion of each question with synonymous sets contributes to obtain better results by the IR-n system.

3 Baseline system: passages improved with Name Entity recognition.

The baseline system is a Passages Retrieval system. Following with the approach of last years, this model is based on passages with new elements which help the interaction with the user. These new elements are Name Entity Recognition (NER) in the passages and the synonyms of the words.

Our aim is to help the user to find the answer of the query. With this aim the most relevant passages are shown and the words of the query are highlight in the text. Furthermore, the entity type of the answer is detected and the words which are of this type are also highlighted. Finally, the synonyms of the query are shown and they are highlighted in the text.

The passages are extracted by IR-n system. IR-n [4] is a passage retrieval system (RP). RP systems [2] study the appearance of query terms in contiguous fragments of the documents (also called passages). One of the main advantages of these systems is that they allow us to determine not only if a document is relevant or not, but also the detection of the relevant part of the document.

DRAMNERI [14] is a knowledge based Named Entity Recognition system that uses rules and gazetteers in order to identify and classify named entities. This is done sequentially by applying several modules which perform different tasks: tokenization, sentence partition, named entity identification and finally named entity classification.

As example a screen of the approach based on passages is shown in Figure 1. This is matched with the question one and the passage three of this question.

Firstly, the question in Spanish is presented and following the synonyms of this question which has been obtained by means of the method that is explained in the section 2. Next to the synonyms, there is a checkbox which allows the user to carry out the search with query expansion based on synonyms. Moreover, the words and synonyms of the query (only if user has selected the checkbox to carry out query expansion) are highlighted in blue color.

Under the synonyms, this approach lets the user to select the entity type that is expected as an answer. Because of that, a list containing all types of entities that NER detects is shown. The entity which NER has detected as entity type of the answer is selected from the list. Furthermore, the distinct entities detected by NER are shown in the passages. They are highlighted in red color.

When NER is applied to a query, on one hand the entity type of the answer is returned and, on the other hand, all the entities of this type in the text are highlighted. This could be useful for the user because he doesn't need to read all the passage. Firstly it could see if the request is in the marked entities, otherwise the whole passage will be read.

Moreover, as it is shown in the figure 1, this year it has also been included an option that allows to see the whole document. This will be useful if the request is not in the passage but it is in the document.

Figure 1: HTML interface with passages

4 Experimental system: A Question Answering System based on syntactic-semantic similarity.

The experimental system is a Question Answering (QA) System that follows a linguistic oriented approach based on deep linguistic knowledge.

Our objective is to show user the minimum context necessary to evaluate the correction/utility of the answer. The context is the clause: the set of words related with a verb in a sentence. A clause is formed by one or more nominal or prepositional phrases. Therefore, the system shows user the possible answer plus the words/phrases that form the clause.

According to their syntactic relations, there are two kinds of clauses: principal clauses (if the verb is the main one of the sentence), and subordinate clauses (if the verb is subordinated).

The intuitive idea behind this approach is that between the question and the answer exists a deep semantic relation: a question is formed by a clause (or more, in complex questions) and the answer appears inside another clause. The objective is to calculate the syntactic-semantic similarity between the question and the clause in which the possible answer appears.

Both the question and the possible answers are formally represented as syntactic-semantic patterns. Basically, the syntactic-semantic pattern of a clause is the subcategorization frame of the verb. It is formed by the next components [9] [10]:

1. The verb: each verb forms a syntactic semantic pattern. It is represented by means of its lemma and its sense[4].

2. The complements of the verb: the set of complements (argumentals -subcategorizated- and adjuncts) that appears with the verb. They are represented by the lemma of the head of the phrase and its sense (or senses, if it is not possible an automatic disambiguation of the ambiguous head nouns). These head nouns are common nouns or proper nouns.

The input of the system is the output of the Passage Retrieval Sistem IR-n. All the passages returned by IR-n system are processed with a PoS_tagger (Tree-tagger [13]) and a syntactic parser (SUPAR [11]). From this, the system extracts patterns (one for each verb) and stores them in a database of syntactic-semantic patterns. Then all the senses of each head noun and each verb is extracted from EuroWordNet ([16]).

[4]The sense of the verb of the query has been disambiguated (section 2). The sense of the verb of each possible answer is represented by all the possible senses that provide WordNet.

A pattern is extracted from the question too. In this case, the sense of nouns and verbs has been automatically disambiguated.

Once all the patterns are extracted, the system calculates the syntactic-semantic similarity between the question pattern and all the patterns extracted from the passages. This process has two steps:

1. A filter of proper nouns:

   If a proper noun appears in the question, it must appear in the answer. If it does not appear in a pattern, it is not the correct one[5]. At least a proper name of the question must appear in the answer pattern. If, for example, a question asks about "Thomas Mann", the system accepts all patterns with the proper noun "Thomas", the proper noun "Mann" or both.

2. A syntactic-semantic measure of similarity:

   The system calculates the syntactic-semantic similarity between the question patterns (Pq) and the possible answer pattern (Pa) (the patterns that have been selected in the previous filter), according to the next formula:

   $$Sim(Pq, Pa) = 2(SimVpq, Vpa) + \frac{NumA_qa + NumPN_qa}{2}$$

   where

   - `SimVpq - Vpa` represents the semantic similarity between the verb of the query pattern and the verb of the answer pattern. It is computed by the D. Lin's formula ([3] [1]) [6].
   - `NumA_qa` represents the number of equal arguments between the query pattern and the answer pattern.
   - `NumPN_qa` represents the number of equal proper names between the query pattern and the answer pattern.

   This similarity is semantic and syntactic, because it uses semantic information of verbs and syntactic information of arguments and proper nouns.

   The main component is the semantic similarity between both verbs. The idea behind this formula is that the semantic of the verb establishes the semantic framework of the complete pattern (the subcategorization frame). So both patterns (the question pattern and the answer pattern) must be semantically related mainly by the verb sense. Then, this general semantic relation is specified by the numbers of equal arguments, both commons nouns and proper nouns.

The output of the system is a rank list of patterns, from the most similar with the question pattern up to the less one. For the interactive process, the system shows user the clause related with each syntactic-semantic pattern. The user must check each clause, until finding the correct answer (Figure 2).

The system uses deep knowledge: it does not operate with the clause or superficial patterns. Instead of this, it operates with the syntactic semantic pattern behind the clause: an abstract pattern. However, the user only interacts with clauses, not with the syntactic semantic patterns.

5 Results

In general, the results show that it is better a large context than a small one. That is, the users locate correct answers better with a passage retrieval system (plus name entity recognition) than with a more specific QA system that shows only clauses (Figure 3).

Three users locate more correct answers with experimental system (small context), and five with baseline system (large context) (Figure 5).

Figure 2: HTML interface with clauses

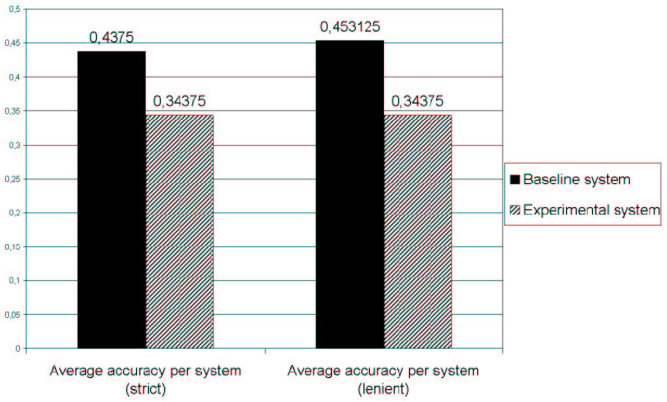

Figure 3: General results

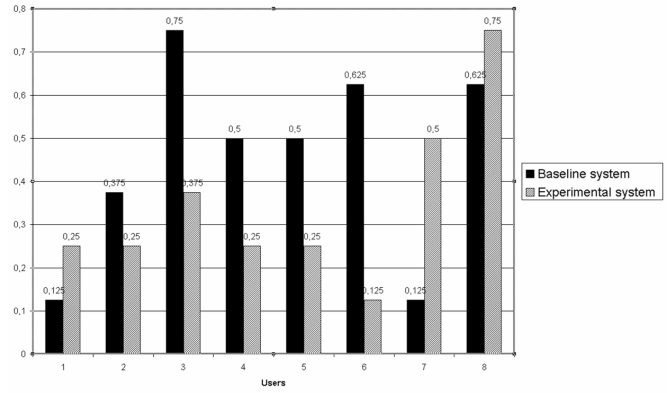

Figure 4: Results user by user: lenient

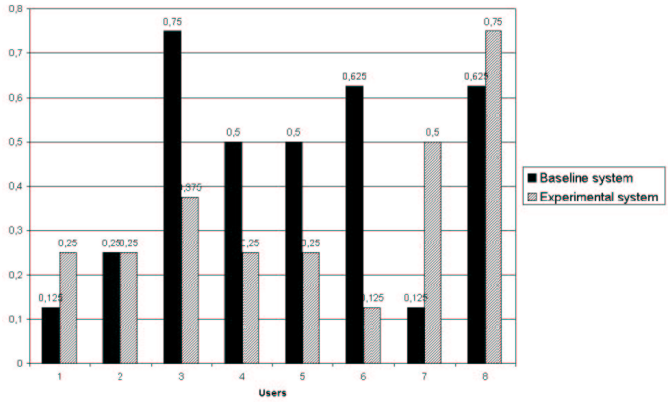

Figure 5: Results user by user: strict

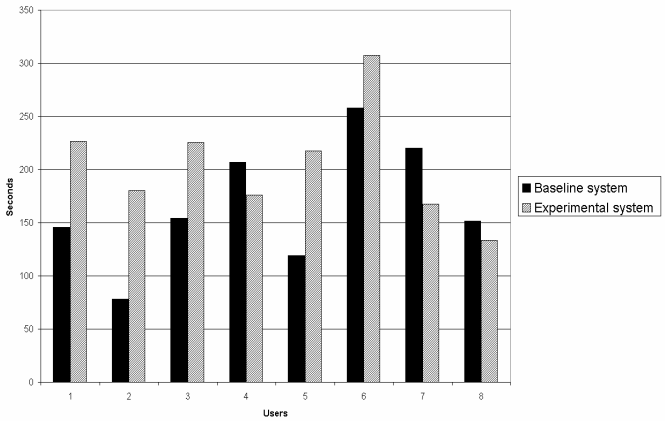

Figure 6: Time consuming by each user.

However, the better results are achieved with both systems: user 3 and user 8. With these results, we think that the improvement of the QA system based on syntactic-semantic patterns will improve the interaction process.

According to the English knowledge of the users, users with low knowledge have reported that they prefer the experimental system, based on clauses. One of them (user 7) has located correct answers with the clauses (0.5 strict accuracy), better than with passages (0.125 strict accuracy).

Comparing the time consumed by each user (Figure 6), the user that has located correct answers with experimental system (clauses) is the one that has consumed less time (user 8). In general, users have spent much time looking for correct answers, because they tried to find more context in the complete document. In these cases, the context shown by both systems is not sufficient (for example, user 6).

The use of a Name Entity Recognition system has been really useful during the interaction process. All users, except one, report that to know the name entities of the passage and the possible answer helped them during the localization of the correct answer.

However, users did not use the synonyms and the expansion of the query during the interaction process. Only one user (5) said that the synonyms were really useful to locate the correct answer. The use of synonyms is shown in Figure 7.

[5]or the user will not be able to decide if it is the correct one, because the context doesn't provide enough information in order to decide about the correctness of the clause.

[6]We have used the T. Pedersen's implementation: http://search.cpan.org/~tpederse/

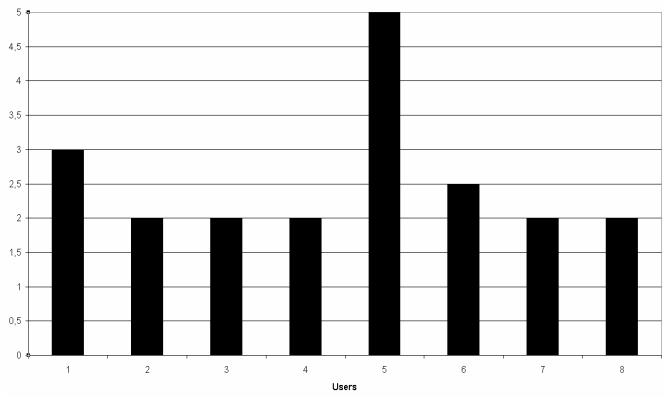

Figure 7: Use of synonyms.

6 Conclusions

It is difficult to establish a fixed context useful for Interactive Question Answering. According to the results of this experiment, for an interactive user interface it is more useful passages, in which more context appear, than a simple clauses, in which the contexts is formed by few words: between a large context or a short context, users prefer the large one. However, for users with poor knowledge of the langauge of the answer, it is more useful (and fast) to interact with short context.

So, for an interactive approach to QA, it is important not only the precision of the system, but also the amount of information that the system shows to the user. This is the information that users need to decide about the correctness or usefulness of the answer.

The use of a name entity recognition system that show user the possible answer of a passage is really a useful tool for an optimum interaction. However, the use of synonyms in the interaction process is not useful at all. It is more useful during the automatic expansion of the query.

7 Future work

This experiment has been a preliminary evaluation of a QA system based on syntactic-semantic patterns. On one hand, this system has two main problems that will be improved in future:

1. The system shows the correct answer, but the user is not able to detect that this is the correct one. In these cases the clause is formed only by few words (mainly in subordinate clauses). It is not sufficient context to detect the correct answer. To solve this problem it is necessary to show a complete sentence. A sentence has a complete sense, but the clause has not.

2. The system does not show the correct answer. There are some questions in which the system has not detected the correct clause in the first 50 patterns. That is, non of the 50 patterns semantically similar to the question pattern has the correct answer. To solve this problem, first, it is necessary to improve the clause splitter; and, second, the syntactic-semantic measure of similarity.

On the other hand, this system will be improved in several aspects: for example, the use of manually annotated corpus to improve the quality of the patterns. Also, some experiments will be done in order to specify the best syntactic-semantic similarity.

For future iCLEF, our idea is the combination of these two approaches during the interaction process: to show the clause and the passage at the same time.

8 Acknowledge

This paper has been partially supported by Spanish Government: R2D2 project (TIC 2003-07158-C04-01), CES-ECES project (HUM2004-21127-E) and REXIN project (FIT-340100-2004-14); and Valencian Government: CLASITEX project (GV04B-276) and RECENT project (GV04B-268).

Thank you very much to the users: Pilar, Silvia, Lorena, Sandra, Norberto, Sergio, Oscar and Mr. Talking Head.

References

[1] A. Budanitsky and G. Hirst. Semantic Distance in WordNet: An Experimental, Application-oriented Evaluation of Five Measures. In *Workshop on WordNet and Other Lexical Resources. North American Chapter of the Association for Computational Linguistics (NAACL-2001)*, Pittsburgh, 2001.

[2] M. Kaskziel and J. Zobel. Passage retrieval revisited. In *Proceedings of the 20th annual International ACM Philadelphia SIGIR*, pages 178–185, 1997.

[3] D. Lin. An information-theoretic definition of similarity. In *Proc. 15th International Conf. on Machine Learning*, pages 296–304. Morgan Kaufmann, San Francisco, CA, 1998.

[4] F. Llopis. *IR-n: Un Sistema de Recuperación de Información Basado en Pasajes*. PhD thesis, University of Alicante, 2003.

[5] B. Magnini and G. Cavaglia. Integrating Subject Field Codes into WordNet. In M. Gavrilidou, G. Crayannis, S. Markantonatu, S. Piperidis, and G. Stainhaouer, editors, *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation*, pages 1413–1418, Athens, Greece, 2000.

[6] M. T. Maybury, editor. *New Directions in Question Answering*. AAAI Press - MIT Press, 2004.

[7] B. Navarro, F. Llopis, and MA Varó. Comparing syntactic semantic patterns and passages in Interactive Cross Language Information Access (iCLEF at University of Alicante). *Comparative Evaluation of Multilingual Information Access Systems: 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003*, Lecture Notes in Computer Science 3491, Springer-Verlang, 2004.

[8] B. Navarro, L. Moreno-Monteagudo, S. Vázquez, F. Llopis, A. Montoyo, and MA. Varó. Improving interaction with the user in Cross-Language Question Answering through Relevant Domains and Syntactic-semantic patterns. *Workshop of Cross-Language Evaluation Forum (CLEF 2004)*, Lecture Notes in Computer Science 3237, Springer-Verlang, 2005.

[9] B. Navarro, M. Palomar, and P. Martínez-Barco. A General Proposal to Multilingual Information Access based on Syntactic Semantic Patterns. In Anje Dsterhft and Bernhard Thalheim, editor, *Natural Language Processing and Information Systems - NLDB 2003*, pages 186–199. Lecture Notes in Informatics, GI-Edition, Bonn, 2003.

[10] B. Navarro, M. Palomar, and P. Martínez-Barco. Automatic extraction of syntactic semantic patterns for multilingual resources. In *4th International Conference on Language Resources and Evaluation (LREC)*, Lisbon, 2004.

[11] M. Palomar, A. Ferrández, L. Moreno, M. Saiz-Noeda, R. Mu noz, P. Martínez-Barco, J. Peral, and B. Navarro. A robust partial parsing strategy based on the slot unification grammars. In *Proceedings of Sixth Conference on Natural Language Processing (TALN)*, pages 263–272, Corsica (France), 1999.

[12] G. Rigau, E. Agirre, and J. Atserias. Combining unsupervised lexical knowledge methods for word sense disambiguation. In *Proceedings of joint 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics ACL/EACL'97*, Madrid, Spain, 1997.

[13] H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings International Conference on New Methods in Language Processing.*, pages 44–49, Manchester, UK, 1994.

[14] A. Toral. DRAMNERI: a free knowledge based tool to Named Entity Recognition. In *Proceedings of the 1st Free Software Technologies Conference*, 2005. Accepted.

[15] S. Vázquez, A. Montoyo, and G. Rigau. Using relevant domains resource for word sense disambiguation. *IC-AI'04 International Conference*, II, 2004.

[16] P. Vossen, L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, and W. Peters. The eurowordnet base concepts and top ontology. Deliverable d017, d034, d036, eurowordnet (le 4003), University of Amsterdam, 1997.