Automatically-Extracted Thesauri for Cross-Language IR: When Better is Worse

Ralf D. Brown
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890
ralf+@cs.cmu.edu

Abstract

A statistical algorithm for extracting bilingual term dictionaries (thesauri) from parallel text is presented, along with refinements for improving their size and accuracy. Somewhat paradoxically, increasing the accuracy of the extracted thesaurus can in fact *reduce* the performance of an IR system using it to perform query translation for cross-language information retrieval.

1 Introduction

One of the query translation methods for Cross-Language Information Retrieval (CLIR) compared in (Carbonell et al., 1997a; Yang et al., 1997) was a corpus-based bilingual term-substitution thesaurus, called EBT (for example-based term substitution), which substantially outperformed a general-purpose dictionary such as used in (Hull and Grefenstette, 1996; Ballesteros and Croft, 1997). The algorithm for generating this thesaurus (Brown, 1997) was originally developed for use in word-level alignment in an example-based machine translation (EBMT) engine (Brown, 1996). After some additional refinements were added to the extraction algorithm to enhance EBMT performance, CLIR was re-tested using the improved thesaurus (Yang et al., 1998). Although the newer corpus-based terminology dictionary is substantially better – containing far fewer erroneous translations – its performance fails to match that of the earlier one, created with a much simpler and considerably less computationally-intensive method.

2 Thesaurus Extraction

The basic thesaurus extraction algorithm consists of two phases: counting co-occurrences and filtering the counts using a threshold function. The newer refinement is to use the thesaurus generated by the basic algorithm as a seed to a portion of the EBMT word-level alignment method, yielding an improved thesaurus.

2.1 Basic Thesaurus Extraction

The counting phase of the thesaurus extraction creates a large co-occurrence table, which is simply a two-dimensional matrix indexed by source-language words in one dimension and by target-language words in the other, where every entry in the table indicates the number of times that the given pair of words co-occurred in the entire corpus. For each sentence pair in the parallel corpus, increment all of the elements of the co-occurrence table indicated by the cross-product of the words in the source-language sentence of the pair and the words in the target-language sentence. For example, if the example consists of the phrases "of the" and "de la", the following four entries are incremented:

```
(of,de) (of,la) (the,de) (the,la)
```

The filtering phase sets all entries in the co-occurrence table which do not pass at least one of three thresholding tests to zero; any remaining non-zero entries are then used to construct the thesaurus. The tests are based on the Mutual Conditional Probability between the two words of a word pair: if the conditional probability of encountering the source-language word in the same sentence pair as the target-language word and the conditional probability of encountering the target-language word given the source-language word are both "sufficiently" high, then the two words are considered to be mutual translations of each other.

Mathematically: for a given word pair W_s and W_t (source- and target-language words), which co-occur C times, the pair is considered a translation if

$$P(W_t|W_s) \geq thr[C] \text{ and } P(W_s|W_t) \geq thr[C]$$
$$\text{or}$$
$$P(W_t|W_s) \geq thr_1[C] \text{ and } P(W_s|W_t) \geq thr_2[C]$$
$$\text{or}$$
$$P(W_t|W_s) \geq thr_2[C] \text{ and } P(W_s|W_t) \geq thr_1[C]$$

where $thr[C]$ is the symmetric threshold and $thr_1[C]$ and $thr_2[C]$ are the asymmetric threshold (the latter attempts to account for words

Count	0.10	0.10$\sharp$	0.15	0.15$\sharp$
1	1.00	1.00	1.00	1.00
2	0.50	0.50	0.50	0.50
3	0.67	0.33	0.67	0.33
4	0.50	0.25	0.50	0.25
5	0.40	0.20	0.40	0.20
6	0.33	0.17	0.33	0.17
7	0.29	0.14	0.29	0.15
8	0.25	0.13	0.25	0.15
9	0.22	0.11	0.22	...
10	0.20	0.10	0.20	
11	0.18	0.10	0.18	
12	0.17	...	0.17	
13	0.15		0.15	
14	0.14		0.15	
15	0.13		...	
16	0.13			
17	0.12			
18	0.11			
19	0.11			
20	0.10	0.10	0.15	0.15

Table 1: Sample Filtering Threshold Functions

which are polysemous in one language but not the other). The thresholds are a function of the absolute number of co-occurrences of the word pair because frequent words may still co-occur by chance a small number of times, while rare words which co-occur the same number of times are likely to be translations of each other. Thus, setting a high threshold for low values of C as in Table 1 allows correct definitions to be induced for rare words, including singletons, while lower thresholds for larger C permit more frequent words to have multiple translations.

After filtering, any remaining non-zero entries (i.e. those word pairs which co-occurred sufficiently often to pass the thresholding tests) are added to the thesaurus, along with the co-occurrence count for each word pair. Figure 1 shows a random sample of an automatically-generated English-Spanish thesaurus, where the number associated with each Spanish translation is the number of times that it co-occurred with the English word.

2.2 Thesaurus Refinement

To refine the "seed" thesaurus generated by the basic extraction algorithm, a portion of the EBMT engine's word-level alignment algorithm is used. Using the observation that in European languages, word order is typically locally consistent (i.e. constituents in one language are likely to correspond to a coherent unit in another even when the large-scale word order differs dramatically), a constrained correspondence table be-

```
(ABI (ABI 4)(BEVERAGE 2)(AMALGAMATED 2))
(ALMAHDI (AL-SADIQ 1)(AL-MAHDI 1))
(BIOTOPES (BIOTOPOS 2))
(BLEACH (LEJÍA 1))
(COMPLEMENTARITY (COMPLEMENTARIEDAD 77))
(D-1 (D-1 91)(D-2 43))
(DEEPEN (PROFUNDIZAR 17))
(DYNAMICS (DINÁMICA 77))
(EBW (HAZ 6)(ELECTRONES 6)(SOLDADORA 6))
(EXTRACONTINENTAL (EXTRACONTINENTALES 1))
(GEOSYSTEMS (GEOEX-1986 1)(GEOSISTEMAS 1))
(KALPAGÉ (KALPAGÉ 2))
(KG (KILOGRAMOS 16)(KG 10))
(MILITARY-IDEOLOGICAL
   (MILITAR-IDEOLÓGICA 1))
(MOHAMMAD (MOHAMMAD 12))
(MONASTERY (MONASTERIO 2))
(NON-NUCLEAR-WEAPON (POSEEDORES 78))
(ORCI (OIRI 8))
(PASHTU (PASHTU 1)(BRISTISH 1))
(PYAT (PYAT 1)(CABARET 1)(PERODIN 1)
   (VILLARD 1))
(RAVANDI (RAVANDI 1)(KATCHOUI 1))
(REDISCOVER (REENCONTRAR 1))
(SCENES (ESCENAS 5))
(SECRECY (SECRETO 53))
(SHANKANGA (SHANKANGA 1))
(TECNOLÓGICO (UCMM 1))
(XIAODI (HU 2)(XIAODI 2))
(XXVI (XXVI 86))
(|1506TH| (|1506A| 8))
```

Figure 1: Sample Thesaurus Entries

tween a sentence and its translation can be produced. This constrained table is then used to drive the count-and-filter algorithm as shown in Figure 2.

The constrained correspondence table is generated by creating a table of possible correspondences between the source and target halves of the sentence pair (a bi-text map which may be incomplete as well as ambiguous), pruning outliers, and then filling in small gaps. The initial bi-text mapping marks each source/target word pair as a possible correspondence if the target word is one of the candidate translations listed in the seed thesaurus (or whatever bilingual dictionary may be available).

Outliers in the bi-text map are pruned by determining an "expected" range of positions for the translation of each word, and removing any elements of the bi-text map which are outside of the expected range if there are also possible correspondences within that range. For each word in the source-language sentence, the expected range is the segment of the input text between the earliest and latest possible correspondences for the preceding and following word, expanded by N words (two for these experiments) in each

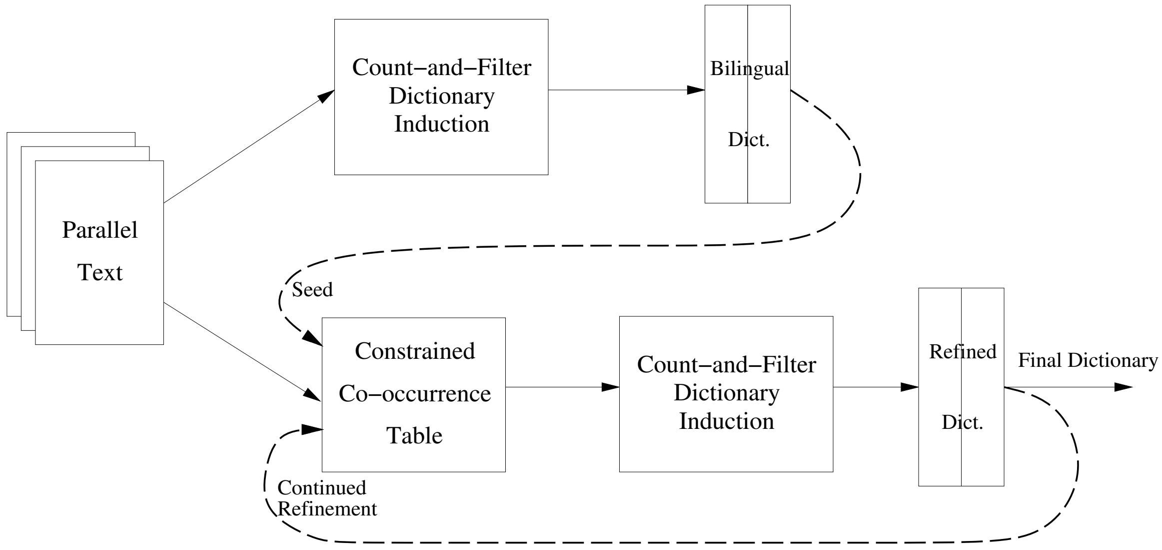

Figure 2: Flowchart of Refinement Algorithm

direction to allow for word-order variation. If either the preceding or following word has no known correspondences in the bi-text map, the allowable range is considered to be the entire target-language sentence. The first and last words of the sentence assume that the allowable range is the range of possible correspondences of the adjacent word, expanded by $2N$ words in each direction. After processing the entire source-language sentence in this manner, the target-language sentence is processed in a like manner, and the entire procedure repeated until no more outliers can be removed. Figure 3 shows the effect of removing outliers for a particular sentence pair; black squares are correspondences which survive the pruning, while shaded squares are the outliers which are removed in a total of two passes along each axis. Note that while pruning is most effective when the overall word order is roughly the same (as in this example), the technique is robust enough to handle large-scale reordering of constituents between the sentence and its translation, since the "expected" range encompasses both adjacent words, regardless of where they occur in the translation.

Once the bi-text map has been pruned, single-word gaps are filled, and the mapping is pruned a second time if necessary. Gap-filling consists of marking, for each source-language word which has no known correspondences but where both the preceding and following word *do* have correspondences, all target-language words without known correspondences. This augmentation is based on the assumption that, if only a small number of words on each side are missing from the bi-text mapping, they most likely correspond to each other. The four circles in Figure 3 indicate the correspondences which are marked when the one-word gaps in the original bi-text are filled. A second round of pruning will then remove the lower left circle and the top four marked squares in the third column from the right, as shown in Figure 4. Unlike the original bi-text, the final mapping clearly shows that the word order in both the Spanish and English versions of the sentence is essentially the same, but does contain some small-scale inversions.

After the bi-text mapping has been pruned and gaps have been filled, the result is used in the counting phase of thesaurus extraction. For each marked entry in the bi-text mapping, increment the appropriate word pair in the main correspondence table used for extraction; once all sentences have been processed, filter the correspondence table as before. A simple way to summarize the difference between the initial thesaurus generation and the refined thesaurus is that the initial generation is equivalent to using a bi-text matrix with every single entry marked, rather than merely those for which a bilingual dictionary and coherence assumption indicate a possible translation. In other words, the initial generation step is completely structure-blind, while the refinement step makes use of the structure of the aligned sentences, albeit in a limited and language-independent manner.

3 Corpus Reduction

The thesauri used for the CLIR experiments reported here and in (Carbonell et al., 1997a; Yang et al., 1997) are extracted from a Spanish-English parallel corpus which consists primarily

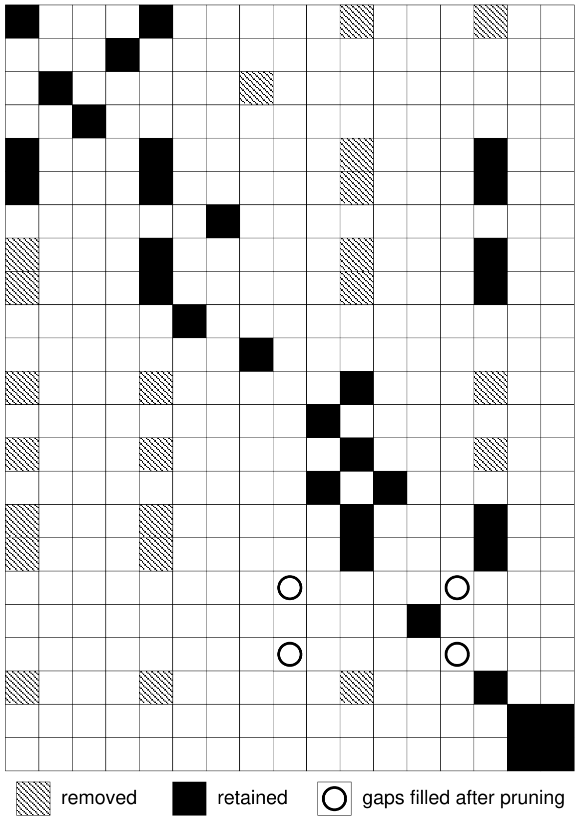

Figure 3: Pruning Outliers from Bi-Text Map

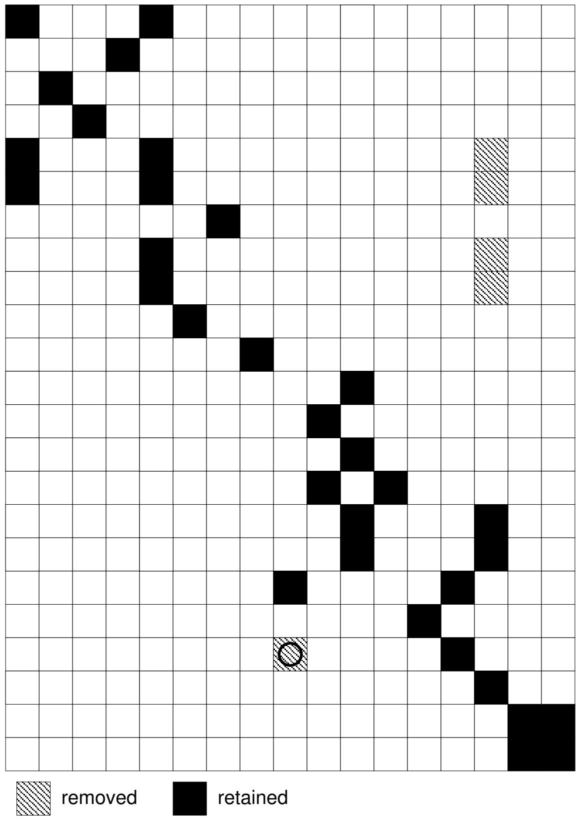

Figure 4: Re-Pruning After Filling Gaps

of sentence-aligned text from the UN Multilingual Corpus available from the Linguistic Data Consortium (Graff and Finch, 1994). In addition to a partially manually-cleaned version of the Spanish and English portions of the UN Corpus[1] totalling about 250 megabytes, the parallel corpus contains some three megabytes of text from the Pan-American Health Organization (PAHO), 200K of text from ARPA evaluations, and about 85K (1300 entries) from a phrasebook.

Two methods of corpus reduction were used to test performance when smaller amounts of parallel text are available. The first method is simply to use a contiguous segment of specified size from the corpus; the second is to select the set of least-redundant source-language sentences (and their translations) which total the desired size.

For the first method, the last N megabytes of text in the corpus are used. In the corpus, the UN text occurs first, then the PAHO text, ARPA texts, and finally the phrasebook, so initially only UN texts are removed as the corpus size is decreased.

For the second method, the algorithm to select sentences is a form of Maximal Marginal Relevance (Carbonell et al., 1996; Carbonell et al., 1997b). The relevance measure is the average of the inverse occurrence frequency (in all prior sentences) of the bigrams in each sentence; no similarity measure is used since the desired result is to find the maximally-diverse set of sentences. Selection requires at least two passes over the corpus: in the first pass, an approximate threshold is determined, which is then used to decide which sentences to retain during the second pass. During the second pass, bigram frequencies are only updated for sentences which are selected; as a result, scores are higher during the second pass, which will cause more than the desired number of sentences to be retained. This may necessitate additional iterations using the already-reduced corpus as input instead of the original corpus.

4 Results

The various thesauri generated using the methods previously described were tested in a cross-language information retrieval setting. The

[1] Numerous sentential alignments have been corrected, and a number of idiosyncrasies in the text, such as British spellings and embedded cross-references to other UN documents, have been removed.

Corpus	Thr	Vocab	Defs	Avgp	TL/ML
full	0.15	43575	1.97	0.4067	86.1%
full	0.20	42905	1.95	0.4316	91.4%
full	0.22	42171	1.93	0.4294	91.0%
full	0.25	42232	1.93	0.4111	87.0%
full	0.27	41501	1.92	0.4318	91.4%
full	0.30	41095	1.92	0.3992	84.5%
refined	0.01	47445	3.11	0.3926	83.1%
refined	0.02	44809	2.54	0.3969	84.0%
refined	0.05	42389	1.64	0.4189	88.7%
refined	0.08	40825	1.38	0.4172	88.3%
refined	0.10	40095	1.31	0.4220	89.3%
refined	0.12	39342	1.27	0.4167	88.3%
refined	0.15	34914	1.12	0.4117	87.2%
refined	0.20	33606	1.08	0.3858	81.7%
small	0.08	24423	4.18	0.4314	91.3%
small	0.10	24250	3.61	0.4358	92.3%
small	0.11	24183	3.42	0.4334	91.8%
small	0.12	24094	3.27	0.4347	92.0%
small	0.13	23992	3.15	0.4379	92.7%
small	0.14	23832	3.00	0.4378	92.7%
small	0.15	22410	2.60	0.4347	92.0%
small	0.20	21570	1.42	0.4177	88.4%
small	0.27	20797	1.37	0.4025	85.2%
sm-ref	0.05	22333	3.37	0.4109	87.0%
sm-ref	0.08	21837	2.51	0.4250	90.0%
sm-ref	0.10	21679	2.20	0.4319	91.4%
sm-ref	0.11	21580	2.09	0.4338	91.8%
sm-ref	0.12	21497	2.00	0.4310	91.2%
sm-ref	0.13	21395	1.92	0.4291	90.8%
sm-ref	0.15	19910	1.54	0.4299	91.0%
sm-ref	0.20	19590	1.40	0.4190	88.7%
sm-ref	0.22	19250	1.35	0.4079	86.4%
mmr75	0.08	37587	2.61	0.4183	88.6%
mmr75	0.10	37419	2.18	0.4262	90.3%
mmr75	0.12	37252	2.04	0.4209	89.2%
mmr75	0.15	35359	1.65	0.4147	87.8%
mmr75	0.20	35027	1.55	0.4151	87.9%
Monolingual baseline				0.4721	100.0%

Table 2: EBT Performance Variation

same methodology, queries, training and test data[2] were used for the experiments discussed here as for those reported in (Carbonell et al., 1997a; Yang et al., 1997): the English queries were translated into Spanish using the thesaurus and then supplied to the SMART v11.0 information-retrieval system (Salton, 1989) to find the relevant documents. Since the relative frequencies of the various candidate translations in the thesaurus are known, they are used to weight the terms in the translated query. SMART weights the terms in a query by the number of times they occur, so an approximate weighting is performed by outputting each term an appropriate number of times, such that the total number of words produced for each word in the query totals 20 (as early tests found no further improvement from increasing the total weighting beyond 20).

Rather than translating the entire corpus, query translation (Davis and Dunning., 1996) is used, because in large-scale practical systems it is usually infeasible to do the former.

The monolingual baseline for all thesaurus-based tests is the SMART vector space model without relevance feedback (SMART.basic), using the "ntc.ntc" weighting scheme (cosine vector difference, with simple TF*IDF weighting). Table 2 shows the performance of the various thesauri both in terms of absolute 11-point average precision (Avgp) and as the ratio between translingual (cross-language) and monolingual performance (TL/ML); the thesauri will be described in more detail below. Average precision is defined as usual – the average of the observed precisions at 0%, 10%, ..., 100% recall (where precision at 0% recall is interpolated) – and the ratio is simply the 11-point average precision for translingual retrieval divided by the baseline monolingual average precision, to indicate the amount of degradation caused by crossing from one language to another. In addition, Table 2 indicates the vocabulary of the thesaurus (the number of English words for which it provides Spanish correspondences) in the column labeled "Vocab" and the average number of Spanish terms for each English word in the column labeled "Defs".

Five different corpora (named "full", "small", "refined", "sm-ref", and "mmr75" in Table 2) were used to generate thesauri. For each corpus, multiple thesauri were generated using different threshold functions, identified in the second column of Table 2.

The "full" corpus contains the entire available sentence-aligned Spanish-English text (UN multilingual corpus, PAHO corpus, ARPA evaluation texts, and phrasebook; nearly 260 megabytes of text). The "refined" corpus is actually identical to the "full" corpus, but the thesauri are built via three iterations of the refinement algorithm described previously, using an initial seed thesaurus generated from the "full" corpus with a minimum threshold of 0.05. The "small" corpus consists of a contiguous segment, totalling about 33 megabytes (including all of the non-UN text), and "sm-ref" is built from the same text using the refinement algorithm. Finally, the "mmr75" corpus contains the 75,000

[2] The parallel corpus, queries, and relevance judgements are available to members of the Linguistic Data Consortium; please contact the author for details.

least-redundant sentences selected as described above, comprising aproximately 23.5 megabytes of text.

The threshold functions used to filter the co-occurrence table generated during the count phase all start at 1.00 (perfect co-occurrence) for a co-occurrence count of 1 (except the "0.01" function, which begins at 0.50), and decrease more or less rapidly to the minimum value by which they are identified; see Table 1 for example functions.

As can be seen from Table 2, for each corpus, the overall performance first improves and then eventually begins to decrease again as the threshold function (identified in the column labeled "Thr") is lowered. The "small" corpus performs as well as the "full" corpus[3], while the three corpora intended to produce more accurate thesauri fail to match the performance of the two corpora to which the simple, basic extraction algorithm is applied. These results will be discussed in the next section.

5 Discussion

The varying performance of the thesauri built from the "full", "refined", and "small" corpora can be explained in terms of the presence/absence of query terms in the thesaurus and the degree of query expansion performed by the thesaurus. Clearly, any query terms which are absent from the thesaurus can not be translated, and thus will contribute to the translated query which is actually used to perform the retrieval only if they are names which appear in the document collection. Therefore, the larger the thesaurus, the fewer terms are omitted from use in retrievals; this factor indicates that more lenient thresholds should be used in thesaurus extraction.

Query expansion has been known to be beneficial for some time (Buckley et al., 1995). Due to the method by which the thesaurus is extracted, three expansions can occur: multiple inflections, multiple synonyms, and collocated terms. Depending on usage patterns in the corpus (and the target language in general), a particular surface form in the source language may correspond to multiple inflected forms in the target language. Provided that the extraction thresholds are not too strict, all of the inflected variants may be listed as possible translations for the word. Similarly, if the corpus translates a particular word using various synonyms, all of the synonyms may be captured by the thesaurus. Finally, the automatically-extracted thesauri include the translations of other terms which are frequently collocated in the source language – what (Melamed, 1997) calls "indirect associations" (e.g. "Burkina" always collocates with "Faso" because both are part of the name of the country "Burkina Faso", and thus each of these words is linked with both by the thesaurus). The basic extraction algorithm does not take word position into account, so the translations of highly-collocated words will have a Mutual Conditional Probability nearly as high as the translations of the term itself. While the refinement algorithm can remove many collocated terms, it is only rarely able to distinguish between two words which always appear adjacent to each other, not just in the same sentences.

For some concrete examples, refer back to Figure 1. One can see that the abbreviation "kg" for "kilogram" is translated as both the abbreviation and as the spelled-out word. Similarly, the acronym "ABI" also translates as two of the words of its expansion, and the name "Xiaodi" finds the full name "Hu Xiaodi".

Since more lenient thresholds also allow more collocated terms to be listed for each word, one would expect that lower thresholds should be used in thesaurus extraction for this reason as well. However, if an excessive number of translations are listed for each term, retrieval precision will suffer (due to irrelevant terms in the translated query), and thus lower the 11-point average precision. The combination of these factors produces the peak in performance at some intermediate value – this is the point at which excessive expansion begins to reduce precision more than it enhances recall.

One can now readily see why the performance of the "refined" thesauri peaks at a slightly lower value: the reduced query expansion (as indicated by the reduced average number of translations for a term) in the refined thesaurus outweighs the additional coverage afforded by the increased vocabulary.

The redundancy-reduced corpus was used to give an additional data point, since it had been created anyway. In practice, one would normally use all available parallel text rather than artificially reducing the training corpus. An examination of Table 2 shows that the redundancy-removal algorithm does indeed preserve most or all of the significant information in the parallel text – the extracted thesauri are only slightly smaller than those extracted from the full text, and much larger than those ex-

[3] In fact, the absolute numbers are slightly higher for "small", but that is due to recent experimentation with the threshold functions which has not yet been applied to the "full" corpus due to the increased amount of processing time required for the larger amount of text. Prior to the recent experiments, the "small" corpus had an identical peak performance of 0.4318.

tracted from a contiguous section of the parallel text which actually contains more text. Performance is slightly lower than for either the "full" or "small" corpora, but interestingly, the curve of threshold versus performance is noticeably flatter, yielding good performance over an even larger range of threshold values than either of the unprocessed corpora.

Because about one-sixth of the words defined in the "mmr75" thesaurus do not appear in the largest ("refined-0.01") thesaurus, merging these terms into the larger thesaurus may be beneficial for EBMT word-level alignment, but this still needs to be investigated.

A careful examination of the average-precision values for the various thesauri reveals that all training corpora except "mmr75" show two local maxima with a small dip in between. Determining the reason for this will require further investigation. One hypothesis is that there are some words in the queries for which the thesauri do not include definitions until the performance on the other words has already passed its peak. Once these words are present in the thesauri, they provide an additional boost until the number of unrelated definitions rise even further.

6 Conclusion

Sometimes "better" is actually worse. Query expansion (if not taken to excess) helps retrieval performance by enhancing recall, and the statistical nature of the automatic thesaurus extraction provides query expansion in the form of collocated terms. Improving the accuracy of the thesaurus in this instance at best resulted in unchanged performance and at worst reduced performance by limiting the expansion peformed by the thesaurus. At least for cross-lingual retrieval, the 20-fold greater computation required to refine the thesaurus is unnecessary and possibly even counter-productive.

References

L. Ballesteros and B. W. Croft. 1997. Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval. In *UMass Tech Report: IR-104*.

R. D. Brown. 1996. Example-Based Machine Translation in the PANGLOSS System. In *Proceedings of the Sixteenth International Conference on Computation Linguistics*, pages 169–174.

R. D. Brown. 1997. Automated Dictionary Extraction for "Knowledge-Free" Example-Based Translation. In *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-97)*.

C. Buckley, G. Salton, J. Allan, and A. Singhal. 1995. Automatic Query Expansion Using SMART: TREC 3. In *Overview of the Third Text REtrieval Conference (TREC-3)*, pages 69–80.

J. G. Carbonell, J. Goldstein, and Y. Geng. 1996. Query-Relevant Document Summarization. Technical report, Carnegie Mellon University, Pittsburgh, PA.

J. G. Carbonell, Y. Yang, R. E. Frederking, R. D. Brown, Y. Geng, and D. Lee. 1997a. Translingual Information Retrieval: A Comparative Evaluation. In *Proceedings of IJCAI-97*, August.

Jaime Carbonell, Yibing Geng, and Jade Goldstein. 1997b. Automated Query-Relevant Summarization and Diversity-Based Reranking. In *AI in Digital Libraries - Moving from Chaos to (More) Order*, pages 9–14, August. IJCAI-97 Workshop Series.

M. Davis and T. Dunning. 1996. A TREC Evaluation of Query Translation Methods for Multi-Lingual Text Retrieval. In *The 4th Text Retrieval Conference (TREC-4)*.

D. Graff and R. Finch. 1994. Multilingual Text Resources at the Linguistic Data Consortium. In *Proceedings of the 1994 ARPA Human Language Technology Workshop*. Morgan Kaufmann.

D.A. Hull and G. Grefenstette. 1996. Querying Across Languages: a Dictionary-based Approach to Multilingual Information Retrieval. In *19th Ann Int ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96)*, pages 49–57.

I. Dan Melamed. 1997. A Word-to-Word Model of Translational Equivalence. In *35th Annual Meeting of the Association for Computational Linguistics (ACL'97)*, pages 490–497.

G. Salton. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, Pennsylvania.

Y. Yang, R. D. Brown, R. E. Frederking, J. G. Carbonell, Y. Geng, and D. Lee. 1997. Bilingual-corpus Based Approaches to Translingual Information Retrieval. In *Proceedings of MULSAIC'97*, August.

Y. Yang, J.G. Carbonell, R.D. Brown, and R.E. Frederking. 1998. Translingual Information Retrieval: Learning from Bilingual Corpora. *Artificial Intelligence Journal*. Special issue: Best of IJCAI-97.