# Can Monolingual Users Create Good Multilingual Queries without Machine Translation?
## (Working notes)

Bill Ogden & Bo Du
Computing Research Lab
New Mexico State University
ogden@crl.nmsu.edu

Our interest in Interactive and Cross Language Text Retrieval has led to the design of a user interface for the cross language task. While many automatic techniques for automatic query term translation and disambiguation have been proposed and tested, little work has involved the evaluation of a system in combination with its user.  We and others have proposed designing an interface that allows the user to help disambiguate terms provided by a system by providing "back-translations" of the system selected terms from which a monolingual user can select the appropriate meanings.  The MULINEX system (http://mulinex.dfki.de) provides a query assistant feature with just such an interface. For our CLEF experiment we wanted to see if these types of interfaces would help a monolingual user create good multilingual queries with minimal automatic translation capabilities.  We compared the retrieval performance of manual queries produced by our monolingual user, to the performance of automatic translated queries produced by the "bablefish" AltaVista web site as generated by the Systran MT system (http://world.altavista.com/).

The experimental system is designed to allow users to query in English with targets in many languages. The only required translation resources are target language to English bilingual dictionaries, which in many cases can be obtained freely on the WEB.

Queries for our CLEF experiment were generated by a single native Chinese speaker, who knew English but had little to no experience with German, French, or Italian.  For each topic, the user would read the English title, description, and narrative, and select the English terms from these sections judged to be the best query terms.  They were not allowed to select terms not in the original topic.  Then for each of the other target languages the system showed extended English definitions of query terms and phrases alongside their dictionary translations. The system only shows query terms and phrases that occur in the target data.  The user then selected query terms with English definitions most accurately reflecting the intention in the original query. The query terms selected for each language is used to retrieve and rank documents for that language and the results for all languages are merged into the final ranked list for our experimental CLEF run.

 We produced a comparison run (not judged) using queries generated by submitting the title section of each English topic to the translation feature available on the WEB from Alta Vista and the resulting translations for German, French, and Italian were collected separately. These queries were then used to retrieve and rank documents in the same manner used for our experimental run.

The retrieval system was an internally developed UNICODE retrieval system that uses standard IDF/TDF weighted document ranking  (http://crl.nmsu.edu/Research/Projects/tipster/ursa). Results from the separate languages were merged using a simple method that normalized the document relevance scores to Z-score for each language/topic and then sorted the different languages results using the normalized scores.

The two methods produced similar results.  Average Precision for our monolingual user was .1136 and for the Systran system was .1212.  Detail analysis will be produced for the full paper, but it appears that both systems had trouble with many of the topics and that one did better on some topics and one on others. Clearly a careful item analysis will be necessary

This was a preliminary experiment designed to test the feasibility of our approach. As usual the quality of the bilingual dictionaries will have a strong effect on the outcome.  Some good query terms just were not present in the dictionaries we used.  In addition, our retrieval and ranking software could be better tuned to take advantage of the forms of the dictionary entries and phrases.  Follow up analysis will compare the results obtained by the monolingual user to the results obtained with the hand-translated queries provided for the cross language topics. We will also compare the performance monolingual user queries to the hand-translated ones for each language in individual monolingual runs.

These preliminary results can at least be used to conclude that MT may not be necessary to produce good cross-language queries when you can involve the user. This is important for language pairs that do not have MT systems available but have good online bilingual dictionaries. Thus, an interactive system could work for many more language pairs than those automatic systems.

In this CLEF experiment, the monolingual user receives no feedback concerning effects of their query term selections. This is unfortunate given that we are actually interested in the whole interaction, including the identification and selection of relevant documents. We are working on several techniques, again using limited translation resources, that will help a monolingual user judge the relevance of documents in languages the cannot read. The CLEF evaluation task is not suited for the type of document selection study that could help us determine how successful these techniques may be. However the CLEF relevance judgments data can help us design and evaluate such studies and we are planning to conduct interactive document selections studies in the near future.