UNIVERSITY OF TAMPERE

Authors:	Lehtokangas Raija, Airio Eija, Järvelin Kalervo
Name of article:	Transitive dictionary translation challenges direct dictionary translation in CLIR
Year of publication:	2004
Name of journal:	Information Processing and Management
Volume:	40
Number of issue:	6
Pages:	973-988
ISSN:	0306-4573
Discipline:	Natural sciences / Computer and information sciences
Language:	en
School/Other Unit:	School of Information Sciences

URN: http://urn.fi/urn:nbn:uta-3-737
DOI: http://dx.doi.org/10.1016/j.ipm.2003.10.005

Transitive dictionary translation challenges direct dictionary translation in CLIR

Raija Lehtokangas [a], Eija Airio [a], Kalervo Järvelin [b]
[a] Department of Information Studies, University of Tampere, Finland
[b] Center for Advanced Study, University of Tampere, Finland
email: {raija.lehtokangas, eija.airio, kalervo.jarvelin}@uta.fi

Corresponding author
Kalervo Järvelin
Center for Advanced Study
FIN-33014 University of Tampere, Finland

tel. : +358-3-215 6953
fax : +358-3-215 6560
email: kalervo.jarvelin@uta.fi

Abstract

The paper reports on experiments carried out in transitive translation, a branch of cross-language information retrieval (CLIR). By transitive translation we mean translation of search queries into the language of the document collection through an intermediate (or pivot) language. In our experiments, queries constructed from CLEF 2000 and 2001 Swedish, Finnish and German topics were translated into English through Finnish and Swedish by an automated translation process using morphological analyzers, stopword lists, electronic dictionaries, *n*-gramming of untranslatable words, and structured and unstructured queries. The results of the transitive runs were compared to the results of the bilingual runs, i.e. runs translating the same queries directly into English. The transitive runs using structured target queries performed well. The differences ranged from –6.6 to +2.9 % units (or –25.5 to + 7.8 %) between the approaches. Thus transitive translation challenges direct translation and considerably simplifies global CLIR efforts.

Keywords: Cross-language information retrieval; Query translation; Structured queries; Transitive translation

1. Introduction

The amount of accessible electronic information has exploded in recent years thanks to Internet and other international networks. There is a great diversity in the languages texts are written in. The more languages there are, the more there are language barriers to be crossed. Thus it is understandable that cross-language information retrieval (CLIR) has become an important area in both research and practice. For overviews of CLIR, see Oard and Diekema (1998); Pirkola, Hedlund, Keskustalo and Järvelin (2001).

Information retrieval is traditionally based on matching the words of a query with the words of a document. In CLIR, this kind of direct matching is impossible because the query and the document collection are in different languages. Translation is needed: either the query has to be translated into the language of the documents or the documents have to be translated into the language of the query. Translating the whole document collection is more demanding, as it requires more resources, which is why query translation is more common in CLIR. The query in one language (called *source* language) is translated into the language of the documents (called *target* language). The basic methods in query translation are machine translation, corpus-based translation and dictionary-based translation. (Hull & Grefenstette, 1996) There are also methods bypassing direct translation of query words or documents. See, e.g., a recent work based on language models (Lavrenko, Choquette & Croft, 2002).

Machine translation is not an ideal method of translating queries unless the queries are formulated in grammatically correct sentences. On the other hand, parallel or comparable corpora are seldom available in the topic areas of all queries. Dictionaries that can be used in CLIR are easier to find. These are usually bilingual machine readable dictionaries (MRD) designed for a human reader, and converted for CLIR purposes by removing superfluous material. Bilingual or multilingual thesauri have also been developed for CLIR purposes (see, e.g., Gilarranz, Gonzalo & Verdejo, 1997). Translation in CLIR is a simpler process than what is normally meant by translation: query words are most often translated separately, one by one, without taking into consideration their relations to each other. When using an MRD, each word of a query is simply replaced with all of its translation equivalents in the target language. All the translation equivalents are taken into the final CLIR query. (Ballesteros, 2000; Pirkola, 1998) Dictionary-based translation in CLIR has been used by a number of researchers, among them Ballesteros (2000), Ballesteros and Croft (1996, 1997, 1998), Gollins (2000), Gollins and Sanderson (2001), Hedlund, Keskustalo, Pirkola, Sepponen and Järvelin (2001), Hull and Grefenstette (1996), Pirkola (1998), Pirkola et al. (2000, 2001), Pirkola, Puolamäki and Järvelin (2003).

However, it is not always easy to find suitable MRDs between languages. There are not always good dictionaries even between common European languages. Direct translation from language A into language B may therefore not be possible. However, there might be a dictionary between language A and language C, and one between language C and language B, which means that translation would be possible first from A into C and then from C into B. This kind of translation through an intermediate (also a pivot) language is called *transitive translation*.

One of the basic problems associated with MRD translation is translation ambiguity (Ballesteros, 2000; Pirkola et al., 2000). Natural language words often have more than one sense. When a word is translated, most often all the senses are automatically taken into the translated query even though not all of them are relevant. In dictionary-based CLIR, methods for choosing between senses to translate have been explored but have not yet proven effective (Sperer & Oard, 2000). If we translate, for example, the Swedish 2001 CLEF title “Reservat för valar” (“Reserve for whales”) into English, the inflected word form ’valar’ is first normalized into base form ‘val’ (noun singular nominative) which has three senses: 1) election 2) whale 3) selection, choice. If all these senses are included in the dictionary, we might have the following words in the target language query: *election poll whale choice choosing selecting selection*, of which only *whale* is correct.

Ambiguity may occur at every stage of the translation process because the query words in both source, pivot and target language may be ambiguous. The number of irrelevant words in a query is likely to increase every time a translation is performed. It is easy to imagine that ambiguity would be a problem of transitive translation in particular because of the additional translation phases needed.

In most cases transitive translation has performed worse than bilingual translation. This is probably because of the ambiguity introduced by double translation. Some techniques have been experimented with to improve the performance of transitive translation. Gollins and Sanderson (2001) tried to solve the problem of ambiguity in transitive translation through triangulation, i.e. by using several translation routes. They used several pivot languages and merged the translation results from the different routes. This indeed had a favourable effect. On the whole, the effectiveness was low, mainly because of the poor translation resources used. Ballesteros (2000), for her part, reduced the ambiguity of transitive translation by query structuring and various expansion techniques. In this paper we study how well transitive translation performs compared to the baseline direct bilingual translation when morphological analyzers, electronic dictionaries, stopword lists and *n*-gramming of untranslatable words are used in the translation process. The choice of languages to be used as source and pivot languages and the use / non-use of structured target queries are the variables tested in this study. The effect of triangulation is also tested.

This paper does not introduce new techniques to CLIR. However, the present combination of techniques has not been used before in transitive CLIR (only in direct CLIR). In particular, compound splitting and component translation, as well as *n*-gram translation of problem words have not been used as transitive CLIR techniques. They are, however, interesting components in the process since

- compound splitting in the transitive process gives rise to much ambiguity, which needs to be managed
- problem words (proper names) are translated directly from the source language into the target language by *n*-gram matching in the target language index, i.e. avoiding transitive translation; this is both necessary (there is no pivot word list) and effective.

We are able to

- confirm earlier results for new language pairs
- show that transitive CLIR is a competitive technique, with the present mix of tools, at a much higher performance level than reported previously
- show that at high performance levels triangulation is only useful in the case of unstructured queries. With structured queries it is not helpful.

This article is organized as follows: methods and data used are presented in Section 2, and findings in Section 3. In Section 4 the findings are further discussed, and some suggestions for future research are given. Section 5 concludes the paper.

2. Methods and data

2.1. The test database

As a test collection we used the English collection of CLEF[1], which contains newspaper articles from the Los Angeles Times and consists of 113,005 indexed documents. CLEF provided 33 test topics in the year 2000 campaign and 47 test topics in the 2001 campaign. These topics – the Finnish, Swedish and German versions – together with CLEF relevance assessments against the Los Angeles Times collection were used in the tests. The two topic sets behave differently, the year 2000 set being harder. Therefore the sets are treated separately in this study.

2.2. The InQuery retrieval system and queries

As a retrieval system we used InQuery, a probabilistic retrieval system provided by the Center for Intelligent Information Retrieval (CIIR)[2] at the University of Massachusetts (Broglio, Callan & Croft, 2000). InQuery[3] queries are either natural language queries (e.g. English sentences) or structured queries. Structured queries are constructed by using, e.g., the operator *syn*, which treats all of its arguments as instances of one search key. All operators are preceded by the hash sign #, and the arguments are delimited by parentheses, e.g. #syn(ship vessel boat). The proximity operator *uwn* (unordered window, n stands for an integer) requires that all of its arguments occur within the window of n words, e.g. #uw3(information retrieval). If no operator is given, the operator *sum* is used as default. This treats all of its arguments as having an equal influence on the result. Pirkola (1998), Pirkola et al. (2001), and Hedlund et al. (2001) describe the use of InQuery query language in our CLIR systems.

2.3. UTACLIR translation process

The transitive translation processes used in this study were constructed on the basis of already existing bilingual processes that use English as target language and Finnish, Swedish and German as source languages. They are automated processes in which target language queries are constructed from source language request sentences by using bilingual electronic dictionaries, morphological analysis programs and stopword lists. Dictionaries for Finnish-English and Swedish-English translation are by Kielikone plc, Finland and include respectively 110,000 and 60,000 entries. The dictionary used for German-English translation is by Oxford Duden and includes 260,000 entries. The morphological analyzers are SWETWOL, FINTWOL, GERTWOL and ENGTWOL by Lingsoft plc, Finland. Hedlund et al. (2001) describe this process.

In direct bilingual processes, source language words are translated one by one, not considering their relations to one another (e.g. phrasal connections). All the translation equivalents which a dictionary gives for a source language word are accepted into the (final) target query. Queries are either structured or unstructured: in the case of structured target queries, all the translation equivalents of a source word are combined by the *syn* operator of the InQuery retrieval system, using the Pirkola Method (for *syn*-based query structuring, see Pirkola, 1998; Kekäläinen & Järvelin, 1998). Special attention is paid to the processing of compound words because all the source languages used are rich in compounds (by a compound word we mean two or more words written together, e.g. *sunbeam*, *horseshoe*). Compound words that cannot be translated as such are split and translated as components; translation equivalents of the components of a compound are combined in the final

[1] http://www.clef-campaign.org, http://www4.eurospider.ch/CLEF/
[2] http://ciir.cs.umass.edu
[3] http://ciir.cs.umass.edu/irdemo/inqinfo/inqueryhelp.html

query by a proximity operator. *n*-Gram techniques are applied to words (and compound components) that do not translate (e.g., proper names). (Hedlund et al., 2001)

2.4. UTACLIR transitive translation process

In this study, we experimented with four transitive processes: Swedish-Finnish-English, Finnish-Swedish-English, German-Finnish-English and German-Swedish-English. The bilingual processes (Section 2.3) were used as baselines. In translations Swedish-Finnish, Finnish-Swedish and German-Finnish, the corresponding Motcom dictionaries by Kielikone plc, Finland, were used. In translations from German into Swedish, a bilingual wordlist compiled from a German-Swedish-German dictionary, Norstedts Tyska Ordbok by Norstedts Ordbok AB, Sweden (127,000 words and phrases), was used. All the translation equivalents of a source query word were accepted into the final target query. We experimented with query structuring using the Pirkola (1998) Method: all the target words derived from the same source word were combined by the *syn* operator of the InQuery retrieval system. The performance of the structured queries was compared to the performance of the unstructured queries, the only distinction between the approaches being the use / non-use of the *syn* operator in the target queries.

As test topics we used the CLEF 2000 and 2001 topics (33 and 47 topics respectively). As source language queries were used the *title* and the *description* fields of the topics. Throughout the process, words were processed one by one, i.e., the phrase structure of the source or the pivot language was not preserved. Only in translations into the target language (English), was the proximity operator *uwn* applied to translation equivalents consisting of more than one word (e.g. #uw6(land property)).

The transitive process can in general terms be described as follows: topic words are first normalized, then stopwords are removed. After translation into the pivot language, the translated words are normalized and stopwords removed. After translation into the target language, the translated words are again normalized and stopwords removed. There is a slight difference in the translation processes, however. In the processes Swedish-Finnish-English and German-Finnish-English there is no word form normalization after translation into Finnish because normalization at that stage, besides being unnecessary, proved harmful as it introduces irrelevant words into the process (due to inflectional homography).

In handling compounds that were not translated as a whole word, an approach slightly different from the original UTACLIR process was applied: we did not use the proximity operator *uwn* to combine the translation equivalents of the components, but these were combined by the *syn* operator only. For example, German source word ‘Weltwetter’ (‘world weather’) was translated into English via Finnish as follows: *#syn(earth world globe weather better)*. (In order to make the results of the transitive and bilingual runs comparable, the same approach was also applied to the bilingual runs in our experiments.)

To give an example of the (structured) target queries, the direct and transitive translations of the Finnish CLEF title 071 ‘Vihannekset, hedelmät ja syöpä’ (‘Vegetables, fruit and cancer’) are given below (assuming that the topic consists of these words only):

Finnish-English:
#sum(#syn(vegetable) #syn(fruit) #syn(bite capture eat #uw6(eat away) #uw6(eat) #uw6(feed) take cancer carcinoma))

Finnish-Swedish-English:

#sum(#syn(vegetable *elegy clymer leggy @beymer @heymer @meryle*) #syn(fruit) #syn(food nourishment #uw7(give birth) feed support consume spend consume eat drink cut stab *pinta int intact @inta @intal @intar* consume bite hit strike beat hit beat beat cut mow beat dial roll eat eat cancer crayfish cancer #uw6(crab)))

In the transitive example above, the words in italics come from the *n*-gramming process for untranslated words, in this case Swedish words not translated into English. In the process, the six best matching words are taken from the database index, three of them from among the index words not recognized by the morphological analyzer (indicated by the sign @). (For *n*-gram techniques, see Pirkola, Keskustalo, Leppänen, Känsälä and Järvelin (2002).)

The bilingual runs Swedish-English, Finnish-English and German-English were used as the baseline runs. The results of the transitive runs were compared to those of the bilingual runs to find out how much is lost in effectiveness when translation is performed through an intermediate language. Moreover, experiments on triangulation (Gollins & Sanderson, 2001) were conducted to discover whether the performance of the original transitive runs German-Finnish-English and German-Swedish-English could be further improved using this approach. For each German source query word, only translation equivalents common to both translation routes were accepted into the final target query. For example, for the German query word 'Unfälle' (CLEF topic 018), the following words were accepted into the triangulated target query:

#syn(accident disaster misfortune calamity),

originating from these two sets of words:
#syn(accident disaster misfortune calamity accident) (Ger-Swe-Eng);
#syn(accident shipwreck wreck crash accident calamity disaster misadventure misery misfortune accident catastrophe disaster accident accident #uw6(send)) (Ger-Fin-Eng).

Unless otherwise specified, the Wilcoxon signed ranks test was applied for runs having Swedish or Finnish, and the Friedman two-way analysis of variance by ranks for runs having German as source language. The significance level was 0.01, unless otherwise specified.

3. Findings

Altogether eight transitive runs were carried out, two using Swedish, two Finnish and four German as the source language. The effectiveness results of the runs are presented in Tables 1 and 2. The results were evaluated as average precision over 10 recall points (10-100 %), using the *deval* evaluation program of InQuery. The results of the baseline direct translations and the monolingual English runs are also given, as well as the differences between the bilingual and the transitive runs. The results from the triangulated runs are likewise presented. Besides being compared to the bilingual results, the triangulated results are also compared to the results of the original transitive runs. In Table 3, the performance of the structured and unstructured target queries is compared. The number of words in target language queries for the different translation routes is given in Table 4, and the percentage of source query words translated in the target queries in Table 5.

The structured transitive runs, on the average, performed well compared to the bilingual runs, the difference between transitive and bilingual translation varying from +3 to –7 % units (or +8 to –25 %), depending on the language combination. There was, however, great variation among the results of individual topics. For some topics, bilingual and transitive queries performed approximately equally, for some the bilingual query performed markedly better, for some the transitive. Among the 2001 Swedish topics, the greatest difference between the runs was 99.4 % (0.6 % for the bilingual run, 100.0 for the transitive run). In all but one transitive run, dictionaries from the same producer were used in both translations. This undoubtably had a favourable effect on the results. On the other hand, even if dictionaries from two different producers were used in the run German-Swedish-English, effectiveness was not affected. - Figures 1-3 present the recall-precision curves for the year 2001 runs, using structured target queries.

Differences between the structured bilingual and transitive runs having respectively Swedish or Finnish as source language were not statistically significant by the Wilcoxon signed ranks test at the 0.01 level. The Friedman two-way analysis of variance by ranks was applied for runs having German as source language. Even there, no statistical significance at the 0.01 level was found between the structured runs German-English, German-Finnish-English, German-Swedish-English and German triangulated.

The effect of query structuring is clearly seen in Tables 1-3. In all the runs, the structured queries outperform the unstructured queries. Contrary to the structured runs, differences between the unstructured bilingual and transitive runs having respectively Swedish or Finnish as source language were statistically significant, excluding the Swedish 2000 run. Likewise, there was a statistically significant difference between the unstructured German 2001 runs. Differences between the unstructured German 2000 runs, on the other hand, were not significant.

Table 3 presents how much effectiveness is improved when structured queries are used instead of unstructured queries. For bilingual runs, the improvement ranges from 13 to 21 %, for transitive runs from 31 to 105 %. In each case, the unstructured and structured results were compared, and the Wilcoxon signed ranks test was applied for each case. As for the 2001 results, all the differences were statistically significant at the 0.05 level. Among the 2000 results differences between the structured and unstructured results relating to the processes Swedish transitive, Finnish bilingual and transitive, and German transitive via Swedish were significant at the 0.05 level. Our results show that query structuring is particularly important for transitively translated target queries. This is consistent with the findings of Ballesteros (2000) and Gollins (2000). - Figures 5 and 6 are presented as examples of the effect of query structuring on direct and transitive translation respectively.

As for structured queries, the effectiveness of the transitive runs could not be further improved by triangulation: in three out of four runs there was an impairment and in one run only a minor improvement of effectiveness. For unstructured queries, the situation was somewhat different: there was an improvement of 32 to 73 % compared to the original transitive runs. According to our tests, triangulation is unnecessary, even harmful, if *syn* operator query structuring is used in the target queries. These findings corroborate at a high performance level the finding by Gollins (2000) that triangulation has only a marginal effect when transitive translations are structured. According to our tests, unstructured target queries could instead derive great benefit from it. It is presumable that the target queries of Gollins and Sanderson (2001) were also unstructured. If this is the case, our findings corroborate theirs, too. - See Figures 3 and 4 for the contribution of triangulation in the case of structured / unstructured target queries.

Table 4 gives the number of words in target language queries to give an idea of the increase of keyword numbers by translations. The increase in number of words from direct to transitive translation varied from 95 to 432 %. Increased number of words means increased ambiguity.

However, there is no straightforward relationship between the increase in number of words and loss of effectiveness when direct and transitive translations are compared. For example, among the 2000 topics the greatest increase of words (432 %) was in the translation route Swedish-Finnish-English, yet this route showed the least decrease in effectiveness (- 1.7 %). In all the translation routes, the unstructured queries were more affected by increase in number of words. This is in line with what was observed about the importance of query structuring for transitively translated queries.

The rate of source query words translated in the target queries is given in Table 5. The two main reasons for a word not to be translated are: 1) the word is not recognized by the morphological analyzer and thus not normalized, or 2) the word is not included in the dictionary as an entry. As in our translation process, words were studied one by one. No distinction between common nouns and proper names was made because both are handled equally in the translation process, too. Stopwords, numerals and one letter words in the source queries were excluded from the calculations. A source query word was considered translated if there was at least one translation equivalent in the target query for the word. For compound words translated as components, at least one translation equivalent for each component was required in the target query. The rate of translated source query words was high, ranging from 78.0 to 91.9 %. However, it was not checked if the translation equivalents were appropriate or not. Some proper names, for example, were incorrectly interpreted as common nouns in the source or pivot language. For example, German surname 'Schneider' was translated into the English word 'tailor', and *el* and *il* in the names 'el Nino', 'Kim il Sung' were interpreted as native words of the source language Swedish, denoting 'electricity' and 'gust of wind', respectively. Even here, there was no straightforward relationship between the rate of translated words and the effectiveness values: for all translation routes, the former were higher among the 2000 topics, whereas the latter were higher among the 2001 topics.

4. Discussion

Transitive translation, i.e. translation through an intermediate language, may be the only means of translation between two languages when there is a lack of suitable translation resources between the languages. Secondly, it may reduce the number of translation routes needed when translations have to be performed between a large number of languages. If there are, for example, 50 languages and a translation system is needed between each pair of these, there will be no less than 2,450 translation pairs and systems. If there is a single pivot language that can be used in all the translations, only 98 translations will be needed (49 translations from the source languages into the pivot language, and 49 from the pivot language into the target languages). The more languages there are, the more abruptly the number of translations can be reduced by using a pivot. This is a theoretical approach to the problem, however. In real situations one pivot language would hardly be enough but, nevertheless, the number of translations needed would be greatly reduced compared to the original situation. This approach, of course, requires that the performance level of the transitive translation systems is not markedly inferior to that of the direct translation systems. The results of this study, however, are encouraging in this respect.

In this study, the performance of the queries based on transitive translation was good - on the average it did not differ much from the performance of the baseline direct translations. It should be noted that this comparable performance was achieved by the basic translation process, using only query structuring - no other measures, such as triangulation (Gollins & Sanderson, 2001), were needed. However, for the first time the process used compound word splitting and problem word *n*-gramming in transitive CLIR. The findings indicate that although compound splitting tends to increase query ambiguity, it still is beneficial in transitive CLIR because query structuring controls ambiguity. Problem word (proper name) *n*-gramming directly into the target language bypasses transitive translation and may greatly improve effectiveness. It would not make sense to try a transitive process on such words – nor, in most cases, to put them intact into the target queries. The present study also shows what mix of components is needed in order to approach the average precision figures of direct translation when the overall performance is 20 to 40 % (average precision).

In the CLEF campaign, the best performing systems of bilingual retrieval throughout the years have achieved a performance level of 41 to 49 % MAP (mean average precision), see e.g. Jones (2002) and Savoy (2002). These systems have employed varying approaches (e.g. machine translation) and a range of techniques (e.g. query biased summarization and pseudo relevance feedback), some of which depend heavily on language-pair specific translation tools and systems. The UTACLIR performance levels reported above are slightly lower but based on a simple dictionary translation approach. Some of the techniques employed in related research, e.g., query biased summarization, pseudo relevance feedback, are applicable in the UTACLIR framework while not part of the present study. The present study shows that the simple dictionary translation approach, with structured queries, generalizes well to the transitive case. By adding other components, further improvements are within reach.

The good results of this study can partly be explained by the dictionaries: in all but one transitive translation route, dictionaries from the same producer were used in both translations, and this undoubtably had a favourable effect. This approach, however, is not uncommon in transitive translation studies. For example, the dictionaries used by Gollins and Sanderson (2001) and Ballesteros (2000) were from the same producer. In previous studies, transitive translations have performed much worse compared to the baseline direct translations. The transitive translations of Ballesteros (2000) using simple word-by-word translation from Spanish into French through English achieved an average precision 91 % below the baseline direct translation. When query structuring by the *syn* operator was applied to the transitive translation, it achieved an average precision 34 % below that of direct translation. Gollins and Sanderson (2001) used in their study, among other things, transitive runs German-Spanish-English and German-Dutch-English. The average respective precision of these runs was 81% and 92 % below the bilingual run. However, through triangulation, i.e. by merging the translation results of the two translation routes, they could raise the effectiveness at its best to 21 % below the bilingual run. When a third translation route was added, the performance was raised to the level of the bilingual translation. Altogether, the performance level of the experiments was low, the average precision of the bilingual run being 5.5 %.

The importance of target query structuring for CLIR has been shown in a number of previous studies (Ballesteros, 2000; Ballesteros & Croft, 1998; Gollins, 2000; Hedlund et al., 2001; Meng et al., 2000; Oard & Wang, 2001; Pirkola, 1998; Pirkola et al., 2000, 2003; Sperer & Oard, 2000). Our study confirms these results and shows that query structuring is particularly important when transitive translation is used. This is natural, since query structuring has a clear disambiguating effect (see, e.g., Ballesteros, 2000; Ballesteros & Croft, 1998; Pirkola, 1998), reducing the ambiguity introduced by the additional translation phases in transitive translation. Besides, our results suggest that triangulation is unnecessary for transitive translation when target query structuring is used.

The results of this study encourage us to undertake further research into transitive translation and its limits, for example: Between what kind of languages is transitive translation successful? What kind of languages should be selected as pivots? What should the dictionaries be like? In translations between a large number of languages, which pivot languages would meet these criteria: a) the language can perform as a pivot between as many languages as possible b) the quality of the translations using this pivot is as good as possible ?

The results of this study were achieved using English as a target language. Whether the same performance level could be reached using other target languages is an interesting question, but is to be investigated later. The languages we were experimenting with were all European languages. This means that they are fairly similar in their conceptual structures, which certainly facilitates translation. Transitive translation studies should also be concerned with languages not having so much in common because only then would translation systems be tested to the utmost (for problems of cross-cultural translation, see Cosijn, Pirkola, Bothma & Järvelin, 2002). In translations between languages conceptually different, e.g. European and indigenous African languages, it might be advantageous to use one (African) pivot language into which all the (European) source languages would first be translated. If the problems on the conceptual level could be solved at this stage, the second translation into the individual (African) target languages might be more straightforward.

According to this study, the results of the bilingual runs were on the average not much affected by the addition of an additional translation phase to the process, i.e. by the introduction of a pivot language. This suggests a further research problem: what if, instead of one pivot language, there were two or more in sequence - could the high level of performance still be retained? Using more than one pivot language has a motivation in reality, too. In some cases when there is a shortage of translation resources between languages, one pivot language may not be enough, but two or more may be needed. Of course a pure academic interest plays a role here, too.

5. Conclusions

In this study, transitive translations were carried out using three source languages, Swedish, Finnish and German, two pivot languages, Finnish and Swedish, and English as a target language. The results of the transitive translations were compared to the results of the direct translations between the three source languages and the target language. The transitive translations performed better than expected given the results of previous transitive translation studies. The difference from the baseline varied at its best from –6.6 to +2.9 % units (or –25.5 to + 7.8 %), the overall performance level being high: 20 to 40 % (average precision). Since the results can be as promising as these, it is worthwhile to further research transitive translation and its limits. Some new research problems emerged in the course of the study, e.g. using two or more pivot languages instead of one. The paper also suggests a number of tools and techniques that are needed to pursue high performance levels, none previously unheard of but new in combination in transitive translation.

The results of this study show that transitive translation is effective. This is encouraging because there is a growing need for translations between languages for which direct translation is not possible. Besides being a necessity when translation resources are scarce, translating through an intermediate language can bring advantages, too. Using a suitable pivot language can, for example, reduce the number of individual translations needed when translations have to be performed between a large number of languages. We hope that some problems related to translations between languages conceptually very different might thereby also be resolved.

Acknowledgements

The InQuery search engine was provided by the Center for Intelligent Information Retrieval at the University of Massachusetts.

ENGTWOL (Morphological Transducer Lexicon Description of English): Copyright (c) 1989-1992 Atro Voutilainen and Juha Heikkilä.
FINTWOL (Morphological Description of Finnish): Copyright (c) Kimmo Koskenniemi and Lingsoft plc. 1983-1993.
GERTWOL (Morphological Transducer Lexicon Description of German): Copyright (c) 1997 Kimmo Koskenniemi and Lingsoft plc.
SWETWOL (Morphological Transducer Lexicon Description of Swedish): Copyright (c) 1998 Fred Karlsson and Lingsoft plc.
TWOL-R (Run-time Two-Level Program): Copyright (c) Kimmo Koskenniemi and Lingsoft plc. 1983-1992.
MOT Dictionary Software was used for automatic word-by-word translations. Copyright (c) 1998 Kielikone plc, Finland.

This research was partly funded by the research project Query Structures and Dictionaries as Tools in Concept-based and Cross-lingual Information Retrieval funded by the Academy of Finland (Research Projects 44703; 49157) and the CLARITY project (Information Society Technologies Programme, IST-2000-25310).

The authors wish to thank Dr. Jaana Kekäläinen for her assistance in preparing this article.

References

Ballesteros, L. A. (2000). Cross language retrieval via transitive translation. In Croft, W. B. (Ed.) *Advances in information retrieval: Recent research from the CIIR* (pp. 203-234). Kluwer Academic Publishers.

Ballesteros, L. & Croft, B. (1996). Dictionary methods for cross-lingual information retrieval. In *Proceedings of the Seventh International Conference on Database and Expert Systems Applications (DEXA)* (pp. 791-801). Berlin: Springer.

Ballesteros, L. & Croft, W. B. (1997). Phrasal translation and query expansion techniques for cross-language information retrieval. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 84-91). New York: ACM Press.

Ballesteros, L. & Croft, W. B. (1998). Resolving ambiguity for cross-language retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 64-71). New York: ACM Press.

Broglio, J., Callan, J. & Croft, W. B. (1994). INQUERY system overview. In *Proceedings of the TIPSTER Text Program (Phase)* (pp. 47-67).

Cosijn, E., Pirkola, A., Bothma, TJD & Järvelin, K. (2002). Information access in indigenous languages: a case study in Zulu. In Bruce, H. et al. (Eds.) *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)* (pp. 221-238). Seattle, WA, USA: Libraries Unlimited.

Gilarranz, J., Gonzalo, J. & Verdejo, F. (1997). An approach to conceptual text retrieval using the EuroWordNet multilingual semantic database. In *Working Notes of AAAI Spring Symposium on Cross-Language Text and Speech Retrieval* (pp. 51-57). Stanford, CA.

Gollins, T. J. (2000). Dictionary based transitive cross-language information retrieval using lexical triangulation. Master of Science Thesis. University of Sheffield.
Available: http://dis.shef.ac.uk/mark/cv/publications/dissertations/Gollins2000.pdf

Gollins, T. & Sanderson, M. (2001). Improving cross language retrieval with triangulated translation. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 90-95). New York: ACM Press.

Hedlund, T., Keskustalo, H., Pirkola, A., Sepponen, M. & Järvelin, K. (2001). Bilingual tests with Swedish, Finnish and German queries: dealing with morphology, compound words and query structure. In Peters, C. (Ed.) *Cross-Language Information Retrieval and Evaluation*: *Proceedings of the CLEF 2000 Workshop, Lecture Notes in Computer Science*, 2069 (pp. 210-223). Berlin: Springer.

Hull, D. & Grefenstette, G. (1996). Querying across languages: a dictionary-based approach to multilingual information retrieval. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 49-57). New York: ACM Press.

Jones, G. J. F & Lam-Adesina, A. (2002). Exeter at CLEF 2001: experiments with machine translation for bilingual retrieval. In Peters et al. (Eds.) *Evaluation of Cross-Language Information Retrieval Systems: Proceedings of the CLEF 2001 Workshop, Lecture Notes in Computer Science*, 2406 (pp. 59-77). Berlin: Springer.

Kekäläinen, J. & Järvelin, K. (1998). The impact of query structure and query expansion on retrieval performance. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 130-137). New York: ACM Press.

Lavrenko, V., Choquette, M. & Croft, W. B. (2002). Cross-lingual relevance models. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 175-182). New York: ACM Press.

Meng, H., Chen, B., Grams, E., Khudanpur, S., Lo, W-K., Levow, G-A, Oard, D., Schone, B., Tang, K., Wang, H-M., and Wang, J.Q. (2000). Mandarin-English information (MEI): investigating translingual speech retrieval. *Final Report of the Mandarin-English Information Group at Workshop 2000,* Center for Language and Speech Processing, Johns Hopkins University, Baltimore, Maryland, USA. Available: http://www.clsp.jhu.edu/ws2000/final_reports/mei/ws00mei.pdf

Oard, D. & Diekema, A. R. (1998). Cross-language information retrieval. In Williams, M. E. (Ed.) *Annual Review of Information Science and Technology*: *Vol. 33* (pp. 223-256). Medford, NJ: Information Today.

Oard, D. & Wang, J. (2001). NTCIR-2 experiments at Maryland: comparing structured queries and balanced translation. *The Second NTCIR Workshop*, 7-9 March, Tokyo, Japan.
Available: http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/jianqiang.pdf

Pirkola, A. (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* (pp. 55-63). New York: ACM Press. Available: http://www.info.uta.fi/tutkimus/fire/archive/the_effects_of_query.pdf

Pirkola, A., Hedlund, T., Keskustalo, H. & Järvelin, K. (2000). Cross-lingual information retrieval problems: methods and findings for three language pairs. *ProLISSa Progress in Library and Information Science in Southern Africa. First Biannual DISSAnet Conference.* Pretoria, 26-27 October 2000. Available: http://www.info.uta.fi/tutkimus/fire/archive/cross_lingual.pdf

Pirkola, A., Hedlund, T., Keskustalo, H. & Järvelin, K. (2001). Dictionary-based cross-language information retrieval: problems, methods, and research findings. *Information Retrieval*, *4*, 209-230.
Available: http://www.info.uta.fi/tutkimus/fire/archive/dictionary_based.pdf

Pirkola, A., Keskustalo, H., Leppänen, E., Känsälä, A.-P. & Järvelin, K. (2002). Targeted s-gram matching: a novel *n*-gram matching technique for cross- and monolingual word form variants. *Information Research, 7*(2). Available: http://informationr.net/ir/7-2/paper126.html

Pirkola, A., Puolamäki, D. & Järvelin, K. (2003). Applying query structuring in cross-language retrieval. *Information Processing & Management, 39*(3), 391-402.

Savoy, J. (2002). Report on CLEF-2002 experiments: combining multiple sources of evidence. In Peters, C. (Ed.) *Working notes for the CLEF 2002 Workshop* (pp. 31-46). Rome, 19-20 September 2002. Available: http://clef.iei.pi.cnr.it:2002/workshop2002/WN/3.pdf

Sperer, R. & Oard, D. (2000). Structured translation for cross-language information retrieval. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 120-127). New York: ACM Press.

Table 1 **Effectiveness of structured target queries: average precision for bilingual, transitive and triangulated translation, using CLEF 2000 and 2001 Swedish, Finnish and German topics**

	2000 topics (n=33)			2001 topics (n=47)		
Translation route	Average precision	Difference	Difference %	Average precision	Difference	Difference %
Eng monol	36.1			48.0		
Swe-Eng	23.6			37.3		
Swe-Fin-Eng	23.2	-0.4	-1.7 %	40.2	+2.9	+7.8 %
Fin-Eng	26.0			40.5		
Fin-Swe-Eng	20.4	-5.6	-21.6 %	35.9	-4.6	-11.4 %
Ger-Eng	23.5			38.9		
Ger-Fin-Eng	17.5	-6.0	-25.5 %	32.3	-6.6	-17.0 %
Ger-Swe-Eng	19.5	-4.0	-17.0 %	32.6	-6.3	-16.2 %
Ger triang	18.5	-5.0	-21.3 %	31.0	-7.9	-20.3 %
Triang vs. Ger-Fin-Eng		+1.0	+5.7 %		-1.3	-4.0 %
Triang vs. Ger-Swe-Eng		-1.0	-5.1 %		-1.6	-4.9 %

Table 2 **Effectiveness of unstructured target queries: average precision for bilingual, transitive and triangulated translation, using CLEF 2000 and 2001 Swedish, Finnish and German topics**

	2000 topics (n=33)			2001 topics (n=47)		
Translation route	Average precision	Difference	Difference %	Average precision	Difference	Difference %
Eng monol	36.1			48.0		
Swe-Eng	19.9			31.6		
Swe-Fin-Eng	15.7	-4.2	-21.1 %	19.6	-12.0	-38.0 %
Fin-Eng	21.6			33.5		
Fin-Swe-Eng	12.0	-9.6	-44.4 %	22.1	-11.4	-34.0 %
Ger-Eng	20.8			34.3		
Ger-Fin-eng	13.4	-7.4	-35.6 %	16.0	-18.3	-53.4 %
Ger-Swe-Eng	14.5	-6.3	-30.3 %	18.7	-15.6	-45.5 %
Ger triang	19.2	-1.6	-7.7 %	27.7	-6.6	-19.2 %
Triang vs. Ger-Fin-Eng		+5.8	+43.3 %		+11.7	+73.1 %
Triang vs. Ger-Swe-Eng		+4.7	+32.4 %		+9.0	+48.1 %

Table 3 **Effectiveness of structured vs. unstructured target queries**

	2000 topics (n=33)				2001 topics (n=47)			
Translation route	Unstr	Struct	Diff	Diff %	Unstr	Struct	Diff	Diff %
Swe-Eng	19.9	23.6	+3.7	+18.6 %	31.6	37.3	+5.7	+18.0 %
Swe-Fin-Eng	15.7	23.2	+7.5	+47.8 %	19.6	40.2	+20.6	+105.1 %
Fin-Eng	21.6	26.0	+4.4	+20.4 %	33.5	40.5	+7.0	+20.9 %
Fin-Swe-Eng	12.0	20.4	+8.4	+70.0 %	22.1	35.9	+13.8	+62.4 %
Ger-Eng	20.8	23.5	+2.7	+13.0 %	34.3	38.9	+4.6	+13.4 %
Ger-Fin-Eng	13.4	17.5	+4.1	+30.6 %	16.0	32.3	+16.3	+101.9 %
Ger-Swe-Eng	14.5	19.5	+5.0	+34.5 %	18.7	32.6	+13.9	+74.3 %
Ger triang	19.2	18.5	-0.7	-3.6 %	27.7	31.0	+3.3	+11.9 %

Table 4 **Number of words in the target language queries**

	2000 topics (n=33)			2001 topics (n=47)		
Translation route	Number of words (all topics)	Number of words per topic	Increase: Trans vs. Dir	Number of words (all topics)	Number of words per topic	Increase: Trans vs. Dir
Eng monol	272	8.2		485	10.3	
Swe-Eng	918	27.8		1438	30.6	
Swe-Fin-Eng	4884	148.0	432 %	6491	138.1	351 %
Fin-Eng	1080	32.7		2051	43.6	
Fin-Swe-Eng	2604	78.9	141 %	4261	90.7	108 %
Ger-Eng	1219	36.9		1863	39.6	
Ger-Fin-Eng	5814	176.2	377 %	9967	212.1	435 %
Ger-Swe-Eng	3234	98.0	165 %	3625	77.1	95 %

Table 5 Rate of translated source query words in (structured) target queries (%)

	2000 topics (n=33)		2001 topics (n=47)	
Translation route	Average precision	Rate of translation	Average precision	Rate of translation
Swe-Eng	23.6	87.0	37.3	78.0
Swe-Fin-Eng	23.2	88.4	40.2	82.0
Fin-Eng	26.0	91.9	40.5	81.7
Fin-Swe-Eng	20.4	89.9	35.9	81.4
Ger-Eng	23.5	89.7	38.9	82.6
Ger-Fin-Eng	17.5	86.4	32.3	80.8
Ger-Swe-Eng	19.5	86.6	32.6	79.0

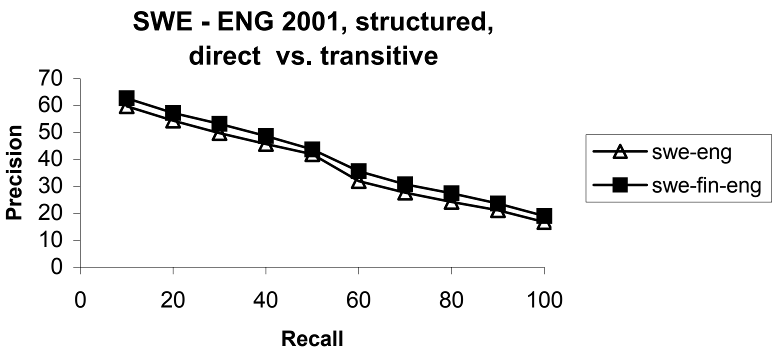

Fig. 1. Precision-recall curves for direct and transitive translation, using CLEF 2001 Swedish topics and structured target queries

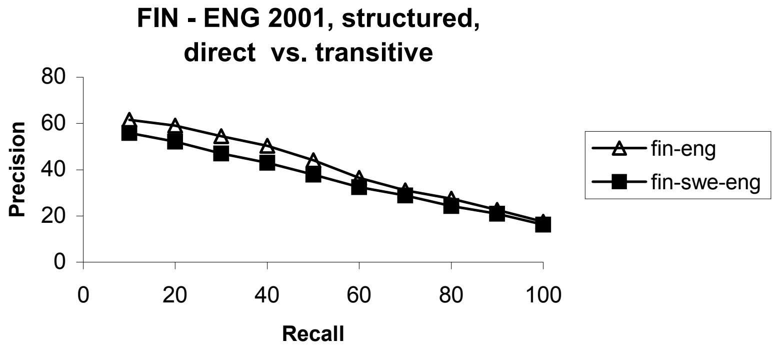

Fig. 2. Precision-recall curves for direct and transitive translation, using CLEF 2001 Finnish topics and structured target queries

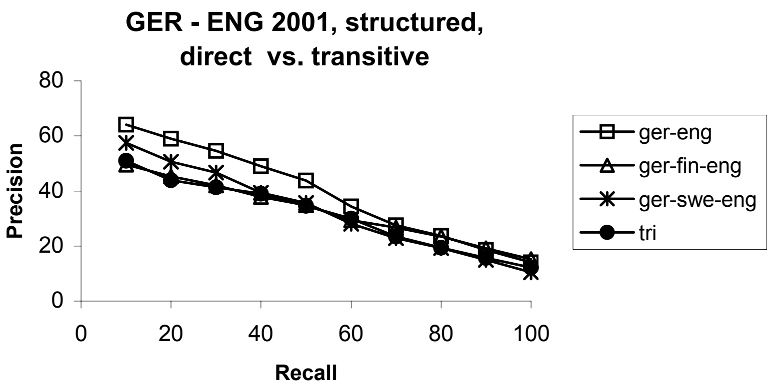

Fig. 3. Precision-recall curves for direct, transitive and triangulated translation, using CLEF 2001 German topics and structured target queries

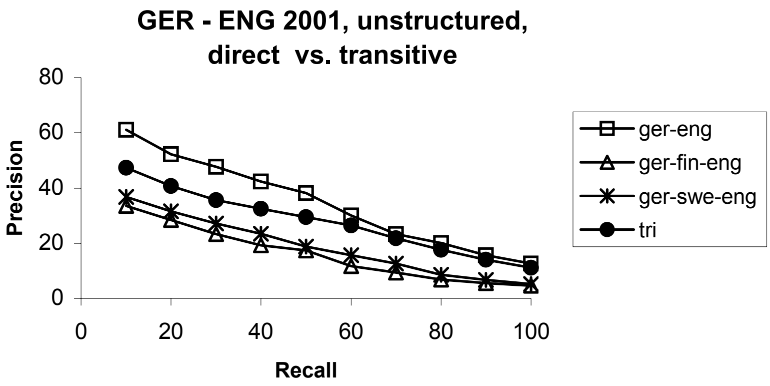

Fig. 4. Precision-recall curves for direct, transitive and triangulated translation, using CLEF 2001 German topics and unstructured target queries

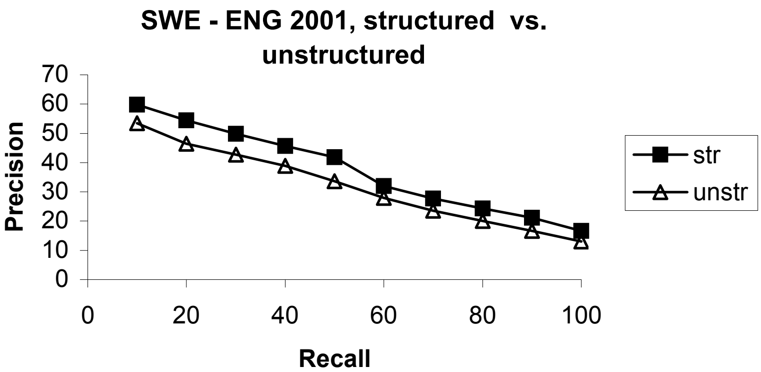

Fig. 5. Precision-recall curves for direct translation, using CLEF 2001 Swedish topics, and structured and unstructured target queries

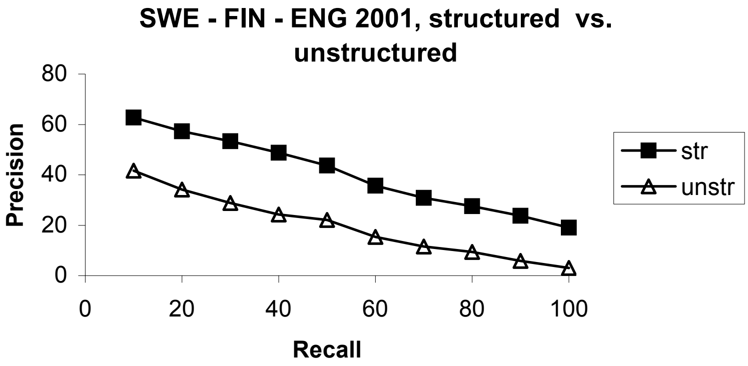

Fig. 6. Precision-recall curves for transitive translation, using CLEF 2001 Swedish topics, and structured and unstructured target queries