Finding Terminology Translations from Non-parallel Corpora

Pascale Fung
Dept. of Electrical and Electronic Engineering
University of Science & Technology (HKUST)
Clear Water Bay, Hong Kong
pascale@ee.ust.hk
Kathleen McKeown
Computer Science Department
Columbia University
New York, NY 10027

Abstract

We present a statistical word feature, the Word Relation Matrix, which can be used to find translated pairs of words and terms from non-parallel corpora, across language groups. Online dictionary entries are used as seed words to generate Word Relation Matrices for the unknown words according to correlation measures. Word Relation Matrices are then mapped across the corpora to find translation pairs. Translation accuracies are around 30% when only the top candidate is counted. Nevertheless, top 20 candidate output give a 50.9% average increase in accuracy on human translator performance.

1 Introduction

Despite a surge in research using parallel corpora for various machine translation tasks (Brown et al. 1993),(Brown et al. 1991; Gale & Church 1993; Church 1993; Dagan & Church 1994; Simard et al. 1992; Chen 1993; Melamed 1995; Wu & Xia 1994; Wu 1994; Smadja et al. 1996), the amount of available bilingual parallel corpora is still relatively small in comparison to the large amount of available monolingual text. It is unlikely that one can find parallel corpora in any given domain in electronic form. This is a particularly acute problem in language pairs such as Chinese/English or Japanese/English where there are fewer translated texts than in European language pairs. While we should make use of any existing parallel corpora as lexical translation resources, we should not ignore the even larger

amount of monolingual text. However, using non-parallel corpora for lexical translation has been a daunting task, considered much more difficult than that with parallel corpora.

In this paper, we present an initial algorithm for translating technical terms using a pair of non-parallel corpora. Evaluation results show translation precisions at around 30% when only the top candidate is considered. While this precision is lower than that achieved with parallel corpora, we show that top 20 candidate output from our algorithm allows translators to increase their accuracy by 50.9%. In the following sections, we first describe a pair of non-parallel corpora we use for experiments, and then we introduce the Word Relation Matrix (WoRM), a statistical word feature representation for technical term translation from non-parallel corpora. We evaluate the effectiveness of this feature with two sets of experiments, using English/English, and English/Japanese non-parallel corpora.

2   Related work

Few attempts have been made to explore non-parallel corpora of monolingual texts in the same domain. Early work uses a pair of non-parallel texts for the task of lexical disambiguation between several senses of a word (Dagan 1990).

This basic idea extends to choosing a translation among multiple candidates (Dagan & Itai 1994) given collocation information. A similar idea is later applied by (Rapp 1995) to show the plausibility of correlations between words in non-parallel text. He proposed a matrix permutation method matching co-occurrence patterns in two non-parallel texts, but noted that computational limitations hamper further extension of this method. Using the same idea, (Tanaka & Iwasaki 1996) demonstrated how to eliminate candidate words in a bilingual dictionary.

All the above works point to a certain discriminatory feature in monolingual texts — context and word relations. However, these works remain in the realm of solving ambiguities or choosing the best candidate among a small set of possibilities. It is argued in (Gale & Church 1994) that feature vectors of 100,000 dimensions are likely to be needed for high resolution discriminant analysis. It is so far questionable whether feature vectors of lower dimensions are discriminating enough for extracting bilingual lexical pairs from non-parallel corpora with a large number of candidates. Is it possible to achieve bilingual lexicon translation by looking at words in relation to other words? In this paper, we hope to shed some light on this question.

3   Two pilot non-parallel corpora

In our experiments, we use two sets of non-parallel corpora: (1) Wall Street Journal (WSJ) from 1993 and 1994, divided into two non-overlapping parts. Each resulting English corpus has 10.36M bytes of data. (2) Wall Street Journal in English and Nikkei Financial News in Japanese, from the same time period. The WSJ text contains 49M bytes of data, and the Nikkei 127M bytes. Since the Nikkei is encoded in two-byte Japanese character sets, the latter is equivalent to about 60M bytes of data in English.

The English Wall Street Journal non-parallel corpus gives us an easier test set on which to start. The output of this corpus should consist of words matching to themselves as trans-

lations. It is useful as a baseline evaluation test set providing an estimate on performance.

The WSJ/Nikkei corpus is the most non-parallel type of corpus. In addition to being written in languages across linguistic families by different journalists, WSJ/Nikkei also share only a limited amount of common topic. The Wall Street Journal tends to focus on U.S. domestic economic and political news, whereas the Nikkei Financial News focuses on economic and political events in Japan and in Asia. Due to the large difference in content, language, writing style, we consider this corpus more difficult than others. However, the result we obtain from this corpus gives us a lower-bound on the performance of our algorithm.

4 An algorithm for finding terminology translations from non-parallel corpora

Bilingual lexicon translation algorithms for parallel corpora in general make use of fixed correlations between a pair of bilingual terms, reflected in their frequent co-occurrences in translated texts, to find lexicon translations. We use correlations both between monolingual lexical units, and between bilingual or multilingual lexical units, to find a consistent pattern which is represented as statistical word features for translation.

We illustrate the possible correlations using the word *debentures* in the two different parts of WSJ. Figure 1 shows segments from both texts containing the word *debentures*.

Figure 1: Part of the concordances of the word *debenture* in WSJ_1 and WSJ_2

Universal said its 15 3/4% debentures *due* Dec
sold $75 *million* of 6% debentures priced at par and *due* Sept
sold $40 *million* of 6 1/4% *convertible* debentures priced at par and *due* March 15
GTE offered a $250 *million* issue of 8 1/2% debentures *due* in 30 years
$250 *million* of notes *due* 1997 and $250 *million* of debentures *due* 2017
sold $300 *million* of 7 1/2% *convertible* debentures *due* 2012 at par
said it agreed to issue $125 *million* Canadian in *convertible* debentures
senior *subordinated* debentures was offered through Drexel
said it completed the redemption of all $16 *million* of its 9% *subordinated* debentures *due* 2003
Moody's assigned a Baa-3 rating to a proposed $100 *million* *convertible* *subordinated* debenture is
and its 12 1/2% senior *subordinated* debentures at par
$20 *million* of *convertible* debentures *due* June 1
issues of $110 *million* of senior notes *due* 1997 and $115 *million* of *convertible* debentures *due*
said it reached an agreement with holders of $30 *million* of its *convertible* *subordinated* debentu
downgraded the *subordinated* debentures of Bank of Montreal
common shares and $35 *million* of *convertible* debentures *due* 2012
$35 *million* of *convertible* debentures *due* May 15
financed with $450 *million* of new Western Union senior secured debentures to be placed by Dre
Commission to issue as much as $125 *million* of 30-year debentures packaged with common stoc
to redeem its entire $55 *million* face amount of 8 3/4% *convertible* *subordinated* debentures *due*

Figure 1 shows that:

1. *debentures* co-occurs most frequently with words such as *million* and *due* in both texts.

2. *debentures* is less related to *engineering*, which does not appear in any segments containing *debentures*.

3. Given all words in one text, *debentures* is closely correlated with a subset of all words in the texts. In Figure 1, this subset consists of *million, due, convertible, subordinated*, etc.

Following the above observations, we propose the following algorithm for finding word or term translation pairs from non-parallel corpora:

1. Given a bilingual list of known translation pairs (i.e. seed words)

2. For every unknown word or term e in language 1, find its *correlation1* with every word in the seed word list in language 1 $\Rightarrow$ relation vector $WoRM1$

3. Similarly for unknown words c in language 2, find its *correlation1* with every word in the seed word list in language 2 $\Rightarrow$ relation vector $WoRM2$

4. Compute *correlation2*$(WoRM1, WoRM2)$; if it is high, e and c are considered as a translation pair.

We use online dictionaries to provide the it seed word lists. To avoid problems of polysemy and nonstandardization in dictionary entries, we choose a more reliable, less ambiguous subset of dictionary entries as the seed word list. This subset contains dictionary entries which occur at mid-range frequency in the corpus so that they are more likely to be content words. They must occur in both sides of the non-parallel corpora, and have fewer number of candidate translations. Such seed words serve as the textual anchor points in non-parallel corpora. For example, we obtained 1,416 entries from the Japanese/English online dictionary EDICT using these criteria.

5 A word in relation to seed words

Word correlations are important statistical information which has been successfully employed to find bilingual word pairs from parallel corpora. Word correlations $W(w_s, w_t)$ are computed from general likelihood scores based on the co-occurrence of words in common segments. Segments are either sentences, paragraphs, or string groups delimited by anchor points:

$$
\begin{aligned}
\Pr(w_s = 1) &= \frac{a + b}{a + b + c + d} \\
\Pr(w_t = 1) &= \frac{a + c}{a + b + c + d} \\
\Pr(w_s = 1, w_t = 1) &= \frac{a}{a + b + c + d}
\end{aligned}
$$

where a = number of segments where both words occur
b = number of segments where only w_s occur
c = number of segments where only w_t occur
d = number of segments where neither words occur

All correlation measures use the above likelihood scores in different formulations. In our Word Relation Matrix (WoRM) representation, we use the correlation measure $W(w_s, w_t)$ between a seed word w_s and an unknown word w_x. a, b, c and d are computed from the segments in the monolingual text of the non-parallel corpus.

$W(w_x, w_s)$ is the weighted mutual information in our algorithm since it is most suitable for lexicon compilation of mid-frequency technical words or terms:

$$\Pr(w_x = 1, w_s = 1) \log_2 \frac{\Pr(w_x = 1, w_s = 1)}{\Pr(w_x = 1)\Pr(w_s = 1)}$$

Given n seed words $(w_{s1}, w_{s2}, \ldots, w_{sn})$, we thus obtain a Word Relation Matrix for w_x:

$$(W(w_x, w_{s1}), W(w_x, w_{s2}), \ldots, W(w_x, w_{sn}))$$

As an initial step, all $\Pr(w_s = 1)$ are pre-computed for the seed words in both languages. We have experimented with various segment sizes, ranging from phrases delimited by all punctuations, a sentence, to an entire paragraph.

From our experiment results, we conclude that the right segment size is a function of the frequency of the seed words:

$$\text{segment size} \propto \frac{1}{\text{frequency}(W_s)}$$

If the seed words are frequent, and if the segment size is as large as a paragraph size, then these frequent seed words could occur in every single segment. In this case, the chances for co-occurrence between such seed words and all new words are very high, close to one. With large segments, such seed words are too biasing and thus, smaller segment size must be used. Conversely, we need a larger segment size if seed word frequency is low.

Consequently, we use the paragraph as the segment size for our experiment on Wall Street Journal/Nikkei Corpus since all the seed words are mid-frequency content words. We computed all binary vectors of the 1,416 seed words w_s where the i-th dimension of the vector is 1 if the seed word occurs in the i-th paragraph in the text, zero otherwise.

We use a smaller segment size – between any two punctuations – for the segment size for the Wall Street Journal English/English corpus since many of the seed words are frequent.

Next, $Pr(w_x = 1)$ is computed for all unknown words x in both texts. The WoRM vectors are then sorted according to $W(w_x, w_{si})$. The most correlated seed word w_{sj} will have the top scoring $W(w_x, w_{sj})$.

As an example, using 307 seed word pairs in the WSJ/WSJ corpus, we obtain the following most correlated seed words with *debentures* in two different years of Wall Street Journal as shown in Figure 2. In both texts, the same set of words correlate with *debenture* closely.

WoRM plots of *debentures* and *administration* are shown in Figures 3 and 4 respectively. The horizontal axis has 307 points representing the seed words, the vertical axis has the value of the correlation scores between these 307 seed words and our example words. These figures show that WoRMs of the same words are similar to each other, but WoRMs are different between different words.

6 Matching Word Relation Matrices

When all unknown words are represented in WoRMs, a matching function is needed to find the best WoRM pairs as bilingual lexicon entries. There are many metrics we can use to measure the closeness of two WoRMs.

When matching vectors are very similar such as those in the WSJ English/English corpus, a simple metric like the Euclidean Distance could be used to find those matching pairs:

$$\mathcal{E} = \sqrt{\Sigma_{1 < i < n}(w_{s_i} - w_{t_i})^2}$$

seed word	corr1(text1)	seed word	corr1(text2)
amount	1083.35	amount	1083.35
July	695.58	offered	646.30
offered	646.30	preferred	551.50
Canadian	596.42	July	695.58
preferred	551.50	June	393.14
June	393.14	exchange	387.16
exchange	387.16	issue	373.80
issue	373.80	notes	229.45
notes	229.45	gas	158.60
gas	158.60	Capital	157.64

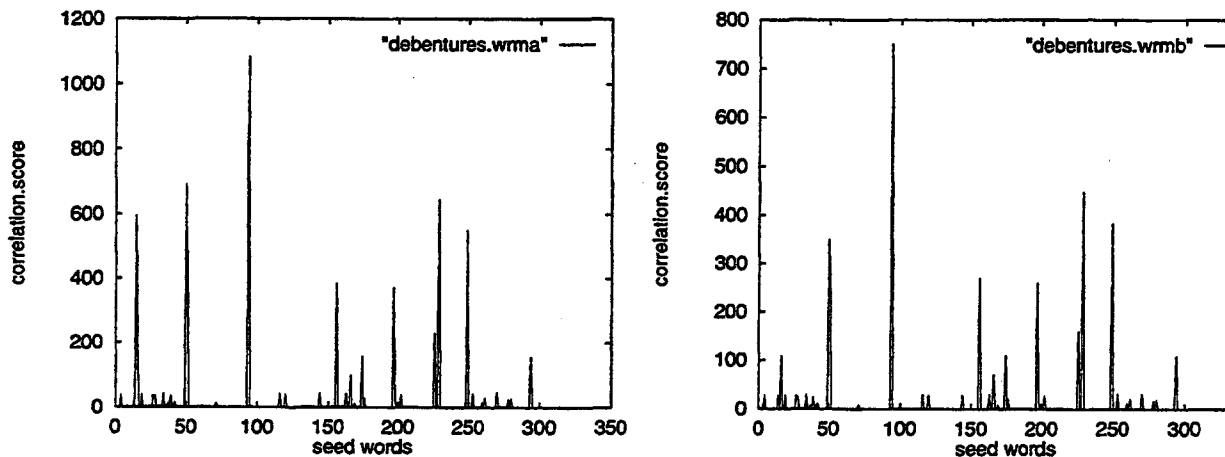Figure 2: Most correlated seed words with *debentures*

Figure 3: Word relation matrix for *debenture* in both texts

However, most word pairs in truly non-parallel bilingual corpus are less similar than those in Figure 3. The y value of a new word is high when there is a x-th seed word which co-occurs with it significantly often. If a pair of bilingual words are supposed to be translations of each other, they should share the most significant y values. In this case, the Cosine Measure would be more appropriate where:

$$C = \frac{\Sigma_{1 < i < n}(w_{s_i} \cdot w_{t_i})}{\sqrt{\Sigma w_{s_i}^2 \cdot \Sigma w_{t_i}^2}}$$

The Cosine Measure will give the highest value to vector pairs which share the most non-zero y values. Therefore, it favors word pairs which share the most number of closely related seed words. However, the Cosine Measure is also directly proportional to another parameter, namely the actual $(w_{s_i} \times w_{t_i})$ values. Consequently, if w_s has a high y value everywhere, then the Cosine Measure between any w_t and this w_s would be high. This violates our assumptions in that although w_s and w_t might not correlate closely with the same set of seed words, the matching score would be nevertheless high. This is another supporting reason for choosing mid-frequency content words as seed words.
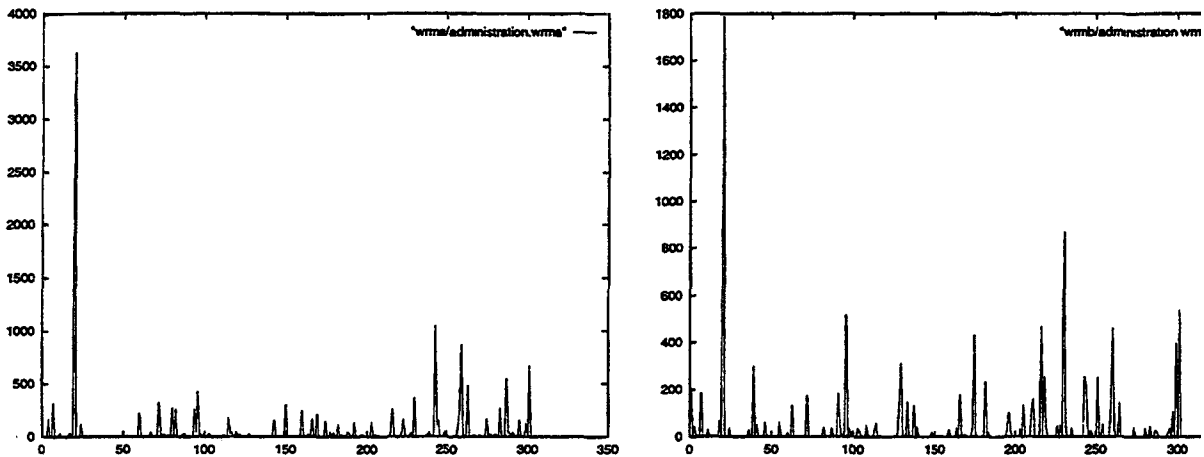
197

Figure 4: Word relation matrix for *administration* in both texts

7 Evaluation 1: Matching English words to English

The evaluation on the WSJ/WSJ English/English corpus is intended as a pilot test on the discriminative power of the Word Relation Matrix. This non-parallel corpus has minimal content and style differences. Furthermore, using such an English/English test set, the output can be evaluated automatically—a translated pair is considered correct if they are identical English words.

307 seed words are chosen according to their occurrence frequency (400-3900) to minimize the number of function words. However, a frequency of 3900 in a corpus of 1.5M words is quite high. As a result, a segment delimited by two punctuations is used as the context window size. Furthermore, the frequent nature of the seed words led to our choice of the Euclidean Distance, instead of the Cosine Measure. The choices of segment size, seed words, and Euclidean Distance measure are all direct consequences of the atypical nature of the English/English pilot test set.

We selected a test set of 582 (set A) by 687 (set B) single words with mid-range frequency from the WSJ texts. We computed the WoRM feature for each of these test words and computed the Euclidean Distance between every word in these sets. We then calculated the accuracy by counting the number of words whose top one candidate is identical to itself, obtaining a precision of 29%.

By allowing N-top candidates, the accuracy improves as shown in the graphs for 582 words output in Figure 5 (i.e. a translation is correct if it appears among the first N candidates). If we find the correct translation among the top 100 candidates, we obtain a precision of around 58%. N-top candidates are useful as translator aids.

Meanwhile, precisions for translating less polysemous content words are higher. If only the 445 content words (manually selected) are kept from the 582-word set, the precisions at different top N candidates for the 445-word set are higher as shown in Figure 5 by the dotted line. We believe the accuracy would be even higher if we only look at really unambiguous test words, such as an entire technical term. It is well known that polysemous words usually have only one sense when used as part of a collocation or technical term (Yarowsky 1993).

8 Evaluation 2: Matching Japanese terms to English

Evaluations are also carried out on the Wall Street Journal and Nikkei Financial News corpus, matching technical terms in Japanese to their counterpart in English. This evaluation is a difficult test case because (1) the two languages, English and Japanese, are across language groups; (2) the
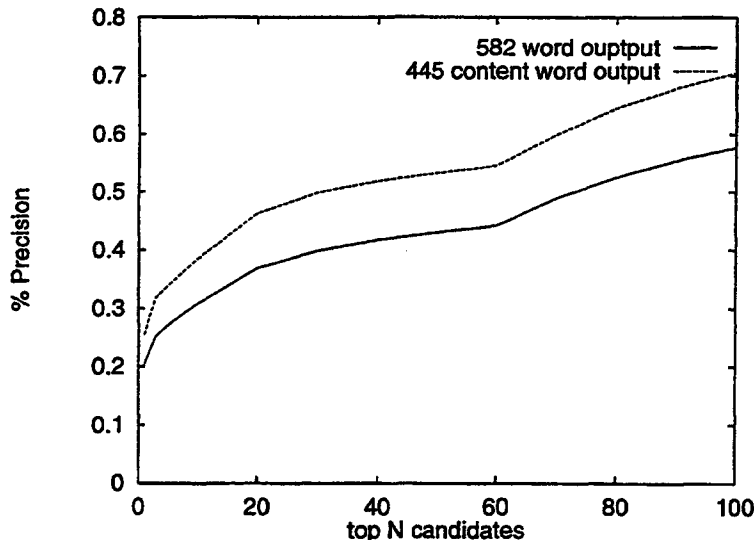
Figure 5: Evaluation results of WoRM in 1993/94 Wall Street Journal

two texts, Wall Street Journal and Nikkei Financial News, do not focus on the same topics; and (3) the two texts are not written by the same authors.

1,416 entries from the Japanese/English online dictionary EDICT with occurrence frequencies between 100 and 1000 are chosen as seed words. Since these seed words have relatively low frequencies compared to the corpus size of around 7 million words for the WSJ text, we chose the segment size to be that of an entire paragraph. For the same reason, the Cosine Measure is chosen as a matching function.

For evaluation, we need to select a test set of known technical term translations. We hand-translated a selected set of technical terms from the Nikkei Financial News corpus and looked them up in the Wall Street Journal text. Among these, 19 terms, shown in Figure 6, have their counterparts in the WSJ text.

Three evaluations were carried out. In all cases, a translation is counted as correct if the top candidate is the right one. Test I tries to find the correct translation for each of the nineteen Japanese terms among the nineteen English terms. To increase the candidate numbers, test II is carried out on 19 Japanese terms with their English counterparts plus 293 other English terms, giving a total of 312 possible English candidates. The third test set III consists of the nineteen Japanese terms paired with their translations and 383 single English words in addition. The accuracies for the three test sets are shown in Figure 7; precision ranges from 21.1% to 52.6%.

Figure 8 shows the ranking of the *true translations* among all the candidates for all 19 cases for the purpose of a translator-aid. Most of the correct translations can be found among the top 20 candidates.

8.1 Translator-aid results

The previous two evaluations show that the precision of best-candidate translation using our algorithm is around 30% on average. While it is far from ideal, this is the first result of terminology translation from non-parallel corpora. Meanwhile, we have found that the correct translation is often among the top 20 candidates. This leads us to conjecture that the output from this algorithm can be used as a translator-aid.

public investment	公共投資
trade negotiation	貿易交渉
nuclear inspection	核査察
price competition	価格競争
cost cut	コスト削減
economic growth	経済成長
U.S.-Japan trade	日米経済
U.S.-Japan trade	日米貿易
economic policy	経済制
NTT	日本電信電話
environmental protection	環境保全
free trade	自由貿易
economic reform	経済改革
NAFTA	北米自由貿易協定
world trade	世界貿易
consumption tax	消費税率
European Union	欧州連合
tax reform	税制改革
credit guarantee	信用保証
budget deficit	財政赤字

Figure 6: The 19 term test set for the WSJ/Nikkei corpus

Test set	I (19)	II (312)	III (402)
Precision of best candidate	10/19=52.6%	4/19=21.1%	6/19=31.6%

Figure 7: Precisions for the best candidate translation in the WSJ/Nikkei corpus

To evaluate this, we again chose the nineteen English/Japanese terms from the WSJ/Nikkei non-parallel corpus as a test set. We chose three evaluators who are all native Chinese speakers with bilingual knowledge in English and Chinese. Chinese speakers are able to recognize most Japanese technical terms since they are very similar to Chinese. We asked them to translate these nineteen Japanese terms into English without using dictionaries or other reference material. The translators have some general knowledge of international news. However, none of them specializes in economics or finance, which is the domain of the WSJ/Nikkei corpus. Their output is in SET A. Our system then proposes two sets of outputs: (1) for each Japanese term, our system proposes the top-20 candidates from the set of 312 noun phrases. Using this candidate list, the translators again translate the nineteen terms. Their output based on this information is in SET B; (2) for each Japanese term, our system proposes the top-20 candidates from the set containing 383 single words plus the nineteen terms. The result of human translation based on this candidate list is in SET C. Sets A, B and C are all compared to the original translation in the corpus. If the translation is the same as in the corpus, then it is judged as correct. The results are shown in Figure 9. Evaluators on average are able to translate 8 terms out of 19 by themselves, whereas they can translate 18 terms on average with the aid of our output. Translation precision increases on the average by 50.9%.

9 Conclusion

We have described a statistical word signature feature, the Word Relation Matrix, that can be used to find matching pairs of content words or terms in a pair of same-domain non-parallel bilingual

English translation	rank in test I (19)	rank in test II (312)	rank in test III (402)
public investment	14	128	61
trade negotiation	4	8	5
nuclear inspection	12	139	76
price competition	11	54	16
cost cut	2	2	2
economic growth	2	2	2
U.S.-Japan trade	4	17	5
U.S.-Japan trade	4	17	5
economic policy	1	1	1
NTT	4	18	5
environmental protection	4	17	5
free trade	4	16	5
economic reform	12	139	83
NAFTA	1	1	1
world trade	1	4	1
consumption tax	1	4	1
European Union	1	1	1
tax reform	1	1	1
credit guarantee	3	11	4
budget deficit	1	1	2

Figure 8: Rank of the correct translations for WSJ/Nikkei evaluations

Translator	SET A	without aid	SET B		SET C		improvement
A	10/19	52.6%	18/19	94.7%	18/19	94.7%	42.1%
B	7/19	36.8%	17/19	89.5%	18/19	94.7%	57.9%
C	7/19	36.8%	17/19	89.5%	17/19	89.5%	52.7%
Average	8/19	42.1%	17.3/19	91.2%	17.7/19	93.0%	50.9%

Figure 9: Translator improvement on term translation

texts. Evaluation shows a precision of about 30%. We showed that humans are able to translate more than twice as many Japanese technical terms into English when our system output is used, compared to translating a random set of 19 Japanese terms without aid. It is also a significant initial result for lexical translation from truly non-parallel corpora, particularly across language groups.

For future work, the quality of seed words can be improved by using a training algorithm to select seed words according to their discriminative power. The dimensionality of WoRM vectors we have chosen is not optimal. A high dimensionality of vectors is usually favorable (Gale & Church 1994). On the other hand, high dimensionality can also lead to noise, Therefore, dimensionality reduction methods such as the Singular Value Decomposition (Shütze 1992) or clustering is often used. In our case, this means that we should choose a large subset of highly discriminative seed word pairs. Additionally, the Word Relation Matrix could be used in combination with other word signature features for non-parallel corpora.

In addition to the evaluation results, we have also discovered that the content words in the same segment with a word or term all contribute to the occurrence of this word. This feature represents

some of the long-distance relations between the word and multiple other words which are not its immediate neighbors. The information can be used in language modeling in addition to the currently popular N-gram models and word trigger pairs.

References

BROWN, P., J. LAI, & R. MERCER. 1991. Aligning sentences in parallel corpora. In *Proceedings of the 29th Annual Conference of the Association for Computational Linguistics*.

BROWN, P.F., S.A. DELLA PIETRA, V.J. DELLA PIETRA, & R.L. MERCER. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

CHEN, STANLEY. 1993. Aligning sentences in bilingual corpora using lexical information. In *Proceedings of the 31st Annual Conference of the Association for Computational Linguistics*, 9–16, Columbus, Ohio.

CHURCH, KENNETH. 1993. Char_align: A program for aligning parallel texts at the character level. In *Proceedings of the 31st Annual Conference of the Association for Computational Linguistics*, 1–8, Columbus, Ohio.

DAGAN, IDO. 1990. Two languages are better than one. In *Proceedings of the 28th Annual Conference of the Association for Computational Linguistics*, Berkeley, California.

DAGAN, IDO & KENNETH W. CHURCH. 1994. Termight: Identifying and translating technical terminology. In *Proceedings of the 4th Conference on Applied Natural Language Processing*, 34–40, Stuttgart, Germany.

DAGAN, IDO & ALON ITAI. 1994. Word sense disambiguation using a second language monolingual corpus. In *Computational Linguistics*, 564–596.

GALE, WILLIAM A. & KENNETH W. CHURCH. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102.

GALE, WILLIAM A. & KENNETH W. CHURCH. 1994. Discrimination decisions in 100,000 dimensional spaces. *Current Issues in Computational Linguisitcs: In honour of Don Walker*, 429–550.

MELAMED, I. DAN. 1995. Automatic evaluation and uniform filter cascades for inducing N-best translation lexicons. In *Proceedings of the 3rd Annual Workshop on Very Large Corpora*, Boston, Massachusettes.

RAPP, REINHARD. 1995. Identifying word translations in non-parallel texts. In *Proceedings of the 35th Conference of the Association of Computational Linguistics, student session*, 321–322, Boston, Mass.

SHÜTZE, HINRICH. 1992. Dimensions of meaning. In *Proceedings of Supercomputing '92*.

SIMARD, M., G FOSTER, & P. ISABELLE. 1992. Using cognates to align sentences in bilingual corpora. In *Proceedings of the Forth International Conference on Theoretical and Methodological Issues in Machine Translation*, Montreal, Canada.

SMADJA, FRANK, KATHLEEN MCKEOWN, & VASILEIOS HATZSIVASSILOGLOU. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, 21(4):1–38.

TANAKA, KUMIKO & HIDEYA IWASAKI. 1996. Extraction of lexical translations from non-aligned corpora. In *Proceedings of COLING 96*, Copenhagan, Danmark.

WU, DEKAI. 1994. Aligning a parallel English-Chinese corpus statistically with lexical criteria. In *Proceedings of the 32nd Annual Conference of the Association for Computational Linguistics*, 80–87, Las Cruces, New Mexico.

WU, DEKAI & XUANYIN XIA. 1994. Learning an English-Chinese lexicon from a parallel corpus. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, 206–213, Columbia, Maryland.

YAROWSKY, D. 1993. One sense per collocation. In *Proceedings of ARPA Human Languag Technology Workshop*.