Automating the Acquisition of Bilingual Terminology

Pim van der Eijk

Digital Equipment Corporation
Kabelweg 21
1014 BA Amsterdam
The Netherlands
eijk@cecehv.enet.dec.com

Abstract

As the acquisition problem of bilingual lists of terminological expressions is formidable, it is worthwhile to investigate methods to compile such lists as automatically as possible. In this paper we discuss experimental results for a number of methods, which operate on corpora of previously translated texts.
Keywords: parallel corpora, tagging, terminology acquisition.

1 Introduction

In the past several years, many researchers have started looking at bilingual corpora, as they implicitly contain much information needed for various purposes that would otherwise have to be compiled manually. Some applications using information extracted from bilingual corpora are statistical MT ([Brown et al., 1990]), bilingual lexicography ([Cati-zone et al., 1989]), word sense disambiguation ([Gale et al., 1992]), and multilingual information retrieval ([Landauer and Littmann, 1990]).

The goal of the research discussed in this paper is to automate as much as possible the generation of bilingual term lists from previously translated texts. These lists are used by terminologists and translators, e.g. in documentation departments. Manual compilation of bilingual term lists is an expensive and laborious effort, hence the relative rarity of specialized, up-to-date, and manageable terminological data collections. However, organizations interested in terminology and translation are likely to have archives of previously translated documents, which represent a considerable investment. Automatic or semi-automatic extraction of the information contained in these documents would then be an attractive perspective.

A bilingual term list is a list associating source language terms with a ranked list of target language terms. The methods to extract bilingual terminology from parallel texts were developed and evaluated experimentally using a bilingual, Dutch-English corpus. There are two phases in the process:

1. Process the texts to extract terms. The definition of the notion 'term' will be an important issue of this paper, as it is necessary to adopt a definition that facilitates comparison of terms in the source and target language. Section 4 will show some flaws of methods that define terms as words or nouns. Terminologists commonly use full noun phrases[1] as terms to express (domain-specific) concepts. The NP level is shown to be a better level to compare Dutch and English in sections 5.1 and 5.2.

   This phase acts as a linguistic front end to the second phase. The various techniques used to process the corpus are described in section 2.

2. Apply statistic techniques to determine correspondences between source and target language. In section 3 we will introduce a simple algorithm to select and order potential translations for a given term. This method will subsequently be compared to two other methods discussed in the literature.

The usual benefits of modularity apply because the two phases are highly independent.

---

[1]To some extent, a particular domain will also have textual elements specific to the domain that are not NPs. We will ignore these, but essentially the same methods could be used to create bilingual lists of e.g. verbs.

This paper is structured as follows. Section 2 introduces the operations carried out on the evaluation corpus. Section 3 describes the translation selection method used. Section 4 discusses initial experiments which use words, resp. only nouns, as terms. Section 5 contains an evaluation of a larger experiment in which NPs are used as terms. Related research is discussed in [Gaussier et al., 1992], [Gale and Church, 1991a] and [Landauer and Littmann, 1990]. Section 6 compares our method with these approaches. Section 7 summarizes the paper, and compares our approach to related research.

2 Text preprocessing

A number of experiments were carried out on a sample bilingual corpus, viz. Dutch and English versions of the official announcement of the ESPRIT programme by the European Commission, the Dutch version of which contains some 25,000 words. The texts have been preprocessed in several ways.

Lexical Analysis Word and sentence boundaries were marked up in SGML. This involved taking into account issues like abbreviations, numerical expressions, character normalization. No morphological analysis (stemming or lemmatization) was applied.

Alignment The experiments were carried out on parallel texts aligned at the sentence level, i.e. the texts have been converted to corresponding *segments* of one, or a few, sentences. Reliable sentence alignment algorithms are discussed in [Brown et al., 1991] and [Gale and Church, 1991b]. For our experiments we used the Gale-Church method, which is implemented by Amy Winarske, ISSCO, Geneva. Figure 1 is a display of two aligned segments.

Figure 1: Aligned text segments

Een hardnekkige weerzin tegen vroegtijdige standaardisatie verhindert een wisselwerking tussen produkten	↔	A persisting aversion to early standardisation prevents an inter-working of products

Tagging In order to investigate the role of syntactic information, the texts have been tagged. A tagged version of the English text was supplied by Umist, Manchester. The Dutch version was tagged automatically using a tagger inspired on the English tagger described in [Church, 1988]. This tagger uses as contextual information a trigram model constructed using a previously tagged corpus, viz. the "Eindhovense corpus". The system furthermore uses as lexical information a dictionary derived from a subset of the Celex lexical database, which contains information about the possible categories and relative frequencies of about 50,000 inflected Dutch word forms.

Figure 2 shows the tagged aligned segments.

Figure 2: Tagged aligned text segments

Een_d hardnekkige_a ↔ A_d persisting$_a$ aversion$_n$ weerzin$_n$ tegen$_p$ to$_p$ vroegtijdige$_a$ standaar-disatie$_n$ verhindert$_v$ een$_d$ early$_a$ standardisation$_n$ prevents$_v$ and inter-wisselwerking$_n$ tussen$_p$ working$_n$ of$_p$ products$_n$ produkten$_n$

Parsing On the basis of previous tagging, the texts are superficially parsed by simple pattern matching, where the objective is to extract a list of term noun phrases. The following grammer rule, where "w" is a marked up word, expresses that English term NPs consist of zero or more words tagged as adjectives followed by a one or more words tagged as nouns.

$$\text{np} \rightarrow \text{w}_a^* \text{ w}_n^+$$

The grammar rule doesn't take postnominal complements and modifiers into account, because the lexicon lacks information to disambiguate PP attachment. We will later see (section 5.3) that this causes problems in relating Dutch and English NPs. Figure 3 shows the result of parsing, with recognized NPs in bold face. Texts can be parsed in linear time using finite state techniques.

Figure 3: Parsed aligned text segments

Een **hardnekkige weerzin** tegen **vroegtijdige standaardisatie** verhindert een **wisselwerking** tussen **produkten**	↔	**A persisting aversion** to **early standardisation** prevents an **inter-working** of **products**

3 Translation selection

A number of variants of bilingual term acquisition algorithms have been implemented that operate on parallel texts. These methods use the output of the operations in section 2, then build a database of "translational co-occurrences", determine and order target language terms for each source language term, (optionally) apply filtering using threshold values, and write a report.

The selection and ordering technique used is similar to another well-known ranking method, viz. mutual information. We will compare experimental results based on our method and on mutual information in section 6.1.

Co-occurrence In conducting our experiments, a simple statistic measure was used to rank the probability that a target language term is the translation of a source language item. This measure is based on

the intuition that the translation of a term is likely to be more frequent in the subset of target[2] text segments aligned to source text segments containing the source language term than in the entire target language text.

The method consists in building a "global" frequency table for all target language terms. Furthermore, for each source language term, a "sub-corpus" of target text segments aligned to source language segments containing that source language term is created. A separate, "local" frequency table of target language terms is built for each source language term. Candidate translation terms tl for a source language term sl are ranked by dividing the "local" frequency by their "global" frequency, and select those pairs for which the result ≥ 1.

$$\frac{freq_{local}\ (tl|sl)}{freq_{global}\ (tl)}$$

Threshold An important drawback of this definition is that very low-frequent target language terms, which just happen to occur in an aligned segment will get unrealistically high scores. To eliminate these, we imposed a threshold by removing from the list those target language terms whose local frequency was below a certain threshold. The threshold is defined in terms of the global frequency of the source language term.

$$\frac{freq_{local}\ (tl|sl)}{freq_{global}\ (sl)} \geq threshold$$

The default threshold used was 50%. However, this restriction does not improve results for those source language terms that are infrequent themselves. The effects of variation of this threshold on precision and recall are discussed in section 5.2, where it will be shown that the threshold, as a parameter of the program, can be modified by the user to give a higher priority to precision or to recall.

Similar filters could be established by defining a threshold in terms of the global frequency of the target language term. One could also require minimal absolute values[3].

Position-sensitivity An option to the selection method is to calculate the "expected" position of the translation of a term (using the size[4] of source and target fragments and the position of the source term in the source segment). For the target language terms, the score is decreased proportionally to the

---

[2]It should be noted that we are comparing two translationally related texts; there need not be an actual directional source → target relation between the texts.

[3]For example, [Gaussier et al., 1992] selected source language terms co-occurring more than six times with target language terms.

[4]Size and distance are measured in terms of the number of words (or nouns, NPs) in the segments.

distance from the expected position, normalized by the size of the target segment[5].

4 Word and noun-based methods

4.1 Experiment

In the word and noun-based methods, a test suite of 100 Dutch words which were tagged as a noun was selected at random. In the word-based method, the frequencies being compared are the frequencies of the word forms. In the noun-based method, only frequencies of nouns are compared. Figure 4 shows the result of some experiments. The quality of the methods can be measured in recall -whether or not a translation of a term is found- and precision. We define precision as the ability of the program to assign the translation, given that this translation has been found, the highest relevance score.

Figure 4: Word and noun-based methods

Term	Position	Recall	Precision
word	no	52%	33%
word	yes	52%	77%
noun	no	48%	49%
noun	yes	43%	77%

The experiments demonstrate that position-sensitivity results in a major improvement of precision. The size of the segments of the aligned program is still fairly large (on average, over 24 words per segment in the test corpus), therefore there will in general be a lot of candidate translations for a given term. Especially in the case of a small corpus such as ours, this results in a tendency to return a number of terms as *ex aequo* highest scoring items. Apparently, there is little distortion in the order of terms in the corpus.

Another conclusion that can be drawn from the examples is that use of categorial information alone does not improve precision, even though the number of candidate translations is greatly reduced. Position-sensitivity is a much more effective way to achieve improved precision. One factor explaining this lack of succes is the error rate introduced by text tagging, which the word-based method does not suffer from. As expected, there is an inherent reduction in recall because nouns do not always translate to nouns.

Figure 5 shows an example of the output of the position-sensitive, word-based system. The word *industry* occurs 88 times globally (fourth output column) in the corpus, twice locally, in segments aligned

---

[5]This option introduces a complication in that local scores are no longer simple co-occurrence counts, whereas global scores still are. This is partly responsible for lower recall in figures 4 and 9.

to segments containing *industrietak*. This local frequency is adapted to *1.8315..* (the third output column), because of position-sensitivity.

Figure 5: Example output

Found 2 matches for *industrietak* in 912 segments
13.073232323232324 *industry* 1.8315151515151515 88
3.5176684881602913 *is* 1.376969696969697 244
2.331223628691983 *in* 1.7727272727272727 474

4.2 Evaluation

The real concern raised by the results of the four methods discussed is the very low recall. There are various categories of errors common to all methods, which will be discussed in more detail in the evaluation of a much larger experiment in section 5.3.

However, a more fundamental problem specific to the word and noun-based methods is the inability to extract translational information between higher-level units such as noun phrases or compounds. The English compound *programme management* is related to a single Dutch word, viz. *programmabeheer*, and even more complex sequences such as *high speed data processing capability* are translations of *snelle gegevensverwerkingscapaciteit*, where *high speed* is mapped to the adjective *snel* and *data processing capability* to *gegevensverwerkingscapaciteit*. The compound problem alone represents 65% of the errors, and is a general problem which comes up in comparing languages like German or Dutch to languages like French or English.

Although the compound problem can also be addressed by morphological decomposition of compounds, there are two other advantages to compare the languages at the *phrasal* rather than at the (tagged) lexical level.

Sometimes, an ambiguous noun is disambiguated by an adjective, e.g. *financial statement*, where the adjective imposes a particular reading on the head noun. A phrasal method is then based on less ambiguous terms, and will therefore yield more refined translations.

Furthermore, the method implicitly lexicalizes translation of collocational effects between adjectives and head nouns.

5 Phrase-based methods

5.1 Evaluation of phase-based methods

Initial experiments with a phrase-based method showed a small quality increase. However, in order to evaluate the performance of the phrase-based methods in more detail, a much larger and representative collection of NPs was selected. This collection consisted of 1100 Dutch NPs, which is 17% of the total number of NPs in the Dutch text.

A list associating these terms to their correct translations was compiled semi-automatically, by using some of the methods described in this paper and checking and correcting the results manually. 61 NPs were removed from the collection because the translation of some occurrences of these terms turned out to be incorrect, very indirect, simply missing from the text, or because they suffered from low-level formatting errors or typing errors. Also, a program to automate the evaluation process was implemented. The remaining set was divided in two groups.

1. One group contained 706 pairs of NPs which the extraction algorithms should be able extract from the text, because they occur in correctly aligned segments, and are tagged and parsed correctly.

2. The other group consists of 334 NPs which it would not be able to extract because of one or a combination of errors in one of the preprocessing steps. Section 5.3 contains a detailed analysis of these errors.

It is important to note that due to these errors, the extraction algorithms will not be able to achieve recall beyond 68%. Nevertheless, the acquisition algorithms, when operating on NPs instead of words or nouns, perform markedly better, cf. figure 6. The recall of both methods is 64%, which is much better than word and noun-based methods. When only taking into account the group of 706 items which didn't have any preprocessing errors, recall is even 94%. Finally, precision again improves considerably by applying position-sensitivity. Section 5.4 discusses attempts to further improve precision.

Figure 6: Phrase-based methods

Position	Recall	Precision
no	64% (94%)	35%
yes	64% (94%)	68%

5.2 Tunability

The threshold is defined in terms of the source language term frequency. As can be expected, a high threshold results in relatively higher precision and relatively lower recall. Figure 7 shows some figures of varying thresholds with the position-sensitive method. As in figure 6, the score in parentheses is the recall score when attention is restricted to the set of 706 NPs. The 50% threshold is the default for the experiments discussed in this paper, cf. the second row of table 6.

The threshold value of our method is a parameter that can be changed, so that an appropriate threshold can be selected, depending on the desired priority of precision and recall.

116

Figure 7: Effects of variation of threshold value

Threshold	Recall	Precision
100%	15% (23%)	100 %
95%	31% (45%)	96%
90%	42% (62%)	88%
75%	54% (79%)	76%
50%	64% (94%)	68%
25%	66% (97%)	64%
10%	66% (97%)	59%

5.3 Analysis of errors affecting recall

The errors can be classified and quantified as follows. There are four classes of technical problems caused by the various preprocessing phases, and two classes of fundamental counter-examples. These are the four classes of errors due to preprocessing.

1. Incorrect alignment of text segments accounts for 6% of the errors.

2. In 15% of the errors part of a term is tagged incorrectly. This is often due to lexicon errors. An incompatibility between lexical classification schemes accounts for another 7% of the errors. The Dutch tagger also has no facility to deal with occasional use of English in Dutch text (4%).

3. The tagger (and its dictionary) currently doesn't recognize multi word units, hence e.g. *with respect to* wrongly yields the term *respect* (6%).

4. In many cases the syntactic structures of the terms in the two languages do not match. This is the main source of errors (47%). The pattern matcher ignores postnominal PP arguments and modifiers in both languages. However, a Dutch postnominal PP argument often maps to the first part of an English noun-noun compound, as in the following example, where *markt* maps to *market* and *versplintering* to *fragmentation*.

> versplintering$_n$ van$_p$ $\leftrightarrow$ market$_n$
> de$_d$ markt$_n$ fragmentation$_n$

The majority of errors (85%) is therefore due to errors in text preprocessing, where there are still many possible improvements. The remaining two classes are fundamental counter-examples.

1. In a number of cases (15%), NPs do not translate to NPs, e.g. the following Dutch sentence contains the equivalent of *careful management*.

> snelle$_a$ maar$_c$ $\leftrightarrow$ needs$_v$
> zorgvuldige$_a$ leiding$_n$ to$_m$ be$_v$ rapid$_a$ but$_c$
> vraagt$_v$ carefully$_{adv}$
>   managed$_v$

2. In two cases (1%), the solution of a genuine ambiguity by the tagger did not correspond to

the interpretation imposed by the translation. In the following example, the deverbal meaning of *vervaardiging* imposes the interpretation of *manufacturing* as a gerund.

> hoofdaccent$_n$ op$_p$ de$_d$ $\leftrightarrow$ main$_a$ emphasis$_n$ on$_p$
> vervaardiging$_n$ van$_p$ manufacturing$_{n/v}$?
> elementen$_n$ elements$_n$

However, these two classes affect only 5% of all terms. The theoretically maximal recall, assuming that the alignment program, tagger and NP parser all perform fully correctly, is 95%. Since the parser is currently extremely simplistic, we expect that major improvements can be readily achieved[6].

5.4 Improving precision

The results in figure 6 and 7 show an important improvement in recall. One factor impeding better precision is the small size of the corpus. In our corpus, 71% of the Dutch NPs is unique in the corpus, and precision suffers from sparsity of data. Still, it is useful to investigate ways to improve precision.

One obvious option we explored was to exploit compositionality in translation. The Dutch terms in figure 8 all contain the 'subterm' *schakelingen*, the English terms the subterm *circuits*. This evident regularity is not exploited by any of the discussed methods. We experimented with an approach where co-occurrence tables are built of terms as well as of heads of terms[7] and where this information is used in the selection and ordering of translations. Surprisingly, this improved results for non-positional methods, but not for positional methods. We do expect these regularities to emerge with much larger corpora.

There are some other possibilities which could be explored. The terms could lemmatized, so that information about inflectional variants can be combined. There may also be a correlation in *length* of terms and their translations. Finally, the alignment program provides a measure of the quality of alignment, which is not yet used by the program.

6 Related Research

In this section we compare our work with two other methods reported on in the literature. In section 6.1 we compare our work to work discussed in [Gaussier et al., 1992], which is based on mutual information. Section 6.2 discusses [Gale and Church, 1991a], which is based on the ϕ^2 statistic.

---

[6]It is conceivable to partly automate the acquisition of the necessary lexical knowledge, viz. determining which nouns are likely to take PP complements, but our corpus is too small for this type of knowledge acquisition.

[7]In fact, it turned out to be better to use final substrings (e.g. six or seven characters) of the head noun of the NP instead of the head itself to avoid the compound problem discussed in section 4.2.

Figure 8: Terms containing *circuits*

geintegreerde opto-electronische schakelin- gen	↔	integrated optoelectric circuits
snelle logische schake- lingen	↔	high speed logic circuits
geintegreerde schakelingen	↔	integrated circuits

A third method to extract bilingual terminology is the use of latent semantic indexing, cf. [Landauer and Littmann, 1990]. Latent semantic indexing is a vector model, where a term-document matrix is transformed to a space of much less dimensions using a technique called singular value decomposition. In the resulting matrix, distributionally similar terms, such as synonyms, are represented by similar vectors. When applied to a collection of documents and their translations, terms will be represented by vectors similar to the representations of their translations. We have not yet compared our method to this approach.

6.1 Mutual information

The selection and ranking method is not based on the concept of mutual information (cf. [Church and Hanks, 1989]), though the technique is quite similar. The mutual information score compares the probability of observing two items together (in aligned segments) to the product of their individual probabilities.

$$I(sl, tl) = \log_2 \frac{P(sl, tl)}{P(sl)P(tl)}$$

The difference is that in our method the global frequency of the source language term is only used in the threshold, and is not used for computing the translational relevance score. Mutual information is used for translation selection and ranking in [Gaussier *et al.*, 1992]. For comparison, the evaluation was repeated using mutual information as selection and ordering criterium. The first two rows in figure 9 show mutual information achieves improved recall when compared to figure 6, but at the expense of reduced precision[8].

In [Gaussier *et al.*, 1992] a filter is used which eliminates all candidate target language terms that do not provide more information on any other source language term. The last two rows in figure 9 show results from our implementation of that technique.

---

[8]It is possible to select only pairs with a mutual information score greater than some minimum value, which reduces recall and improves precision. However, reducing recall to the level in figure 6 still leaves precision at a level much below the precision level given there.

In both cases, the threshold results in a huge improvement of precision, at the expense of recall. The position-sensitive result is comparable to the 90% row in table 7.

Figure 9: Phrase-based methods using mutual information

Position	Filter	Recall	Precision
no	no	66% (98%)	25%
yes	no	66% (98%)	58%
no	yes	55% (82%)	38%
yes	yes	40% (59%)	89%

6.2 The ϕ^2 method

In [Gale and Church, 1991a], another association measure is used, viz. ϕ^2, a χ^2-like statistic. In the following formula, assume a is the co-occurrence frequency of a source language term sl and a target language term tl, b the frequency of sl minus a, c the frequency of tl minus a, and d the number of regions containing neither sl, nor tl.

$$\phi^2 = \frac{(ad - bc)^2}{(a + b)(a + c)(b + d)(c + d)}$$

As in the other methods, the co-occurrence frequency can be modified to reflect position-sensitivity. We incorporated this measure into our system and evaluated the performance. This result is similar to the 25% threshold in figure 7.

Figure 10: Results using ϕ^2-statistic

Position	Recall	Precision
no	66% (97%)	37%
yes	66% (97%)	64%

7 Discussion

In this paper a number of methods to extract bilingual terminology from aligned corpora were discussed. The methods consist of a linguistic term extraction phase and a statistic translation selection phase.

The best term extraction method (in terms of recall) turned out to be a method that defines terms as NPs. NPs are extracted from text using part of speech tagging and pattern matching. Both tagging and NP-extraction can still be improved considerably. Precision is improved by preferring terms at 'similar' positions in target language segments.

The translation selection method selects and orders translations of a term by comparing global and

local frequencies of the target language terms, subject to a threshold condition defined in terms of the frequency of the source language term. The threshold is a parameter which can be used to give priority to precision or recall.

The re-implementation of the algorithms discussed in [Gaussier et al., 1992] and [Gale and Church, 1991a] results in precision/recall figures comparable to our method. It should be noted that these studies establish correspondences between words rather than phrases. We have shown a phrasal approach yields improved recall in the Dutch-English language pair. These studies dealt with an English-French corpus. To some extent, the mismatch due to compounding may be less problematic for this language pair, but the example of the translation of the English expression *House of Commons* to *Chambre des Communes*[9] shows this language pair would also benefit from a phrasal approach. These are lexicalized phrases and are described as such in dictionaries[10].

Another difference is that position-sensitivity in ranking potential translations is not taken advantage of in the earlier proposals. Tables 9 and 10 show these methods also benefit from this extension. Both proposals also have no direct analog to our threshold parameter, which allows for prioritizing precision or recall (cf. section 5.2).

One aspect not covered at all in our proposal is the technical problem of memory requirements which will emerge when using very large corpora. This issue is discussed in [Gale and Church, 1991a]. Future experiments should definitely concentrate on experiments with much larger corpora, because these would allow us to carry out realistic experiments with techniques such as mentioned in section 5.4. We also expect precision to improve in larger corpora, because most NPs are unique in the small corpus we used so far.

Acknowledgements

---

[9]Discussed in [Landauer and Littmann, 1990, page 34] and [Gale and Church, 1991a, page 154].

[10]This example again pinpoints the need for improved NP-recognition, because the PP *of Commons* would not be attached to the NP by the NP rule in section 2.

References

[Brown et al., 1990] P.F. Brown, J. Cocke, S.A. DellaPietra, V.J. DellaPietra, F. Jelinek, J.D. Lafferty, R.L. Mercer, and P.S. Roossin. A statistical approach to machine translation. *Computational Linguistics*, 16:85–97, 1990.

[Brown et al., 1991] P. Brown, J. Lai, and R. Mercer. Aligning sentences in parallel corpora. In *29th Annual Meeting of the Association for Computational Linguistics*, pages 169–176, 1991.

[Catizone et al., 1989] R. Catizone, G. Russel, and S. Warwick. Deriving translation data from bilingual texts. In Uri Zernik, editor, *Proc. of the First Int. Lexical Acquisition Workshop*, Detroit, 1989.

[Church and Hanks, 1989] K. Church and P. Hanks. Word association norms, mutual information, and lexicography. In *27th Annual Meeting of the Association for Computational Linguistics*, pages 76–83, 1989.

[Church, 1988] K. Church. A stochastic parts program and noun phrase parser for unrestricted text. In *2nd Conference on Applied Natural Language Processing (ACL)*, 1988.

[Gale and Church, 1991a] W. Gale and K. Church. Identifying word correspondences in parallel texts. In *4th Darpa Workshop on Speech and Natural Language*, pages 152–157, 1991.

[Gale and Church, 1991b] W. Gale and K. Church. A program for aligning sentences in bilingual corpora. In *29th Annual Meeting of the Association for Computational Linguistics*, pages 177–184, 1991.

[Gale et al., 1992] W. Gale, K. Church, and D. Yarowsky. Using bilingual materials to develop word sense disambiguation methods. In *Fourth International Conference on theoretical and methodological issues in machine translation*, pages 101–112, Montréal, 1992.

[Gaussier et al., 1992] E. Gaussier, J-M Langé, and F. Meunier. Toward bilingual terminology. In *Joint ALLC/ACH Conference*, Oxford, 1992.

[Landauer and Littmann, 1990] T. Landauer and M. Littmann. Fully automatic cross-language document retrieval using latent semantic indexing. In *Proceedings of the 6th Conference of the UW Centre for the New Oxford English Dictionary and Text Research*, pages 31–38, 1990.