

Priberam's question answering system in a cross-language environment

Adán Cassan, Helena Figueira,
André Martins, Afonso Mendes, Pedro Mendes, Cláudia Pinto, Daniel Vidal

Priberam Informática
Alameda D. Afonso Henriques, 41 - 2º Esq.
1000-123 Lisboa, Portugal
Tel.: +351 21 781 72 60
Fax: +351 21 781 72 79

{ach, hgf, atm, amm, prm, cp, dpv}@priberam.pt

Abstract

Following last year's participation in the monolingual question answering (QA) track of CLEF, where Priberam's QA system achieved state-of-the-art results, this year we decided to take part in both Portuguese and Spanish monolingual tasks, as well as in two bilingual (Portuguese-Spanish and Spanish-Portuguese) tasks of QA@CLEF. The architecture of our QA system relies on previous work done for a multilingual semantic search engine developed in the framework of TRUST¹, where Priberam was responsible for the Portuguese module. Unlike TRUST, however, which used a third party indexing engine, our QA system is based on the indexing technology of LegiX, Priberam's legal information tool², whose indexing engine was adapted to index semantic information, ontology domains, question categories and other specificities for QA. Given the multilingual platform where the system has been developed and tested, as well as the results obtained so far, it seemed natural to assess its language independence. To that intent, we have extended it to another language, Spanish, thus demonstrating the applicability of the system architecture. This paper describes the improvements and changes implemented in Priberam's QA system since last CLEF participation, detailing the work involved in its cross-lingual extension and discussing the results of the runs submitted to evaluation.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2 [Database Management]: H.2.3 Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation, Languages

Keywords

Question answering, Questions beyond factoids

¹TRUST – Text Retrieval Using Semantic Technologies – was a European Commission co-financed project (IST-1999-56416), whose aim was the development of a multilingual semantic search engine capable of processing and answering natural language questions in English, French, Italian, Polish and Portuguese.

²See <http://www.legix.pt> for more information.

1 Introduction

Priberam took part in the 2005 CLEF campaign, introducing its QA system for Portuguese in the monolingual task. This year, encouraged by last year's results [1], we decided to participate in the same track, but we extended the participation to the Spanish monolingual and bilingual (Portuguese-Spanish and Spanish-Portuguese) tasks.

The architecture of our QA system remains roughly the same, as detailed in [2, 3]: after the question is submitted, it is categorised according to a question typology. An internal query retrieves a set of potentially relevant documents, containing a list of sentences related with the question. Sentences are weighted according to their semantic relevance and similarity with the question. Next, through specific answer patterns, these sentences are examined once again and the parts containing possible answers are extracted and weighted. Finally, a single answer is chosen among all candidates.

This year, we focused on the handling of temporally restricted questions, the addition of another language, and the adaptation of the system to a cross-language environment. Priberam relied on the scalability of the system's architecture to cope with the inclusion of another language module, Spanish. Currently, two M-CAST³ partners, the University of Economics, Prague (UEP) and TiP, are also using this framework to develop Czech and Polish language modules. The purpose of this year's participation in QA@CLEF was to evaluate the language independence of the system, as well as its performance in a bilingual context.

The next section gives an overview of the work done for the inclusion of another language, describing the development of the Spanish module. Section 3 addresses the various improvements in Priberam's QA system architecture and depicts the cross-language scenario. Section 4 discusses both the monolingual and bilingual results of the system at this year's QA@CLEF campaign, and section 5 concludes with future guidelines.

2 Addition of Spanish

Taking advantage of the company's natural language processing (NLP) technology and workbench [3, 4], it was somewhat straightforward to build a new language module for Spanish, similar to the Portuguese one. The QA system architecture was designed to be language independent, and by using the same software tools, like Priberam's SintaGest, new language modules can easily be implemented and tested. This means that only the language resources (lexicon, thesaurus, ontology, QA patterns) have to be adapted or imported to be in conformity with the existing NLP tools. With a team of four people, the work on the lexicon took us about three months to complete, because manual work was involved, and the adaptation and development of the QA rules took about two months, as it had to be tested while it was being developed.

The Spanish ontology was added to the common multilingual ontology, in a joint work of Priberam and Synapse Développement. As detailed in [2], this multilingual taxonomy groups over 160 000 words and expressions through their conceptual domains organised in a four level tree with 3 387 terminal nodes.

Priberam started by acquiring a Spanish lexicon and converting it to the same format and specifications of the Portuguese one. After loading the existent lexical information in a database, we had to establish equivalencies between POS categories, uniform them and classify all the lexical entries, so the Spanish lexicon could be used with the tools that were used to build the Portuguese module. New entries in the lexicon were also inserted, mainly proper nouns, such as toponyms and anthroponyms. This was particularly important for the recognition of named entities (NEs).

Unlike the Portuguese lexicon, the Spanish one does not contain any semantic features, nor sense definitions connected to the ontology levels. The semantic classification has already been

³M-CAST – Multilingual Content Aggregation System based on TRUST Search Engine – is an European Commission co-financed project (EDC 22249 M-CAST), whose aim is the development of a multilingual infrastructure enabling content producers to access, search and integrate the assets of large multilingual text (and multimedia) collections, such as internet libraries, resources of publishing houses, press agencies and scientific databases (<http://www.m-cast.infovide.pl>).

started but we still have to define senses for each entry in the lexicon and link each one of them to the levels of the ontology. This future work in the Spanish lexicon will hopefully lead to similar results of both language modules. There is also still no thesaurus for Spanish, which may influence the retrieval of documents and sentences that contain synonyms of the question's keywords.

Some relations between words were semi-automatically established, such as derivation relations (e.g. *caracterizar/caracterización*), names of toponyms and their gentilics (e.g. *Australia/australiano*), currencies (e.g. *Grecia/euro*) and cities (e.g. *España/Madrid*). Other relations, such as hyperonymy, hyponymy, antonymy and synonymy, still need to be implemented in the Spanish lexicon and improved in the Portuguese one.

The system's overall design remained unchanged for Spanish; the question patterns, answer patterns and question answering patterns [3] were adapted to fit the new language module. The typology of the 86 question categories was not altered, since it is language independent. Due to the syntactical similarities between the two Romance languages, many of the Portuguese patterns remained applicable. Groups of semantically related words and question identifiers had to be translated and revised, which was very helpful to improve and fix some errors of the Portuguese language module.

For the Spanish module, some of the Portuguese contextual rules, such as the ones for morphological disambiguation and for NEs recognition, were rewritten and adapted as well. Again, the similarity between Portuguese and Spanish allowed us to adapt much of the work done previously for Portuguese without having to start everything from scratch. Here the work was mostly done with constants and entity identifiers [3], whereas the detection rules were just adapted in some cases.

Like in Portuguese, Spanish morphological disambiguation is done in two stages: first, the contextual rules defined in SintaGest are applied; then, remaining ambiguities are suppressed with a statistical POS tagger based on a second-order hidden Markov model [5, 6]. For training, we used a corpus previously disambiguated with SVMTool⁴ [7].

3 Cross-language and improvements of the system architecture

The five main steps of Priberam's QA system are:

- The *indexing process*, in which a large set of files in text format is analysed and index keys for morphologically disambiguated lemmas, question categories and ontology domains are created;
- The *question analysis*, in which a question is parsed, categories for the question are determined and pivots and other search keys are extracted;
- The *document retrieval*, in which a query is made to the index database and a set of document sentences is retrieved;
- The *sentence retrieval*, in which each sentence previously retrieved is parsed and given a score to express its likelihood of containing an answer;
- The *answer extraction*, in which a unique answer is selected from the best-scored sentences, by means of extraction patterns.

This year some minor changes were introduced regarding: (i) the handling of temporally restricted questions, (ii) the final validation of the extracted answers, and (iii) the adaptation of the system to a cross-language environment. The next subsections describe these three topics in more detail.

⁴SVMTool is developed by TALP Research Center NLP group, of Universitat Politècnica de Catalunya (<http://www.lsi.upc.es/~nlp/SVMTool/>).

Additionally, other small improvements were made, such as the implementation of a priority based scheme to make question categorisation more assertive, and the use of inexact matching techniques (based on the Levenshtein distance) for partial matching of proper nouns in the sentence retrieval module. The latter was done adapting existing technology for the Portuguese spell checker⁵. Future work will address extending the use of inexact matching techniques in the document retrieval module, which will prevent the exclusion of documents where a proper noun is misspelled or spelled differently from the question, which is included in the causes of failure in the cross-language task (cf. section 4).

3.1 Handling of temporally restricted questions

Improving the system’s ability to handle temporally restricted questions was one of the goals for this year’s QA@CLEF. The organisation distinguishes three types of temporal restrictions:

- Restriction by *date*: e.g., “Who was the US president in 1962?”;
- Restriction by *period*: e.g., “How many cars were sold in Spain between 1980 and 1995?”;
- Restriction by *event*: e.g., “Where did Michael Milken study before enrolling in the University of Pennsylvania?”.

Our approach focuses on the two first types of restrictions. We take into account two sources of information: (i) the dates of documents, and (ii) temporal expressions in the documents’ text.

The documents’ dates in the collections *EFE* (for Spanish), *Público* (for European Portuguese) and *Folha de São Paulo* (for Brazilian Portuguese) are an instance of metadata. To exploit this source, we provided our system with the ability to deal with metadata information. This is a common requirement of real-world systems in most domains (the Web, digital libraries and local archives). Our procedure to deal with dates is extensible to other kinds of metadata, such as the document’s title, subject, author information, etc. This is currently being done for M-CAST project, applied to digital libraries. In many cases, questions with restrictions require a hybrid search procedure, using both metadata Boolean-like queries and NLP techniques. The same hybrid behaviour is required to handle temporally restricted questions. During indexation, our system finds the adequate piece of metadata containing the date of each document which is then indexed. This makes the system able to handle Boolean-like date queries (such as “select all documents dated above 1994-10-01 and below 1994-10-15”).

A more difficult issue is the recognition of temporal expressions in natural language text. While *absolute* dates like “25th April, 1974” are easy to recognise and convert to a numeric format, the same does not happen when the dates are *incomplete* (“25th April”, “April 1974”, “1974”), when the temporal expressions are less conventional (“spring 1974”, “2nd quarter of 1974”), or when we consider *temporal deictics* (like “yesterday”, “last Monday”, “20th of the current month”) that require knowledge about the discourse date.

As a first approach, we were only concerned with temporal expressions that refer to possibly incomplete absolute dates or periods. To answer temporally restricted questions, we need to perform operations over numeric representations of dates and periods, like *date comparison* (is 1974-04-25 greater, equal or lower than 1974-05-01?), *translation of time units* (add 8 days to 1974-04-25), or *checking if a date lies in a period* (is 1974-04-28 in the interval [1974-04-25, 1974-05-01]?). Although with a full numeric representation of dates these operations are exact and well defined, the need to deal with incomplete dates requires some fuzziness.

The procedure for answering a temporally restricted question is as follows: during question analysis, the temporal expression corresponding to the restriction is recognised and converted to the above numeric format. The document retriever module *relaxes* the restriction into a larger period (starting a few days before, and ending a few days after), and applies a metadata query to

⁵The Portuguese spell checker is included in FLiP, together with a grammar checker, a thesaurus, a hyphenator, a verb conjugator and a translation assistant that enable different proofing level – word, sentence, paragraph and text – of European and Brazilian Portuguese. An online version is available at <http://www.flip.pt>.

retrieve a set T_1 of those documents whose dates are relevant. This “relaxation” approach seems reasonably appropriate for newspaper corpora, since it works both for news about events that already occurred, and for announcements of incoming events. As an example, consider the question “Contra que clube jogou o Marítimo a 2 de Abril de 1995?” [Against what team did Marítimo play on the 2nd April 1995?]. Here T_1 will contain all documents from the 29th March until the 5th April 1995. After this, another query is applied to retrieve a set T_2 of documents containing temporal expressions that match the original (non-relaxed) temporal restriction of the question. Here, fuzzy matches are accepted (although having a lower score) to deal with incomplete dates. An OR-operation of these two queries retrieves a set $T = T_1 \cup T_2$ of documents that possibly satisfy the restriction either because of their date or because of the temporal expressions they contain. Finally, the usual document retrieval procedure continues as in the unrestricted case, with the difference that the top-30 retrieved documents are constrained to belong to the set T . These documents are then analysed at sentence level by the sentence retrieval module and only those sentences that either have some temporal expression satisfying the restriction, or belong to a properly dated document, are kept for answer extraction.

3.2 Answer validation

Last year the absence of a strategy for final answer validation was considered one of the major causes of failure of our system. This year we have made a naïve approach to this problem. In order to strengthen the match between the question and the sentence containing the answer, we demand the following: (i) the answering sentence must match (at least partially) all the proper nouns and NEs in the question, and (ii) it must match a given amount of nominal and verbal pivots in the question. Here, a match is any correspondence of lemmas, heads of derivation, or synonyms. If a candidate answer is extracted from a sentence that fails one of these two criteria, it is discarded, although it can still be used for coherence analysis of other answers.

Notice that this approach ignores any relation among the question pivots and the extracted answer; it merely checks for matches of each pivot individually. Current work is addressing a more sophisticated strategy for answer validation through syntactic parsing. The idea is to capture the argument structure of the question, by setting a syntactical role to each pivot or group of pivots in the sentence, requiring that the answer to be extracted is contained in a specific phrase, and checking if the answering sentence actually offers enough support.

3.3 Adaptation of the system to a cross-language environment

Multilingual question answering has been introduced in this year Priberam’s participation, namely in the Portuguese-Spanish and Spanish-Portuguese tasks. The adaptation of the system to a cross-language environment required few modifications. Moreover, unlike most approaches, it is self-contained, in the sense that it does not require using any external software or Web searches to perform translations. Instead, we use an ontology based direct translation, refined by means of a statistical corpora based approach.

The central piece of our system for the cross-language tasks is the multilingual ontology, also used in last year’s CLEF for monolingual [3] and bilingual [8] purposes. The combination of the ontology information of all TRUST languages provides a bidirectional word/expression translation mechanism. Some language pairs are directly connected, as is the case of Portuguese and Spanish; others are connectable using the English language as an intermediate. This allows operating in a cross-language scenario for any pair of languages in the ontology (among Portuguese, Spanish, French, English, Polish and Czech⁶). For each language pair that is directly connected, translation scores are used to reflect the likelihood of each translation. By using this method, the system is capable of selecting the preferential equivalent among the available translations. For instance, in the case of the Spanish word *hijo* in the ontological domain [family/lineage], which has the Portuguese translations *filho* (son) and *criança* (child), the selected translation is *filho*, since it has

⁶The French, Polish and Czech parts of the ontology are property respectively of Synapse Développement, TiP and University of Economics, Prague (UEP).

a higher score. These scores are computed in a background task using the Europarl parallel corpus [9], a paragraph aligned corpus in the official languages of the European Union, containing the proceedings of the European Parliament. After a preliminary sentence alignment step, each aligned sentence is processed simultaneously for the two languages under consideration. Once processing is finished, words from one language are associated to their translation in the second language corresponding sentence. The number of times one word is associated to another is recorded and then used to calculate the translation score. This parallel corpus was also used to improve the ontology translation database. Translations were extracted based on the co-occurrence of words in aligned sentences using likelihood and Chi-squared criteria and then added to the database. Presently, we are also exploiting Europarl for other purposes, like word sense disambiguation.

In a cross-language environment, our QA system starts by instantiating independent language modules for each language (in this case, Portuguese and Spanish). Suppose that a question is asked in language X , while language Y is chosen as target. First, the question analyser performs pivot extraction and question categorisation, using the language module for X . Then, if the ontology has a direct connection from X to Y , each pivot is translated to language Y . Otherwise, each pivot is translated to English, and then the language module for Y is used to translate from English to Y . Whenever multiple translations are available, the one with the highest translation score is chosen as default, while the others are kept as synonyms (Table 1 illustrates this process). This is done for pivots’ lemmas, heads of derivation and synonyms; in the end, equivalent lemmas, heads of derivation and synonyms are selected for language Y . Notice that this strategy does not translate the whole question from X to Y , which would discard alternative translations of some words. At this point, the language module for Y has pivots in its own language, and question categories, which are language independent. It only needs to activate the question answering patterns (QAPs) to be later used by the answer extraction module [3]. Remember that in a monolingual environment, the QAP activation is done via question patterns (QP), which allows taking profit of finer relations between the question and the answer structures that go beyond question categories. However, this is not possible in a cross-language environment, since QPs are connected to the language module for X and QAPs to the one for Y . For this reason, our current approach does cross-language QAP activation via question categories: we look in language Y for all the QAPs associated with those categories. Finally, the system proceeds as in the Y monolingual environment, with the document retrieval module, the sentence extractor and the answer extractor.

Original pivots	Translated pivots	Other translations kept as synonyms
recorde	récord	
mundial	mundial	universal
salto em altura	salto de altura	
salto	salto	brinco, salto mortal, sobresalto, carrerilla, bronco, voltereta, rebote, tacón, talón, cambio brusco
altura	momento	porte, tamaño, talla, estatura, nivel, niveles, altitud, cumbre, cima, tope, pináculo, grandeza, altura, instante, segundo

Table 1: Example of the Portuguese-Spanish translation for the question “Qual é o recorde mundial do salto em altura?” [What is the world record for high jump?]. Notice that the ontology includes the expression *salto em altura/salto de altura*, and that although *momento* [moment] is chosen as the preferential translation for the polysemic word *altura*, the most correct *altura* and *altitud* are also taken into account as pivot synonyms.

4 Results

This year the CLEF organization did not provide the sets of questions classified according to the CLEF question typology (factoid, definition, temporally restricted factoid and list). Therefore,

we divided the sets of 200 questions into five major categories regarding the type of the expected answer: *denomination/designation* (DEN), *definition* (DEF), *location/space* (LOC), *quantification* (QUANT) and *duration/time* (TEMP).

Ans. →	Right				Wrong				Inexact				Unsupported				Total				Accuracy (%)			
Quest. ↓	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES	PT	ES
DEN	69	49	33	29	29	44	62	68	2	1	0	2	0	2	1	1	100	96	96	100	69.0	51.0	34.4	29.0
DEF	19	18	9	14	8	9	18	14	4	0	1	1	0	0	0	2	31	27	28 [†]	31	61.3	66.7	32.1	45.1
LOC	19	13	10	8	9	11	15	20	1	0	0	0	1	3	1	2	30	27	26 [†]	30	63.3	48.2	38.5	26.7
QUANT	11	11	9	9	7	13	15	9	0	0	0	0	0	0	0	0	18	24	24	18	61.1	45.8	37.5	50.0
TEMP	16	14	11	7	5	9	13	13	0	3	2	1	0	0	0	0	21	26	26	21	76.2	53.8	42.3	33.3
Total	134	105	72	67	61	86	123	124	7	4	3	4	1	5	2	5	200	200	200	200	67.0	52.5	36.0	33.5

[†]The differences in the number of categorized questions between ES and PT-ES are due to an inexact translation of question 178, in which the Spanish LOC question “¿Dónde está el Hermitage?” [Where is the Hermitage located?] was incorrectly translated to a Portuguese DEF question “O que é o Hermitage?” [What is the Hermitage?].

Table 2: Results by category of question.

The general results of Table 2 show that there was a slight increase regarding last year’s Portuguese monolingual task. As for Priberam’s first participation in the Spanish monolingual task, we achieved better results than last year’s best performing system [1]. As for the cross-language environment, the accuracy is significantly lower, and relatively similar for the Portuguese-Spanish and Spanish-Portuguese tasks.

Despite the fairly satisfying results in the Spanish monolingual task, there is still a considerable gap between the performance of the Spanish and the Portuguese modules. A few reasons contributed to this. The semantic classification, the addition of proper nouns and the NEs detection rules for Spanish were just in an early stage, which led to a higher percentage of errors related to the extraction of candidate answers. The absence of a thesaurus also played a part in the document and sentence retrieval stages. Finally, one has to bear in mind that, since the addition of the Spanish language module is recent, it is not so extensively tested and fine-tuned.

Table 3 displays the distribution of errors along the main stages of our QA system, both in the monolingual and bilingual runs.

Question →	W+X+U				Failure (%)			
Stage ↓	PT	ES	PT	ES	PT	ES	PT	ES
			ES	PT			ES	PT
Document retrieval	21	16	28	27	31.8	16.8	21.9	20.3
Extraction of candidate answers	23	41	19	26	34.9	43.2	14.8	19.5
Choice of the final answer	15	30	36	37	22.7	31.6	28.1	27.8
NIL validation	7	8	14	12	10.6	8.4	10.9	9.0
Translation	–	–	29	28	–	–	22.7	21.1
Other	0	0	2	3	0.0	0.0	1.6	2.3
Total	66	95	128	133	100.0	100.0	100.0	100.0

Table 3: Reasons for wrong (W), inexact (X) and unsupported (U) answers.

In the monolingual runs, most errors occur during the extraction of candidate answers. This is related with several issues: QAPs that are badly tuned or erroneously applied due to errors in question categorisation, difficulties in writing QAPs to extract the right answer in long sentences, and a few errors due to the handling of morphological and lexical relations (for instance, no connection was found between *festival de cinema* of PT question 82 “Que festival de cinema atribui o ‘Urso de Ouro’?” [What film festival awards the ‘Golden Bear’?] and the NE *Festival de Cinema de Berlim* of the retrieved sentence), as well as anaphoric relations (e.g. the ES question 80 “¿En qué año murió Bernard Montgomery?” [In what year did Bernard Montgomery die?] did not retrieve the answer “(...) su muerte, ocurrida *en 1976*, motivó un auténtico duelo nacional”). The inclusion of other kinds of metadata besides the date may improve the answer extraction performance, as suggested by PT question 133 “Onde ganharam os Abba o Festival

da Eurovisão?” [Where did Abba win the Eurovision Song Contest?], whose answer, *Brighton*, should be inferred from the title of the news.

As for the document retrieval stage, many errors are connected with long questions with many pivots, which may have the effect of filtering out relevant documents while keeping others that contain more but less important pivots. There were also issues of matching proper nouns and NEs, which are usually the core pivots (e.g. in PT question 53 “Quem venceu a Volta a França em 1988?” [Who won the Tour of France in 1988?] the right document was not retrieved, because the sentence of the answer contained *Tour* instead of *Volta a França*); tokenization issues (e.g. in PT question 166 “Qual foi o resultado do Itália-Nigéria no Campeonato do Mundo de 1994?” [What was the score of Italy-Nigeria in the World Cup of 1994?] *Itália-Nigéria* was considered a single token, hence not matching *Itália* or *Nigéria*); difficulties with ambiguous NEs (e.g. in PT question 189 “Quem escreveu ‘A Capital’?” [Who wrote ‘A Capital’?], ‘A Capital’ occurs frequently in the corpus as the name of a newspaper and less often as the title of a book). Furthermore, some questions could only be answered if a more sophisticated strategy of indexation was made (e.g. the answer to ES question 124 “¿Qué zar ruso murió en 1584?” lies in a document with a list of events: “EFEMERIDES DEL 17 DE MARZO (...) Defunciones (...) 1584.- Ivan ‘El Terrible’, zar ruso.”, and could only be retrieved with a different indexation strategy to deal with items in a list). Table 3 shows also that the Spanish document retrieval stage outperformed the Portuguese one, although we were expecting the opposite behaviour since the system did not use a Spanish thesaurus. The only explanation we could find is that the Portuguese questions were, on average, harder to retrieve than the Spanish ones, more of them requiring anaphora resolution and difficult matches between NEs (e.g., *Isabel II/Elizabeth 2^a*).

Our simple strategy for answer validation led to fair results. Although this year’s CLEF organization did not provide us with the information of which questions were considered NIL, we estimated our system’s NIL recall/precision as respectively 65%/43% (PT), 60%/34% (ES), 30%/29% (PT-ES) and 30%/18% (ES-PT).

The work done for temporally restricted questions, described in the subsection 3.1, allowed the system to retrieve the right answer in 40.7% of the questions, in the Portuguese run, and 32.1%, in the Spanish one. The justification for these results is related with several different issues, mostly in the document retrieval and answer selection stages, and not necessarily with our procedure to handle this type of restriction. For instance, in PT question 155 “Contra que clube jogou o Marítimo a 2 de Abril de 1995?” [Against what team did Marítimo play on the 2nd April 1995?] the answer was not extracted because, due to tokenization, we could not retrieve the document containing the answer, *Marítimo-Tirsense*, although the system correctly matched the date of the document within the relaxed interval set by the temporal restriction.

In the bilingual runs, one of the major causes of failure is related with translation errors, mostly with proper nouns and NEs equivalencies. For instance, in the questions “Quantos Óscares ganhou ‘A Guerra das Estrelas’?”; “O que é a Eurovisão?” and “Cuándo se suicidó Cleopatra?”, the system missed the translations *A Guerra das Estrelas/La guerra de las Galaxias*, *Eurovisão/Eurovision* and *Cleópatra/Cleopatra*. Since NEs and proper nouns are generally the most important elements for the right extraction of answers, if one fails to translate them correctly, they are bound to have a negative impact in the results. Notice that the ontology was not designed to contain proper nouns and NEs except those belonging to closed domains, like names of countries and other geographic entities. Frequently (like in the two last examples above) there are only slight spelling differences between a word and its translation. This suggests dealing with this issue by extending the use of inexact matching techniques to the document retrieval module, hence preventing the exclusion of documents by misspelling of proper nouns or NEs. This technique may also improve the monolingual usage, since in “real-world” systems it is frequent to have different spellings of proper nouns in the question and in the target documents.

5 Conclusions and Future Work

Despite the positive results presented in the previous section, there are still a few issues that need to be solved, in order to improve the general performance of the system. With that purpose in mind, we are currently enriching the Spanish lexicon with the inclusion of proper nouns, semantic features, synonyms and lexical relations.

In addition, we are implementing syntactical processing to question categorisation to capture the argument structure of the question. This will allow a more tuned answer extraction, by matching the specific roles of the question arguments in the answer. This strategy will also lead to an improvement of the cross-language performance, taking profit of the language independence of that argument structure.

Current work is being done in M-CAST project for dealing with other metadata information besides dates, like the document title, its author, and the subject. A task to be addressed in the future is the application of inexact matching techniques, currently used only in the sentence retrieval stage, in the document retrieval module as well. Of course, this would imply some changes in the indexation scheme. As said above, we expect this to improve not only the translation of proper nouns and NEs in cross-language environments, but also the monolingual robustness. Finally, research is being done to address word sense disambiguation. Europarl corpus [9] is being used with the aim of building a Portuguese sense disambiguated corpus, taking profit of the multilingual ontology, which enables domain specific translation, hence allowing discriminating word senses.

Acknowledgements

Priberam Informática would like to thank the partners of the NLUC consortium⁷, especially Synapse Développement, the CLEF organization and Linguateca. We would also like to acknowledge the support of the European Commission in TRUST (IST-1999-56416) and M-CAST (EDC 22249 M-CAST) projects.

References

- [1] A. Vallin, B. Magnini, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Penas, M. de Rijke, B. Sacaleanu, D. Santos, and R. Sutcliffe. Overview of the CLEF 2005 multilingual question answering track. In *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (Vienna, Austria, 21-23 September)*, 2005. Available at http://clef.isti.cnr.it/2005/working_notes/WorkingNotes2005/vallin05.pdf.
- [2] C. Amaral, D. Laurent, A. Martins, A. Mendes, and C. Pinto. Design and Implementation of a Semantic Search Engine for Portuguese. In *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, 26-28 May*, volume 1, pages 247–250, 2004. Also available at <http://www.priberam.pt/docs/LREC2004.pdf>.
- [3] C. Amaral, H. Figueira, A. Martins, A. Mendes, P. Mendes, and C. Pinto. Priberam's question answering system for Portuguese. In *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (Vienna, Austria, 21-23 September)*, 2005. Available at http://clef.isti.cnr.it/2005/working_notes/WorkingNotes2005/amaral05.pdf.
- [4] C. Amaral, H. Figueira, A. Mendes, P. Mendes, and C. Pinto. A Workbench for Developing Natural Language Processing Tools. In *Pre-proceedings of the 1st Workshop on International Proofing Tools and Language Technologies (Patras, Greece, 1-2 July)*, 2004. Also available at <http://www.priberam.pt/docs/WorkbenchNLP.pdf>.

⁷See <http://www.nluc.com>.

- [5] S.M. Thede and M.P. Harper. A second-order hidden Markov model for part-of-speech tagging. In *Proceedings of the 37th Annual Meeting of the ACL, Maryland: College Park*, pages 175–182, 1999.
- [6] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing (2nd printing)*. The MIT Press, Cambridge, Massachusetts, 2000.
- [7] Jesús Giménez and Lluís Màrquez. SVMTool: A general POS tagger generator based on Support Vector Machines. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04) (Lisbon, Portugal)*, 2004.
- [8] D. Laurent, P. Séguéla, and S. Nègre. Cross Lingual Question Answering using QRISTAL for CLEF 2005. In *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (Vienna, Austria, 21-23 September)*, 2005. Available at http://clef.isti.cnr.it/2005/working_notes/WorkingNotes2005/laurent05.pdf.
- [9] P. Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit X (Phuket, Thailand)*, pages 79–86, 2005. Also available at: <http://www.iccs.inf.ed.ac.uk/~pkoehn/publications/europarl-mtsummit05.pdf>.