MT-Based Query Translation CLIR Meets Frequent Case Generation

Kimmo Kettunen

ACKNOWLEDGMENT

This work was supported by the Academy of Finland grant number 1124131. I wish to thank Ms. Eija Airio, Dept. of Information Studies, University of Tampere, for implementing all the Unix scripts for the query processes

ABSTRACT

The paper introduces the evaluation results of Cross Language Information Retrieval(CLIR) for three target languages, Finnish, German and Swedish using English as the source language. Our CLIR approach is based on machine translation of topics and usage of the Frequent Case Generation (FCG) method for management of query term variation in translated topics and retrieval in inflected indexes. Retrieval results of more standard query term variation management approaches, such as stemming and lemmatization of translated topics, are also shown. Results of the paper show, that when machine translation of queries are combined with FCG, results can be at best very promising. The best Machine Translation (MT) programs seem to translate standard laboratory type Information Retrieval (IR) topics quite well at least from the query performance point of view. Few times the translated queries perform as well as or slightly better than the monolingual baseline. Many times differences to monolingual baseline are small.

* Kymenlaakso University of Applied Sciences, Finland, Kkettun4@welho.com

PAPER TYPE: Research Paper

KEYWORDS

Machine Translation; Cross Language Information Retrieval (CLIR); Frequent Case Generation (FCG); Information Retrieval (IR); Stemming; Lemmitization

INTRODUCTION

Cross Language Information Retrieval (CLIR) has become one of the research areas in information retrieval during the last 10+ years **(Kishida, 2005)**. The development of WWW has been one of the key factors that increased interest in retrieval tasks where the language of the queries is other than that of the retrieved documents. There are vast amounts of textual data in various languages available electronically and the textual and linguistic abundance increases constantly. Thus there is and will be a social need for retrieval systems, where user can state his/her search request in native language and get the documents in another language that he/she is capable of understanding to the extent that some information need is satisfied. The CLIR system will take care of the translation of the user's search request and thus the user does not have to think about search terms in a foreign language he/she is capable of understanding quite well in reading but in which he/she might not be that fluent in producing especially if the search task involves special terminology. In this paper we evaluate retrieval results of machine translated queries, which have been combined to a novel approach of query term variation management in monolingual retrieval, Frequent Case Generation (FCG). Machine translation of queries has been evaluated for many languages, such as Spanish English (Fiquerola et al., 2000; Xu and Weichsedel) English {Spanish, Chinese, Arabic} **(Xu and**

Weichsedel, 2004), German English (Oard and Hackett, 1997), {French, German, Italian, Spanish, Japan, Chinese} English **(Jones and Lam-Adesina, 2001; Lam-Adesina and Jones, 2002, 2003)** and English {French, Portuquese}**(Monz, 2006)**. FCG, on the other hand, has been quite recently introduced to management of monolingual query term variation **(Kettunen 2008; Kettunen, Airio & Järvelin 2007)**. It has proven quite successful in management of query term variation for morphologically very or somehow complex languages, such as Finnish, German and Swedish, and so it is of interest to verify, if the method can be used in CLIR of these same languages. **Airio and Kettunen (2009)** have tried FCG quite successfully in CLIR with English, Finnish and Swedish, but in this context it was used with a dictionary-based query translation tool, Utaclir **(Hedlund, 2003; Hedlund et al., 2004)**. We shall report results of MT-based query translation CLIR from English to three languages: Finnish, German and Swedish. The translation direction is chosen deliberately from a morphologically simple language to more complex ones. No translations to English from the three languages have been done, because this would not give results of great interest, as the monolingual retrieval results of English do not differ almost at all regardless of the used query term variation management method **(Kettunen, 2008)**. Translations to three morphologically more complex target languages will show better differences in keyword variation management methods.

We use different query term variation management techniques in our experiments for comparison. This means also that different types of textual indexes are used for retrieval. As will be seen in the Results section, three different types of query term variation management methods are used **(Kettunen, 2009)**. Lemmatization is reduction of words to a linguistic base-form, and stemming is

reduction of words to a stem form with a simple stemmer program. FCG produces full inflected forms of keywords. Besides these three methods also plain unprocessed word forms are used as a comparison. **Figure 1** clarifies relationship of query term management methods to the form of the textual index.

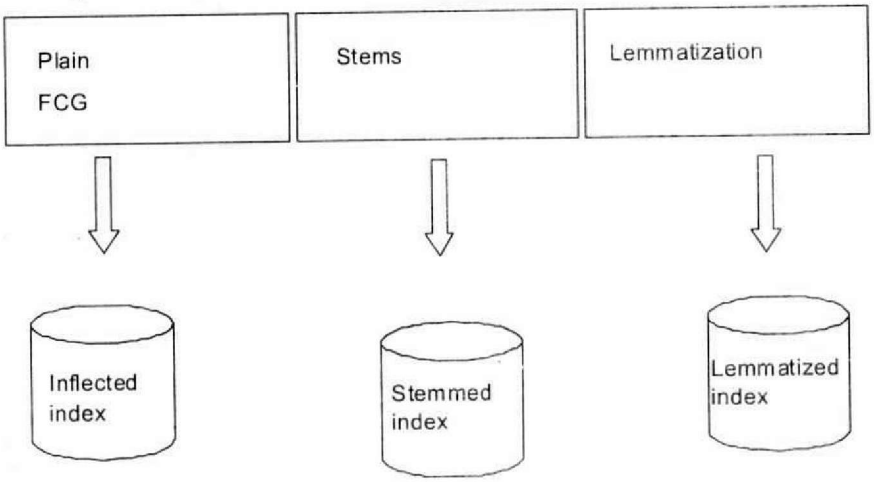


Figure 1 Query Term Management Methods and Indexes

FCG enables retrieval in non-normalized indexes i.e. in an index where no stemming, lemmatization or other similar techniques have been applied, and which are common for example in the Web **(Rasmussen, 2003)**. Thus, application of the FCG methods in CLIR context offers a practical method to perform bilingual retrieval in an inflected word form index.

We study the following research questions in this paper: does FCG bring anything useful to CLIR? Can it solve some of the problems of CLIR and could it be a useful approach? This is our main research question, and as smaller topic we discuss whether MT based

CLIR is getting any more feasible, especially in comparison to dictionary-based CLIR, with present state MT?

The paper is arranged as follows. First we shall introduce some related CLIR research and basic concepts and problems of CLIR. After that we introduce our materials and methods and machine translation programs we have used. Section 4 shows results for the three languages and section 5 discusses the results and draws conclusions.

CLIR APPROACHES AND PROBLEMS OF CLIR

CLIR has had many approaches. The most fundamental distinction in CLIR has been target of translation: one may either translate only the user queries or whole documents in the textual database. The former has been more used due to cost reasons and simplicity of the approach. When queries are translated, different methods can be used: either the queries are translated with electronic dictionaries or word lists, with machine translation programs or using large parallel corpora and statistical methods as translation's knowledge source. All these (query) translation methods have been successful and widely used and they can also be mixed. **(cf. Abusalah et al., 2005; Kishida, 2005; Kraaij, Nie and Simard, 2003; Levow et al., 2005; Oard and Diekema, 1998)**. In this chapter we introduce briefly the three different methods for query translation as all of the different translation approaches are somehow present in our experiments.

- **Query translation using dictionaries or word lists**

Dictionary-based query translation systems have been popular, because they are usually easily implemented from available bi- or multilingual machine readable dictionaries. A query is translated by replacing each query term with its translation equivalents from the

dictionary or word list. Translation in such a system is merely a bag of words, where every possible translation is taken out from the dictionary, even if the translation is not correct in the context. Effects of false translation equivalents are usually minimized by query structuring **(Pirkola, 1998)** or by query term weighting **(Levow, Oard and Resnik, 2005)**. Limits of the approach include out-of-vocabulary words (OOVs), i.e. words missing from the dictionaries, ambiguous translations and failure to translate multiterm concepts (phrases) **(Ballesteros and Croft, 1997; Kishida 2005; Lewow, Oard and Resnik, 2005)**.

A typical dictionary-based query translation system is Utaclir that is described in detail in **Hedlund (2003)** and **Hedlund et al. (2004)**. Utaclir translates queries between English and {Finnish, German, Swedish} and uses bi- and multilingual machine readable dictionaries as its translation resource. Its capabilities are enhanced for example by compound word handling, phrase construction and n-gram techniques for problem word translation. In its basic form it typically achieves about 65-79 % of the performance of monolingual baseline. Compound word handling boosts the performance slightly, and about 67-82 % of the monolingual performance is achieved then. Transitive translations via a third language also boost performance of Utaclir **(Hedlund et al. 2004)**.

Lehtokangas, Keskustalo and Järvelin (2008) used a slightly different type of dictionary-based query translation. Their method was based on the use of a transitive dictionary translation, where translations from source to target language are made through a third language (pivot). In their case English and Swedish were used as pivot languages and Finnish as the target language. This method enables translations between languages that do not have electronic sources for direct translation. The authors reported good results, and

the transitive runs achieved on average 66-72 % of the monolingual and 85-93 % of the direct translation performance. Pseudo-relevance feedback raised the performance of transitive translation further.

Lewov, Oard and Resnik (2005) describe their system, which translates between English and {French, Mandarine Chinese, German, Arabic}. Their CLIR results were between 69-101 % of the monolingual baseline. The approach is based on the use of wordlists obtained from the Web.

➢ **Machine Translation**

Machine Translation programs have been more readily available during the last years, and their quality has also become better. Many of the programs are available as free web services with some restrictions on the number of words to be translated, and many standalone workstation programs can be obtained with evaluation licenses. Thus MT programs are good candidates for query translation. CLIR can also be considered a good application area for "crummy MT" **(Church and Hovy, 1993).**

The following problems of machine translation with respect to CLIR are usually listed **(Kishida, 2005; Oard and Hackett, 1997)**:

- queries are usually short and do not provide enough context for translation
- many times queries are only bags of words, not proper sentences thus undermining the benefits of MT
- translation ambiguity might be problematic: MT programs produce usually one translation, and if the ambiguity resolution of the program fails, the translation may be wrong; also, the translation will not benefit from possible query expansion effects of synonyms or related words as a translations of a dictionary-based query translation system do.

If these "problems" are considered more carefully, at least two of

them do not hold in a laboratory IR type of retrieval evaluation. Typical topics of e.g. CLEF collection are quite long and usually also contain full sentences or at least well-formed nominal phrases, so their translation by MT programs should not be very problematic. The third problem exists, but also its effect might not be that severe: one of our test programs (Tolken99) had a mode for multiple translations of words, but this option did not produce almost any gain in retrieval. Most of the multiple translations given by the program were actually conjunctions and other function words, which do not enhance retrieval.

- **Translations based on usage of parallel corpora**

Statistical machine translation (SMT) has become popular during the last years. The main idea of statistical translation is that translation probabilities for words or phrases are sought for from parallel corpora, equal texts in different languages. For every source (S) and target (T) sentence pair in the corpus, there is a probability P(T|S), "i.e. the probability that T is the target sentence, given that S is the source." Thus "the translation procedure is a question of finding the best value for P(T|S)" **(Somers 2004)**. Cross-language Information Retrieval using parallel corpora as translation resource has become one of the latest developments in CLIR, as parallel corpora are more and more available especially in the Web. Below some CLIR results using parallel corpora as translation resource are discussed.

Monz (2006) reports results of statistical phrase-based machine translation for query translation from English to French and Portuguese. The Europarl parallel corpus was used to estimate the phrase translation probabilities. The results of runs were quite good, although no baseline comparison was given. The main advantage of

phrase-based machine translation is that it produces substantially better translations than word-based MT. Phrase-based translation models use translation probabilities for sequences of consecutive words instead of individual words, and thus more contextual information is captured **(Monz, 2006)**.

Kraaij (2001) showed that mining of translation equivalents from web corpus yielded as good results as Systran MT program for En Fr retrieval. **Kraaij et al. (2003)** further developed a web-mining system for acquisition of parallel corpora and trained translation models of English French and English Italian on these corpora. These translation models were then utilized in translation of CLEF topics from years 2000-2002. Their results showed that at best the parallel corpora approach was able to outperform "a good MT system (Systran)", especially when different parallel translation techniques were combined. They emphasize, that the noisiness of parallel corpora does not affect CLIR, while IR is quite error tolerant. One of the benefits of their approach was also that query translation and retrieval was integrated **(Nie, 2003)**.

MATERIALS AND METHODS, FCG AND MACHINE TRANSLATION PROGRAMS

This section introduces materials and methods that have been used in the experiments. We also discuss briefly machine translation programs involved in the research setting, because their origins and nature vary.

- **Materials and methods**

CLEF collections for Finnish, German and Swedish were utilized in this study. In Table 1, the number of documents and topics with relevant documents in each collection is shown. There are 60 topics to be translated from English to each target language, but each

language has a different number of topics that have relevant documents.

Table 1 Collections used in the study

Language	Collection	Collection size (docs)	Topics	IR system
FI	CLEF 2003	55 344	45	Lemur
DE	CLEF 2003	294 809	56	Lemur
SV	CLEF 2003	142 819	54	Lemur

The retrieval system was Lemur. Lemur combines an inference network retrieval model with language models, which are thought to give more sound estimates for word probabilities in documents **(Metzler and Croft, 2004; Grossman and Frieder, 2004)**.

➢ **Query translations**

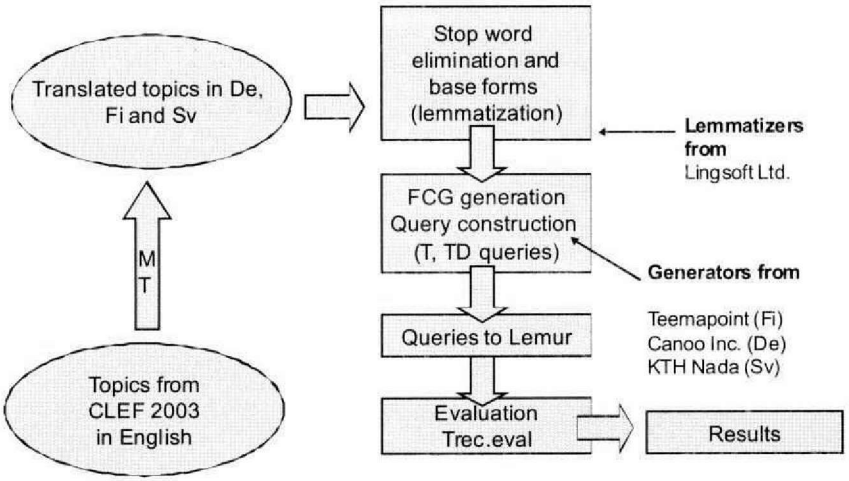
The process of query translation and retrieval is arranged as follows in our experiments.

1) English CLEF 2003 topics are first translated to target languages with available machine translation programs for each language. We translated separately title and title and description fields from the topics. Some of the used MT programs are free web versions, some commercial programs that have been used under test license. En Fi topics were translated with Sunda's MT program (www.sunda.fi), Google Translate Beta (http://translate.google.com/translate_t) and Teemapoint's MT program (www.teemapoint.fi, version 1.3). Programs used for En Sv translation are Systran's web translator (http://www.systran.co.uk/), Google Translate Beta (http://translate.google.com/translate_t) and

Tolken99 (http://www.tolken99.net/, version 4.2), a MT program for PCs. EnDe translations were performed with four different MT engines: Google Translate Beta, Promt Reverso's MT program (http://translation2.paralink.com/), Babelfish's web translator (http://babelfish.yahoo.com/translate_url), which is Systran's program **(Yang and Lange, 2003)** and Translate It!, an older PC workstation MT program. All translations were performed in May and early June of 2008. The translation phase can be characterized as a black box, the results of which are used in the retrieval phase as such, and thus translation and retrieval are separated, which is quite typical to MT query translation approach **(Kraaij, Nie and Simard, 2003)**.

2) After translation the translated topics are normalized morphologically with Lingsoft's FINTWOL, SWETWOL and GERTWOL lemmatizers respectively. Lemmatized translated topics are sent to FCG procedures that generate variant keyword forms for nouns and adjectives of each language's queries with word form generators obtained from different sources **(cf. listing and references in Kettunen, 2008)**. The final translated FCG queries are run in the inflected textual database of the target language using the Lemur query engine and results are evaluated with *trec.eval*. For comparison also IR results of lemmatized, stemmed and plain query translations are shown. For Finnish and Swedish we have also available comparable Utaclir results.
The whole process of the query translation and retrieval is shown in

Figure 2.
Figure 2 Topic Translation and Query Construction for FCG Queries



The description of process in **Figure 1** is slightly simplified in the following way:

- Translation of topics was done to topics that were first stripped from XML structure and query numbers (pre-processing)
- After translation compound sign '-' was replaced by space in translations, while Lemur's indexing does not allow '-' and this also boosts retrieval results (post-processing)
- Some of the translation programs marked un-translated words with tagging or angle brackets. These markings were removed from translations and un-translated words were left in the translation (post-processing)
- A rudimentary tagging <topic>Topic text</topic>was added to the translations so that the query construction procedures could

process the queries correctly (post-processing).

All these changes were made either programmatically or with a text editor's search and replace function.

➢ **The FCG method**

The FCG method has been first presented for management of morphological variation of query words with Finnish, Swedish, German and Russian monolingual queries in **Kettunen and Airio (2006)** and **Kettunen and colleagues (2007)**. **Kettunen (2008)** has used the method for English, Finnish, German, and Swedish retrieval with off-the-shelf word form generators in a fully automatic query generation process. The FCG method and its language specific evaluation procedure are characterized as follows **(Kettunen, 2008)**:

1) The distribution of nominal case/other word forms is first studied through corpus analysis for a language. The corpus can be quite small, because variation at this level of language can be detected even from smaller corpuses. Variation in textual styles may affect slightly the results, so a style neutral corpus is the best.

2) After the most frequent (case) forms for the language have been identified through corpus statistics, the IR results of using only these forms for noun and adjective keyword forms are tested in a known test collection. As a comparison the best available keyword and index management method (lemmatization or stemming) is used, if such is available. The number of tested FCG retrieval procedures depends on the morphological complexity of the language: more procedures can be tested for a complex language, only a few for a simpler one.

3) After evaluation, the best FCG procedure with respect to morphological normalization is usually distinguished. The testing procedure will probably also show that more than one FCG procedure is giving quite good results, and thus a varying number of keyword forms can be used for different retrieval purposes, if necessary.

It should be noted, that the FCG method does not usually outperform the gold standard, usage of a lemmatizer, for morphologically complex languages. It provides, however, a simple and usually easily implementable alternative for lemmatization for languages that might lack language technology tools for information retrieval.

In this paper FCG is utilized in CLIR for keyword variation management. As shown in Figure 1, the translated queries are first morphologically normalized and after that variant forms of nouns and adjectives are generated in a FCG query generation procedure. These queries are then run in the inflected index of the target language. A more detailed description of the FCG generation procedure and strategy is given in **Kettunen (2008)**.

- **Machine translation programs**

Availability of MT programs for most common (European) languages can be considered quite good at present. Even small languages like Finnish and Swedish have several machine translation programs available for translations from English. We were able to use three machine translation programs for Finnish and Swedish and four for German. For German it would be easy to use many more programs to translate the queries, but it was not thought necessary. For Finnish and Swedish a few more programs are known to exist, but they were either not available or of so low quality, that their use was not feasible. All the used MT programs differ slightly from each other in some respects. The oldest translator was the English German version of Translate It!, a DOS program from the 1990s based on a word-by-word translation model. Google Translate Beta is the latest development of all the programs using statistical methods and parallel corpora in translation. One of the Finnish translators, Sunda's web service, is a commercial large scale translator using the transfer approach in translation. Teemapoint's translator is a workstation

program for PCs that is still under development. Swedish Translator Tolken99 is also a PC workstation program of a small company. Systran's Swedish web translator and all the German translators can be considered as large scale commercial translators, although Translate It! is a rather old program.

RESULTS

This section shows results of the CLIR evaluations for the three languages. The main evaluation figure given in the result tables is non-interpolated mean average precision for all the runs of different methods given by *trec.eval*. Comparison to monolingual baseline is shown as percentage of the monolingual result in parentheses.

➢ Results of Finnish queries

In Tables 2 and 3 results of Finnish long and short queries are shown. Meanings of the keyword variation management method columns for all the languages are as follows: column *lemmatization* shows results of runs, where translated queries have been lemmatized (i.e. words reduced to base forms) and run in a lemmatized index where compound words have been split to composite parts. The column *plain* shows results of translated queries with no further linguistic processing. The column *stems* shows results where translated queries have been stemmed with the Snowball stemmer for each language. The *FCG* columns show results from Frequent Case Generation method. Figures after the FCG show number of the nominal forms used for each FCG procedure in the language. The last column (if present at all), *Comb,* shows results from Utaclir's translations where FCGs have been combined with n-gramming for unknown words (cf. Airio and Kettunen, 2009). These figures are shown for comparison purposes for Finnish and Swedish. The

monolingual row in each table gives a monolingual baseline for every method in a similar monolingual retrieval setting. The monolingual results are from Kettunen (2008).

Table 2 Results of Finnish TD Queries, Mean Average Precisions in Per Cent.

	Lemmati-zation	Stems	Plain	FCG_12	FCG_9	Comb
Monolingual queries	50.7	46.2	37.5	48.0	47.3	N/A
Sunda's MT program	39.7 (78.3 %)	32.7 (70.8 %)	26.3 (70.1 %)	33.7 (70.2 %)	33.9 (71.7 %)	N/A
Teemapoint's MT progam	44.5 (87.8 %)	32.4 (70.1 %)	22.5 (60.0 %)	36.9 (76.9 %)	36.6 (77.4 %)	N/A
Google Translate Beta	40.2 (79.3 %)	39.3 (85.1 %)	31.2 (83.2 %)	38.6 (80.4 %)	38.5 (81.4 %)	N/A
UTACLIR	34.1 (67.3 %)	N/A	11.2 (29.9 %)	32.5 (67.7 %)	32.4 (68.5 %)	37.4 (N/A)

Table 3 Results of Finnish T Queries, Mean Average Precisions in Per Cent.

	Lemmati-zation	Stems	Plain	FCG_12	FCG_9
Monolingual queries	45.3	38.4	30.4	40.3	40.2
Sunda's MT program	25.7 (56.7 %)	22.4 (58.3 %)	16.2 (53.3 %)	21.8 (54.1 %)	21.6 (53.7 %)
Teemapoint's MT progam	22.7 (50.1 %)	15.1 (39.3 %)	14.5 (47.7 %)	27.4 (68.0 %)	26.9 (66.9 %)
Google Translate Beta	33.1 (73.1 %)	26.4 (68.8 %)	20.5 (67.4 %)	27.9 (69.2 %)	27.3 (67.9 %)

Finnish results show that the best results are achieved with Google's Translate Beta both in TD and T queries. Google's translations yield retrieval results that are 79-85 % of the monolingual results of TD queries depending on the term variation management method. The best results are obtained with lemmatization of translations, but also FCGs and stemming yield good results that outperform plain translations with about 8 absolute per cent. This was expected, as the Finnish plain monolingual queries also perform much worse than any of the term variation management methods. Other translations programs perform fairly well, ranging from 70 to 88 % of the monolingual performance. Teemapoint's program succeeds well with lemmatized TD queries being the best here. Noticeable is also that all the MT programs yield better results than Utaclir without combined results of N-grams, and Google's MT results outperform also the combined results of Utaclir. Plain Utaclir queries perform badly because Utaclir produces baseform translations directly out of dictionaries and retrieval was performed in an inflected index **(Airio and Kettunen, 2009)**.

With short queries the results are worse, so that the MAPs are much lower and the percentage of the monolingual runs achieved with translations is also lower. Google performs well and evenly also here, but some of the other MT results are quite low. Teemapoint's translator does not perform well with stemmed and plain queries but succeeds well with FCGs. Sunda's translations do not perform very well with any of the methods.

- **Results of German queries**

In Tables 4 and 5 we show results of German long and short queries. The μ value below the method's name shows the used μ parameter value, which gave the best result for runs - default being 2500 with the Dirichlet smoothing **(Metzler and Croft, 2004)**.

Table 4 Results of German TD Queries, Mean Average Precisions in Per Cent.

	Lemmatization	Stems	Plain	De_FCG_4	De_FCG_2
	(μ= 1500)	(μ= 2000)	(μ= 2400)	(μ= 700)	(μ= 700)
Monolingual queries	44.6	43.3	38.4	41.6	39.4
Promt Reverso MT program	35.0 (78.5 %)	35.1 (81.1 %)	27.5 (71.6 %)	28.0 (67.3 %)	25.1 (63.7 %)
Google Transla te Beta	45.1 (101.1 %)	46.5 (107.4 %)	39.9 (103.9 %)	41.8 (100.5 %)	36.1 (91.6 %)
Babelfish MT progam	35.6 (79.8)	36.8 (85.0 %)	30.3 (78.9 %)	31.0 (74.5 %)	26.6 (67.5 %)
Translate It! MT program	35.4 (79.4 %)	32.2 (74.4 %)	26.1 (68.0 %)	27.9 (67.1 %)	26.4 (67.0 %)

Table 5 Results of German T Queries, Mean Average Precisions in Per Cent.

	Lemmatization	Stems	Plain	De_FCG_4	De_FCG_2
	(μ= 2500)	(μ= 2000)	(μ= 2300)	(μ= 1800)	(μ= 2800)
Monolingual queries	35.2	33.5	28.5	29.6	30.3
Promt Reverso MT program	26.5 (75.3 %)	27.0 (80.6 %)	21.4 (63.9 %)	21.8 (73.6 %)	22.1 (72.9 %)
Google Translate Beta	35.4 (100.6 %)	35.8 (106.9 %)	30.1 (105.6 %)	29.6 (100.0 %)	30.6 (101.0 %)
Babelfish MT progam	29.1 (82.7 %)	29.5 (88.1 %)	24.2 (84.9 %)	23.7 (80.1)	22.1 (72.9 %)
Translate It! MT program	26.1 (74.1 %)	20.3 (60.6 %)	20.5 (71.9 %)	17.9 (60.5 %)	17.7 (58.4 %)

German results are partly astonishing. Results of Google Translate are almost always slightly better than or equal to monolingual results both in TD and T queries. Only with TD queries De_FCG_2 performs worse than monolingual queries, but achieves still 91.6 % of the monolingual MAP. Babelfish is the second best,

and Promt Reverso and Translate It! perform worst. Babelfish and Promt perform relatively better with T queries than with TD queries, and Translate It! is clearly the worst with T queries.

➢ **Results of Swedish queries**

Tables 6 and 7 show results of Swedish long and short queries. The μ value below the method's name shows the used μ parameter value, which gave the best result for runs - default being 2500 with the Dirichlet smoothing **(Metzler and Croft, 2004)**.

Table 6 Results of Swedish TD Queries, Mean Average Precisions in Per Cent.

	Lemmatization	Stems	Plain	Sv_FCG_4	Sv_FCG_2	Comb
	(μ= 800)	(μ= 1500)	(μ= 2500)	(μ= 500)	(μ= 1100)	
Monolingual queries	45.1	41.5	37.4	39.1	36.4	N/A
Tolken99 MT progam	33.7 (74.7 %)	28.0 (67.5 %)	20.0 (53.5 %)	25.8 (66.0 %)	23.0 (63.2 %)	N/A
Systran MT program	26.2 (58.1 %)	22.9 (55.2 %)	18.2 (48.7 %)	21.1 (53.1 %)	17.7 (48.6 %)	N/A
Google Translate Beta	44.5 (98.7 %)	40.5 (97.6 %)	36.0 (96.3 %)	39.7 (101.5 %)	38.7 (106.3 %)	N/A
UTACLIR	37.6 (83.4 %)	N/A	18.1 (48.4 %)	27.3 (69.8 %)	25.1 (69.0 %)	27.8 (N/A)

Table 7 Results of Swedish T Queries, Mean Average Precisions in Per Cent.

	Lemmatization	Stems	Plain	Sv_FCG_4	Sv_FCG_2
	(μ= 800)	(μ= 2500)	(μ= 900)	(μ= 900)	(μ= 900)
Monolingual queries	39.0	36.2	29.5	36.2	34.7
Tolken99 MT progam	32.1 (82.3 %)	21.4 (59.1 %)	16.3 (55.3 %)	21.3 (58.8 %)	18.7 (53.9 %)
Systran MT program	17.3 (44.4 %)	15.4 (42.5 %)	11.2 (38.0 %)	13.6 (37.6 %)	13.9 (41.0 %)
Google Translate Beta	31.3 (80.3 %)	27.9 (77.1 %)	24.3 (82.4 %)	29.8 (82.3 %)	28.8 (83.0 %)

Swedish results for Google are once again outstanding. With TD queries Google achieves 96-106 % of the monolingual baseline, and with T queries the results are 77-83 % of the baseline. Tolken99 and Systran do not perform very well, and Systran is clearly the worst of all. Tolken99 succeeds quite well when translations are lemmatized, but otherwise its performance is not very good, as stems and FCGs achieve only 63-67 % of the monolingual baseline in TD queries and 54-59 % in T queries. Only Google is able to outperform Utaclir, while Systran and Tolken99 perform clearly worse than plain Utaclir.

DISCUSSION AND CONCLUSION

We stated the following research questions for the paper. The first one was our main research question and the second a minor topic.

- Does FCG bring anything considerably new to CLIR? Can it solve some of the problems of CLIR and could it be a useful approach?
- Is MT based CLIR getting any more feasible, especially in comparison to dictionary-based CLIR, with present state MT?

As stated in the methods section, we had 10 different MT programs for three languages in use. MT programs were chosen mainly with respect to availability and/or quality (Swedish and Finnish), not with respect to large scale coverage of all possible programs for the language (especially German). Programs were from different producers and sometimes even from different periods of MT (German). With Swedish and Finnish we had available comparable CLIR results from a dictionary-based CLIR system, Utaclir.

Our main research question is partly hard to answer. MT+FCG performed quite well in all the languages when the MT program was good (Google Translate in all the languages and all Finnish MT

programs). With Finnish MT+FCG performed clearly better than raw MT without any further linguistic processing, MAPS being about 7-14 % better in Finnish long queries and about 5.5-13 % better with short queries. Long German MT+FCG queries did not perform very well, and the best results of German FCGs were only about 2 % better than plain translations, and sometimes De_FCG_2 performed worse than plain translations. Short German FCG queries did not perform well. The biggest improvement to plain translations was only about 0.7 %, and there were dips of 0.5-2.5 % with respect to plain translations. With Swedish the situation was again better. MT+FCG performed 3-5.5 % better with long queries and about 2.5-5.5 with short queries. Best performance was usually given when MT queries were further lemmatized, and also stemming of the translations was beneficial. Overall it seems, that the monolingual FCG results of **Kettunen (2008)** are predictive of the behavior of MT+FCG in CLIR: morphologically most complex language, Finnish, gains most from the method also in CLIR, Swedish somehow and German least. This might be partly due to the used word form generators: Finnish and Swedish word form generators are lexicon-less and thus quite flexible in generation, but the German generator uses a 100 K lexicon that evidently lacks many of the topical words. **Kettunen (2008)** reports, that 19 % of the words in German monolingual TD queries were left without any generations. To sum up the effects of MT+FCG in CLIR: the method seems promising and should be evaluated more with CLIR of morphologically complex enough languages that have available good quality MT resources.

The answer to our second research question is definitely positive: many of our MT programs achieved good retrieval results. When compared to results of Utaclir, Swedish and Finnish MT translations gave better performance with all the three Finnish MT

programs and with one Swedish (Google Translate Beta). Plain Finnish FCG results were better than those of Utaclir with all the three MT programs, and Google Translate Beta outperformed also "boosted" Utaclir FCG that was combined with N-grams. The same happened in Swedish with Google Translate's results. At best the Finnish MT translations achieved about 80 % of the monolingual MAPs with FCGs and even more when lemmatization was used (88 % at best with Teemapoint's translator). MT programs performed better with long queries than with short ones, where they were able to achieve between 50 73 % of the monolingual baseline. This is in accordance with differences in results of long and short queries in monolingual retrieval. The length of the queries with respect to MT is not an issue here, because translations of T queries are almost always the same as the beginnings of TD queries even when T parts are translated separately.

For German we did not have comparable dictionary-based translation results, so our only comparison is the monolingual baseline. With German Google Translate Beta performed very well regardless of the method of linguistic processing of the translation. Translated queries with Google Translate outperformed results of monolingual baseline slightly in four cases out of five both in long and short queries, as was seen in Tables 4 and 5. Other MT programs gave performance that was about 76-80 % of the monolingual baseline, which is quite good. With short queries German MT programs performed also well.

Airio and Kettunen (2009) compared three different methods of query term variation management in non-normalized indexes, n-gramming, FCG and combination of both methods. Combination was usually the best method, but as they mention, it is also quite resource consuming. Our experiments showed that MT-based query

translation combined with the FCG method works also well, but at the same time it can also be considered resource consuming, because query translations need to be first lemmatized for FCG generation. On the other hand, extra processing of translated queries is needed also with stemming and lemmatization and besides also the target indexes need processing with these query term variation management methods. If the indexes are inflected, FCG seems to offer a possible solution in CLIR of morphologically complex target languages.

REFERENCES

Abusalah, Mustafa., Tait, John., & Oakes, Michael. (2005). Literature Review of Cross Language Information Retrieval. *Transactions on Engineering, Computing and Technology,* V. 4, pp. 175177.

Airio, Eija., & Kettunen, Kimmo. (2009). Does Dictionary based bilingual retrieval work in non-normalized index? Information Processing and Management (to appear).

Ballesteros, Lisa., & Croft, W. B. (1997). Phrasal translation and query expansion techniques for cross-language information retrieval. In *Proceedings of the 20th ACM SIGIR conference on research and development in information retrieval*, 8491.

Braschler, Martin. (2004). Combination Approaches for Multilingual Text Retrieval. *Information Retrieval*, 7 (1-2), 183-204.

Church, Kenneth W., & Hovy, E.H. (1993). Good application for crummy machine translation. *Machine Translation,* 8(4), 239-258.

G. Figuerola, C., Alonso Berrocal, J. L., Zazo, A. F., & Gómez-Díaz, R. (2000). Retrieval of bilingual Spanish-English information by means of a standard automatic translation system. *Working Notes for CLEF Workshop*. Retrieved August 14, 2008 from http://clef.isti.cnr.it/DELOS/CLEF/salamanca.pdf

Grossman, David A. and Frieder, Ophir. (2004). *Information Retrieval. Algorithms and Heuristics (2nd ed.),* Springer: Netherlands.

Hedlund, Turid. (2003). Dictionary-based Cross-language Information Retrieval. *Acta Universitatis Tamperensis,* 962.

Hedlund, Turid., Airio, E., Keskustalo, H., Lehtokangas, R., Pirkola, A., & Järvelin, Kalervo. (2004). Dictionary-based cross-language information retrieval: learning experiences from CLEF 2000-2002. *Information Retrieval* 7 (1-2), 99-119.

Jones, G.J.F., & Lam-Adesina, A.M. (2001). *Exeter at CLEF 2001: Experiments with Machine Translation for Bilingual Texts.* Retrieved August 19, 2008 from http://www.ercim.org/publication/ws-proceedings/CLEF2/jones.pdf

Kettunen, Kimmo. (2008). Automatic Generation of Frequent Case Forms of Query Keywords in Text Retrieval. In Nordström, B. and Ranta, A. (eds.), *Advances in Natural Language Processing (pp.222-236)*, GoTAL 2008, LNAI 5221. Springer Verlag.

Kettunen, Kimmo. (2009). Reductive and Generative Approaches to Management of Morphological Variation of Keywords in Monolingual Information Retrieval - an Overview. *Journal of Documentation*, 65(2), 267290.

Kettunen, Kimmo., & Airio, Eija. (2006). Is a morphologically complex language really that complex in full-text retrieval? In T. Salakoski et al. (eds.), *Advances in Natural Language Processing (pp.411-422)*, LNAI 4139. Berlin Heidelberg: Springer-Verlag.

Kettunen, Kimmo., Airio, Eija., & Järvelin, Kalervo. (2007). Restricted Inflectional Form Generation in Management of Morphological Keyword Variation. *Information Retrieval 10(4-5), 45.*

Kishida, Kazuaki. (2005). Technical issues of cross-language information retrieval: a review. *Information Processing & Management,* 41 (3), 433455.

Kraaij, Wessel. (2001). TNO at CLEF-2001: Comparing Translation Resources. In *Working Notes for the CLEF 2001 Workshop.* Retrieved August 21, 2008 from http://www.ercim.org/publication/ws-proceedings/clef2/kraaij.pdf.

Kraaij, Wessel., Niey, J.-Y., & Simard, M. (2003). Embedding Web-Based Statistical Translation Models in Cross-Language Information Retrieval. *Computational Linguistics,* (29) 3, 381-419.

Lam-Adesina, A.M. and Jones, G.J.F. (2002). *Exeter at CLEF 2002: Experiments with Machine Translation for Monolingual and Bilingual Texts*. Retrieved August 19, 2008 from http://clef.isti.cnr.it/workshop2002/wn/7.pdf

Lam-Adesina, A. M., & Jones, G.J.F. (2003). *Exeter at CLEF 2003: Experiments with Machine Translation for Monolingual, Bilingual and Multilingual Texts*. Retrieved August 19, 2008 from http://clef.isti.cnr.it/2003/WN_web/18b.pdf

Lehtokangas, Raija., Keskustalo, Heikki., & Järvelin, Kalervo. (2008). Experiments with Transitive Dictionary Translation and Pseudo-Relevance Feedback Using Graded Relevance Assessments. *Journal of the American Society for Information Science and Technology*, 59(3), 476-488.

The Lemur Toolkit for Language Modeling and Information. (2008). Retrieval, Retrieved August 19, 2008 from http://www.lemurproject.org/

Levow, Gina-Anne., Oard, D. W. and Resnik, P. (2005). Dictionary-based techniques for cross-language information retrieval. *Information Processing & Management* 41 (3), 523547.

McNamee, Paul and Mayfield, James. (2002). Comparing Cross-Language Query Expansion Techniques by Degrading Translation Resources. In *Proceedings of Sigir'02, Tampere, Finland,* 159-166.

Metzler, Donald., & Croft, W. Bruce. (2004). Combining the Language Model and Inference Network Approaches to Retrieval. *Information Processing and Management.* Special Issue on Bayesian Networks and Information Retrieval 40 (5), 735750.

Monz, Christof. (2006). Statistical Machine Translation and Cross-Language IR: QMUL at CLEF 2006. http://clef.isti.cnr.it/2006/working_notes/workingnotes2006/monzCLEF2006.pdf

Nie, Jian-Yun. (2003). Query Expansion and Query Translation as Logical Inference. *Journal of the American Society for Information Science and Technology,* 54 (4), 335-346.

Oard, Douglas W., & Hackett, Paul. (1997). Document Translation for Cross-Language Text Retrieval at the University of Maryland. *The Sixth Text Retrieval Conference (TREC 6).* Retrieved August 19, 2008 from http://trec.nist.gov/pubs/trec6/papers/umd.ps.gz

Oard, Douglas W., & Diekema, Anne R. (1998). Cross-language information retrieval. In Martha E. Williams (ed.), *Annual Review of Information Science and Technology* (ARIST), V. 33, pp. 223256.

Pirkola, Ari. (1998). The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval. In: *Proceedings of the 21st Annual International ACM Sigir Conference on Research and Development in Information Retrieval*, Melbourne, August 24-28. New York: ACM, 55-63.

Pirkola, Ari,. Hedlund, Turid,. Keskustalo, Heikki., & Järvelin, Kalervo. (2001). Dictionary-based cross-language information retrieval: problems, methods, and research findings. *Information Retrieval,* 4(3/4), 209230.

Rasmussen, E. M. (2003). Indexing and Retrieval for the Web. In: Cronin, B. (ed.), *Annual Review of Information Science and Technology*, V. 37, pp. 91124.

Somers, Harold. (2004). Machine Translation: latest developments. In Mitkov, Ruslan (ed.), *The Oxford Handbook of Computational Linguistics (pp.512-528).* Oxford, New York: Oxford University Press.

Xu, Jinxi and Weichsedel, Ralph. 2004. Empirical studies on the impact of lexical resources on CLIR performance. *Information Processing & Management,* V. 41, pp. 475-487.

Yang, Jin and Lange, Elke. 2003. Going live on the Internet. In Somers, H. (ed.), *Computers and Translation. A translator's guide (pp.191-211).* Amsterdam ; Philadelphia: John Benjamins Publishing Company.