BRILI, an English-Spanish Cross-Lingual Question Answering System⋆

Sergio Ferrández, Antonio Ferrández, Sandra Roger, Pilar López-Moreno and Jesús Peral

Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante, Spain
{sferrandez,antonio,sroger,jperal}@dlsi.ua.es, P.Lopez@ua.es

Abstract. This paper describes the BRILI Cross-lingual Question Answering (CL-QA) system. The BRILI system is capable to answer English questions from Spanish documents. It has participated in the 2006 edition of the Cross-Language Evaluation Forum (CLEF). Some new characteristics in BRILI are described, especially the strategy used for the question processing module in which the Inter Lingual Index (ILI) Module of EuroWordNet (EWN) is used with the aim of reducing the negative effect of question translation on the overall accuracy. The experimental evaluation on the CLEF 2004 set of questions reveals a precision of 34%, which implies an improvement of around 20% in precision in comparison with the monolingual QA system using machine translation. These experiments prove that our system obtains better results than using machine translation and other current bilingual QA systems.

1 Introduction

Information retrieval (IR) is the science of searching for information in documents. Question Answering (QA) is not a simple task of IR. The aim of a QA system is to localize the exact answer to a question in a non-structured collection of documents.

In a Cross-Lingual (CL) framework, the question is formulated in a language different from the one of the documents, which increases the difficulty. Nowadays, multilingual QA systems have been recognized as an important issue for the future of IR because of the multilingual nature of the available information. The multilingual QA tasks were introduced in the Cross-Language Evaluation Forum (CLEF) 2003 for the first time.

This paper describes the development stages of the English–Spanish BRILI system (Spanish acronym for "Question Answering using Inter Lingual Index Module"), at the same time presenting the results it achieves for the official questions of CLEF 2004 (the prototype architecture of the BRILI system was previously presented in [7]). BRILI is fully automatic, including the modules of language identification and question processing.

The main goal of this paper is to describe the used strategy in order to face the English-Spanish QA bilingual task, which has been developed using the Inter Lingual Index (ILI) module of EuroWordNet (EWN).

The rest of the paper is organized as follows: section 2 describes the state of the art studying the effects of question translation on the precision of most current CL-QA systems. Afterwards, the architecture of BRILI is shown and analyzed in section 3. Section 4 presents and discusses the results obtained using the English questions of QA CLEF 2004 and searching for the answer in the Spanish corpus. And finally, section 4 details our conclusions and future work.

2 State of the art

The overall accuracy of a QA system is directly affected by its ability to correctly analyze the question it receives. In the case of a CL-QA system, the precision of the system depends on the correct translation and analysis of the questions that are received as input. An imperfect translation of the question causes a negative impact on the overall accuracy of the systems. As Moldovan [15] stated, Question Analysis phase is responsible for 36.4% of the total of number of errors in open-domain QA.

Currently, four different approaches are used by CL-QA systems in order to solve the bilingual task in which, the question and the documents are in different languages. The first one [3, 5, 9] uses an automatic Machine Translation (MT) system to translate the question into the language in which the documents are written. On the other hand, other systems [4, 12] translate the question using bilingual dictionaries. Beside, some systems [2, 10, 11, 16, 19] procese different translations from several MT. Finally, some implementations [20, 21] use a set of pre-processing transformation rules in order to correct the wrong output produced by the automatic MT. Next, these strategies are explained.

The translation of the question into the language of the documents, without any further processing, is the simplest technique available. When compared to one of the precision of around 25% achieved in the Spanish monolingual task, the system loses about 30% of this precision in the CL task.

In the case where several translations are processed, the strategy used in order to detect the optimal translation is based on the use of *trigram chains* for each translation, these chains are submitted to a web search engine obtaining the web count of each trigram, the optimal translation is the one with the highest trigram chain weight. Other strategy for the combination of translations is based on their pertinence to the target document collection rather than on their grammatical correctness. Beside, there is other kind of system in which the translations returned by the machine translation are ranked according to a language model. In this kind of implementations that obtain a precision in monolingual Spanish task of around 33%, the lost of overall accuracy range from 30 to 49% when all the question proposed by CL tasks of CLEF 2005 are used.

In the systems that use bilingual dictionaries in order to face the bilingual QA tasks, the lost of precision is about 40% reaching the 60% when the bilingual task is between Italian and French (which is quite high than in other systems). Notice that the precisi is 64% in the monolingual French QA task, therefore the lost of precision is higher.

Finally, some systems apply rules to the original question before carrying out the translations. They are based on the using of lexical rules (for example: the word *"Dov'e"* is substituted for the words *"Dove é"*) and on the using of a shallow parsing phase on the question in order to look for different types of phrases. Each phrase is independently translated using a translation engine. The overall accuracy of this type of systems suffers a lost of 32% in relation to the monolingual task that obtains a precision of 27.50% when the Bulgarian question of CLEF 2005 are used.

All of the previous strategies described are based on the translation of the question and some of them try to correct the translation mistakes through different implementations as it has been pointed out.

The low quality of MT provides a load of errors inside all the steps of the localization of the answer. This fact was corroborated on the last edition of CLEF 2005 [22] where the precision on CL English–Spanish task of the best system was approximately 33% lower than for monolingual Spanish task. And it can be checked observing the official results from the best CL-QA system, whose precision is 64% for the monolingual French task and 39.5% when the questions are written in English.

For CL-QA, translations are often inexact and quite fuzzy, beside, the MT systems resolve the ambiguity by means of only giving one translation per word. These facts cause an important decrease on the precision of the systems. For instance, MT systems generate errors such as translations of names that should be left untranslated. The impact of this kind of mistakes should be controlled and valued.

The BRILI system introduces two improvements that alleviate the negative effect produced by the MT: (1) BRILI considers more than only one translation per word by means of using the different synsets of each word in the ILI module of EuroWordNet ; (2) unlike the current bilingual English–Spanish QA systems, the question analysis is developed in the original language without any translation. Next, both new techniques are detailed.

3   The BRILI system

3.1   Overview

In this section, the architecture and functionality of our method to open domain CL-QA system are detailed.

BRILI is designed to localize answers from documents, where both answers and documents are written in different languages. The system is based on complex pattern matching using NLP tools [1, 6, 13, 18]. Beside, Word Sense Disambiguation (WSD) is applied to improve the precision of the system (a new proposal of WSD for nouns presented in [8]).

Its architecture is shown in Figure 1. It carries out an indexation phase where all the documents are analyzed in its own language, in which syntactic and semantic information is stored.
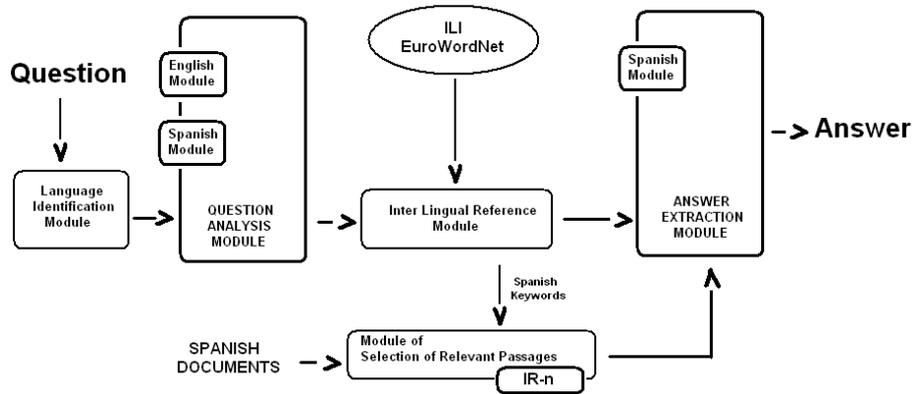
Fig. 1. System architecture

For the first implementation of the BRILI system, we have used the indexation phase of documents and the answer extraction module of our monolingual Spanish QA system [17] which was presented for the first time at the 2005 edition of the CLEF) and it has also participated in the 2006 edition.

The modules of language identification, the phase of Question Analysis, the Inter Lingual Reference (ILR) module of BRILI system, the selection of relevant passages and examples of the extraction of answers are described below.

3.2 The language identification module

The module of language identification have been developed to automatically distinguish the correct language of the question and documents. It is based on two main techniques: the use of dictionaries (specific language stopwords) and the use of part-of-word terminology (for example, "ing" in the case of English). This philosophy presents a good precision [14] in Spanish and English.

3.3 The question analysis nodule

In this section, the Question Analysis of BRILI is described. Previously, some aspect must be commented. Our method is based fundamentally on syntactic analysis of the question and documents. In order to make this task, SUPAR [6] system is used, which works in the output of a PoS tagger [1, 18]. Using the output of SUPAR, BRILI extracts syntactic blocks (SB), such as verb phrase (VP), simple nominal phrase (NP), and simple prepositional phrase (PP). These are our basic syntactic units when defining patterns.

BRILI develops two main tasks in the phase of Question Analysis, these are shown below:

– **To detect the expected answer type.** Therefore, BRILI detects the type of information that the answer has to satisfy to be a candidate of an answer (proper name, quantity, date, . . .)
– **To identify the main SBs of the question.** BRILI extracts the SB that are necessary to find the answers.

In the first task, a taxonomy has been designed, which is based on WordNet Based-Types and EuroWordNet Top-Concepts. Our taxonomy consists of the next categories: person, profession, group, object, place city, place country, place capital, place, abbreviation, event, numerical economic, numerical age, numerical measure, numerical period, numerical percentage, numerical quantity, temporary year, temporary month, temporary date and definition.

The expected answer type is achieved using a set of syntactic patterns. BRILI has about 200 English patterns for the determination of the different semantic category of our ontology. The system compares the SB of the patterns with the SB of the question, the result of the comparison determines the category of the question. The next example shows the behavior of this task:

– **Question 06 at CLEF 2006:** *Which country did Iraq invade in 1990?*
– **Pattern:** [WHICH] [synonym of COUNTRY]
– **Expect answer type:** place country

In the second task, the objective is to select the main SB of the question that make possible to locate the solution in the documents that can contain the answer. Beside, these SB contain the *"keywords"*, words that must be referenced using the ILR module into the language in which the documents are written. The next example shows the identification of the main SB of the question of type *year*:

– **Question 75 at CLEF 2006:** *In which year was the Football World Cup celebrated in the United States?*
– **SBs:**

```
[NP Football World Cup]
[VP to celebrate]
[PP in [NP United States]]
```

– **Keywords to be referenced in the ILR module:** Football World Cup celebrate United States

Furthermore, in this task, some SB can be discarded. The discarded SBs are not useful in order to carry out the searching of the answer. In the next example this behavior is shown, where the words *"can"* and *"be"* are discarded to the extraction of the answer phase:

– **Question 33 at CLEF 2004:** *How can an allergy be treated?*
– **SBs:**

```
[VP to treat]
[NP allergy]
```

– **Keywords to be referenced in the ILR module:** treat allergy

3.4   The Inter Lingual Reference (ILR) module

This section describes the ILR module of the BRILI system, and the adopted methodology to create references between words in different languages.

The ILR module carries out the process of linking the two languages involved in the cross-lingual QA process. It links words in the language of the question into words in the language of the documents. The inputs of the ILR module are the detected keywords in the previous phase of question analysis. The figure 2 shows the created links to the word *"prohibition"*:
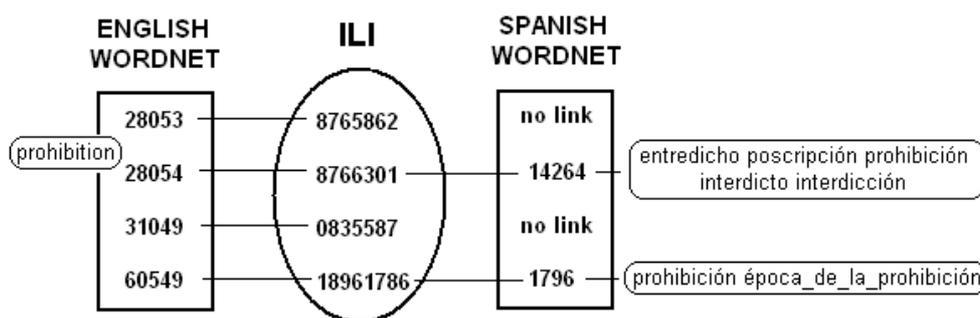

Fig. 2. Links to the word *"prohibition"*

In the previous example, the system finds more than one Spanish equivalents for one English word. The current strategy employed to solve this handicap consists of assigning more value to the word with the highest frequency. In the case of the previous example, the most valuated Spanish word would be *"prohibici"*.

On the other hand, the words that are not in EWN are translated into the rest of the languages using an on-line Spanish Dictionary[1]. Furthermore, BRILI uses gazetteers of organizations and places in order to translate words that have not linked using ILI. Therefore, in order to decrease the effect of incorrect translation of the proper names, the matches using these words in the search of the answer are realized

---

[1] http://www.wordreference.com

using the translated word and the original word of the question. The found matches using the original English word are valuated a 20% less.

Our approach does not achieve a MT of the question, it indexes the words using the ILI of EWN reducing the negative effect of question translation on the overall accuracy.

Next, a example is shown to detail the complete process of the ILR module:

– **Question 101 at CLEF 2004:** *What army occupied Haiti?*
– **SBs:**

```
[NP army]
[VP to occupy]
[NP Haiti]
```

– **Keywords to be referenced in the ILR module:** army occupy Haiti
  • army ↦ ejcito
  • occupy ↦ absorber **ocupar** atraer residir vivir colmar rellenar **ocupar** llenar
  • Haiti ↦ Hait
– **SB to search in Spanish documents:**

```
[NP ejcito]
[VP ocupar]
[NP Hait
```

3.5   The module of Selection of relevant passages

The phase of Selection of relevant passages uses the IR-n system [13]. The inputs of IR-n are the detected keywords and the translated words that are not in EWN. For instance, using the question 92 at CLEF 2006, *Where is Interpol's headquarters?*, IR-n receives as input the words: *"Interpol"* and *"comandancia"*. IR-n returns a list of passages where the system applies the answer extraction module.

3.6   The answer extraction module

The final step of BRILI is the phase of Extraction of the Answer which is composed of three monolingual modules. BRILI uses the syntactic blocks of the question and different sets of syntactic patterns (according to the language) with lexical, syntactic and semantic information to find out the correct answer. Next, an example of a syntactic pattern for Spanish is shown, which captures solution in a Spanish sentence.

– **Question 70 at CLEF 2005:** *Which French president inaugurated the Eurotunnel?*
– **SB:** [NP French president] [VP to inaugurate] [NP Eurotunnel]
– **Document:** *"... el Presidente Francés, **Jacques Chirac**, inauguró el Eurotunel ... "* (... the French President, Jacques Chirac, inaugurated the Eurotunnel ...)
– **Syntactic pattern**:
  [NP (*"French president"*), apposition [NP (SOLUTION)]] + [VP (*"to inaugurate"*)] + [NP *"Eurotunnel"*)]

4   Evaluation

The experiments described in this section have been carried out using the BRILI system and it has been compared with our monolingual Spanish QA system [17] with three Machine Translation system: MT1, MT2 and MT3[2].

In order to evaluate our new proposal of CL-QA system, the CLEF 2004 set of 200 English question was answered by searching the EFE 1994 Spanish corpora.

Table 1 shows the improvement in relation to the using the MT1 (see column 4) and the lost of precision[3] in the answer extraction in relation to the Spanish monolingual QA task (see column 5).

In the experiments using MT services, the errors produced by the question translation (see rows 4, 5 and 6) generate worse results than using the BRILI system (+19.12%, see row 3). Again, the experimental evaluation showed up the negative effect of the MT approach.

These experiments prove that our system obtains better results than using MT and other current bilingual QA systems. Beside, this affirmation can be corroborated checking the official results on the last edition of CLEF 2005 [22] where the precision on English–Spanish CL task of the best system was approximately 33% lower than the same system for monolingual Spanish task (BRILI only a 11.68% lower). On the other hand, the official results from the best English–French CL-QA system (−38.28%), also demonstrate this conclusion.

---

[2] MT1: http://ets.freetranslation.com,
  MT2: http://www.systransoft.com and MT3: http://babelfish.altavista.com
[3] Correct answers return on the first place.

Table 1. Improvement in relation to the using the MT1 system and lost of precision in relation to the Spanish monolingual task

System	Task	Precision %	Improvement %	Lost %
Baseline	Spanish–Spanish	38.5	+28.44	-
BRILI	English–Spanish	34	**+19.12**	**−11.68**
MT1	English–Spanish	27.5	-	−28.44
MT2	English–Spanish	29.5	+6.78	−22.27
MT3	English–Spanish	29.5	+6.78	−22.27

5   Conclusions and Future Work

In this paper, the BRILI CL-QA system is presented and evaluated in its English–Spanish version.

The BRILI system reduces the using of MT services avoiding the negative effect that causes this kind of strategies in CL-QA system by means of the using of the ILI module of EuroWordNet.

Beside, two improvements that alleviate the negative effect produced by the MT systems, are introduced. BRILI considers more than one translation per word by means of using the different synsets of each word in the ILI module of EuroWordNet and unlike the current bilingual English–Spanish QA system, the question analysis is developed in the original language without any translation.

The experiments on the CLEF 2004 English questions prove that the BRILI system generates better results than using MT (+19.12%) and than other current bilingual QA systems (−11.68% with regard to the English–Spanish one with −33%).

In the future, the system will be capable to answer English, Spanish and Catalan questions from English, Spanish and Catalan documents. Beside, Word Sense Disambiguation algorithm will be applied to the questions in order to sort and weight the different meanings (synsets) of the words.

On the other hand, a Name Entity Recognition will be applied to the question in order to detect possible names of persons that will not be translated. For instance, using the question 112 at CLEF 2004, *Who is Bill Clinton?*, the words "Bill Clinton" are not translated.

References

1. S. Acebo, A. Ageno, S. Climent, J. Farreres, L. Padró, R. Placer, H. Rodriguez, M. Taulé, and J. Turno. *MACO: Morphological Analyzer Corpus-Oriented.* ESPRIT BRA-7315 Aquilex II, Working Paper 31, 1994.
2. R. M. Aceves-Pérez, L. Villaseñor-Pineda, and M. Montes y Gómez. *Using n-gram models to combine query translations in cross-language question answering,* In Alexander F. Gelbukh, editor, *CICLing*, volume 3878 of *Lecture Notes in Computer Science*, pages 453–457. Springer, 2006.
3. M. Adriana and Rinawati. *University of Indonesia Participation at Query Answering-CLEF 2005.* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
4. L. Aunimo, J. Makkonen, and R. Kuuskoski. *Cross-Language Question Answering for Finnish.* Proceedings of the Web Intelligence Symposium, Finnish Artificial Intelligence Conference, September 2004.
5. C. de Pablo-Sánchez, A. González-Ledesma, J.L. Martínez-Fernández, J.M. Guirao, P. Martínez, and A. Moreno. *MIRACLE's 2005 Approach to Cross-Lingual Question Answering.* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
6. A. Ferrández, M. Palomar, and L. Moreno. *An Empirical Approach to Spanish Anaphora Resolution.* Machine Translation. Special Issue on Anaphora Resolution In Machine Translation, **14**(3/4):191–216, December 1999.
7. S. Ferrández and A. Ferrández. *Cross-lingual question answering using inter lingual index module of eurowordnet,* Advances in Natural Language Processing. Research in Computing Science. ISSN: 1665-9899, 18:177–182, February 2006.
8. S. Ferrández, S. Roger, A. Ferrández, A. Aguilar, and P. López-Moreno. *A new proposal of word sense disambiguation for nouns on a question answering system,* Advances in Natural Language Processing. Research in Computing Science. ISSN: 1665-9899, **18**:83–92, February 2006.
9. M. A. García-Cumbreras, L. A. Ureña, and F. Martínez-Santiago. *BRUJA: Question Classification for Spanish. Using Machine Translation and an Englsih Classifier.* Workshop on Multilingual Question Answering. EACL 2006 .ISBN: 2-9524532-4-1, April 2006.
10. J. M. Gómez, E. Bisbal, D. Buscaldi, and P. Rosso E. Sanchis. *Monolingul and Cross-Language QA using a QA-oriented Passage Retrieval System,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
11. B. Grau, A. Ligozat, I. Robba, M. Sialeu, and A. Vilnat. *Term Translation Validation by Retrieving Bi-Terms,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.

12. D. Laurent, P. Séguéla, and S. Negre. *Cross Lingual Question Answering using QRISTAL for CLEF 2005,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
13. F. Llopis and J. L. Vicedo. *Ir-n, a passage retrieval system,* In Workshop of Cross-Language Evaluation Forum (CLEF), 2001.
14. T. Martínez, E. Noguera, R. Muñoz, and F. Llopis. *Web track for CLEF2005 at ALICANTE UNIVERSITY,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
15. D. I. Moldovan, M. Pasca, S. M. Harabagiu, and M. Surdeanu. *Performance issues and error analysis in an open-domain question answering system,* ACM Trans. Inf. Syst, **21**:133–154, 2003.
16. G. Neumann and B. Sacaleanu. *DFKI's LT-lab at the CLEF 2005 Multiple Language Question Answering Track,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
17. S. Roger, S. Ferrández, A. Ferrández, J. Peral, F. Llopis, A. Aguilar, and D. Tomás. *AliQAn, Spanish QA System at CLEF-2005,* In Workshop of Cross-Language Evaluation Forum (CLEF), 2005.
18. H. Schmid. TreeTagger — a language independent part-of-speech tagger. *Institut fur Maschinelle Sprachverarbeitung, Universitat Stuttgart*, 1995.
19. R. Strótgen, T. Mandl, and R. Schneider. *A Fast Forward Approach to Cross-Lingual Question Answering for English and German,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
20. R. F. E. Sutcliffe, M. Mulcahy, and I. Gabbay. *Cross-Language French-English Question Answering Using the DLT system at CLEF 2005,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
21. H. Tanev, M. Kouylekov, B. Magnini M. Negri, and K. Simov. *Exploring Linguistic Indices ans Syntactic Structures for Multilingual Question Answering: ITC-irst at CLEF 2005,* In Workshop of Cross-Language Evaluation Forum (CLEF), September 2005.
22. A. Vallin, B. Magnini, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peñas, M. Rijke, B. Sacaleanu, D. Santos, and R. Sutcliffe. *Overview of the CLEF 2005 Multilingual Question Answering Track,* Comparative Evaluation of Multilingual Information Access Systems: 6th Workshop of the Cross-Language Evaluation Forum, CLEF, Trondheim, Norway, August 21-22, 2003, Revised Selected Papers.Lecture Notes in Computer Science, 2005.