Applying Wikipedia's Multilingual Knowledge to Cross–Lingual Question Answering*

Sergio Ferrández, Antonio Toral, Óscar Ferrández, Antonio Ferrández, and Rafael Muñoz

Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante, Spain
{sferrandez,atoral,ofe,antonio,rafael}@dlsi.ua.es

Abstract. The application of the multilingual knowledge encoded in Wikipedia to an open–domain Cross–Lingual Question Answering system based on the Inter Lingual Index (ILI) module of EuroWordNet is proposed and evaluated. This strategy overcomes the problems due to ILI's low coverage on proper nouns (Named Entities). Moreover, as these are open class words (highly changing), using a community–based up–to–date resource avoids the tedious maintenance of hand–coded bilingual dictionaries. A study reveals the importance to translate Named Entities in CL–QA and the advantages of relying on Wikipedia over ILI for doing this. Tests on questions from the Cross–Language Evaluation Forum (CLEF) justify our approach (20% of these are correctly answered thanks to Wikipedia's Multilingual Knowledge).

1 Introduction

Currently, the exponential growth of digital information requires processes capable of searching, filtering, retrieving and classifying this information. Moreover, the information required by the users might be in different languages. Nowadays, one of the most demanded way of accessing multilingual information is to obtain information from sources written in different languages than that of input queries. Obviously, multilinguality is one of the main difficulties that impedes the right acquisition of information.

For this purpose, Computational Linguistics applications such as Information Retrieval (IR) and Question Answering (QA) are used. IR is the science of searching for documents that contain the information required by the user, whereas QA can be defined as the task consisting of answering precise and arbitrary questions formulated by the user. The aim of a QA system is to find the correct answer to user questions in a non-structured collection of documents. In Cross–Lingual (CL) environments, the question is formulated in a different

* This work has been developed in the framework of the project QALL-ME, which is a 6th Framework Research Programme of the European Union (EU), contract number: FP6-IST-033860.

Z. Kedad et al.(Eds.): NLDB 2007, LNCS 4592, pp. 352–363, 2007.

language from that of the documents, which increases the difficulty. As it was revealed in the Cross-Language Evaluation Forum (CLEF) 2006 [15], multilingual tracks of IR and QA tasks have been recognized as an important issue in information access.

In this paper, we present a novel approach for solving the CL–QA task. Our strategy consists of a CL–QA system [9], which performs the references between words in different languages using the Inter Lingual Index (ILI) module of EuroWordNet (EWN) [22] as well as the multilingual relations encoded in Wikipedia[1]. The original contribution of this research consists of the application of Wikipedia's Multilingual Knowledge (WMK) in order to overcome ILI's low recall regarding proper nouns and as an affordable alternative to other approaches that rely on hand–coded dictionaries of proper nouns and therefore avoiding maintenance. Besides, a detailed study justifying the need to translate this kind of nouns within CL–QA is included.

The rest of the paper is organized as follows: section 2 describes the background of current CL–QA systems Afterwards, our CL–QA system based in ILI is presented. This is followed by a detailed description about the integration of WMK. Next, section 5 illustrates a study about the need for translating Named Entities in CL–QA. In section 6, an evaluation regarding CLEF official questions is presented. Finally, section 7 wraps up the paper with our conclusions and future work proposals.

2 Background

The overall accuracy of CL–QA systems is directly affected by their ability to correctly analyze and translate the question that is received as input. An imperfect or fuzzy translation of the question causes a negative impact on the overall accuracy of the systems. According to [17], the Question Analysis phase is responsible for 36.4% of the total of number of errors in open–domain QA.

The last edition of CLEF (2006) [15] has confirmed that most of the implementations of current CL-QA systems [5,13,18,20,21]are based on the use of on-line translation services. However, a recent research [8] presents a study detailing the common errors produced by Machine Translation (MT) based systems and proposes an alternative approach to overcome such errors.

This revision of the state of the art focuses on the bilingual English–Spanish QA task, because the CL–QA system used for the evaluation works in these languages. In CLEF 2006, three different approaches have been presented by CL-QA systems as solutions for the bilingual English–Spanish task.

The first one [6] translates entire documents into the language in which the question is formulated. This system uses a statistical MT system that has been trained using the European Parliament Proceedings Parallel Corpus 1996–2003 (EUROPARL).

The second system [23] uses an automatic MT tool to translate the question into the language in which the documents are written. This strategy is

[1] www.wikipedia.org

the simplest technique available. In this case, when comparing to the Spanish monolingual task, the system loses about 55% of this precision in the CL task.

The third system [12] translates the question using different on-line machine translators and some heuristics. This technique consults several web services in order to obtain an acceptable translation.

The previously described strategies are based on the use of MT in order to carry out the bilingual English–Spanish task, and all of them try to correct the translation errors through different heuristics.

The translations are often inexact and quite fuzzy. Besides, the MT systems resolve the ambiguity by means of only giving one translation per word. These facts cause an important negative impact on the precision of the systems. This can be checked on the last edition of CLEF 2006 where the cross lingual systems obtained less than 50% of correct answers compared to the monolingual task.

For instance, MT systems generate errors [8] such as translations of names that should be left untranslated, translations of polysemous words where the sense translated is not the correct one, syntactic errors in the translation, wrong translations of interrogative particles, incorrect lexical-syntactic category of the translated words and unknown words by the MT and thus left untranslated. The impact of this kind of mistakes should be controlled and evaluated.

In the next sections, our strategy of CL–QA system and the integration of WMK in order to control the references between languages are detailed.

3 System Description

In this section, the architecture and functionality of our method to open domain CL-QA [7] are detailed. A graphic depicting the overall architecture of the system is shown in figure 1.

The system is designed to localize answers from documents, where both answers and documents are written in different languages. The system is based on complex syntactic pattern matching using Computational Linguistics tools [1,14,19]. Also, a new proposal of Word Sense Disambiguation (WSD) for nouns (presented in [11]) is applied to improve the precision of the system.

The fundamental and original characteristic of our approach is the strategy used for the Inter Lingual Reference (ILR) Module in which the ILI Module of EuroWordNet (EWN) [22] is used with the aim of reducing the negative effect of question translation on the overall accuracy. This multilingual knowledge source is used to reference verbs, common nouns and proper nouns (named entities).

Named Entities (NEs) contained in the input questions are identified and classified by the Named Entity Recognition (NER) NERUA system. Four entity types are considered: person (PER), location (LOC), organization (ORG) and miscellaneous[2] (MISC). The recognition of NEs makes the ILR module capable of carrying out a customized treatment for each entity type.

The strategy followed by the ILR module introduces two improvements:

[2] This entity type is assigned when a detected entity cannot be enclosed in any of the remaining ones. E.g. Maastricht treaty (in question 13 of QA–CLEF 2006).

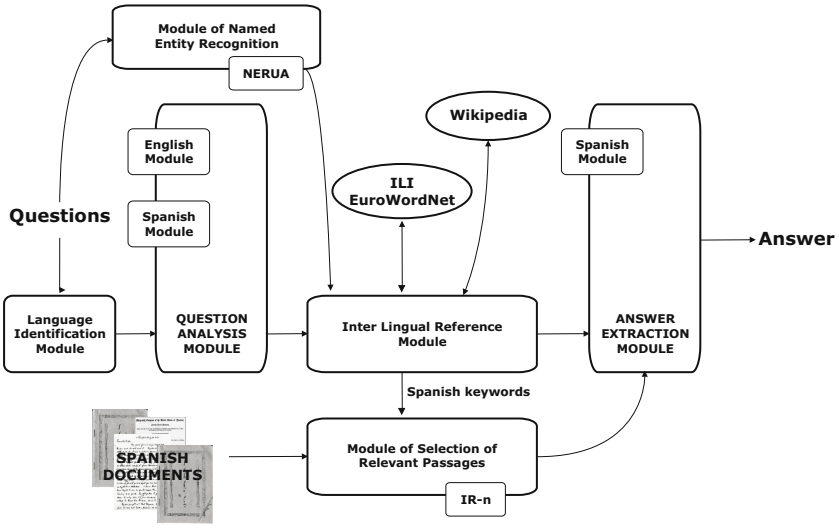

Fig. 1. Architecture of the system

(1) The consideration of more than one translation per word by means of using the different synsets of each word in the ILI module of EWN. Figure 2 shows the references provided by the ILI module for the input word "*president*" in English when the target language is Spanish.

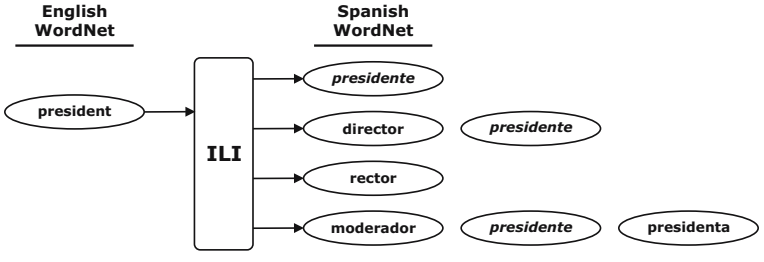

Fig. 2. Links to the word "*president*"

As can be seen in figure 2, in some cases the ILR module obtains more than one Spanish equivalent for each English word. The current strategy employed to get the best translation consists of assigning a weight depending on the frequency of each word in ILI. In this case, the most weighted Spanish word is *"presidente"*. This strategy improves the method commonly followed by MT services in which only one possible translation is given for each word.
(2) Unlike the current bilingual English–Spanish QA systems, the question analysis is developed in the original language without any translation. The

system develops two main tasks in the question analysis phase using a set of syntactic patterns:

- The detection of the expected answer type. The system detects the type of information that the answer has to satisfy to be a candidate of an answer (proper nouns, quantity, date, ...).
- The identification of the main Syntactic Blocks (SB) of the question. The system extracts the SB that are necessary to find the answers.

In order to show the complete process, an example of a question at CLEF 2006 is provided:

- **Question 107 at CLEF 2006:** *How many soldiers does Spain have?*
- **SB:**
  [Noun Phrase *soldier*]
  [Verb Phrase *to have*]
  [Noun Phrase *Spain*]
- **Type:** entity-amount
- **Keywords to be referenced with ILI:** soldier have Spain
  soldier ↦ soldado
  have ↦ estar-enfermo *tener* padecer sufrir causar inducir hacer consumir tomar ingerir experimentar *tener* poseer *tener* recibir aceptar querer constar figurar existir
  Spain ↦ España

On the other hand, the verbs and common nouns that are not referenced in ILI are translated into Spanish using an on-line Spanish Dictionary[3]. Moreover, in order to decrease the effect of incorrect translation of the proper nouns, the matches using these words in the search of the answer are realized using the set of translated words and the original word of the question. The matches found using the original English word are valued at 20% less.

The final step of the CL–QA process is the Extraction of the Answer. The system uses the syntactic blocks of the question and different sets of syntactic patterns (according to the type of the question) with lexical, syntactic and semantic information to find out the correct answer.

In the next section, our novel strategy which integrates multilingual knowledge from Wikipedia in order to translate named entities is presented.

4 Integrating Wikipedia's Multilingual Knowledge in CL-QA

The main drawback of using ILI is that it contains very few proper nouns.[4] In fact, according to [16], WordNet 1.6 contains 3,876 proper nouns. This word

[3] `http://www.wordreference.com`

[4] The word class corresponding to the NE types considered: person, location, organization and miscellaneous.

class is highly evolving, meaning that new proper nouns appear continuously. As ILI is a hand-tagged resource developed by a small number of linguist experts, it becomes obvious that it would be tedious and time-consuming to maintain a considerable amount of proper nouns within its infrastructure.

Exploiting Wikipedia is an appropriate way in order to fill this gap. Wikipedia is an encyclopedia written in a collaborative way[5] that contains a huge amount of proper nouns,[6] and like this word class, this resource is continuously updated. Moreover, it has multilingual links that reference entries in an input language with their equivalents in other languages.

Wikipedia has been already employed within monolingual QA [4]. However, although their multilingual capabilities have been used for tasks such as multilingual corpora creation [3] and discovery of related entries of Wikipedia in different languages [2], to our knowledge, they have not been applied within the CL–QA environment. The following example shows how CL–QA can benefit from the incorporation of this knowledge.

- **Question 186 at CLEF 2006:** *In which town in Zeeland did Jan Toorop spend several weeks every year between 1903 and 1924? ("¿En qué ciudad de Zelanda pasaba varias semanas al año Jan Toorop entre 1903 y 1924?")*

  The question contains two proper nouns: "*Zeeland*" and "*Jan Toorop*". None of them is referenced in ILI. However, both have an entry in the English version of Wikipedia, and both entries contain a reference to their Spanish equivalents: "*Zelanda*" and "*Jan Toorop*" respectively. Furthermore, if this question would have been translated by a MT service, the string "*Jan Toorop*" would have been converted to "*Enero Toorop*" interpreting that Jan states for January.

To incorporate WMK into our CL–QA system, the ILR module performs a special treatment of NEs that depends on the entity type (this decision will be justified in the study presented in section 5). Person entities are directly translated by WMK whereas the remaining entity types are translated by ILI, and if no translation is found in this resource, WMK is used. The hypothesis is that both resources contain complementary information and therefore a combination of them could achieve better CL–QA performance.

In order to include WMK in our system, database dumps[7] provided by the Wikimedia Foundation were downloaded and tailored for our specific needs as well as for efficiency reasons. Besides, an API to access this database and gather the required information was developed. Both this API and utilities to download, import and tailor Wikimedia database dumps are part of the software wiki_db_access, which has been released with a free license with the aim that it could be useful for research purposes.[8]

[5] On 2007/01/16 the English version had 3,247,299 registered users.
[6] The dump used contains 1,496,097 encyclopedic entries.
[7] Available at `http://download.wikimedia.org`
[8] Available at `http://www.dlsi.ua.es/~atoral/`

Table 1. Percentage of questions containing NEs and percentage of NEs that should be translated

Dataset	Questions					
		overall	PER	LOC	ORG	MISC
Questions CLEF 2004	with NEs	81%	23.5%	28%	15%	20.5%
	NEs should be translated	44.89%	2.1%	60.7%	56.7%	48.8%
Questions CLEF 2005	with NEs	93%	34%	25.5%	24%	13.5%
	NEs should be translated	36%	10.3%	50.9%	39.6%	55.5%
Questions CLEF 2006	with NEs	89%	31%	24.5%	22.5%	24%
	NEs should be translated	42.69%	3.2%	65.3%	40%	50%
Average	with NEs	87.7%	29.5%	26%	20.5%	19.3%
	NEs should be translated	41.2%	5.2%	59%	45.4%	51.4%

5 The Need for Translating NEs in CL–QA

This section presents a minute study on the need for translating NEs in CL–QA. The dataset used has been the official 600 English questions of CLEF 2004, 2005 and 2006. The aim of this study is to find out solutions in order to overcome the errors in the translation of NEs between different languages. We provide results on how important it is to translate NEs in CL–QA and how they can be successfully translated.

Table 1 presents the results on our study to find out the percentage of questions that contain NEs and the percentage of these NEs that need to be translated. The percentage of questions with NEs is quite high (81% for 2004, 93% for 2005 and 89% for 2006, i.e. 87.7% on average). From these entities, nearly half of them should be translated (44.89% for 2004, 36% for 2005 and 42.69% for 2006, i.e. 41.2% on average). The remaining percentage of NEs should not be translated (for all of these NEs, no reference is found in ILI[9] while most of them are present in Wikipedia but their name both in the input and target language is the same). Regarding the entity types, it can be seen that it is very important for CL–QA to translate locations, organizations and miscellaneous entities while the impact of not treating person entities would be low.

We have discovered that most of the mistakes regarding wrong ILI references are caused by trying to translate a word that should not be translated (e.g. a person name). Being person entities those with a lower need to be translated, and being ILI a resource with low recall regarding proper nouns, it is for this entity type that ILI obtains the worst performance. Table 2 shows the percentage of person entities that is wrongly translated by ILI in the question sets. Roughly, 30% of person entities are wrongly translated, which has a considerable impact for the CL–QA process.

[9] Even if any of these NEs would be incorrectly translated by ILI, our CL–QA system takes into account as well as the translated NEs, these NEs in the original language.

Table 2. Percentage of wrong translation of type person using the ILI mdule

Dataset	Wrong Translation
Person Entities from CLEF 04	28.6%
Person Entities from CLEF 05	27.6%
Person Entities from CLEF 06	30.3%
Average	28.8%

The following example (see Table 3) shows a case in which ILI fails to translate person NEs whereas WMK provides the correct reference in the target language. This justifies our decision to directly translate person entities by means of using WMK. In this example, the proper noun "*Jan*" is confused with the abbreviation of the month "*January*" by the ILI module of EuroWordNet while WMK provides the correct reference in Spanish. In these cases, the need for some kind of treatment such as NER is clear in order to classify entities and therefore to perform a specialized treatment of NEs depending on the entity type. This will be discussed in the following section.

Table 3. Question 184 CLEF 2006

Language	Question 184 CLEF 2006
English	Who is Jan Tinbergen?
Spanish	Quién es Jan Tinbergen?
Translated Keywords	
using ILI	**enero** Tinbergen
using WMK	Jan Tinbergen

In a nutshell, the study has proved that it is important to translate NEs in CL–QA. It has also been revealed that a specialized treatment should be carried out depending on the entity type. Concretely, ILI's performance for person entities is very low. In fact, the CL–QA system obtains better results if person entities are not translated at all than if they are translated by ILI. However, the idiosyncracies of WMK provide a treatment of person entities that overcome ILI's limitations.

6 Experimental Results

6.1 Evaluation Environment

For carrying out this evaluation, the CLEF 2004, 2005 and 2006 sets of 600 English and Spanish questions and the EFE 1994–1995 Spanish corpora are used. These corpora provide a suitable framework in order to check the CL–QA system precision.

The set of questions is composed of "*factoid questions*" and "*definition questions*". The factoid questions are fact–based questions, asking for the name of a person, a location, the extent of something, the day on which something happened, etc.

Furthermore, with regard to the corpora created for training the NE recognizer, we have carried out the following strategy. We have manually annotated all the question datasets (2004, 2005 and 2006) and in order to apply NER to the 2006 question set, we have used as a training corpora the question sets belonging to 2004 and 2005 editions. For the 2005 question set, the 2004 and 2006 datasets were used as a train, and finally, for the 2004 question set we have merged the 2005 and 2006 question sets in order to create the training corpus.

Regarding WMK, we have used the English database dump provided by Wikimedia (enwiki-20061104) and specifically the page, pagelinks and langlinks data.

6.2 Result Analysis

The aim of these experiments is to evaluate the impact of applying WMK to our CL–QA system. We show the recall performance for translating entities obtained by both ILI and WMK. Besides, we provide the precision of our CL–QA system and compare it with the precision of our monolingual system.

From the NEs that should be translated, table 4 shows the percentage that are translated by using ILI and from the NEs not translated by ILI it shows the percentage that is translated by WMK. Although ILI is able to translate barely half of the NEs (57,3% for LOC, 39,8% for ORG and 59,6% for MISC, i.e. 39.1%

Table 4. NEs translated by ILI and WMK

Dataset		**ILI**	**WMK**
CLEF'04	PER	-	100%
	LOC	54.5%	90%
	ORG	29.4%	75%
	MISC	85%	100%
CLEF'05	PER	-	100%
	LOC	54.5%	93.3%
	ORG	10.5%	94.1%
	MISC	31.3%	90.9%
CLEF'06	PER	-	100%
	LOC	62.8%	84%
	ORG	50%	88.8%
	MISC	62.5%	88.8%
Average	PER	-	100%
	LOC	57.3%	89.1%
	ORG	39.8%	86%
	MISC	59.6%	93.2%
	TOTAL	39.2%	92.1%

in TOTAL), this is overcome by applying WMK (100% for PER, 89,1% for LOC, 86% for ORG and 93,2% for MISC, i.e. 92.1% in TOTAL).

Table 5 shows the precision of our system[10] in the CL scenario (questions in English and documents in Spanish) compared with the monolingual one (questions and documents in Spanish) for the questions sets of CLEF 2004, 2005 and 2006. Regarding the CL scenario, we not only show the total precision, but also provide the percentage of precision that is obtained thanks to the use of WMK (see second row of table 5). The importance of applying WMK to the CL–QA system is corroborated by these results as around 20% of the questions are correctly answered because of the incorporation of this module (18% for 2004, 23.5% for 2005 and 16% for 2006).

Table 5. QA system evaluation

Dataset		**Prec.**
English Questions (CL, % total answered)	CLEF'06	44%
	CLEF'05	42.5%
	CLEF'04	33.5%
English Questions (CL, % answered by using WMK)	CLEF'06	16%
	CLEF'05	23.5%
	CLEF'04	18%
Spanish Questions (monolingual, % total answered)	CLEF'06	50.5%
	CLEF'05	51.5%
	CLEF'04	41.5%

Compared to other state–of–the–art CL–QA systems, our approach obtains better results [10]. In fact, our precision loss of CL with respect to the monolingual run is around 17% whereas in the English–Spanish QA task at CLEF 2006 [15] the precision on English–Spanish CL–QA task was approximately 50% lower than for the monolingual Spanish task.

7 Conclusions

This paper has presented a novel approach that consists of applying multilingual knowledge encoded in Wikipedia to a CL–QA system based on the ILI module of EWN in order to improve the translation of NEs contained in the input questions. This original strategy to use WMK within CL–QA is motivated by two reasons that are proved by the evaluation results presented in the current paper: (i) the small percentage of NEs referenced in ILI (39.2% of NEs that should be translated in CLEF 2004, 2005 and 2006 questions) and (ii) the need to translate NEs in CL–QA environments (41.2%). A study that demonstrates the latter hypotheses has been presented and discussed.

[10] To calculate this value, both correct and the inexact answers that contain more information than that required by the query are considered.

The proposed approach has been evaluated on CLEF 2004, 2005 and 2006 English–Spanish CL–QA questions. For each year question set we provide the percentage of NEs that is translated by using ILI and, from the remaining NEs (those that ILI does not translate), the percentage that gets translated by applying WMK. The results prove that although ILI leaves a considerable percentage of NEs untranslated (ILI successfully translates between 39,8% and 59,6% of the entities), WMK succeeds to translate on average between 86% and 100% of these NEs depending on the entity type. Moreover, around 20% of the input questions are correctly answered by the CL–QA system as a consequence of using WMK. Besides, our CL–QA system has been evaluated by comparing the precision obtained at both CL and monolingual scenarios. The precision loss remains lower (around 17%) than for other state-of-the-art systems (around 50%).

Another contribution of this paper is the release as free software of the software tools used to process and gather information from Wikimedia database dumps.

Finally, as a future work proposal, we would like to take advantage of the knowledge that can be acquired by employing both multilingual resources incorporated in our system (ILI and WMK). In order to do this we plan to study strategies to combine in different ways the knowledge present in both resources.

References

1. Acebo, S., Ageno, A., Climent, S., Farreres, J., Padró, L., Placer, R., Rodriguez, H., Taulé, M., Turno, J.: MACO: Morphological Analyzer Corpus-Oriented. ESPRIT BRA-7315 Aquilex II, Working Paper 31 (1994)
2. Adafre, S., Jijkoun, V., Rijke, M.: The University of Amsterdam at WiQA 2006. In: Workshop of Cross-Language Evaluation Forum (CLEF), Alicante, Spain (2006)
3. Adafre, S., Rijke, M.: Finding similar sentences across multiple languages in wikipedia. In: EACL 2006 Workshop on New Text, Wikis and Blogs and Other Dynamic Text Sources, Trento, Italy (2006)
4. Ahn, D., Jijkoun, V., Mishne, G., Müller, K., de Rijke, M., Schlobachz, S.: Using Wikipedia at the TREC QA Track. In: Proceedings of the Thirteenth Text Retrieval Conference (TREC 2004) (2005)
5. Bos, J., Nissim, M.: Cross-Lingual Question Answering by Answer Translation. In: Workshop of Cross-Language Evaluation Forum (CLEF) (2006)
6. Bowden, M., Olteanu, M., Suriyentrakorn, P., Clark, J., Moldovan, D.: LCC's PowerAnswer at QA@CLEF 2006. In: Workshop of Cross-Language Evaluation Forum (CLEF) (2006)
7. Ferrández, S., Ferrández, A.: Cross-lingual question answering using inter lingual index module of eurowordnet. Advances in Natural Language Processing. Research in Computing Science 18, 177–182 (2006) ISSN: 1665-9899 18
8. Ferrández, S., Ferrández, A.: The Negative Effect of Machine Translation on Cross-Lingual Question Answering. In: Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Processing. LNCS, vol. 4394, pp. 494–505. Springer, Heidelberg (2007)
9. Ferrández, S., Ferrández, A., Roger, S., López-Moreno, P., Peral, J.: BRILI, an English-Spanish Question Answering System. In: Proceedings of the International Multiconference on Computer Science and Information Technology. 1st International Symposium Advances in Artificial Intelligences and Applications (AAIA '06), pp. 23–29 (November 2006) ISSN 1896-7094

10. Ferrández, S., López-Moreno, P., Roger, S., Ferrández, A., Peral, J., Alavarado, X., Noguera, E., Llopis, F.: AliQAn and BRILI QA System at CLEF-2006. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
11. Ferrández, S., Roger, S., Ferrández, A., Aguilar, A., López-Moreno, P.: A new proposal of word sense disambiguation for nouns on a question answering system. Advances in Natural Language Processing. Research in Computing Science 18, 83–92 (2006) ISSN: 1665-9899
12. García-Cumbreres, M.A., Ureña-López, L.A., Martínez-Santiago, F., Perea-Ortega, J.M.: BRUJA System. The University of Jaén at the Spanish task of CLEFQA 2006. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
13. Gillard, L., Sitbon, L., Blaudez, E., Bellot, P., El-Béze, M.: The LIA at QA@CLEF-2006. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
14. Llopis, F., Vicedo, J.L.: Ir-n, a passage retrieval system. In: Workshop of Cross-Language Evaluation Forum (CLEF) (2001)
15. Magnini, B., Giampiccolo, D., Forner, P., Ayache, C., Jijkoun, V., Osevona, P., Peñas, A., Rocha, P., Sacaleanu, B., Sutcliffe, R.: Overview of the CLEF 2006 Multilingual Question Answering Track. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
16. Magnini, B., Negri, M., Preete, R., Tanev, H.: A wordnet-based approach to named entities recognition. In: Proceedings of SemaNet '02: Building and Using Semantic Networks, Taipei, Taiwan, pp. 38–44 (2002)
17. Moldovan, D.I., Pasca, M., Harabagiu, S.M., Surdeanu, M.: Performance issues and error analysis in an open-domain question answering system. ACM Trans. Inf. Syst 21, 133–154 (2003)
18. Sacaleanu, B., Neumann, G.: DFKI-LT at the CLEF 2006 Multiple Language Question Answering Track. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
19. Schmid, H.: TreeTagger — a language independent part-of-speech tagger. Institut fur Maschinelle Sprachverarbeitung, Universitat Stuttgart (1995)
20. Sutcliffe, R.F.E., White, K., Slattery, D., Gabbay, I., Mulcanhy, M.: Cross-language French-English Question Answering using the DLT System at CLEF 2006. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
21. Tomás, D., Vicedo, J.L, Bisbal, E., Moreno, L.: Experiments with LSA for Passage Re-Ranking in Quesion Answering. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)
22. Vossen, P.: Introduction to eurowordnet. Computers and the Humanities 32, 73–89 (1998)
23. Whittaker, E.W.D., Novak, J.R., Chatain, P., Dixon, P.R., Heie, M.H., Furui, S.: CLEF2005 Question Answering Experiments at Tokyo Institute of Technology. In: Workshop of Cross-Language Evaluation Forum (CLEF) (September 2006)