TREC-7 Evaluation of Conceptual Interlingua Document Retrieval (CINDOR) in English and French

Anne Diekema, Farhad Oroumchian
Páraic Sheridan, Elizabeth D. Liddy

TextWise LLC
2-212 Center for Science & Technology
Syracuse, NY 13244

Abstract:

TextWise LLC. participated in the TREC-7 Cross-Language Retrieval track using the CINDOR system, which utilizes a *"conceptual interlingua"* representation of documents and queries. The current CINDOR research system uses a conceptual interlingua constructed around the Princeton WordNet, which we are mapping into French and Spanish. The use of an interlingual representation of documents and queries allows us to perform retrieval on any combination of supported languages, rather than having to rely on pairwise translations, while the use of a resource like WordNet allows us to match equivalent terms (including synonyms) across languages.

Although the analysis of our TREC-7 results is clouded somewhat by the kinds of system errors which inevitably occur in a first-time evaluation over large TREC corpora, our evaluation of the conceptual interlingua approach suggests that it provides highly effective cross-language retrieval performance. In particular, we notice that the CINDOR system achieves cross-language retrieval results equivalent in many cases to corresponding monolingual queries, without the loss in retrieval precision observed in many other approaches to cross-language retrieval.

Future work on the CINDOR system, which was evaluated here in its research prototype form, will focus on improving further the coverage of our conceptual interlingua resources and the efficiency of our document processing modules. We are also investigating the construction of an interlingual resource of proper nouns, using technology from other TextWise products, since proper nouns constitute the largest category of 'out-of-vocabulary' terms with respect to our current conceptual interlingua knowledge base. We will also continue to adapt the CINDOR system to handle more languages.

1. Introduction.

The TREC-7 Cross-Language Retrieval track was set as the arena for the very first evaluation of the CINDOR system, a prototype retrieval system developed by TextWise LLC. with the explicit aim of facilitating language-independent document retrieval based on natural language concepts. The CINDOR system currently supports retrieval over English, French and Spanish with a fourth, yet to be determined, language due for inclusion from the beginning of 1999. In contrast to many approaches to cross-language retrieval, which translate either documents or queries between the languages in question [Oard & Diekema 1998], CINDOR takes the approach of translating *both* documents and queries - into its interlingual representation. This has the advantage of permitting matching and retrieval based on any combination of the languages involved, rather than relying on pairwise translations between language pairs.

The current CINDOR system is built around Princeton WordNet version 1.5 [Miller 1990], with the conceptual interlingua consisting of WordNet synset ids. We use the term "conceptual interlingua" to refer to a knowledge base of language-independent concept representations. Our current conceptual interlingua is a hierarchically organized concept lexicon, following WordNet. Concepts in the hierarchy are considered to be essentially language neutral and are then linked to their relevant terminological instantiations in various languages. A more detailed examination of this conceptual interlingua resource is presented in Section 2.

The use of a conceptual resource like WordNet for document retrieval raises specific issues in the indexing and matching of documents and queries. Although there is the advantage of being able to directly match term synonyms rather than only words with identical surface (or lexical stem) form, this can quickly turn to disadvantage without a mechanism for distinguishing between different word *senses*. While it has been claimed that document retrieval as an application is remarkably robust to word sense differences [Sanderson 1997], we believe that retrieval performance would be unlikely to remain unaffected if multiple word senses were further expanded by synonyms and the whole lot then treated as a 'bag of words' for retrieval. Our initial experiments have confirmed that a naïve approach to WordNet indexing is indeed ineffective and the word sense problem must be tackled in some way. This aspect of our work, together with other retrieval issues we have encountered in our approach, is presented in Section 3.

The CINDOR project is currently at the half way point in a two-year development and evaluation exercise. Work to date has focused on construction of the conceptual interlingua resource, design and implementation of system modules, and first-round benchmark evaluation of the system using the TREC-7 data reported here. Over the next year, we plan to significantly improve system operating performance based on the lessons learned in this evaluation, continue to increase the coverage of our language resources, add a fourth language, and develop a dedicated module for the recognition and processing of proper nouns. Each of these lines of development will end with a benchmark evaluation to measure the improvements gained over the results presented here.

Within the context of the TREC-7 Cross-Language Retrieval track, we have evaluated the CINDOR system against both the English and French data. This constitutes only a subset of the primary track evaluation, which also included data in German and Italian. Although the primary evaluation objective of the track was to evaluate systems against all four of the track languages, the track definition from the beginning included provision for groups which were not interested in working with all four languages. Further, although we did submit merged runs with results from combined English and French document collections, our primary evaluation focus has to date been constrained to single language pairs. Although the CINDOR system will naturally support retrieval from multilingual document collections, given the language-independent nature of its retrieval algorithms, the individual English and French language subcollections of the TREC-7 data were indexed independently and separately in this round of evaluation. We feel that the objective of this, our first evaluation, should be to understand completely the performance of cross-language retrieval based on our conceptual interlingua approach. We can more easily gain insight into this performance by analyzing single language pair retrieval experiments rather than multiple language runs. This objective is reflected in the analysis of our TREC-7 performance which we report in Sections 4 and 5, followed by our conclusions in Section 6.

2. Conceptual Interlingua.

The use of a lexical resource such as WordNet for document retrieval has been extensively investigated, based on a belief that such a resource can provide language knowledge with the potential for increased retrieval performance. This potential has been identified by [Gonzalo *et al* 1998] as:

- the possibility to discriminate word senses in documents and queries. This would prevent matching *'spring'* in its "*metal device"* sense with documents mentioning *'spring'* in the sense of *"springtime"*, and then retrieval accuracy could be improved.

- WordNet provides the possibility of matching semantically related words. For instance, *'spring' , 'fountain', 'outflow', 'outpouring',* in the appropriate senses, can be identified as occurrences of the same concept, *'natural flow of ground water'*. And beyond synonymy, WordNet can be used to measure semantic distance between occurring terms to get more sophisticated ways of comparing documents and queries.

Much of the research on WordNet for information retrieval, has in fact focused on the second source of potential; query expansion and semantic relatedness. Investigation of the usefulness of WordNet for word sense disambiguation has taken place in the context of much research in the information retrieval community which has concluded that retrieval is in fact robust to a large degree of sense ambiguity, for example most recently by [Sanderson:97]. This situation is exacerbated when WordNet sense disambiguation experiments lead to the conclusion that, "missing correct matches because of incorrect sense resolution has a much more deleterious effect on retrieval performance than does making spurious matches" [Voorhees:93]. The obvious conclusion is that if WordNet is to be used for sense disambiguation in a retrieval application, then one must proceed carefully.

The conclusions to be drawn from research on the use of WordNet for query expansion suggest an equal amount of caution when pursuing this potential, especially when working in the TREC environment [Voorhees:94],[Voorhees:93]. While there was some suggestion of a query expansion benefit in small document collections, the experimental results from the TREC environment suggested that query expansion made little difference, presumably because the TREC topic statements already provided such a rich description of the information being sought. WordNet also contributed to a loss of retrieval effectiveness in experiments reported by [Richardson & Smeaton 1995]. In an application involving short documents and queries however, specifically with image captions, Smeaton & Quigley [1996] gained an improvement in retrieval performance through the use of a conceptual distance measure based on WordNet 1.4. One important aspect of this work though, was that both image captions and user queries were manually disambiguated with respect to WordNet senses.

The study reported in [Gonzalo *et al* 1998] set out to determine whether retrieval based on properly disambiguated WordNet synonym sets, '*synsets'*, would perform better than a standard system using word forms. The authors concluded that WordNet synsets provided a 29% increase in retrieval performance against the baseline when queries and documents are manually disambiguated. The results of this study are not, however, presented in a manner which would allow us to speculate as to the likelihood that this improvement would be observed in other test environments.

The use of a lexical/conceptual resource such as WordNet acquires a whole new dimension however, when one moves from single language retrieval applications, in which the above reported experiments were conducted, to an environment where multiple languages are involved. Research has indicated, for example, that cross-language retrieval performance is much more sensitive to ambiguity issues than monolingual retrieval [Ballesteros & Croft 1998]. In particular, the conceptual nature of WordNet synsets allows one, in a multilingual environment, to achieve cross-language matching at the conceptual rather than at the lexical level. Further, when synset labels or ids are used as the indexing vocabulary, the indexing language can be considered an *interlingua*; assuming the equivalent terms of each language are mapped to the corresponding synsets, one can match all combinations of languages, without relying on language-to-language translations in all directions between all possible language pairs.

The CINDOR project represents an attempt to build an information retrieval system specifically aimed at providing a language-independent model of retrieval at the conceptual level. The conceptual interlingua consists of the WordNet hierarchy of synset labels. Each synset label (concept) is linked in turn to a set of words or phrases which instantiate that concept in each of the languages supported. For example, the concept of *"elasticity:* the tendency of a body to return to its original shape after it has been stretched or compressed", which has the label *131186,* is instantiated in English and French as follows:

131186 spring, give, springiness
131186 elasticite, flexibilite, moelleux

This means that any document or query term which is identified as an instantiation of the concept of *'elasticity'* is indexed to the concept label *131186*. Whether the term occurs in an English, French, or Spanish document or query, the label will be the same and retrieval will be enabled.

On the other hand, the issues inherent in the use of WordNet which were identified in experiments in a single language, are if anything exacerbated when multiple languages are involved. Word sense ambiguity remains. For example, the term *'spring'* is indexed with the multiple labels:

spring 313842, 109405, 127451, 131186, 154459, 154460...

Some of the ambiguity can be resolved with fairly straightforward approaches, such as the use of part-of-speech analysis to eliminate verb senses when *'spring'* occurs in a role which is obviously a noun. In addition to the within-language ambiguities however, there is an additional layer of across-language ambiguity introduced, not just with respect to translation, but even with words which are spelled identically in different languages but which relate to different concepts (*'false cognates'*). We have discovered in TREC-7 for example, that *'Concorde'* in English should be related to supersonic jet, but in French belongs primarily to the concept of unity. Our participation in the TREC-7 evaluation has provided us with the first set of test data to use in teasing apart the issues of ambiguity, cross-language matching, concept identification, translation etc. which are inherent in our use of a conceptual interlingua for retrieval.

3. CINDOR Retrieval.

Although the CINDOR system is designed to be particularly effective in performing cross-language retrieval based on matching at the conceptual level, we are conscious of the fact that we must pay careful attention not only to performance based on cross-language versus equivalent monolingual queries, but also on the *absolute* performance achieved. This means that, while a great deal of the focus of our work is on research ensuring concept-level matching across languages, we must be acutely aware of the need to also solve within language issues of concept-based retrieval and word sense ambiguity in order to ensure an acceptable level of baseline performance. In this respect we are fortunate to be able to build upon a substantial body of work in cross-language retrieval which has already identified many of the issues involved [Oard & Diekema 1998].

The two issues of cross-language retrieval which we take to be the most important, apart of course from the obvious necessity for translation coverage, are ambiguity [Ballesteros 1998] and the treatment of phrases [Hull & Grefenstette 1996]. We believe the issue of ambiguity is particularly sensitive in our approach since the many different word senses are further expanded with lists of synonyms for each sense.

We explicitly tested the necessity for word sense disambiguation in the CINDOR system by running a test evaluation in which all document and query terms were indexed by all of the applicable concept labels and a straightforward vector space retrieval algorithm was used in matching. Apart from the likelihood of retrieving documents which match an incorrect sense of query terms, what transpired from this experiment was that the highest ranked documents were those that had the most occurrences of the query term with the most different senses! For example, the TREC-6 query '*automobile air pollution'*, retrieved documents about the *'Air Force'* in the highest rank positions. This is explained by the fact that the term *air* belongs to 36 different synsets in WordNet, while *automobile* and *pollution* each have only 3 different possible senses. When document and query indexing is based on synset labels without any disambiguation effort, the effect is to multiply the weight of each term by its degree of ambiguity so the most ambiguous terms receive the highest weighting. Though this effect could probably have been neutralized directly by a sense-based weight adjustment, we felt that the more appropriate approach lay in the direction of tackling the word sense disambiguation problem.

While a definite goal of our ongoing research, the CINDOR system does not currently include specific processing for word sense disambiguation. Seeing the need to address the sense problem described above, while cognizant of the fact that 'incorrect sense resolution has a much more deleterious effect on retrieval performance than does making spurious matches' [Voorhees 1993], we have therefore aimed for the middle ground. We retain all possible term senses in both document and query indexing, but we ensure that only one sense matches and contributes to the score for a retrieved document. Our queries are therefore evaluated as if they had a form similar to:

[automobile1 *OR* automobile2 *OR* automobile3] +
[air1 *OR* air2 *OR* air3 *OR* air4 *OR* ... *OR* air36] +
[pollution1 *OR* pollution2 *OR* pollution3] .

The second important issue of cross-language retrieval is in fact tackled in a similar way. The CINDOR system does not yet include the capability to recognize and process multi-word terms, though again this is a feature that is foreseen in the near future, particularly with respect to multi-word proper nouns. So although the Princeton WordNet contains many entries composed of multi-word entities, these are currently not directly utilized in CINDOR. Instead, indexing is as above, with each single document and query term being assigned a concept label for each synset to which it belongs, and phrase matches are then *scored* at retrieval time as if queries had a form similar to:

[automobile1 *OR* automobile2 *OR* automobile3] *AND*
[air1 *OR* air2 *OR* air3 *OR* air4 *OR* ... *OR* air36] *AND*
[pollution1 *OR* pollution2 *OR* pollution3] .

Other aspects of CINDOR retrieval are related to our experience in matching documents and queries in multiple languages using the conceptual interlingua, though the solutions again benefit from the experience of past research:

> Document terms which are not found to instantiate any concepts in our current conceptual interlingua, essentially out-of-vocabulary words, are added to a table as a new concept with a corresponding new concept label. Since a large portion of these terms will relate to proper nouns not in WordNet, this table is language-independent. This has the effect that identical cognates across languages will match, even though not explicitly present in our conceptual interlingua.
>
> Although we recognize and tag the language of documents, the conceptual interlingua is not consulted on a language-dependent basis. Concept labels are assigned to all document and query terms from all language terms. For example, whether the term *'concorde'* occurs in a French document or an English query, or anywhere else, it will be assigned the concept labels for the English concept related to jets, the French concept of unity, and any others in which the word is found. This ensures the broadest possible set of matches.

CINDOR applies morphological analysis to identify term stems in each language, and performs part-of-speech tagging to distinguish between the four categories used in WordNet; noun, verb, adjective and adverb. Retrieval is based on the Vector Space Model with retrieved documents ranked according to relevance score.

While endeavoring to ensure that we covered the issues of ambiguity, phrases, out-of-vocabulary words, and other cross-language problems while also processing the English and French document and queries by the TREC-7 deadline, it seems, as our analysis of the evaluation results will show, that we overlooked some of the more basic issues of this exercise, including the normalization of terms for case and the normalization of accented characters. While both of these processes are trivial on their own, we have seen that they can be of immense importance in ensuring that query terms and document terms are assigned the right concept labels for matching.

4. Analysis of TREC-7 Results.

Our participation in the TREC-7 Cross-Language Retrieval track was limited to the English and French document subcollections, which are as follows:

English	Associated Press	1988-1990	242,918 docs (760MB)
French	Schweizerischen Depeschen Agentur (SDA)	1988-1990	141,656 docs (250MB)

CINDOR indexing used WordNet version 1.5 augmented by semi-automatically assigned translations of the English terminology in French. The French data was initially derived from a large proprietary multilingual concept lexicon, which was then manually controlled for quality and further augmented by manual additions. Manual additions were independent of TREC queries but were driven by a frequency-sorted list of terms from the relevant TREC document collections, in order to strive for the greatest coverage of term occurrences. Our manual efforts were also focused on nouns over other parts of speech. Analysis of conceptual interlingua coverage shows that slightly less than 20% of the English WordNet has been translated into French so far, while our analysis shows that 70% of non-stopword term occurrences in a French corpus will be present in our conceptual interlingua.

	WordNet 1.5	**French**	**% Coverage**
Terms	168,135	72,473	
Synsets	91,591	16,340	17.8%

Four official runs were submitted to NIST for evaluation, all using only the topics' '*Title'* field. Given the fact that the central objective of the evaluation from our point of view was not an exact match with the track requirements (we did not cover all four languages), only one of our runs was in fact evaluated officially by NIST:

Run	**Query Lang.**	**Doc. Lang**	
TW1E2EF	English	English + French	*evaluated!*
TW2F2EF	French	English+French	
TW3E2F	English	French	
TW4F2E	French	English	

Our analysis of the official evaluation of run *TW1E2EF* indicates that it did not perform well compared to other groups who submitted equivalent cross-language retrieval runs with this combination of English queries and both English and French documents. Having analyzed the results on a query-by-query basis however, we have discovered that the formula used in merging results from different document subcollections, was inconsistent with the formula used in computing CINDOR retrieval scores to take account of word sense selection and phrase matching. Merging was used in this run because the English and French document collections were indexed separately, so the run was compiled by a combination of results from an English-English run and an English-French run. Fixing the inconsistency in merging results increased average precision for this combination from 0.1570 to 0.1967, an increase of 25%, as shown in Figure 1 below.

The officially evaluated English against English and French run also suffered further from the simple implementation issues referred to at the end of Section 3. We believe it will be clearer however if we focus on the single language pair runs between English and French to illustrate the affect of these errors on the performance of our system, rather than trying to disentangle the issues related to individual languages and issues of merging individual runs.

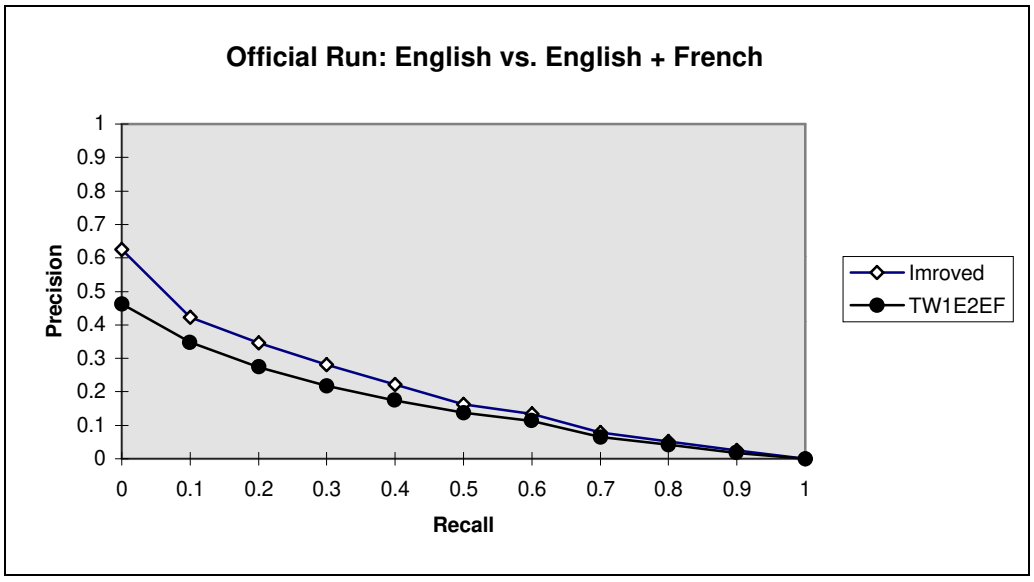

Figure 1
Official run, plus improved merging run

5. Analysis of English and French runs.

Apart from the English against French and French against English runs which we officially submitted to NIST, we have also run monolingual English and monolingual French runs to act as a baseline for our cross-language evaluation.

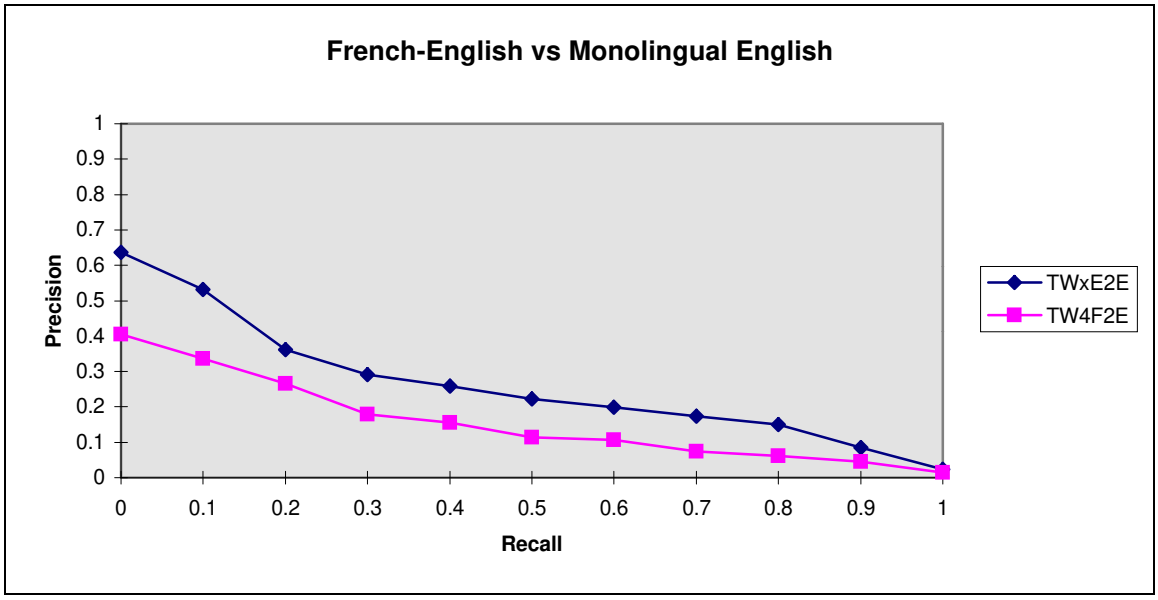

Figure 2.
Comparing French-English Cross-Language to English-only

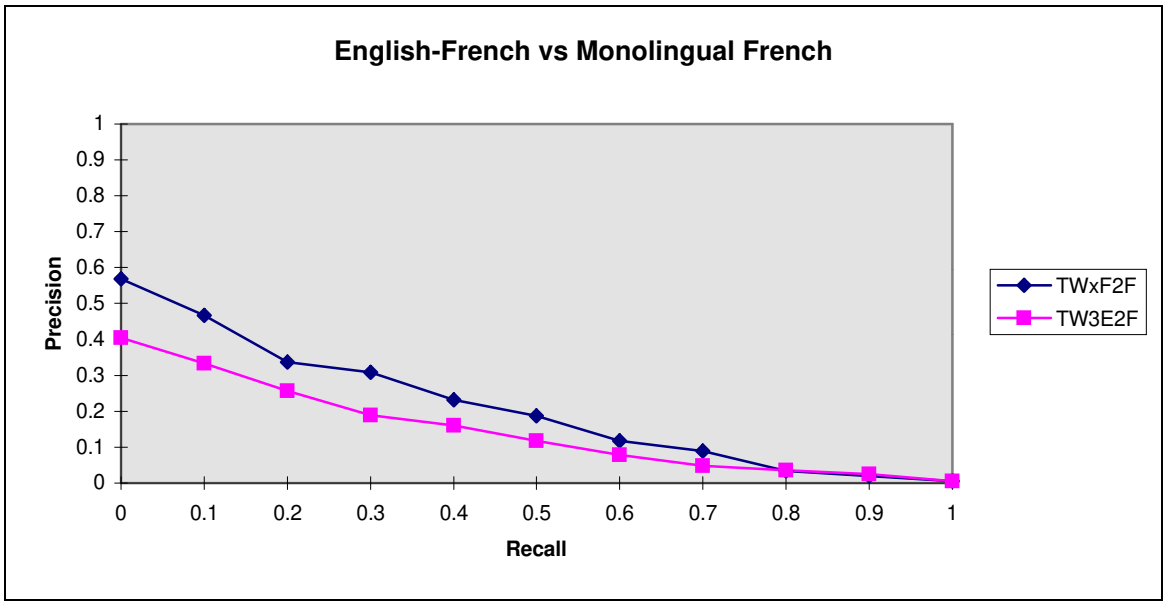

Figure 3.
Comparing English-French Cross-Language to French-only

Runs *TW3E2F* and *TW4F2E*, as presented in Figures 2 and 3, were completed before the official TREC deadline and submitted as official runs to NIST. The monolingual runs, *TWxE2E* and *TWxF2F*, were not submitted because of a limit of four submissions to the cross-language track.

While it is clear from the precision-recall graphs of Figures 3 and 4 that CINDOR cross-language retrieval in each direction is less effective than the equivalent monolingual queries (see Table 1 below), it is instructive to examine performance on a query-by-query basis in order to determine exactly the reasons for this behavior. At an aggregate view, French-English retrieval performs at 58% of monolingual English on the measure of average precision, while English-French achieves 69% of the average precision score of monolingual French.

Query \ Doc	**English**	**French**
English	0.2511	0.1377
French	0.1461	0.1992

Table 1
Average Precision: Cross-Language versus Monolingual

A query-by-query analysis of CINDOR performance is facilitated by Figures 5 and 6 below, which show side-by-side analysis of average precision for cross-language versus monolingual retrieval for each of the 28 test topics over both English (Figure 5) and French (Figure 6) document collections.

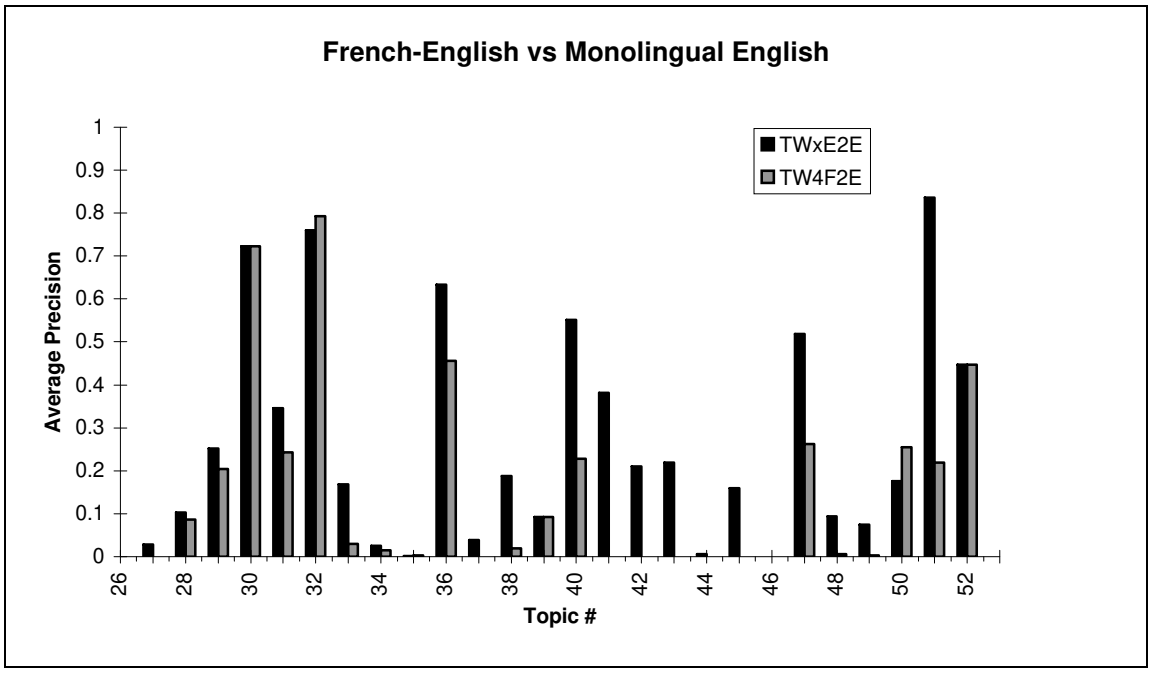

Figure 5
Query-by-query analysis against English documents

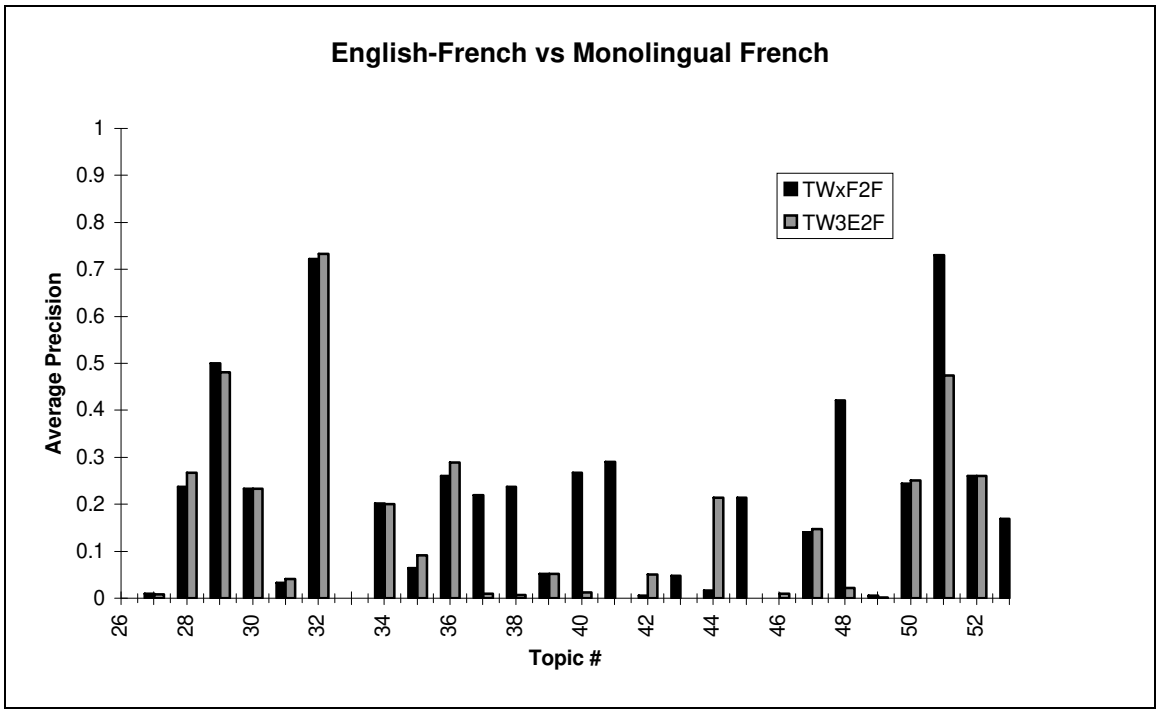

Figure 6.
Query-by-query analysis against French documents

While the initial query-by-query views presented in Figures 5 and 6 indicate specific queries over which the CINDOR system has performed poorly, a detailed analysis of individual query performance serves to illustrate certain classes of problem which have negatively impacted our overall performance on precision and recall. Many of these problems are simple items which one would, but should not, assume would be accounted for as a matter of course in building a cross-language retrieval system:

Implementation issues:

> *accents:* The normalization of accented words was not included as part of CINDOR processing. This meant that words which are accented in one language but not in another, did not match. Words like *'Lótschberg'*, which is not in our current conceptual interlingua but would have matched with *'Lotschberg'* or *'Lötschberg'* had we normalized the accents, were lost.
> *caps:* Although the normalization of capitalized words was included as a step in query processing, this step was mistakenly omitted in processing the French conceptual interlingua resources. Query terms such as *'Pologne'* and *'Allemagne'* were therefore reduced to *'pologne'* and *'allemagne'* during query processing, but did not receive the correct concept label because the conceptual interlingua contained the terms with their uppercase initial character intact.
> *hyphen:* Our lexical analysis module did not segment hyphenated words so *'Franco-Allemande'* was treated as a unit.

General issues:

> *gap / pnoun:* Although listed as two classes in the table above, both refer to particular gaps in our conceptual interlingua coverage. This is not an error, but rather is the nature of our approach (indeed every approach) that there will always be gaps in the resources used.
> *stemming:* When the term *'kidnapping'* was analyzed from a French query using French analysis, it was returned in the form *'kidnapping'*. English analysis would however have given *'kidnap'*, so the match was missed.
> *phrase:* In several cases, such as *'tremblement de terre'* and *'swiss confederation',* phrases were actually present in our conceptual interlingua, but were not recognized as phrases by our query processing.
> *query:* In some queries, particularly English topic 53, the topic title underspecifies the information need. This is indeed normal in many TREC topic descriptions, but here there is an inconsistency across languages. French topic 53 specifies that *'Europe'* is of interest, while the English version does not. A more general example of underspecification is query 33, *'genetic engineering'*, which the description elaborates to genetic engineering only in the agricultural field.

On the other hand, when the queries with these specific errors are taken out of consideration, the picture becomes altogether more positive. Ideally we would like to be able to fix these problems, re-run our tests and report the improved results directly. In the absence of the time necessary to do this however, we are limited to examining the query sub-set which has escaped unaffected by the issues outlined above. Given this limitation, we believe a fair picture of CINDOR's potential strength is presented in Figure 7 below, which presents this query subset in the direction of English-to-French versus French monolingual. This figure illustrates that CINDOR achieves cross-language retrieval performance equivalent to the monolingual level, without the loss of precision normally expected when querying across languages. This is where we believe, based on our notion of a conceptual interlingua, that CINDOR's particular strength lies. It is encouraging that this data supports our belief, though we acknowledge that the full query set must be re-run for verification.

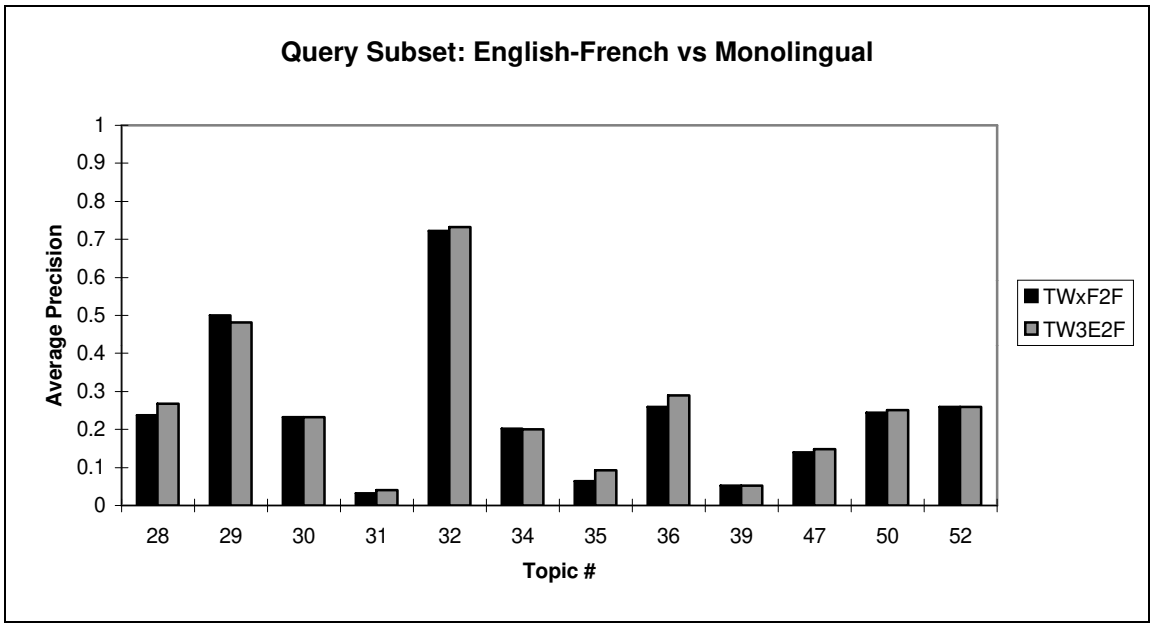

Figure 7.
Performance over queries unaffected by specific errors.

6. Conclusion.

Our query-by-query analysis of CINDOR's cross-language performance in both directions between English and French, outside of specific implementation errors, shows a remarkable number of queries for which there is either no loss of precision, or even a slight gain, when comparing cross-language performance against equivalent monolingual queries. When we examine these queries with respect to both the precision and recall achieved we see that in these cases CINDOR retrieves very often the same number of relevant documents with the same level of precision. Though we have not yet checked if exactly the same actual documents are being retrieved in these cases, it would seem to be that this is so. In several queries we also see a slight re-shuffling of the rank positions of documents returned from cross-language searches, that slightly increases the overall precision of the results.

It is therefore all the more unfortunate that our overall performance was so negatively affected by such simple implementation issues as non-matching capitalized versus non-capitalized words and accented versus unaccented characters. Such is the nature of the task of building a new retrieval system and evaluating it in the TREC environment, all under time pressure, that these kinds of mistakes are often made.

While we would like to present to the TREC community the conclusion that our conceptual interlingua approach to cross-language retrieval is obviously effective, we believe the data, clouded by our errors as it is, only suggests that this is the case. The fact that only 15 of 28 queries were left unaffected by errors (in the direction English-to-French) does not leave enough data for any solid conclusions. It is the fact that on 12 of those 15 queries we have achieved cross-language retrieval performance which closely matches the equivalent monolingual runs, which we believe is more indicative of the true potential of our approach.

While we have examined here the relative performance of cross-language retrieval compared to equivalent monolingual experiments, we have refrained from commenting on the absolute levels of performance achieved. We believe, in the absence of information about the performance of other systems which have submitted runs based on this test collection, we do not have sufficient data to comment competently on this aspect of our results. We do see however, that many of the implementation issues addressed in the previous sections will have negatively impacted our absolute level of performance compared to other groups, and this must be borne in mind when the comparison is finally presented.

References:

[Ballesteros & Croft:98]
L. Ballesteros and W. B. Croft, "Resolving Ambiguity for Cross-Language Retrieval", Proceedings of the 21st Annual Conference on Research and Development in Information Retrieval, SIGIR'98, Melbourne, Australia, 1998.

[Ballesteros & Croft:97]
L. Ballesteros and W. B. Croft, "Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval", Proceedings of the 20th Annual Conference on Research and Development in Information Retrieval, SIGIR'97, Philadelphia, PA, pp84-91, 1997.

[Gonzalo *et al*:98]
J. Gonzalo, F. Verdejo, I. Chugur, and J Cigarrán, "Indexing with WordNet Synsets can Improve Text Retrieval", Proceedings of the COLING/ACL Workshop on Using WordNet for Natural Language Processing, Montreal, Canada, 1998.

[Hull & Grefenstette:96]
D. Hull and G. 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, SIGIR'96, Zurich, Switzerland, 1996.

[Miller:90]
G. Miller, "Wordnet: An On-line Lexical Database", International Journal of Lexicography, special issue, Vol. 3, No. 4, 1990.

[Oard & Diekema:98]
D. W. Oard and A. R. Diekema, "Cross-Language Information Retrieval", To appear in: Annual Review of Information Science and Technology (ARIST), Vol. 33, Martha Williams (Ed.), Information Today Inc., Medford, New Jersey, 1998.

[Richardson & Smeaton 1995]
R. Richardson and A. F. Smeaton, "Using WordNet in a Knowledge-Based Approach to Information Retrieval", In proceedings of the BCD-IRSG Colloquium, Crewe, 1995.

[Sanderson:97]
M. Sanderson, "Word Sense Disambiguation and Information Retrieval", PhD Thesis, Dept. of Computer Science, University of Glasgow, Scotland, 1997.

[Smeaton & Quiqley:96]
A. F. Smeaton and A. Quigley, "Experiments on Using Semantic Distances between Words in Image Caption Retrieval", In proceedings of the 19th International Conference on Research and Development in Information Retrieval, SIGIR'96, Zurich, Switzerland, 1996.

[Voorhees:93]
E. Voorhees, "Using WordNet to Disambiguate Word Senses for Text Retrieval", Proceedings of the 16th Annual Conference on Research and Development in Information Retrieval, SIGIR'93, Pittsburgh, PA, 1993.

[Voorhees:93]
E. Voorhees, "Vector Expansion in a Large Collection", In D. K. Harman, editor, Proceedings of the First Text Retrieval Conference (TREC-1), pages 343-351. NIST Special Publication 500-207, March 1993.

[Voorhees:94]
E. Voorhees, "On Expanding Query Vectors with Lexically Related Words", In D. K. Harman, editor, Proceedings of the Second Text Retrieval Conference (TREC-2), 1994.

Acknowledgments:
The authors wish to acknowledge the important contributions made to many aspects of this work by our colleagues, including; Christophe Amice, Kostya Bogatyrev, Dave Gallaro, Delphine Khanna, Severine Meyer, Woojin Paik, Patricia Patché, Christian Sarmoria, and Edmund Yu.