

# Cross-Language French-English Question Answering using the DLT System at CLEF 2004

Richard F. E. Sutcliffe  
Igal Gabbay  
Michael Mulcahy  
Aoife O’Gorman

Documents and Linguistic Technology Group  
Department of Computer Science  
and Information Systems  
University of Limerick  
Limerick, Ireland

+353 61 202706 Tel  
+353 61 202734 Fax

Richard.Sutcliffe@ul.ie Michael.Mulcahy@ul.ie  
Aoife.OGorman@ul.ie, Igal.Gabbay@ul.ie

## 1. Introduction

This article outlines the participation of the Documents and Linguistic Technology (DLT) Group in the Cross Language French-English Question Answering Task of the Cross Language Evaluation Forum (CLEF). Following our experiences last year (Sutcliffe, Gabbay and O’Gorman, 2003), our aim was to improve the system particularly in the early stages of processing, and to make further refinements to other components.

Question Type	Example Question	Translation
what_capital	130 Quelle est la capitale du Vénézuéla?	What is the capital of Venezuela?
company	149 Qui fabrique Invirase?	Who manufactures Invirase?
what_country	37 Dans quel pays européen est située la ville de Galway?	In which European country is the town of Galway located?
mountain	162 Quelle est la plus haute montagne du monde?	What is the highest mountain of the world?
where	166 Où se trouve Halifax?	Where is Halifax?
how_did_die	47 Comment est mort River Phoenix?	How did River Phoenix die?
who	40 Qui a réalisé "Braveheart"?	Who directed "Braveheart"?
when	30 Quand est-ce que le prince Charles et Diana se sont mariés?	When did prince Charles and Diana get married?
unknown	36 Citez une unité de radioactivité	Name a unit of radioactivity

**Table 1: Some of the Question Types used in the DLT system.** The second column shows a sample question for each type. Translations based on submission to WorldLingo are listed in the third column.

## 2. Architecture of the CLEF 2004 DLT System

### 2.1 Outline

The basic architecture of our system is standard in nature and comprises query type identification, query analysis and translation, retrieval query formulation, document retrieval, text file parsing, named entity recognition and answer entity selection.

## 2.2 Query Type Identification

As last year, simple keyword combinations and patterns were used to classify the query. This was accomplished by using the CLEF 03 queries and translated TREC queries (RALI, 2004) as a model.

## 2.3 Query Analysis and Translation

This stage differed greatly from last year. We started off by tagging the query for part-of-speech using XeLDA (2004). We then carried out shallow parsing looking for various types of phrase. Each phrase was then translated using three different methods. Two translation engines and one dictionary were used. The engines were Reverso (2004) and WorldLingo (2004) which were chosen because we had found them to give the best overall performance in various experiments. The dictionary used was the Grand Dictionnaire Terminologique (GDT, 2004) which is a very comprehensive terminological database for Canadian French with detailed data for a large number of different domains. The three candidate translations were then combined – if a GDT translation was found then the Reverso and WorldLingo translations were ignored. The reason for this is that if a phrase is in GDT the translation for it is nearly always correct. It is an excellent resource. For example 'equipe de football' becomes 'football team' and not 'team of football', 'salle d'opera' (Canadian dialect for 'opera') becomes 'opera house' not 'room of opera' and so on. In the case where words or phrases are not in GDT, then the Reverso and WorldLingo translations were simply combined.

The types of phrase recognised were determined after a study of the constructions used in French queries together with their English counterparts. The aim was to group words together into sufficiently large sequences to be independently meaningful but to avoid the problems of structural translation, split particles etc which tend to occur in the syntax of a question, and which the engines tend to analyse incorrectly.

The structures used were number, quote, cap\_nou\_prep\_det\_seq, all\_cap\_wd, cap\_adj\_cap\_nou, cap\_adj\_low\_nou, cap\_nou\_cap\_adj, cap\_nou\_low\_adj, low\_nou\_low\_adj, low\_nou\_prep\_low\_nou, low\_adj\_low\_nou, nou\_seq and wd. These were based on our observations that (1) Proper names usually only start with a capital letter with subsequent words uncapitalised, unlike English; (2) Adjective-Noun combinations either capitalised or not can have the status of compounds in French and hence need special treatment; (3) Certain noun-preposition-noun phrases are also of significance.

As part of the translation and analysis process, weights were assigned to each phrase in an attempt to establish which parts were more important in the event of query simplification being necessary.

## 2.4 Retrieval Query Formulation

The starting point for this stage was a set of possible translations for each of the phrases recognised above. For each phrase, a boolean query was created comprising the various alternatives as disjunctions. In addition, alternation was added at this stage to take account of morphological inflections (e.g., 'go'<->'went', 'company'<->'companies' etc) and European English vs. American English spelling ('neighbour'<->'neighbor', 'labelled'<->'labeled' etc). The reason for this last step was the addition for this year of the Glasgow Herald collection to the existing LA Times. The list of the above components was then ordered by the weight assigned during the previous stage and the ordered components were then connected with AND operators to make the complete boolean query.

## 2.5 Document Retrieval

During document retrieval, the boolean query was submitted to the DTSearch search engine (DTSearch, 2000) which had previously been indexed on the LA Times and Glasgow Herald collections, with each sentence in the collection being considered as a separate document for indexing purposes. This followed our observation that in most cases the search keywords and the correct answer appear in the same sentence.

In the event that no documents were found, the conjunction in the query (corresponding to one phrase recognised in the query) with the lowest weight was eliminated and the search was repeated. Some attempts were made this year to avoid the situation in which the query is inadvertently simplified to something insufficiently selective and highly frequent in the corpus (e.g. United States).

## 2.6 Text File Parsing

This stage is straightforward and simply involves retrieving the matching 'documents' (i.e. sentences) from the corpus and extracting the text from the markup.

## 2.7 Named Entity Recognition

Named Entity recognition was carried out in the standard way using a mixture of grammars and lists. The number of types was increased to 75 by studying previous CLEF and TREC question sets and these were incorporated into the query categoriser also.

## 2.8 Answer Entity Selection

We used the highest-scoring method of answer selection. In this, the named-entity instance is selected which occurs in the vicinity of the maximum number of keywords taken from the translated query, across all document passages. We also experimented with Google re-ordering using a Magnini-type method (Magnini, Negri, Prevete and Tanev, 2002).

## 3. Runs and Results

### 3.1 Two Experiments

We submitted two runs which differed slightly in their term translation strategy.

Query Type	Classif.		Correct Classification								Incorrect Classification								
	C	NC	Run 1				Run 2				Run 1				Run 1				
			R	X	U	W	R	X	U	W	R	X	U	W	R	X	U	W	
animal	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
colour	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
company	3	0	1	0	0	2	1	0	0	2	0	0	0	0	0	0	0	0	0
def_org	9	0	2	1	0	6	2	1	0	6	0	0	0	0	0	0	0	0	0
def_person	10	1	3	2	0	5	3	2	0	5	0	0	0	0	0	0	0	0	1
distance	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
how_did_die	3	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
how_many3	11	2	1	0	0	10	0	0	0	11	0	0	0	2	0	0	0	0	2
how_old	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
name_part	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
nationality	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
pol_party	3	0	0	0	0	3	0	1	0	2	0	0	0	0	0	0	0	0	0
population	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0
team	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0
what_capital	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
what_country	4	0	1	0	0	3	0	0	0	4	0	0	0	0	0	0	0	0	0
what_mountain	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
what_river	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
when	17	0	8	0	0	9	7	0	0	10	0	0	0	0	0	0	0	0	0
when_wk_day	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0
when_year	10	0	1	0	0	9	1	0	0	9	0	0	0	0	0	0	0	0	0
where	17	2	7	0	0	10	4	0	0	13	1	0	0	1	1	0	0	0	1
who	21	2	4	0	0	17	3	0	0	18	0	0	0	2	0	0	0	0	2
unknown	47	23	3	1	0	43	3	1	0	43	1	3	0	19	1	2	0	0	20
<b>Totals</b>	<b>170</b>	<b>30</b>	<b>35</b>	<b>4</b>	<b>0</b>	<b>131</b>	<b>28</b>	<b>5</b>	<b>0</b>	<b>137</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>24</b>	<b>2</b>	<b>2</b>	<b>0</b>	<b>26</b>	

**Table 2: Results by Query Type.** The columns C and NC show the numbers of queries of a particular type which were classified correctly and not correctly. Those classified correctly are then broken down into Right, inExact, Unsupported and Wrong for each of the two runs Run 1 and Run 2. Finally, those classified incorrectly are broken down in the same way.

### 3.2 Results

Results are summarised by query type in Table 2. Concerning query classification it shows for each query type the number of queries assigned to that type which were correctly categorised along with the number incorrectly

categorised. The overall rate of categorisation success was 85% which is identical to the one achieved in TREC (Sutcliffe, Gabbay, Mulcahy and White, 2004). The number of queries classified as unknown was 70.

The performance of question answering in Run 1 can be summarised as follows. Out of the 170 queries classified correctly, 35 were answered correctly. Out of the remaining 30 queries classified incorrectly a further three were answered correctly. Overall performance was thus 38 / 200 i.e. 19%. Results for Run 2 were as follows. 28 of the 170 queries were answered correctly along with two of the 30 queries giving a total of 30 / 200 i.e. 15%. In both runs 63 questions were answered NIL.

### **3.3 Platform**

We used a Dell PC running Windows 2000 and having 256 Mb RAM. The whole system was ported this year to SICStus Prolog 3.11.1 (SICStus, 2004) which is much faster than Quintus.

## **4. Conclusions**

The overall performance this year was 19% compared to the 11.5% we achieved last year. We can attribute this improvement mostly to a superior translation strategy although much further work on this is required.

If we exclude query types which were represented by just one example, the best performance (100%) was on `how_did_die` queries. However, there were only three queries of this type. On the more common types the best performance was achieved when answering 'when' and 'where' queries (47% and 41%, respectively). The performance on the relatively common types `how_many` and `when_year` was poor (9%, 10%, respectively). A more sophisticated answer selection strategy may improve performance on these. It is hard to assess performance on many of the other query types due to the small number of each.

Query categorisation for this year stood at 85% compared to 79.5% last year. The reduction in the number of (correctly classified) unknown queries from 58 last year to 47 in the current run may reflect an improvement in the coverage of categorisation. However, among the unknown queries several categories emerged which should be added in future systems. These include queries which are similar in nature to list questions but ask for a single hyponym (e.g., Query 17 'name a cetacean'), queries about materials (e.g. Query 70 'What are fibre-optic cables made of?'), queries of the type 'What does company X sell/produce?' and queries about diseases (treatment, way of transmission), newspaper names, wars, and musical bands. This year the test set included 'how' and 'why' queries (e.g., Query 8 'How is the pope?', Query 87 'Tell me a reason for teenage suicide', Query 107 'How does acupuncture work?') which our system cannot answer

A significant proportion of the queries are likely to remain unknown in increasingly difficult evaluations. The poor performance on unknown queries highlights the need to develop an alternative to our current strategy of answering such queries (i.e., finding a sequence of capitalised words). Less than a quarter of the correctly classified unknown queries this year can be answered by the current strategy.

Simple heuristics may improve performance on existing categories. For example, year numbers could often be eliminated from answers to `how_many` queries.

Our boolean search query formulation strategy was a big improvement on last year but was not without its problems. In particular the combination for each phrase of translation alternatives, inflection alternatives and spelling alternatives could result on occasion in highly complex queries which were a problem for our relatively lightly engineered search engine.

## **5. References**

DTSearch (2000). [www.dtsearch.com](http://www.dtsearch.com)

GDT (2004) [http://w3.granddictionnaire.com/btml/fra/r\\_motclef/index1024\\_1.asp](http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp)

Magnini, B., Negri, M., Prevete, R., & Tanev H. (2002). Is it the Right Answer? Exploiting Web Redundancy for Answer Validation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia*, 425-432.

RALI (2004) <http://www-rali.iro.umontreal.ca/LUB/qabilingue.en.html>

Reverso (2004) <http://grammaire.reverso.net/textonly/default.asp>

SICStus (2004) <http://www.sics.se/isl/sicstuswww/site/index.html>

Sutcliffe, R. F. E., Gabbay, I., & O'Gorman, A. (2003). Cross-Language French-English Question Answering using the DLT System at CLEF 2003. *Proceedings of the Cross Language Evaluation Forum, CLEF 2003*, , August 21-22, 2003, Trondheim, Norway, 373-378.

Sutcliffe, R. F. E., Gabbay, I., Mulcahy, M., & White, K. (2004). Question Answering using the DLT system at TREC 2003. In E. M. Voorhees and L. P. Buckland (Eds) *Proceedings of the Twelfth Text Retrieval Conference, TREC 2003, November 18-21, 2003, Gaithersburg, Maryland*. NIST Special Publication 500-255. Gaithersburg, MD: Department of Commerce, National Institute of Standards and Technology, 686-692.

WorldLingo (2004) [http://www.worldlingo.com/products\\_services/worldlingo\\_translator.html](http://www.worldlingo.com/products_services/worldlingo_translator.html)

XeLDA (2004) <http://www.temis-group.com/temis/XeLDA.htm>