

Cross Lingual Question Answering using QRISTAL for CLEF 2008

Dominique Laurent, Patrick Séguéla, Sophie Nègre

Synapse Développement

33 rue Maynard,

31000 Toulouse, France

{dlaurent, p.seguela, sophie.negre }@synapse-fr.com

Abstract

QRISTAL [10], [13] 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 [4]) and in first rank in French for CLEF 2005, 2006 and 2007 [11], [12], [14]. This article describes the improvements of the system since last year. Then, it presents our benchmarked results for the CLEF 2008 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). 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 2008, the French, English and Portuguese modules were used for question analysis. Only the French module was used for answers extraction. The French and English modules are developed by Synapse Développement, modules for other languages are developed by different companies, for example Priberam for Portuguese [1], [2], [3], [5]. These different modules share 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 [10], [11], [12], [13], [14]). 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 is changing 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. 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 [1], [2], [3], [5]. 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).

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*.

3 Improvements since CLEF 2007

For CLEF 2008, we used our same technology and system, in mono and multilingual mode [9], but with some improvements.

Last year, we participated only in monolingual task French-French. Our results were good (54% of right answers) but we seen that, if the results were comparable to precedent years for news corpus (65%), the results on Wikipedia corpus were not good (32%). The reasons of these bad results with Wikipedia corpus are :

- no redundancy in the Wikipedia corpus. For example the area of a country or a region can be found generally in only one article and one time.
- a complex format, nearest of database than news or classical Web pages. Many important information is given in tables with a specific coded format.
- the titles of the pages, with the Redirect system of Wikipedia pages, are often far from the named entities relative to these pages.

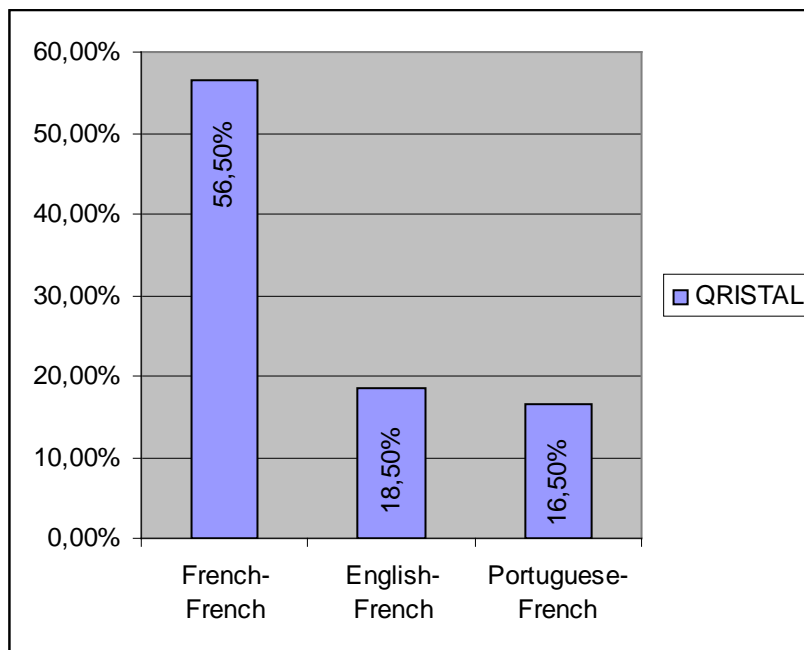
For CLEF 2007, we made an error with the elimination of all the Redirect pages. For CLEF 2008, we indexed these Redirect pages and managed the links between these Redirect pages and the redirected pages. Knowing that sequences of questions have often the answers in the same page, we give a higher score to the first best ranked pages of the first question for the other questions of the sequence. Finally, we reduced the importance of redundancy, decreasing the score of similar answers in different pages.

For CLEF 2008, we also revised our management of Named Entities and Anaphora. For Named Entities, we enhanced our dictionary of Proper Nouns synonyms (in fact Named Entities synonyms). For Anaphora, we

improved our management of possible references, keeping more information about the semantic characteristics of these possible references.

4 Results for CLEF 2008

QRISTAL was evaluated for CLEF 2008 for French to French, English to French and Portuguese to French. That is 1 monolingual and 2 multilingual campaigns. For each one of these tasks, we processed only one run.



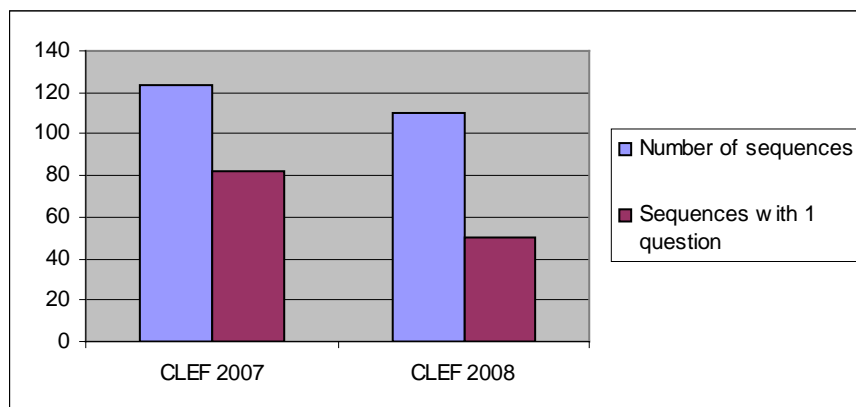
INCORPORER

For French to French, these results are a little better than those we obtained for the CLEF 2007. For English-French and for Portuguese-French, the results are bad and we will give far away the possible reasons.

5 Comparing CLEF 2008 to CLEF 2007

In theory, the CLEF 2008 campaign was to be similar to the campaign of the previous year. In fact, at least for French, the CLEF 2008 evaluation was different and, finally, sharply more difficult than CLEF 2007 !

Firstly, if we look the number of sequences for CLEF 2007 and CLEF 2008, we see that the percentage of unique questions in a sequence was 41% last year and 25% this year, and the number of sequences was 124 in 2007 and 110 this year. This means that a system which don't manage anaphora have a potential optimum of 62% last year and 55% this year.



These simple percentages show the higher difficulty of CLEF 2008 evaluation for French. But many other data can complete this first impression. If you look to the questions of CLEF 2007, you can see that many questions in sequences have in fact no reference to precedent questions in the sequence and don't integrate any anaphora. In CLEF 2007, we have 76 questions inside sequences (in 2nd, 3rd or 4th position) and only 40 anaphora, with only one implicit anaphora (question 57 : *Qui était considéré comme le chef du commando ?*). In CLEF 2008, we have 90 questions inside sequences (in 2nd, 3rd or 4th position) and 69 anaphora, with 5 implicit anaphora (questions 112, 154, 155, 156, 189). So the number of anaphora increased from 20% to 34,5% of the questions (in fact, about 70% more of anaphora !)

The lists are another big difference between CLEF 2007 and CLEF 2008. In CLEF 2007, there were 9 list questions (4,5%), all with the number of answers given in the questions (5, 41, 47, 67, 81, 115, 120, 194, 196). In CLEF 2008, we have 29 list questions (14,5%) and only two questions with the number of answers given in the question (26, 163). So, the percentage of list questions is more than 3 times higher than last year with a greatest difficulty coming from the absence of number of elements...

The questions with temporal restrictions were 39 in 2007 (7, 10, 17, 19, 32, 33, 35, 36, 38, 39, 40, 41, 45, 46, 48, 52, 69, 70, 83, 86, 88, 101, 102, 106, 114, 120, 125, 127, 138, 139, 140, 145, 147, 148, 149, 167, 169, 175, 191) and, in these questions, no one includes an anaphora. This year, the questions with temporal restrictions were 69 and 22 include at least one anaphora !

But the biggest difference between CLEF 2007 and CLEF 2008 is the corpus of the answers ! If the corpora used for CLEF 2007 and CLEF 2008 were exactly the same (news and Wikipedia), the corpus where the answers can be found are not the same :

	CLEF 2007	CLEF 2007 (%)	CLEF 2008	CLEF 2008 (%)
News	96	50,3 %	43	22,9 %
News + Wikipedia	21	11,0 %	14	7,4 %
Wikipedia	74	38,7 %	131	69,7 %
NIL	9		12	

If we imagine a system which manages only news corpus, his higher possible score was 61,3 % last year and only 30,3 % this year ! Because we know that Wikipedia is a very difficult corpus for Question-Answering, this strongly contributed to increase the difficulties from 2007 till 2008. If we look the size of Wikipedia corpus by comparison with news corpus, the percentage of answers in Wikipedia corpus is probably representative this year of the respective sizes, but, by comparison with CLEF 2007, the difficulty is higher.

6 Conclusion

By comparison with CLEF 2007, our results in French-French are good, knowing that the answers were needed to be found essentially in Wikipedia pages, with a high proportion of lists questions (14,5 %), anaphora resolution (38 %) and temporal restrictions (34,5%). For English-French and Portuguese-French, our results are

bad, but all the characteristics described above (many anaphora, list questions, temporal restrictions, etc.) are very penalizing for our system and, more generally, for multilingual Question-Answering.

For us, this type of evaluation has no real sense in multilingual. Translate the question, then translate the possible answers to obtain elements to integrate in the next questions of a sequence is very far of the reality, where no one use a system made for a language with another language and with sequences of questions and anaphora. So, finally, CLEF 2008 and CLEF 2007 have only signification in monolingual tracks.

The improvements made on our system since one year (from CLEF 2007 till CLEF 2008) have been useful, even if the results are not really superior, because the complexity of the questions is bigger and the extraction of answers is more difficult in Wikipedia than in news. But we are not sure that some improvements, specially to take into account the Redirect pages have a sense outside of the CLEF evaluations !

Our participation to Quaero project changes now our point of view and we prepare in this project new QA evaluations using millions pages from the Web as corpus and real requests of users to test our systems. Our intention is to evaluate our technologies with real user cases and in real context, measuring not only the quality of the answers but also, for example, the response time.

Acknowledgments

The authors thank all the engineers and linguists that took part in the development of QRISTAL. They also thank the Portuguese company Priberam for allowing them to use their module for question analysis in Portuguese. They finally thank the European Commission which supported our development efforts through TRUST and M-CAST projects, and the AII and Oseo for support the present development efforts through QUAERO and OpenSem projects.

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] AMARAL C., CASSAN A., FIGUEIRA H., MARTINS A., MENDES A., MENDES P., PINTO C., VIDAL D. (2007). Priberam's question answering system in QA@CLEF 2007, *Working Notes for the CLEF 2007*, 19-21 september 2007, Budapest, Hungary.
- [4] 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.
- [5] CASSAN A., FIGUEIRA H., MARTINS A., MENDES A., MENDES P., PINTO C., VIDAL D. (2007). Priberam's question answering system in a Cross-lingual environment, CLEF 2006, *Working Notes for the CLEF 2006*, 20-22 september 2006, Alicante, Spain.
- [6] GRAU B.. (2004), L'évaluation des systèmes de question-réponse, *Évaluation des systèmes de traitement de l'information*, TSTI, p. 77-98, éd. Lavoisier.
- [7] GRAU B., MAGNINI B. (2007), Préface, *Réponses à des questions*, Traitement automatique des langues, volume 46 – n°3/2005, Hermès, Lavoisier, Paris, 2007.

- [8] 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)*.
- [9] 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.
- [10] 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.
- [11] LAURENT D., SÉGUÉLA P., NÈGRE S. (2005), Cross-Lingual Question Answering using QRISTAL for CLEF 2005, *Working Notes for the CLEF 2005*, 21-23 september 2005, Wien, Austria.
- [12] LAURENT D., SÉGUÉLA P., NÈGRE S. (2005), Cross-Lingual Question Answering using QRISTAL for CLEF 2006, *Working Notes for the CLEF 2006*, 20-22 september 2006, Alicante, Spain.
- [13] LAURENT D., NEGRE S., SEGUELA P, (2005), QRISTAL, le QR à l'épreuve du public. Traitement automatique des langues, volume 46 – n°3/2005, Hermès, Lavoisier, Paris, 2007.
- [14] LAURENT D., SEGUELA P, NEGRE S. (2005), Cross-Lingual Question Answering using QRISTAL for CLEF 2007, *Working Notes for the CLEF 2007*, 19-21 september 2007, Budapest, Hungary.
- [15] LAURENT D. (2006), Industrial concerns of a Question-Answering system ?, *EACL 2006, Workshop KRAQ*, April 3 2006, Trento, Italia.
- [16] LAURENT D., SÉGUÉLA P, NÈGRE S. (2006), QA better than IR ?, *EACL 2006, Workshop MLQA'06*, April 4 2006, Trento, Italia.
- [17] VOORHEES E. M.. (2003), Overview of the TREC 2003 Question Answering Track, NIST, 54-68 (http://trec.nist.gov/pubs/trec12/t12_proceedings.html).