Cross Lingual Question Answering using QRISTAL
for CLEF 2006

Dominique Laurent, Patrick Séguéla, Sophie Nègre

Synapse Développement
33 rue Maynard,
31000 Toulouse, France
{dlaurent, p.seguela, sophie.negre }@synapse-fr.com

Abstract

QRISTAL [9] is a question answering system making intensive use of natural language processing both for indexing documents and extracting answers. It ranked first in the EQueR evaluation campaign (Evalda, Technolangue [3]) and in CLEF 2005 for monolingual task (French-French) and multilingual task (English-French and Portuguese-French). This article describes the improvements of the system since last year. Then, it presents our benchmarked results for the CLEF 2006 campaign and a critical description of the system. Since Synapse Développement is participating to Quaero project, QRISTAL is most likely to be integrated in a mass market search engine in the forthcoming years.

1   Introduction

QRISTAL (French acronym for "Question Answering Integrating Natural Language Processing Techniques") is a cross lingual question answering system for French, English, Italian, Portuguese, Polish and Czech. It was designed to extract answers both from documents stored on a hard disk and from Web pages by using traditional search engines (Google, MSN, AOL, etc.). Qristal is currently used in the M-CAST European project of E-content (22249, Multilingual Content Aggregation System based on TRUST Search Engine). Anyone can assess the Qristal technology for French at www.qristal.fr. Note that the testing corpus for the testing web page is the grammar handbook proposed at http://www.synapse-fr.com.

For each language, a linguistic module analyzes questions and searches for potential answers. For CLEF 2006, the French, English and Portuguese modules were used for question analysis. Only the French module was used for answers extraction. The linguistic modules were developed by different companies. They share however a common architecture and similar resources (general taxonomy, typology of questions and answers and terminological fields).

For French, our system is based on the Cordial technology. It massively uses NLP tools, such as syntactic analysis, semantic disambiguation, anaphora resolution, metaphor detection, handling of converses, named entities extraction as well as conceptual and domain recognition. As the product is being marketed, the linguistic resources need to be permanently updated and it required a constant optimization of the various modules so that the software remains extremely fast. Users are now accustomed to obtain something that looks like an answer within a very short time, not exceeding **two** seconds.

2 Architecture

The architecture of the Qristal system is described in different articles (see [1], [2], [8], [9], [10], [11], [12]). Qristal is a complete engine for indexation and answers extraction. However, it doesn't index the Web. Indexing is processed only for documents based on disks. Web search uses a meta-search engine we have implemented. As we will see in the conclusion, our participation to Quaero project should change this way of use by tagging semantically the Web pages.

Our company is responsible for the indexing process of Qristal. Moreover, it ensures the integration and interoperability between all linguistic modules. Both English and Italian modules were developed by Expert System Company. The Portuguese module was developed by the Priberam Company which also takes part in CLEF 2005 for Portuguese monolingual and in CLEF 2006 for Spanish and Portuguese monolingual, and for Spanish-Portuguese and Portuguese-Spanish multilingual tasks. The Polish module was developed by the TiP Company. The Czech module is developed by the University of Economics of Prague (UEP). These modules were developed within the European projects TRUST [8] (Text Retrieval Using Semantic Technologies) and M-CAST (Multilingual Content Aggregation System based on TRUST Search Engine).

2.1 Multicriteria indexing

While indexing documents, the technology automatically identifies the document language of and the system calls the corresponding language module. There are as many indexes as languages identified in the corpus. Documents are treated per blocks. The size of each block is approximately 1 kilobyte. Block limits are settled on the end of sentences or paragraphs. This size of block (1 kb) appeared to be optimal during our tests. Some indexes relate to blocks like fields or taxonomy whereas other relate to words, like idioms or named entities.

Each linguistic module processes a syntactic and semantic analysis for each block to be indexed. It fills a complete structure of data for each sentence. This structure is passed to the general processor that uses it to increment the various indexes. This description is accurate for the French module. Other language modules are very close to that framework but don't always include all its elements. For example, English and Italian modules do not include an indexing based on heads of derivation.

Texts are converted into Unicode. Then, they are divided into one kilobyte blocks. This reduces the index size as only the number of occurrences per block is stored for a given lemma. This number of occurrences is used to infer the relevance of each block while searching a given lemma in the index. In fact we here use lemmas but the system stores heads of derivation and not lemmas. For example, *symmetric, symmetrical, asymmetry, dissymmetrical* or *symmetrize* will be indexed in the same entry : *symmetry.*

Each text block is analyzed syntactically and semantically. Considering results of this analysis, 8 different indexes are built for:

- heads of derivation. A head of derivation can be a sense for a word. In French, the verb *voler* has 2 different meanings (*to steal* or *to fly*). The meaning "dérober" (*to steal*) will lead to *vol* (*robbery*), *voleur* (*thief*) or *voleuse* (*female thief*). The second meaning, "se mouvoir dans l'air" (*to fly*), will lead to *vol* (*flight*), *volant* (*flying* as an adjective), *voleter* (*to flutter*) or *envol* (*taking flight)* and all its forms.
- proper names. If they appear in our dictionaries.
- idioms. Those idioms are listed in our idioms dictionaries. They encompass approximately 50 000 entries, like *word processing, fly blind* or *as good as your word.*
- named entities. Named entities are extracted from texts. *George W. Bush* or *Defense Advanced Research Project Agency* are named entities.
- concepts. Concepts are nodes of our general taxonomy. 2 levels of concepts are indexed. The first level lists 256 categories, like "visibility". The second level, actually the leaves of our taxonomy, lists 3387 subcategories, like "lighting" or "transparency",
- fields. 186 fields, like "aeronautics", "agriculture", etc.,

- question and answer types for categories like "distance", "speed", "definition", "causality", etc.,
- keywords of the text.

For each language, the indexing process is similar. Extracted data are the same. Thus, the handling of those data is independent of their original language. This is particularly important for cross language question answering.

For the French language, the rate of correct grammatical disambiguation (distinction between name-verb-adjective-adverb) is higher than 99%. The rate of semantic disambiguation is approximately 90% for 9 000 polysemous words and approximately 30 000 senses for these words. Note that this number of senses is markedly inferior to the Larousse one (Larousse is one of the most famous French dictionaries). Note however that our idioms dictionary covers a large number of the senses mentioned in this kind of dictionaries. The indexing speed varies between 200 and 400 Mo per hour with a Pentium 3 GHz, according to the size and number of indexed files.

Indexing question types is undoubtedly one of the most original aspects of our system. While the analysis of the blocks is being made, possible answers are located. For example, a name of function for a person (like *baker, minister, director of public prosecutions)*, a date of birth (like *born on April 28, 1958*), a causality (like *due to snow drift* or *because of freezing*), a consequence (like *leading to serious disruption* or *facilitating the management of the traffic*). This caused the block to be indexed like being able to provide an answer for a given question type.

Currently, our question typology includes 86 types of questions. Those types are divided into two subcategories: factual types and non factual types. Factual types are dimension, surface, weight, speed, percentage, temperature, price, number of inhabitants or work of art. Nonfactual types are form, possession, judgement, goal, causality, opinion, comparison or classification. For CLEF 2006, results were as follows:

	French	English 1	English 2	Portuguese
Good choice	96.5 %	93.5 %	83.0 %	91.0 %

Figure 1. Success rate for question type analysis

These rates are very close to CLEF 2005 results, because we only improved the French module and elaborated a new English module. The "English 1" corresponds to the Synapse English module and the "English 2" corresponds to the Expert Systems English module.

Building a keyword index for each text is also peculiar to our system. Dividing text into blocks made it compulsory. Isolated blocks cannot explicitly mention main subjects of the original text although sentences of these blocks relate to these subjects. The keyword index makes it possible to add contextual information about the main subjects of the text for blocks. Keywords can be a concept, a person, an event, etc.

2.2   Answer extraction

After the user has keyed in his/her question, it is syntactically and semantically analyzed by the system. Question type is inferred. We would like here to draw the attention to the fact that questions are shorter than texts. This lack of context makes the semantic analysis of the question more dubious. That's why the semantic analysis processed on the question is more comprehensive than the analysis processed on texts. Moreover, users have the possibility to interactively force a sense. This possibility, however, was not used for CLEF as the entire process was automatic.

The result of the semantic analysis of the question is a weight for each sense of each word recognized as a pivot. For example, sense 1 is recognized with 20%, sense 2 with 65% and sense 3 with 15%. This weight, together with synonyms, question and answer types or concepts, is considered while searching the index. Thus all senses of a word are taken into account during the index search. This prevents from dramatic consequences due to errors in the semantic disambiguation while making the most of good analysis.

After question analysis, all indexes are searched and the best ranked blocks are analyzed again. As one can notice on figure 2, the analysis of the selected blocks is close to the analysis processed while indexing or question analyzing. On top of this "classic" analysis, a weight for each sentence is inferred. This weight is based on the number of words, synonyms and named entities found in this sentence, the presence of an answer corresponding to the question type and a correspondence between the fields and domain.

After this analysis, sentences are ranked. Then, an additional analysis is processed to extract named entities, idioms or lists that match the answer. This extraction relies on the syntactic characteristics of those groups.

For a question on a corpus located on a hard disk, the response time is approximately 1,3 seconds with a Pentium 3 GHz. On the Web, first answers are provided after 2 seconds. Then the system computes a progressive refining during ten seconds, according to user's parameters like the number of words, the number of analyzed pages, etc.

We tested many answer justification modules, mostly implemented from Web [4], [7] or [15]. Our technology enables, as an option, to use such a module of justification. It consists in searching the web with the words of the question looking for potential answers the system inferred. However this process is seldom selected by users as it increases the response time of a few seconds. It was not used in CLEF 2006 either. The only justification module we used was an internal module which makes the most of the semantic information for proper names enclosed in our dictionaries. For more than 40 000 proper names, we possess information about the country of origin, the year of birth and death, the function for people, country, the area and population for a city, etc. We think this justification module is at the origin of some "unjustified" answers. As a matter of fact, it caused the system to rank first a text including the answer even if the system did not find any clear justification of that answer in the text.

For cross language question answering, English is used as pivot language. The fact that most users are only interested in documents in their own language and English motivated that choice. Thus, for cross language answering, the system processes generally only one translation. For this evaluation, both Portuguese to French and Italian to French runs required two translations: from source language to English and then from English to French. QRISTAL does not use any Web Services for translation because of response time. Only words or idioms recognized as pivots are translated.

3   Improvements since CLEF 2005

For CLEF 2006, we used our same technology and system, in mono and multilingual mode[9], but with a few improvements, such as :

The ontology has been revised all through the year, mainly to emend errors or categorisation. To maintain compatibility with other language ontology, no new category has been added nor deleted.

The dictionaries have been updated, in particular proper names and expressions. This updating effort, while being continuous, has been increased last year, but not the extent to be ready, with other resources, for the said evaluation. This being the case for the dictionary of nominal expressions, leaping from 55 000 expressions to over 100 000 and not integrated in the assessed Qristal version.

A multilingual (French,English,Spanish,Italian,Portuguese) lexicon of translated proper names has been implemented. It includes more than 5 000 proper names and acronyms, mainly toponyms, (countries, provinces, towns) but also name of people, in particular names in Arabic, Russian, Chinese, which spelling differs according to the languages. This lexicon has played an important role in the improvement of the CLEF 2006 results over CLEF 2005, for the Portugese-French pair, as the English translation was avoided.

The English-French and French-English dictionaries have also been revised and increased with now more than 200 000 translations of words or expressions. The impact of this improvement has been measured in comparing the CLEF 2005 results with the former dictionaries versus the CLEF 2006 results using the improved ones. Only one question in Portuguese to French and two questions in English to French find an additional answer with the new dictionaries.

3.1   Improvement of the algorithms

Our syntactic analyser has not noticeably been improved. However, according to the non-yet final results of another benchmark evaluation, our analyser is given as the most performing and , above all, robust for the French language. It appears that it is still yet very far from the complete detection of the complex syntactic structures. So, for the Subject-Verb relation, it detects exactly the subject and the verb in only 9 cases out of 10. But this must be moderated by the fact , the evaluation is carried out on all types of corpora, including emails and chats, somehow more difficult to analyse.

The module « search of the category of the question» has been improved but this impacts only on the monolingual French-French part. For the M-CAST European project, our engine was enriched of numerous utilities to manage very large volumes of data and to satisfy "client-php server" structures, but these have no impact on the performances.

3.2   New English module

The main difference between the engine assessed at CLEF 2005 and CLEF 2006, is due to the in-house development of a new English language processing module. On the basis of our ancient syntactic analyser and english-based linguistic resources not yet completed, we used a beta version of our new English module. It carries out the syntactic and semantic analysis of the question, determines the type of the question, the pivot-words and synonyms, then transfers the results to the French module that implements the requested translations of the same, to finish with the use of the language independent data (type of the question, categories of the ontology, etc.).

4   Results for CLEF 2006

QRISTAL was evaluated for CLEF 2006 for French to French, English to French (Synapse Module and Expert Systems module), and Portuguese to French. That is 1 monolingual and 3 multilingual campaigns. For each one of these tasks, we processed only one run. Note that results obtained in CLEF 2006 could have been obtained with our June 2006 commercial version of our Qristal software.
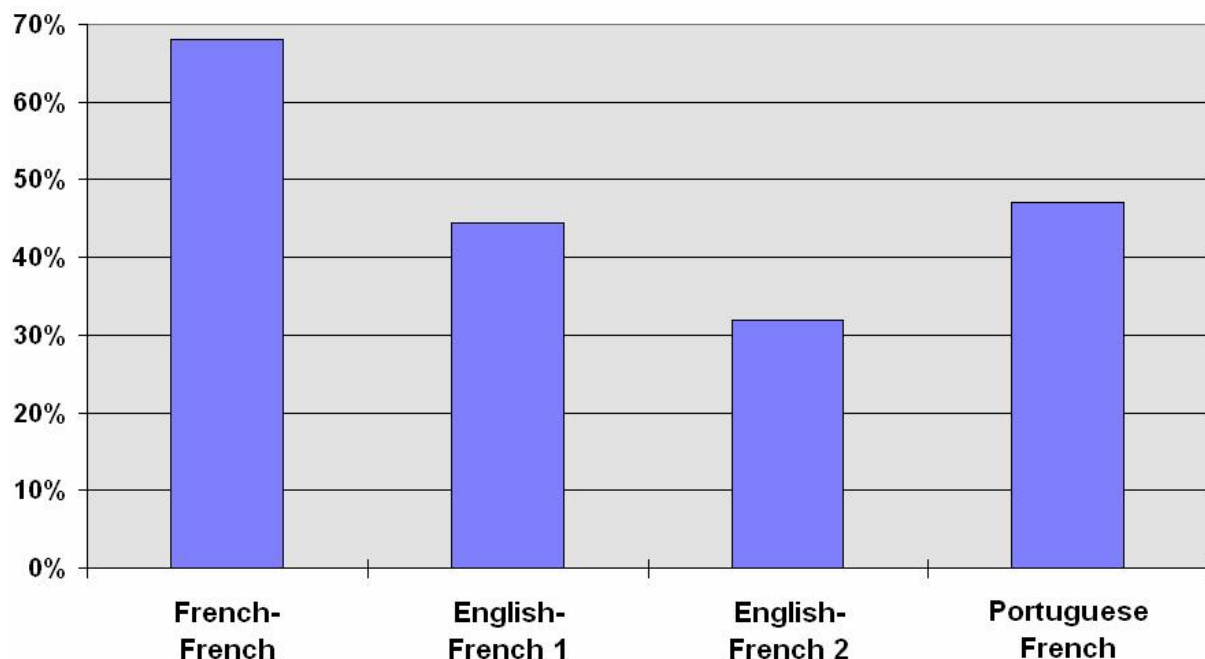

Figure 2. Results of the general task

For French to French and for the pairs English-French and Portuguese-French, these results are better than those we obtained for the CLEF 2005. For French-French, that is 68% for CLEF 2006 and 64% for EQueR 2005. For English-French, that is 44,5 % for CLEF 2006 and 39,5% for CLEF 2005, and for Portuguese-French, that is 47% for CLEF 2006 and 36,5 % for CLEF 2005.
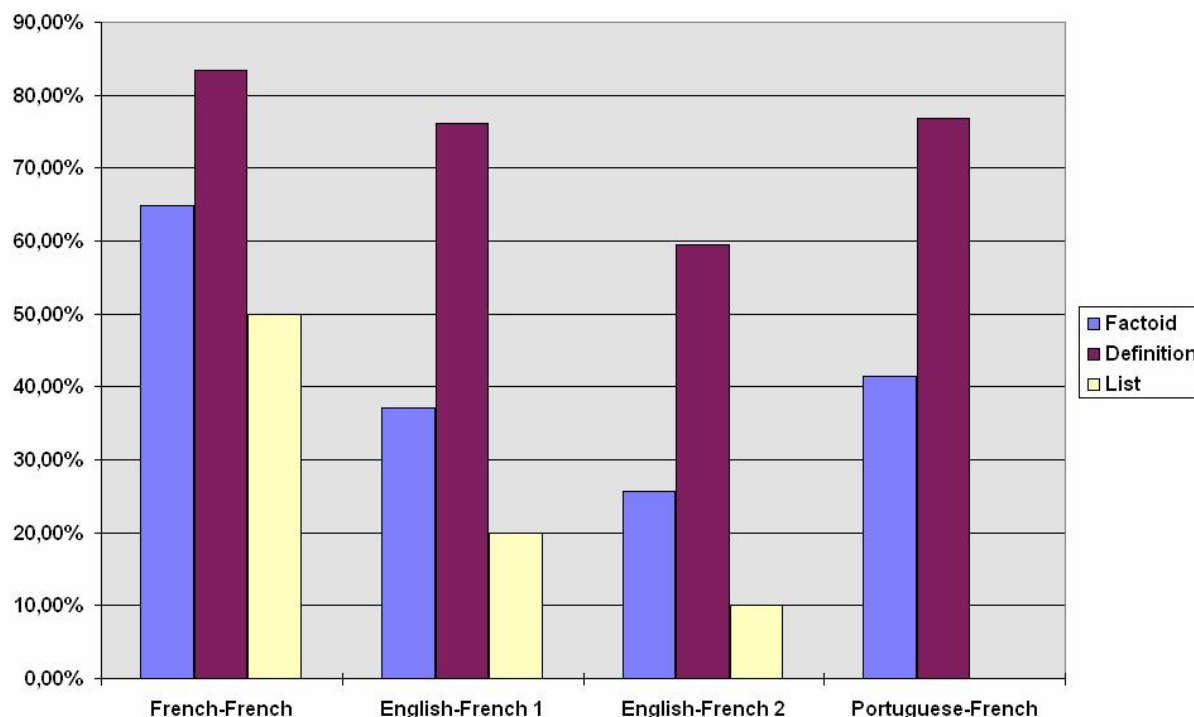
Results per category are as follows:

Figure 3 : Results of our 4 runs for each question type

As last year, our system is highly performing for questions of the type « Definition ». It is to be noted that this type of question , the "loss" of performance in a multilingual context, is less than for the other types of questions. This is due to the fact the "Definition" type of question relates most of the times, to acronyms or surnames of people, for which the translation is simpler and less ambiguous. The contribution of the Portuguese-French proper names lexicon seems obvious as the percentage of found definitions in this pair has moved from 68 % (CLEF 2005) to 77 % (CLEF 2006).

The « list-type » questions were only identified by our French module, thus none of this type of question were exact from the Portuguese, and the questions assessed as exact for English-French were , in fact, lists of one single element. For French, the proportion of exactly identified lists was honourable (50%), but this type of question remains difficult to process by our system.

Specific developments for the NIL questions had been implemented in CLEF 2006. In monolingual mode, the improvement is spectacular since the "precision rate" is now 0.56 versus 0.23 in 2005 and the "recall rate" 0.66 versus 0.25. For English-French they were respectively 0.29 v 0.14 and 0.66 v 0.30. Finally, for Portuguese-French, the "precision rate" was 0.26 v. 0.13 and "recall rate" was 0.70 v 0.15.

Figure 4 presents statistics for answers evaluated as 'R' that stands for right. But CLEF proposed two other qualifications for answers that is 'U' for unjustified and 'X' for inexact. We think 'U' and 'X' answers would be often accepted by users, even 'X' answers if they are presented with their context. For question 57 *Qui est Flavio Briatore ?* (*Who is Flavio Briatore?*), the answer provided by our system was *directeur général de Benetton Formula* (*general manager of Benetton Formula*), whereas the awaited answer was *directeur général de Benetton Formula 1* (*general manager of Benetton Formula 1*). Likewise, for question 96 *A quel parti politique Jacques Santer appartient-il ?* (*Which political party does Jacques Santer belong to ?*), the provided

answer by Qristal was *Parti chrétien-social dès 1966* (*Christian Social Party since 1966*) whereas the awaited answer was *Parti chrétien-social* (*Christian Social party*). This lead us to consider statistics for all answers considered as "not wrong", that is right (R), unjustified (U) or inaccurate (X):

	French-French	English-French 1	English-French 2	Portuguese-French
Not wrong (R+U+X)	159 (79.5%)	97 (48.5%)	71 (35.5%)	101 (50.5%)

Then we had a closer look to questions where the monolingual process finds the answer but the cross language does not. This leads us to the following remark. Questions are defined by reading the corpus and, deliberately or not, people formulating questions tend to reuse words or expressions mentioned in the text of the identified answer. On one hand, this influences the capacity of the system and the importance of each module in the overall process. For example, the use of synonyms is not that important for CLEF as it normally is. On the other hand, for cross language question answering, translations can be fuzzy and potentially quite far from the targeted word or expression especially when one uses English as an intermediate language. In this way, translated words are quite often different from the terms mentioned by both the question and the answer.

For question 1 "*Qu'est-ce qu'Atlantis ?*" ("*What is Atlantis ?*"), the question is translated from the English sentence "*What is Atlantis* ?" but in fact, the word "Atlantis" is normally translated by "Atlantide" in French and "Atlantis" is only kept when you want to speak of the spatial shuttle.

More generally, in comparing the monolingual and multilingual results, one observes that the longer the question is, the less proper names are present, and the less the results are satisfying. The quality loss is estimated to about 15 % for the questions of the "definition" type, and near 50 % for factual questions with temporal anchor. On the 200 questions used for the evaluation, 17 had no proper names and no dates. The following table provides the results of our runs for these questions :

	FR-FR	EN-FR 1	EN-FR 2	PT-FR
18	R	W	R	W
24	R	R	R	R
59	W	W	W	W
64	R	W	W	W
79	R	W	W	W
100	R	W	W	X
104	W	W	W	W
109	W	W	W	W
117	R	W	W	W
118	R	W	W	W
133	X	W	W	W
144	W	R	R	R
164	R	W	W	W
166	R	R	R	R
188	W	W	W	W
189	R	R	W	R
199	W	W	W	W
	58 %	24 %	24 %	24 %

The table shows that for theses questions which quality of the translation is a crucial issue, the results are heavily deteriorated. Question 144, for which only the monolingual run returns an error, was a NIL question for which the French-French module returns, in despite, an answer while the other modules are returning a NIL.

Priberam, the company responsible for the Portuguese module in our engine, participated in CLEF 2006 for Portuguese and Spanish evaluation tracks. It is interesting to note that they obtained results very similar to our results for the Portuguese monolingual run [1] [2] and have similar degradations of results from monolingual to multilingual runs.

5 Outlines

Our CLEF 2006 results are noticeably better than those of our CLEF 2005 campaign, furthermore if one consider that "list" questions had been added this year. Notable too is our English-French run , using our Italian partner's English module, returned less good results in 2006 versus 2005 ((32,5% v 39,5%), this confirming the overall greater difficulty attached to the questions this year.

The following modifications have generated the following improvements:

- Revised processing of the NIL questions. Although this is of little interest to the user, often requesting replies even inexact ones, the revision was implemented for CLEF and its evaluations. The end result has been to reach far better "precision" and "recall"rates for this type of questions.

- Slightly improved translations and primarily the use of the multilingual lexicon of translated proper names and acronyms. Thanks to these dictionaries, the deterioration between the French-French and French-Portuguese pairs has dropped significantly ie 43 % (2005) to 31 % (2006) detailed in the following : (a – CLEF 2005) from 64 % in French-French and 36.5% in Portuguese-French to (b-CLEF 2006) 68% in French-French and 47% in Portugese-French).

- Betterment of the resources and the algorithm to detect the type of the question.

In respect of the deployed efforts, the development costs engaged and the resources upgrading, the global improvement of the results is not astonishing. It seems that the algorithms used by our modules find their limits around 70% of satisfying answers ! However, noticing that only 20 % of the answers are marked "wrong" in French-French, one may think that a revision of the delimitation of the extracted replies could , in the near future, allow the reach of success rates around 80 %.

Numerous other developments and improvements have been incorporated into our system, specially initiated in the framework of our M-CAST European project, but are not visible into the CLEF assessed results. For instance, the speed to return an answer has greatly improved, from an usual 3 seconds in 2005 to less than 1 second in 2006. This was obtained thanks to a preliminary fast analysis of the index returned sentences, which screens more than 80% of the sentences with no pivot-words of the question, hence a very weak probability to contain an answer.

After the CLEF 2005 evaluation, we had identified a few leads for improvement of our system. A few of them have been implemented this year, but a lot remains to be done. We have started the elaboration of a " knowledge base", from a Web-based data extraction, and currently targeting geographical data (country, province, town names) , the whole to be extended, in the forthcoming months, to people names and events. We still not take into account the presentation of the document, and the answer extraction should be revised as still too imprecise.

As for the next coming years, our technology and system should evolve considerably as our firm is a partner of the Franco-German Quaero project, taking in charge the Question Aswering issue for the French and English languages. Within this project, in partnership with the firm –Exalead- having developed the search engine "Eponyme", we should market a general consumer and a professional version of our system, on closed corpora but more rewardingly on the billion Web pages. In this respect, our strategy should evolve towards a semantic tagging of the Web pages with indexation of the tagged items, in view to find the answers to a question within a timeframe of $1/10^{th}$ or $2/10^{th}$ of a second.

6 Conclusion

Despite the introduction of the questions of the « list » type , more difficult to process than the « definition » or « factual » ones, our system improved its results both in mono and multilingual mode. For factual questions, QRISTAL returns around 70 % of exact answers in monolingual mode and almost 50 % in multilingual mode.

A fine-tuned analysis of the results shows that the quality of the answer extraction still can be markedly improved as nearly 10 % of the answers have been assessed as "inexact" ("X") or non-justified ("U") in French-French mode, while the proportion of answers qualified as "false" was barely above 20 % in the same mode.

In the time elapsing between the 2 campaigns, our firm has invested almost 4 man/years in its system, but most developments were carried out on other areas that the quality of the said system. This was true for the answer delivering speed, the multitask operations, the Web accessibility, and the new English language module. We consider the real system improvement as having requested 1 man/year, an important investment for a comparatively reduced improvement.

The incorporation of our system into a Web-focused search engine through the Quaero project, is to necessitate in the coming years, a comprehensive revision of our methods in view to return the most precise answers in less than 2/10th of a second. So, the core of the syntactic and semantic analysis processing will be batch processed and will produce a semantic tagging of the Web pages (or of the documents in closed corpora), along with the indexation of these tags (named entities, possible answers per type of question, key-words, etc.). Furthermore, the presence of a knowledge base fed and updated permanently via the Web, should permit the development of verification procedures of the answers, hence reducing the related errors. These procedures are of the most importance as, beside the delivered improvements, they avoid the delivery of nonsensical or absurd answers that correspondingly generate a user's suspicion on the reliability of the system.

At a time when the Question Answering Systems find their first business use in firms or for the general audience, it is vital for these QA systems to avoid replicating the errors made while introducing the first grammar- checking or voice recognition systems on the markets. As the average quality of these later systems in their initial versions has largely deceived their users, to the extent that any further satisfying development did not compensate for the deception and are still viewed as "unusable and of no interest"!

Acknowledgments

References

[1] AMARAL C., LAURENT D., MARTINS A., MENDES A., PINTO C. (2004), Design & Implementation of a Semantic Search Engine for Portuguese, *Proceedings of the Fourth Conference on Language Resources and Evaluation.*

[2] AMARAL C., FIGUEIRA H., MARTINS A., MENDES A., MENDES P., PINTO C. (2005), Priberam's question answering system for Portuguese, *Working Notes for the CLEF 2005 Workshop,* 21-23 September, Wien, Austria.

[3] AYACHE C., GRAU B., VILNAT A. (2005), Campagne d'évaluation EQueR-EVALDA : Évaluation en question-réponse, *TALN 2005,* 6-10 juin 2005, Dourdan, France, tome 2. – Ateliers & Tutoriels, p. 63-72.

[4] CLARKE C. L. A., CORMACK G. V., LYNAM T. R. (2001), Exploiting Redundancy in Question Answering, *Proceedings of 24th Annual International ACM SIGIR Conference (SIGIR 2001),* p. 358-365.

[5] GRAU B.. (2004), L'évaluation des systèmes de question-réponse, *Évaluation des systèmes de traitement de l'information,* TSTI, p. 77-98, éd. Lavoisier.

[6] HARABAGIU S., MOLDOVAN D., CLARK C., BOWDEN M., WILLIAMS J., BENSLEY J. (2002), Answer Mining by Combining Extraction Techniques with Abductive Reasoning, *Proceedings of The Twelfth Text Retrieval Conference (TREC 2003).*

[7] JIJKUN V., MISHNE G., DE RIJKE M., SCHLOBACH S., AHN D., MÜLLER K. (2004), The University of Amsterdam at QALEF 2004, *Working Notes of the Workshop of CLEF 2004*, Bath, 15-17 september 2004.

[8] LAURENT D., VARONE M., AMARAL C., FUGLEWICZ P. (2004), Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies, *First International Workshop on Proofing Tools and Language Technologies*, Patras, Grèce.

[9] LAURENT D., SEGUELA P. (2005), QRISTAL, système de Questions-Réponses, *TALN 2005*, 6-10 juin 2005, Dourdan, France, tome 1. –Conférences principales, p. 53-62.

[10] LAURENT D., SÉGUÉLA P, NÈGRE S. (2005), Cross-Lingual Question Answering using QRISTAL for CLEF 2005, *CLEF 2005*, 21-23 september 2005, Wien, Austria, .

[11] LAURENT D. (2006), Industrial concerns of a Question-Answering system ?, *EACL 2006, Workshop KRAQ*, April 3 2006, Trento, Italia.

[12] LAURENT D., SÉGUÉLA P, NÈGRE S. (2006), QA better than IR ?, *EACL 2006, Workshop MLQA'06*, April 4 2006, Trento, Italia.

[13] MAGNINI B., VALLIN A., AYACHE C., ERBACH G., PEÑAS A., DE RIJKE M., ROCHA P., SIMOV K., SUTCLIFFE R. (2004), Overview of the CLEF 2004 Multilingual Question Answering Track, *Working Notes of the Workshop of CLEF 2004*, Bath, 15-17 september 2004.

[14] MONZ C. (2003), From Document Retrieval to Question Answering, *ILLC Dissertation Series 2003-4*, ILLC, Amsterdam.

[15] VOORHEES E. M.. (2003), Overview of the TREC 2003 Question Answering Track, NIST, 54-68 (http://trec.nist.gov/pubs/trec12/t12_proceedings.html).