Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval

Lisa Ballesteros and W. Bruce Croft
balleste@cs.umass.edu, croft@cs.umass.edu
Center for Intelligent Information Retrieval
Computer Science Department
University of Massachusetts
Amherst, MA 01003-4610 USA

Abstract

Dictionary methods for cross-language information retrieval give performance below that for mono-lingual retrieval. Failure to translate multi-term phrases has been shown to be one of the factors responsible for the errors associated with dictionary methods. First, we study the importance of phrasal translation for this approach. Second, we explore the role of phrases in query expansion via local context analysis and local feedback and show how they can be used to significantly reduce the error associated with automatic dictionary translation.

1 Introduction

The development of IR systems for languages other than English has focused on building mono-lingual systems. Increased availability of on-line text in languages other than English and increased multi-national collaboration have motivated research in cross-language information retrieval (CLIR) - the development of systems to perform retrieval across languages.

There have been three main approaches to CLIR: translation via machine translation techniques [Rad94]; parallel or comparable corpora-based methods [DD95a, LL90, SB96], and dictionary-based methods [Sal72, Pev72, HG96, BC96]. Each of these approaches has shown promise, but also has disadvantages associated with it. Results suggest that improvements gained via machine translation techniques may not outweigh the cost of linguistic analysis. One disadvantage of methods based on the use of parallel and aligned corpora is lack of resources: parallel corpora are not always readily available and those that are available tend to be relatively small or to cover only a small number of subjects. Performance is also dependent on how well the corpora are aligned. Our work takes the third approach and applies dictionary-based methods.

Automatic machine readable dictionary (MRD) query translation leads to a drop in effectiveness of 40-60% below that of mono-lingual retrieval [HG96, BC96]. This is due primarily to three factors. First, specialized vocabulary not contained in the dictionary will not be translated. Second, dictionary translations are inherently ambiguous and add extraneous terms to the query. Third, failure to translate multi-term concepts as phrases reduces effectiveness.

We are developing strategies for reducing the errors associated with dictionary-based methods and focus on strategies which have a low processing cost and do not require scarce resources. This paper explores the identification of phrases in queries and the effectiveness of simple phrasal translation. In addition, we investigate the role of phrases in query expansion by comparing two approaches, local feedback [AF77] and Local Context Analysis [XC96], to expanding queries at various stages of the "translation" process.

2 Previous Work

Effective systems for mono-lingual information retrieval have been available for several years. Typically, research in the area of multi-lingual information retrieval has focused on incorporating new languages into existing systems to allow them to run in several mono-language retrieval modes. Recently, greater interest in retrieval across languages has motivated more work to study the factors involved in building a CLIR system.

Salton [Sal72] showed early on that with carefully constructed thesauri, cross-language retrieval was nearly as effective as mono-lingual retrieval. This study was good, however the test collection was very small by current standards and it is unrealistic to manually index larger databases.

Landauer and Littman [LL90] have also proposed a method for cross-language retrieval. Latent Semantic Indexing (LSI) [FDD$^+$88] was used to create a multidimensional indexing space for a parallel corpus of English documents and their French translations. Their method has been successful at the task of retrieving a query's translation, in response to that query. However the collection used was small, containing 2482 paragraph-length documents from Canadian Parliamentary proceedings and no results of its effectiveness on the traditional retrieval task have been reported. The method also relies on the use of parallel corpora which are not always readily available.

Another method that relies on parallel and aligned corpora has been suggested by Dunning and Davis [DD93]. Their method is based on the vector space model and involves the linear transformation of the representation of a query in one language to its corresponding representation in another language. The transformation is done by reduction of the document space to generate a translation matrix. They have had some success in efficiently estimating the translation matrix and results of tests to estimate its quality are promising. Further tests of the effectiveness of the method have been limited by its computational complexity.

Davis and Dunning[DD95a, DD95b] have also developed several other approaches to query translation, which they tested on the TREC ISM Spanish queries and collection. Two of these rely on the use of a Spanish-English parallel corpus and one uses evolutionary programming for query optimization. In the first of the parallel corpus approaches English queries were translated by replacing the original query terms with the 100 most frequent terms in the top 100 retrieved documents from the Spanish side of the parallel corpus. The second approach replaces the original query terms with terms found to be statistically significant. The evolutionary programming method starts with a query generated by the high frequency approach. It then modifies queries by randomly adding or deleting query terms. Optimization is done by evaluating query fitness after each round of mutations, and selecting the "most fit" to continue to the next generation. The evolutionary programming approach was the most effective, but results were disappointing, with each of the methods performing well below the word-by-word translation baseline.

More recently, Davis [Dav96] uses part-of-speech tagging to select the best Spanish translations for English query terms. A parallel corpus is then used to further disambiguate the translated queries by choosing the Spanish terms that retrieve documents most like those retrieved for the English query. This approach is more effective than previous ones, achieving up to 73.5% of monolingual performance.

Sheridan and Ballerini [SB96] performed "translations" using co-occurrence thesauri generated from a comparable corpus. Cross-language experiments suggest that using co-occurrence thesauri generated with this type of data yields a translation effect. However, performance measured by average precision is still considerably below that of mono-lingual retrieval. Disadvantages to the approach are that it relies on time-sensitive documents, queries are constrained to referencing specific events, and a strict definition of the notion of relevance. This is a side effect of the way in which the test data was constructed and in theory should not be a problem inherent to the approach, but this has yet to be shown experimentally.

Previous work has been done to recognize and translate phrases in text, for example [SWH96, Kup93]. These approaches identify source language phrases and rely upon the use of parallel corpora to identify the context in which target language translations should be found. Although these approaches work well, we use simple dictionary translation because we are interested in exploring what can be done when scarce resources such as parallel corpora are unavailable.

3 Dictionary Translation and Query Expansion

Previous studies [HG96, BC96] have shown that automatic word-by-word (WBW) translation of queries via MRD results in a 40-60% loss in effectiveness below that of monolingual retrieval. One of the factors causing this drop in effectiveness is ambiguity caused by the transfer of extraneous terms. What may be more important however, is the failure to translate multi-term concepts as phrases. We have shown [BC96] that, despite the loss of phrases, query expansion via "local feedback" could reduce the errors such an approach normally makes. Relevance feedback [SB90] is a method by which a query is modified by the addition of terms found in documents known to be relevant to the query. Local feedback [AF77] differs from classic relevance feedback in that it *assumes* the top retrieved documents are relevant.

Local feedback modification before or after automatic query translation via MRD significantly improves performance. Pre-translation feedback expansion creates a stronger base for translation and improves precision. Local feedback after MRD translation introduces terms which de-emphasize irrelevant translations to reduce ambiguity and improve recall. Combining pre- and post-translation feedback is most effective, and reduces translation error by up to 36%. Improvement appears to be due to the removal of error caused by the addition of extraneous terms via the translation process.

In this paper, we look at another method of query expansion known as local context analysis (LCA)[XC96] to find words and phrases related to each query. LCA is a query expansion method that uses both global and local document analysis, and has been shown to be more effective than simple local feedback. The reason for this study is two-fold. First, we are interested in exploring the effectiveness of simple phrasal translation. Second, we want to compare these two methods of query expansion, local feedback and local context analysis (LCA), for addressing the error associated with dictionary translation of words and phrases.

4 Experiments

The experiments in this study were limited to two languages: Spanish and English. The Spanish queries consisted of

TREC topics SP26-45. Evaluation was performed on the 208 MB TREC ISM (El Norte) Spanish collection with provided relevance judgments. Training data for the pre-translation LCA experiments consisted of the documents in the 301 MB San Jose Mercury News (SJMN) database from the TREC collection.

Each Spanish query has relevance judgments. In order to use these judgments, we need to test the effectiveness of MRD translations to Spanish. To do this, we created base queries by manually translating the Spanish queries to English (herein referred to as BASE). The automatic translations of the base queries could then be evaluated using the relevance judgments of the original queries. The manual translation of the Spanish queries was performed by a bilingual graduate student whose native language is English.

Phrases were identified in BASE queries in the following way. First, queries were tagged with th BBN part-of-speech tagger. Sequences of nouns and adjective-noun pairs were taken to be phrases. Automatic translations were performed by translating individual terms word-by-word and phrases as multi-term concepts. The word-by-word translations were done by replacing query terms in the source language with the dictionary definition of those terms in the target language. Words that were not found in the dictionary were added to the new query without translation. The Collins English-Spanish bilingual MRD was used for the translations. For a more detailed description of this process, see [BC96]. Phrasal translations were performed using information on phrases and word usage contained in the Collins MRD. This allowed the replacement of a source phrase with its multi-term representation in the target language. When a phrase could not be defined using this information, it was translated word-by-word as described above. Stop words and stop phrases such as “A relevant document will” were also removed.

Non-interpolated average precision on the top 1000 retrieved documents is used as the basis of evaluation for all experiments. CLIR would be useful for people who can only afford to have a small number of documents translated or who do not speak a foreign language well enough to formulate a good query, but who can read it well enough to judge a document’s relevance. However it is unrealistic to expect the user to read many retrieved foreign documents to find a relevant one, so in some cases we also report precision at low recall levels. The following sections describe our experiments. In section 5 we analyze and discuss the importance of phrasal translation. Next we present a comparison of LCA and local feedback expansion. Sections 6.1, 6.2, and 6.3 describe how pre-translation, post-translation, and combined pre- and post-translation expansion methods help to improve performance (see Fig. 1 for a flow chart of query processing for the experiments). Finally, section 7 presents conclusions and future work.

All work in this study was performed using the INQUERY information retrieval system. INQUERY is based on the Bayesian inference net model and is described elsewhere[TC91b, TC91a, CCB95].

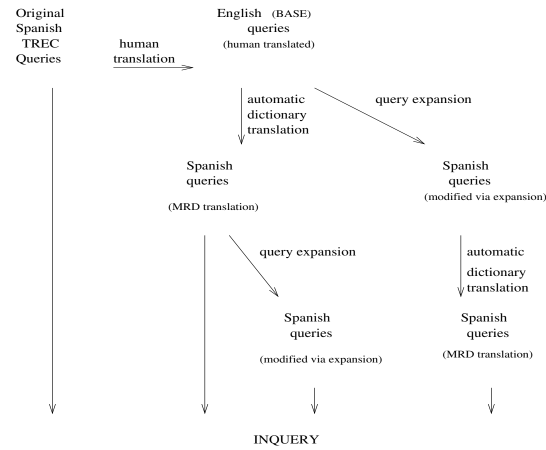

Figure 1: Flow chart of query processing.

5 Phrasal Translation

Failure to translate multi-term concepts as phrases greatly reduces the effectiveness of dictionary translation. In experiments where query phrases were manually translated [BC96], performance improved by up to 25% over automatic word-by-word (WBW) query translation. Our hypothesis is that automatically identifying phrases and defining them as such would improve effectiveness.

To test this hypothesis, we compare performance of automatically translated queries both with and without phrasal identification and translation. Phrasal translations are based on a database of phrasal and word usage information extracted from the Collins Spanish-English MRD. During phrase translation, the database is searched for English phrases. A hit returns the Spanish translation of the English phrase. If more than one translation is found, each of them is added to the query. Table 1 gives some examples of phrasal translations.

Phrase	Translation
united nations	Naciones Unidas Organización de las Naciones Unidas
trade agreement	convenio comercial
south africa	Unión Sudafricana Africa del Sur
member country	los países miembros los países afiliados los países participantes los países pertenecientes

Table 1: Phrasal translations.

The results in Table 2 suggest that in this case, phrasal translation does not improve effectiveness. It gives average precision values for a baseline of automatic WBW translation vs automatic WBW with phrasal translation. A closer look at individual queries reveals that phrasal translation is not ineffective, but that results are sensitive to poor translations. Average precision drops 40% below a baseline of automatic WBW translation for TREC [Har95] query SP30 when phrasal translations are included. However, the problem for this query is that "sports program" is translated as "emisiòn deportiva" meaning televised sports program. When the poor phrasal translation is replaced with a WBW translation, results improve considerably (+150% over the baseline). Table 3 shows 5 representations of SP30: Original, BASE, automatic WBW translation, automatic phrasal + WBW translation, and automatic WBW translation + "good" phrasal translations. Parentheses enclose recognized phrases and brackets enclose phrasal translations. Results for the last three queries are given in Table 4.

	WBW	Phrasal
Avg	0.0823	0.0826

Table 2: Average precision of WBW vs phrasal translation.

programas y intercambios deportivos entre México y los Estados Unidos
(Sports programs) and (exchange programs) between Mexico and the (United States)
deporte caza deporte juego diversión víctima juguete programs canje intercambio programs Méjico México States
[emisión deportiva] cambio canje intercambio programs [Estados Unidos][el coloso del norte] [Estados Unidos de América] Méjico México
deporte caza deporte juego diversión víctima juguete programs cambio canje intercambio programs [Estados Unidos] [el coloso del norte] [Estados Unidos de América] Méjico México

Table 3: Five query representations for SP30: original, BASE, MRD translation of BASE, MRD WBW + phrasal translation of BASE, MRD WBW + "good" phrasal translations of BASE

	WBW	Phrasal	Good Phrasal
Avg	0.0244	0.0148	0.0610
% Change:		-39.3	150.3

Table 4: Average precision for WBW vs two different phrasal translations for query SP30.

These results suggest that well-translated phrases can greatly improve effectiveness, but that poorly translated phrases may negate the improvements. Translation accuracy may be more important for phrases than for terms.

6 Local Context Analysis vs Local Feedback

In experiments similar to those from our earlier work, we translated queries automatically via MRD. Query expansion via LCA was performed either prior to or after translation in the following way. A query set is evaluated and the top ranked passages for each query are retrieved. Queries are then expanded by the addition of the top ranked concepts from the top passages. Recall that concepts may be single or multi-term.

6.1 Pre-translation

In this first set of experiments, we wanted to compare the effectiveness of query expansion prior to automatic translation via LCA to previous results using local feedback. Recall that the queries were manually translated into English, so the Spanish ISM database cannot be used for pre-translation expansion. We chose to use the SJMN database, described above, as a training corpus from which to choose English expansion concepts. Multi-term concepts are translated as phrases. In the event that no phrasal translation is found, phrases are translated WBW. Table 5 shows 4 representations of TREC query SP29. First is the original query, second is the manual translation (BASE) including automatically identified phrases, third is the LCA expanded query, and fourth is the automatic translation of the third. Parentheses surround LCA expansion phrases and phrases automatically identified in the BASE query. Brackets surround the translation of each term or phrase.

las relaciones económicas y comerciales entre México y Canadá
the economic and (commercial relations) between mexico and canada
economic (commercial relations) mexico canada mexico (trade agreement) (trade zone) cuba salinas
[económico equitativo][comercio negocio tráfico industria] [narración relato relación][Méjico México] Canadá [Méjico México] [convenio comercial] [comercio negocio tráfico industria] zona cuba salinas

Table 5: Four query representations: original, BASE (with identified phrases), LCA expanded BASE, WBW + phrasal translation of LCA expanded BASE.

First, we look at the effects of LCA expansion without phrasal recognition in the base query and compare a straight WBW translation of all concepts with a combination of phrasal and WBW translation. We then combine phrasal recognition in BASE with LCA expansion followed by both WBW and phrasal translation. Translations of multi-term LCA concepts were wrapped in the INQUERY #passage25 and #phrase operators. For example, #passage25(#phrase(North American Free Trade Agreement)). Terms within a #phrase operator are evaluated to see whether they co-occur frequently in the collection. If they do, co-occurrences within 3 terms of each other are considered when calculating belief. If not, the terms are treated as having equal influence on the final result in order to allow for the possibility individual occurrences are evidence of relevance. The #passage25 operator looks for the elements to occur within a window of 25. This operator ensures that terms which do not co-occur frequently be found a limited distance apart.

The best results for automatic translations to Spanish are shown in Table 6. Descriptions of query processing for rows 2-7 follow. Row 2 (MRD) is the automatic word-by-word translation of BASE (original TREC queries manually translated). For row 3, phrases were identified in the BASE queries and then WBW translation was augmented by phrasal translation (MRD + Phr). Row 4 shows results for pre-translation LCA expanded BASE queries translated word-by-word (MRD + LCA-WBW). Row 5 represents pre-translation LCA expanded BASE queries translated word-by-word with phrasal translation where possible (MRD + LCA-Phr). In Row 6, after phrase identification in BASE queries, they were expanded via LCA prior to translation. The expanded queries were then translated word-by-word with phrasal translation where possible. Finally, row 7 shows results for pre-translation local feedback expanded BASE queries after word-by-word translation (LF).

Method	Avg	%Change
MRD	0.0823	
MRD+Phr	0.0826	0.3
MRD+LCA-WBW	0.0969	17.7
MRD+LCA-phr	0.1009	22.7
MRD+Phr+LCA-phr	0.1053	27.9
LF	0.1099	33.5

Table 6: Average precision for pre-translation expansion results.

The best results were gained after adding the top 30 concepts from the top 20 documents. They show that LCA expansion is effective, but WBW translation of LCA concepts yields only a 17% increase. This is probably due to the ambiguity introduced through the loss of multi-term concepts. Further improvements are given when phrases are identified in the BASE queries and when multi-term concepts are translated as phrases. If multi-term concepts are translated as phrases, effectiveness goes up by 5%. The addition of phrasal recognition in the BASE queries boosts effectiveness by an additional 5%. These results show that the use of phrasal translation can indeed improve effectiveness.

Pre-translation LCA expansion results are still not as good as those for pre-translation local feedback. This is surprising since comparisons of local feedback and LCA in the mono-lingual environment [XC96] have shown LCA to be more robust for query expansion.

We hypothesized that although most phrases added by LCA appear to be good phrases, they may lose their effectiveness when taken as individual terms. This happens when a phrasal translation fails and we are forced to translate the phrase word-by-word. In addition, poor phrases will also tend to be ineffective when translated word-by-word. To test this, we performed LCA expansion returning only the best single-term concepts. Results in section 5 show that query effectiveness is highly sensitive to the accuracy of phrasal translation. Expansion by individual terms eliminates the negative effects of poor phrasal translations.

We found that in some cases, our hypothesis is supported. However, it is not consistent. Table 7 gives a few examples of LCA expansion with single- and multi-term concepts compared to expansion with only single-term concepts. In this table, each of the expansions was done using the top 20 passages and the top 5 or 30 concepts. Automatic translation is given as a baseline. We believe the inconsistency is related to the types of multi-term concepts that are included in the expansion and on translation accuracy.

Method	Avg prec	%Change
MRD	0.0823	
LCA5-Phrasal	0.0819	-0.5
LCA5-Single	0.1051	27.7
LCA30-Phrasal	0.1053	27.9
LCA30-Single	0.1010	22.7

Table 7: Average precision for multi-term and single-term concept expansion.

Table 8 shows the best pre-translation results for expansion via local feedback and for single-term expansion via LCA. This shows that LCA can be more effective than local feedback when used prior to translation, however the choice of expansion concepts is critical.

	MRD	LF	LCA10-Single
Avg prec	0.0823	0.1099	0.1139
% Change:		33.5	38.5
Precision:			
5 docs:	0.2000	0.2500	0.3100
10 docs:	0.2100	0.2300	0.2750
15 docs:	0.1867	0.2400	0.2600
20 docs:	0.1975	0.2375	0.2350

Table 8: Best pre-translation local feedback and single-term LCA expansion results.

6.2 Post-translation Expansion

In experiments where post-translation LCA expansion was performed, multi-term concepts were wrapped in INQUERY #PHRASE operators. The top ranked concept was added to a query with a weight of 1.0. Each additional concept was down-weighted by 1/100 with respect to the weight given its predecessor. This weighting scheme was shown to be effective in LCA experiments for the TREC5 evaluations [Har96]. Table 10 shows the best results for post-translation expansion via local feedback and LCA. In this table, local feedback expansion was done by addition of the top 20 terms from the top 50 documents. LCA expansion was done by addition of the top 100 concepts from the top 20 passages. Table 9 shows 2 representations of one of these queries. First is the BASE and second the automatic translation of BASE. The last row gives the top 20 expansion concepts that were added to this query, with multi-term concepts in parentheses. Note that all terms are stemmed.

economic commercial relations mexico european countries
comerc narr relat rel econom equit rentabl pai patri camp region tierr mej mex europ
(est un) canada pai europ franci (diversific comerc) mex polit pais alemani rentabl oportun product apoy australi (merc europ) agricultor bancarrot region (comun econom europ)

Table 9: Two query representations for TREC query SP26: BASE and MRD translation of BASE. Row 3 gives the top 20 post-translation LCA expansion concepts for this query.

	MRD	LF	LCA20
Avg prec	0.0823	0.0916	0.1022
% Change:		11.3	24.1
Precision:			
5docs:	0.2000	0.1800	0.2200
10 docs:	0.2100	0.1850	0.2100
15 docs:	0.1867	0.1800	0.2167
20 docs:	0.1975	0.1575	0.2050

Table 10: Best post-translation local feedback and LCA expansion results.

The best post-translation LCA expansion is 11.6% more effective than the best post-translation local feedback expansion. Eleven of 20 queries do better with LCA as compared to 7 which do better with LF. A paired sign test shows this difference to be significant at p = .01. This supports earlier work by Xu which showed LCA to be a more effective query expansion technique than local feedback.

6.3 Combined Pre- and Post-translation Expansion

The combination experiments start with the pre-translation LCA expansion of the BASE queries. After the expanded queries are translated automatically, they are expanded again via LCA multi-term expansion. The base query set for the post-translation expansion phase in these experiments, is the best pre-translation, single-term concept LCA expanded query set, as described in Section 6.1. Table 11 shows 4 representations of one of these queries. First is the original query, second is the manual translation (BASE) including automatically identified phrases, third is the pre-translation LCA single-term expanded query, and fourth is the automatic translation of the third. The last row gives the top 20 expansion concepts that were added to this query, with multi-term concepts in parentheses. Note that all terms are stemmed. Parentheses surround LCA expansion phrases and phrases automatically identified in the BASE query. Brackets surround the translation of each term or phrase.

las relaciones económicas y comerciales entre México y Canadá
the economic and (commercial relations) between mexico and canada
economic (commercial relations) mexico canada mexico free-trade canada trade mexican salinas cuba pact economies barriers
[económico equitativo][comercio negocio tráfico industria] [narración relato relación] [Méjico México] Canadá [Méjico México][convenio comercial] [comercio negocio tráfico industria] zona cuba salinas
canada (libr comerci) trat ottaw dosm (acuerd paralel) norteamer (est un) (tres pais) import eu (vit econom) comerci (centr econom) (barrer comerc) (increment subit) superpot rel acuerd negoci

Table 11: Four query representations: original, BASE (with identified phrases), LCA expanded BASE, WBW + phrasal translation of LCA expanded BASE.

The combined approach is more effective than either pre- or post-translation LCA expansion alone. This was also shown to be the case for local feedback expansion. Table 12 gives results for automatic translation, the best combined pre- and post-translation local feedback expansion, and the best combined LCA expansion. In this experiment, queries were expanded by the top 50 terms from the top 20 passages in the post-translation LCA phase. Fourteen and eleven queries show improvement over MRD translation alone for LCA and LF, respectively. The LCA approach shows a 9% greater improvement than the local feedback approach, but this difference is not statistically significant. When the two methods are compared 9 queries do better with LCA expansion as compared to 10 that do better with LF expansion. However, it is interesting to compare the effects of LCA and local feedback expansion on precision. The LCA expansion has higher precision at low recall levels. This is important in a CLIR environment. The user may not be proficient at reading a foreign language, so could not be expected to look through more than the top retrieved documents.

7 Conclusions and Future Work

Automatic dictionary translations are attractive because they are cost effective and easy to perform, resources are read-

	MRD	LF	LCA20-50
Avg prec	0.0823	0.1242	0.1358
% Change:		51.0	65.0
Precision:			
5 docs:	0.2000	0.2600	0.3700
10 docs:	0.2100	0.2200	0.2850
15 docs:	0.1867	0.2000	0.2767
20 docs:	0.1975	0.2125	0.2600

Table 12: Best combined pre- and post-translation local feedback and LCA expansion results.

Method	Precision	% Monolingual
Monolingual	0.1998	-
MRD	0.0823	41.2
Pre-LF	0.1099	55.0
Pre-LCA	0.1139	57.0
Post-LF	0.0916	45.8
Post-LCA	0.1022	51.1
Comb-LF	0.1242	62.2
Comb-LCA	0.1358	68.0

Table 13: Average precision for all methods.

ily available, and performance is similar to that of other CLIR methods. Ambiguity from failure to translate phrases is largely responsible for the large drops in effectiveness below monolingual performance.

Phrasal translation can greatly improve effectiveness, however improvements are sensitive to the quality of the translations. The effect of one poor translation can counteract any improvement gained by the correct translation of several phrases and may cause additional drops in effectiveness. Certain types of multi-term concepts, such as proper noun phrases, are easily translated via MRD. However, dictionaries do not provide enough context for accurate phrasal translation in other cases.

Query expansion via local feedback and LCA can be used to significantly reduce the error associated with dictionary translation. LCA expansion gives higher precision at low recall levels, which is important in a CLIR environment. Table 13 shows the performance of each method as measured by average precision and percentage of monolingual performance. LCA, which typically expands queries with multi-term phrases, is more sensitive to translation effects when pre-translation expansion is performed. This is because phrases that must be translated WBW, are not as effective when separated into individual terms. Pre-translation LCA expansion with single-term concepts can reduce this problem. Pre-translation LCA expansion with single terms is also more effective than pre-translation local feedback and improves both precision and recall. Post-translation LCA is more effective than post-translation local feedback and tends to improve precision. Combining pre- and post-translation expansion is most effective and improves precision and recall. It can reduce translation error by 45% over automatic translation bringing CLIR performance up from 42% to 68% of monolingual performance. This is still well below a monolingual baseline, but improved phrasal translations should help to narrow the gap.

In this study, we have shown that query expansion techniques can significantly reduce the error associated with dictionary translation. Dictionaries do not provide enough context for accurate translations on a wide range of phrase types, so an alternative must be found. A better phrase translator should not alter our conclusion that query expansion can ameliorate the errors that occur in word-by-word or phrase translation, however further improvements are dependent upon accurate phrasal translation. INFINDER [JC94] is a tool for generating a corpus-based association thesaurus. We are currently exploring its potential for generating a cross-language association thesaurus that would provide enough context for more accurate phrasal translations.

Acknowledgments

We'd like to thank Mark Davis and others at the CRL lab at New Mexico State University for the use of tools such as Norm, which facilitated access to the Collins bilingual MRDs.

This material is based on work supported by the National Science Foundation, Library of Congress, and Department of Commerce under cooperative agreement number EEC-9209623 and by a NASA GSRP grant, #NGT-70358.

References

[AF77] R. Attar and A. S. Fraenkel. Local feedback in full-text retrieval systems. *Journal of the Association for Computing Machinery*, 24:397–417, 1977.

[BC96] Lisa Ballesteros and W. Bruce Croft. Dictionary-based methods for cross-lingual information retrieval. In *Proceedings of the 7th International DEXA Conference on Database and Expert Systems Applications*, pages 791–801, 1996.

[CCB95] J.P. Callan, W.B. Croft, and J. Broglio. Trec and tipster experiments with inquery. *Information Processing and Management*, 31(3):327–343, 1995.

[Dav96] Mark Davis. New experiments in cross-language text retrieval at nmsu's computing research lab. In *To Appear in: Proceedings of the Fifth Retrieval Conference (TREC-5) Gaithersburg, MD: National Institute of Standards and Technology*, 1996.

[DD93] Ted Dunning and Mark Davis. Multi-lingual information retrieval. Technical report MCCS-93-252, Computing Research Laboratory, New Mexico State University, 1993.

[DD95a] Mark Davis and Ted Dunning. Query translation using evolutionary programming for multi-lingual information retrieval. In *Proceedings of the Fourth Annual Conference on Evolutionary Programming*, 1995.

[DD95b] Mark Davis and Ted Dunning. A trec evaluation of query translation methods for multi-lingual text retrieval. In *In Proceedings of the Fourth Retrieval Conference (TREC-4) Gaithersburg, MD: National Institute of Standards and Technology, Special Publication 500-236*, 1995.

[FDD+88] G.W. Furnas, S. Deerwester, S.T. Dumais, T.K. Landauer anmd R.A. Harshman, L.A. Streeter, and K.E. Lochbaum. Information retrieval using a singular value decomposition model of latent semantic structure. In *Proceedings of the 11th International Conference on Research and Development in Information Retrieval*, pages 465–480, 1988.

[Har95] Donna Harman, editor. *Proceedings of the 4th Text Retrieval Conference (TREC-4)*. 1995.

[Har96] Donna Harman, editor. *Proceedings of the 5th Text Retrieval Conference (TREC-5)*. 1996.

[HG96] David A. Hull and Gregory Grefenstette. Querying across languages: A dictionary-based approach to multilingual information retrieval. In *Proceedings of the 19th International Conference on Research and Development in Information Retrieval*, pages 49–57, 1996.

[JC94] Y. Jing and W.B. Croft. An association thesaurus for information retrieval. In *RIAO 94 Conference Proceedings*, pages 146–160, 1994.

[Kup93] Julian M. Kupiec. An algorithm for finding noun phrase correspondances in bilingual corpora. In *Proceedings, 31st Annual Meeting of the ACL*, pages 17–22, 1993.

[LL90] Thomas K. Landauer and Michael L. Littman. Fully automatic cross-language document retrieval. In *Proceedings of the Sixth Conference on Electronic Text Research*, pages 31–38, 1990.

[Pev72] B. Pevzner. Comparative evaluation of the operation of the russian and english variants of the pusto-nepusto-2 system. *Automatic Documentation and Mathematical Linguistics*, 6:71–74, 1972.

[Rad94] K. Radwan. *Vers l'Accés Multilingue en Langage Naturel aux Bases de Données Textuelles. PhD thesis, Université de Paris-Sud, Centre d'Orsay.* 1994.

[Sal72] Gerard Salton. Experiments in multi-lingual information retrieval. Technical report TR 72-154, Computer Science Department, Cornell University, 1972.

[SB90] Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41:288–297, 1990.

[SB96] Paraic Sheridan and Jean Paul Ballerini. Experiments in multilingual information retrieval using the spider system. In *Proceedings of the 19th International Conference on Research and Development in Information Retrieval*, pages 58–65, 1996.

[SWH96] Frank Smadja, Kathleen R. WmKeown, and Vasileios Hatzivassiloglou. Translating collocations fo bilingual lexicons: A statistical approach. *Computational Linguistics*, 22(1):1–38, 1996.

[TC91a] Howard R. Turtle and W. Bruce Croft. Efficient probabilistic inference for text retrieval. In *RIAO 3 Conference Proceedings*, pages 664–661, 1991.

[TC91b] Howard R. Turtle and W. Bruce Croft. Inference networks for document retrieval. In *Proceedings of the 19th International Conference on Research and Development in Information Retrieval*, pages 1–24, 1991.

[XC96] Jinxi Xu and W. Bruce Croft. Querying expansion using local and global document analysis. In *Proceedings of the 19th International Conference on Research and Development in Information Retrieval*, pages 4–11, 1996.