Towards a Simple and Accurate Statistical Approach to Learning Translation Relationships among Words

Robert C. Moore[*]
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
bobmoore@microsoft.com

Abstract

We report on a project to derive word translation relationships automatically from parallel corpora. Our effort is distinguished by the use of simpler, faster models than those used in previous high-accuracy approaches. Our methods achieve accuracy on single-word translations that seems comparable to any work previously reported, up to nearly 60% coverage of word types, and they perform particularly well on a class of multi-word compounds of special interest to our translation effort.

1 Introduction

This paper is a report on work in progress aimed at learning word translation relationships automatically from parallel bilingual corpora. Our effort is distinguished by the use of simple statistical models that are easier to implement and faster to run than previous high-accuracy approaches to this problem.

Our overall approach to machine translation is a deep-transfer approach in which the transfer relationships are learned from a parallel bilingual corpus (Richardson et al., 2001). More specifically, the transfer component is trained by parsing both sides of the corpus to produce parallel logical forms, using lexicons and analysis grammars constructed by linguists. The parallel log-ical forms are then aligned at the level of content word stems (lemmas), and logical-form transfer patterns are learned from the aligned logical-form corpus. At run time, the source language text is parsed into logical forms employing the source language grammar and lexicon used in constructing the logical-form training corpus, and the logical-form transfer patterns are used to construct target language logical forms. These logical forms are transformed into target language strings using the target-language lexicon, and a generation grammar written by a linguist.

The principal roles played by the translation relationships derived by the methods discussed in this paper are to provide correspondences between content word lemmas in logical forms to assist in the alignment process, and to augment the lexicons used in parsing and generation, for a special case described in Section 4.

2 Previous Work

The most common approach to deriving translation lexicons from empirical data (Catizone, Russell, and Warwick, 1989; Gale and Church, 1991; Fung, 1995; Kumano and Hirakawa, 1994; Wu and Xia, 1994; Melamed, 1995) is to use some variant of the following procedure:[1]

- Pick a good measure of the degree of association between words in language L_1 and words in language L_2 in aligned sentences of a parallel bilingual corpus.

[1]The important work of Brown et al. (1993) is not directly comparable, since their globally-optimized generative probabilistic model of translation never has to make a firm commitment as to what can or cannot be a translation pair. They assign some nonzero probability to every possible translation pair.

- Rank order pairs consisting of a word from L_1 and a word from L_2 according to the measure of association.

- Choose a threshold, and add to the translation lexicon all pairs of words whose degree of association is above the threshold.

As Melamed later (1996, 2000) pointed out, however, this technique is hampered by the existence of indirect associations between words that are not mutual translations. For example, in our parallel French-English corpus (consisting primarily of translated computer software manuals), two of the most strongly associated word lemma translation pairs are *fichier/file* and *système/system*. However, because the monolingual collocations *système de fichiers*, *fichiers système*, *file system*, and *system files* are so common, the spurious translation pairs *fichier/system* and *système/file* also receive rather high association scores—higher in fact that such true translation pairs as *confiance/trust*, *parallélisme/parallelism*, and *film/movie*.

Melamed's solution to this problem is not to regard highly-associated word pairs as translations in sentences in which there are even more highly-associated pairs involving one or both of the same words. Since indirect associations are normally weaker than direct ones, this usually succeeds in selecting true translation pairs over the spurious ones. For example, in parallel sentences containing *fichier* and *système* on the French side and *file* and *system* on the English side, the associations of *fichier/system* and *système/file* will be discounted, because the degrees of association for *fichier/file* and *système/system* are so much higher.

Melamed's results using this approach extend the range of high-accuracy output to much higher coverage levels than previously reported. Our basic method is rooted the same insight regarding competing associations for the same words, but we embody it in simpler model that is easier to implement and, we believe, faster to run.[2] As we will see below, our model yields results that seem comparable to Melamed's up to nearly 60% coverage of the lexicon.

A second important issue regarding automatic derivation of translation relationships is the assumption implicit (or explicit) in most previous work that lexical translation relationships involve only single words. This is manifestly not the case, as is shown by the following list of translation pairs selected from our corpus:

> *base_de_données/database*
> *mot_de_passe/password*
> *sauvegarder/back_up*
> *annuler/roll_back*
> *ouvrir_session/log_on*

Some of the most sophisticated work on this aspect of problem again seems to be that of Melamed (1997). Our approach in this case is quite different from Melamed's. It is more general in that it can propose compounds that are discontiguous in the training text, as *rolL_back* would be in a phrase such as *roll the failed transaction back*. Melamed does allow skipping over one or two function words, but our basic method is not limited at all by word adjacency. Also, our approach is again much simpler computationally than Melamed's and apparently runs orders of magnitude faster.[3]

3 Our Basic Method

Our basic method for deriving translation pairs consists of the following steps:

1. Extract word lemmas from the logical forms produced by parsing the raw training data.

2. Compute association scores for individual lemmas.

3. Hypothesize occurrences of compounds in the training data, replacing lemmas constituting hypothesized occurrences of a compound with a single token representing the compound.

---

[2]Melamed does not report computation time for the version of his approach without generation of compounds, but our approach omits a number of computationally very expensive steps performed in his approach.

[3]Melamed reports that training on 13 million words took over 800 hours in Perl on a 167-MHz UltraSPARC processor. Training our method on 6.6 million words took approximately 0.5 hours in Perl on a 1-GHz Pentium III processor. Even allowing an order of magnitude for the differences in processor speed and amount of data, there seems to be a difference between the two methods of at least two orders of magnitude in computation required. Unfortunately Melamed evaluates accuracy in his work on translation compounds differently from his work on single-word translation pairs, so we are not able to compare our method to his in that regard.

4. Recompute association scores for compounds and remaining individual lemmas.

5. Recompute association scores, taking into account only co-occurrences such that there is no equally strong or stronger association for either item in the aligned logical-form pair.

We describe each of these steps in detail below.

3.1 Extracting word lemmas

In Step 1, we simply collect, for each sentence, the word lemmas identified by our MT system parser as the key content items in the logical form. These are predominantly morphologically analyzed word stems, omitting most function words. In addition, however, the parser treats certain lexical compounds as if they were single units. These include multi-word expressions placed in the lexicon because they have a specific meaning or use, plus a number of general categories including proper names, names of places, time expressions, dates, measure expressions, etc. We will refer to all of these generically as "multiwords".

The existence of multiwords simplifies learning some translation relationships, but makes others more complicated. For example, we do not, in fact, have to learn *base_de_données* as a compound translation for *database*, because it is extracted from the French logical forms already identified as a single unit. Thus we only need to learn the *base_de_données/database* correspondence as a simple one-to-one mapping. On the other hand, the *disque_dur/hard_disk* correspondence is learned as two-to-one relationship independently of *disque/disk* and *dur/hard* (which are also learned) because *hard_disk* appears as a multiword in our English logical forms, but *disque* and *dur* always appear as separate tokens in our French logical forms.

3.2 Computing association scores

For Step 2, we compute the degree of association between a lemma w_{L_1} and a lemma w_{L_2} in terms of the frequencies with which w_{L_1} occurs in sentences of the L_1 part of the training corpus and w_{L_2} occurs in sentences of the L_2 part of the training corpus, compared to the frequency with which w_{L_1} and w_{L_2} co-occur in aligned sentences of the training corpus. For this purpose, we ignore multiple occurrences of a lemma in a single sentence.

As a measure of association, we use the log-likelihood-ratio statistic recommended by Dunning (1993), which is the same statistic used by Melamed to initialize his models. This statistic gives a measure of the likelihood that two samples are not generated by the same probability distribution. We use it to compare the overall frequency of w_{L_1} in our training data to the frequency of w_{L_1} given w_{L_2} (i.e., the frequency with which w_{L_1} occurs in sentences of L_1 that are aligned with sentences of L_2 in which w_{L_2} occurs). Since $p(w_{L_1}) = p(w_{L_1}|w_{L_2})$ only if occurrences of w_{L_1} and w_{L_2} are independent, a measure of the likelihood that these distributions are different is, therefore, a measure of the likelihood that an observed positive association between w_{L_1} and w_{L_2} is not accidental.

Since this process generates association scores for a huge number of lemma pairs for a large training corpus, we prune the set to restrict our consideration to those pairs having at least some chance of being considered as translation pairs. We heuristically set this threshold to be the degree of association of a pair of lemmas that have one co-occurrence, plus one other occurrence each.

3.3 Hypothesizing compounds and recomputing association scores

If our data were very clean and all translations were one-to-one, we would expect that in most aligned sentence pairs, each word or lemma would be most strongly associated with its translation in that sentence pair; since, as Melamed has argued, direct associations should be stronger than indirect ones. Since translation is symmetric, we would expect that if w_{L_1} is most strongly associated with w_{L_2}, w_{L_2} would be most strongly associated with w_{L_1}. Violations of this pattern are suggestive of translation relationships involving compounds. Thus, if we have a pair of aligned sentences in which *password* occurs in the English sentence and *mot de passe* occurs in the French side, we should not be surprised if *mot* and *passe* are both most strongly associated with *password* within this sentence pair. *Password*, however, cannot be most strongly associated with both *mot* and *passe*.

Our method carrying out Step 3 is based on

finding violations of the condition that whenever w_{L_1} is most strongly associated with w_{L_2}, w_{L_2} is most strongly associated with w_{L_1}. The method is easiest to explain in graph-theoretic terms. Let the nodes of a graph consist of all the lemmas of L_1 and L_2 in a pair of aligned sentences. For each lemma, add a link to the uniquely most strongly associated lemma of the other language.[4] Consider the maximal, connected subgraphs of the resulting graph. If all translations within the sentence pair are one-to-one, each of these subgraphs should contain exactly two lemmas, one from L_1 and one from L_2. For every subgraph containing more than two lemmas of one of the languages, we consider all the lemmas of that language in the subgraph to form a compound. In the case of *mot*, *passe*, and *password*, as described above, there would be a connected subgraph containing these three lemmas; so the two French lemmas, *mot* and *passe*, would be considered to form a compound in the French sentence under consideration.

The output of this step of our process is a transformed set of lemmas for each sentence in the corpus. For each sentence and each subset of the lemmas in that sentence that has been hypothesized to form a compound in the sentence, we replace those lemmas with a token representing them as a single unit. Note that this process works on a sentence-pair by sentence-pair basis, so that a compound hypothesized for one sentence pair may not be hypothesized for a different sentence pair, if the pattern of strongest associations for the two sentence pairs differ. Order of occurrence is not considered in forming these compounds, and the same token is always used to represent the same set of lemmas.[5]

Once the sets of lemmas for the training corpus have been reformulated in terms of the hypothesized compounds, Step 4 consists simply in repeating step 2 on the reformulated training data.

---

[4]Because the data becomes quite noisy if a lemma has no lemmas in the other language that are very strongly associated with it, we place a heuristically chosen threshold on the minimum degree of association that is allowed to produce a link.

[5]The surface order is not needed by the alignment procedure intended to make use of the translation relationships we discover.

3.4   Recomputing association scores, taking into account only strongest associations

If Steps 1–4 worked perfectly, we would have correctly identified all the compounds needed for translation and reformulated the training data to treat each such compound as a single item. At this point, we should be able to treat the training data as if all translations are one-to-one. We therefore choose our final set of ranked translation pairs on the assumption that true translation pairs will always be mutually most strongly associated in a given aligned sentence pair.

Step 5 thus proceeds exactly as step 4, except that w_{L_1} and w_{L_2} are considered to have a joint occurrence only if w_{L_1} is uniquely most strongly associated with w_{L_2}, and w_{L_2} is uniquely most strongly associated with w_{L_1}, among the lemmas (or compound lemmas) present in a given aligned sentence pair. (The associations computed by the previous step are used to make these decisions.) This final set of associations is then sorted in decreasing order of strength of association.

4   Identifying Translations of "Captoids"

In addition to using these techniques to provide translation relationships to the logical-form alignment process, we have applied related methods to a problem that arises in parsing the raw input text. Often in text–particularly the kind of technical text we are experimenting with–phrases are used, not in their usual way, but as the name of something in the domain. Consider, *Click to remove the View As Web Page check mark*. In this sentence, *View As Web Page* has the syntactic form of a nonfinite verb phrase, but it is used as if it is a proper name. If the parser does not recognize this special use, it is virtually impossible to parse the sentence correctly.

Expressions of this type are fairly easily handled by our English parser, however, because capitalization conventions in English make them easy to recognize. The tokenizer used to prepare sentences for parsing, under certain conditions, hypothesizes that sequences of capitalized words such as *View As Web Page* should be treated as lexicalized multi-word expressions, as discussed in Section 3.1. We refer to this subclass of multiwords as "captoids". The capitalization conventions of French (or Spanish) make it harder to rec-

ognize such expressions, however, because typically only the first word of such an expression is capitalized.

We have adapted the methods described in Section 3 to address this problem by finding sequences of French words that are highly associated with English captoids. The sequences of French words that we find are then added to the French lexicon as multiwords.

The procedure for identifying translations of captiods is as follows:

1. Tokenize the training data to separate words from punctuation and identify multiwords wherever possible.

2. Compute association scores for items in the tokenized data.

3. Hypothesize sequences of French words as compounds corresponding to English multiwords, replacing hypothesized occurrences of a compound in the training data with a single token representing the compound.

4. Recompute association scores for pairs of items where either the English item or the French item is a multiword beginning with a capital letter.

5. Filter the resulting list to include only translation pairs such that there is no equally strong or stronger association for either item in the training data.

There are a number of key differences from our previous procedure. First, since this process is meant to provide input to parsing, it works on tokenized word sequences rather than lemmas extracted from logical forms. Because many of the English multiwords are so rare that associations for the entire multiword are rather weak, in Step 2 we count occurrences of the constituent words contained in multiwords as well as occurrences of the multiwords themselves. Thus an occurrence of *View_As_Web_Page* would also count as an occurrence of *view*, *as*, *web*, and *page*.[6]

The method of hypothesizing compounds in Step 3 adds a number of special features to improve accuracy and coverage. Since we know

[6]In identifying captoid translations, we ignore case differences for computing and using association scores.

we are trying to find French translations for English captoids, we look for compounds only in the French data. If any of the association scores between a French word and the constituent words of an English multiword are higher than the association score between the French word and the entire multiword, we use the highest such score to represent the degree of association between the French word and the English multiword. We reserve, for consideration as the basis of compounds, only sets of French words that are most strongly associated in a particular aligned sentence pair with an English multiword that starts with a capitalized word.

Finally we scan the French sentence of the aligned pair from left to right, looking for a capitalized word that is a member of one of the compound-defining sets for the pair. When we find such a word, we begin constructing a French multiword. We continue scanning to the right to find other members of the compound-defining set, allowing up to two consecutive words not in the set, provided that another word in the set immediately follows, in order to account for French function words than might not have high associations with anything in the English multiword. We stop adding to the French multiword once we have found all the French words in the compound-defining set, or if we encounter a punctuation symbol, or if we encounter three or more consecutive words not in the set. If either of the latter two conditions occurs before exhausting the compound-defining set, we assume that the remaining members of the set represent spurious associations and we leave them out of the French multiword.

The restriction in Step 4 to consider only associations in which one of the items is a mutiword beginning with a capital letter is simply for efficiency, since from this point onward no other associations are of interest.

The final filter applied in Step 5 is more stringent than in our basic method. The reasoning is that, while a single word may have more than one translation in different contexts, the sort of complex multiword represented by a captoid would normally be expected to receive the same translation in all contexts. Therefore we accept only translations involving captoids that are mutually uniquely most strongly associated across the en-

Type Coverage	Mean Count	Token Coverage	Total Accuracy	Single-Word Accuracy	Multiword Accuracy	Compound Accuracy
0.040	1247.23	0.859	0.902–0.920	0.927–0.934	0.900–0.900	0.615–0.769
0.080	670.88	0.923	0.842–0.870	0.922–0.939	0.879–0.879	0.453–0.547
0.120	457.79	0.945	0.801–0.834	0.908–0.924	0.734–0.766	0.452–0.548
0.160	347.58	0.957	0.783–0.820	0.898–0.913	0.705–0.737	0.455–0.562
0.200	280.17	0.964	0.762–0.797	0.893–0.911	0.638–0.688	0.449–0.527
0.240	234.63	0.969	0.749–0.783	0.887–0.904	0.606–0.658	0.431–0.505
0.280	201.89	0.973	0.728–0.767	0.878–0.898	0.575–0.637	0.411–0.487
0.320	177.11	0.975	0.712–0.752	0.875–0.896	0.577–0.643	0.375–0.449
0.360	158.08	0.979	0.668–0.710	0.860–0.884	0.511–0.578	0.340–0.405
0.400	142.45	0.980	0.654–0.696	0.845–0.871	0.486–0.556	0.329–0.391
0.440	129.60	0.981	0.637–0.677	0.844–0.869	0.485–0.550	0.298–0.354
0.480	118.90	0.982	0.641–0.680	0.848–0.872	0.502–0.566	0.297–0.351
0.520	109.83	0.983	0.643–0.681	0.852–0.875	0.511–0.574	0.291–0.344
0.560	102.15	0.984	0.626–0.664	0.839–0.864	0.503–0.564	0.279–0.329
0.600	95.50	0.986	0.596–0.636	0.823–0.852	0.484–0.541	0.255–0.305
0.632	90.87	0.989	0.550–0.595	0.784–0.819	0.429–0.488	0.232–0.286

Table 1: Results for basic method.

tire corpus. To focus on the cases we are most interested in and to increase accuracy, we require each translation pair generated to satisfy the following additional conditions:

- The French item must be one of the mulitwords we constructed.

- The English item must be a multiword, all of whose constituent words are capitalized.

- The French item must contain at least as many words as the English item.

The last condition corrects some errors made by allowing highly associated French words to be left out of the hypothesized compounds.

5 Experimental Results

Our basic method for finding translation pairs was applied to a set of approximately 200,000 French and English aligned sentence pairs, derived mainly from Microsoft technical manuals, resulting in 46,599 potential translation pairs. The top 42,486 pairs were incorporated in the alignment lexicon of our end-to-end translation system.[7] The procedure for finding translations of captoids was applied to a slight superset of the training data for the basic procedure, and yielded 2561 possible translation pairs. All of these were added to our end-to-end translation system, with

---

[7]As of this writing, however, the alignment procedure does not yet make use of the general translation pairs involving compounds, although it does make use of the captoid translation compounds.

the French multiwords being added to the lexicon of the French parser, and the translation pairs being added to the alignment lexicon.

The improvements in end-to-end performance due to these additions in a French-to-English translation task are described elsewhere (Pinkham and Corston-Oliver, 2001). For this report, we have evaluated our techniques for finding translation pairs by soliciting judgements of translation correctness from fluent French-English bilinguals. There were too many translation pairs to obtain judgements on each one, so we randomly selected about 10% of the 42,486 general translation pairs that were actually added to the system, and about 25% of the 2561 captoid pairs.

The accuracy of the most strongly associated translation pairs produced by the basic method at various levels of coverage is displayed in Table 1. We use the terms "coverage" and "accuracy" in essentially the same way as Melamed (1996, 2000). "Type coverage" means the proportion of distinct lexical types in the entire training corpus, including both French and English, for which there is at least one translation given. As with the comparable results reported by Melamed, these are predominantly single lemmas for content words, but we also include occurrences multiwords as distinct types. "Mean count" is the average number of occurrences of each type at the given level of coverage. "Token coverage" is the proportion of the total number of occurrences of items in the text represented by the types included within the type coverage.

Type Coverage	Mean Count	Token Coverage	Single-Word Accuracy
0.040	1628.57	0.791	0.948–0.948
0.080	909.48	0.884	0.938–0.942
0.120	626.84	0.914	0.926–0.943
0.160	480.50	0.934	0.909–0.924
0.200	389.11	0.945	0.896–0.912
0.240	327.03	0.953	0.891–0.910
0.280	281.76	0.958	0.896–0.913
0.320	247.67	0.963	0.876–0.896
0.360	220.62	0.965	0.876–0.898
0.400	199.42	0.969	0.864–0.887
0.440	181.69	0.971	0.846–0.872
0.480	166.64	0.971	0.843–0.868
0.520	153.90	0.972	0.848–0.872
0.560	143.22	0.974	0.844–0.868
0.600	133.87	0.976	0.830–0.859
0.636	127.46	0.984	0.784–0.819

Table 2: Results for single words only.

Type Coverage	Mean Count	Token Coverage	Captoid Accuracy
0.020	50.39	0.149	0.913–0.913
0.040	30.19	0.178	0.902–0.902
0.060	21.67	0.192	0.914–0.914
0.080	17.88	0.211	0.911–0.915
0.100	14.79	0.218	0.901–0.904
0.120	12.61	0.223	0.859–0.864
0.140	11.06	0.228	0.860–0.864
0.160	9.95	0.235	0.858–0.862
0.180	9.04	0.240	0.846–0.851
0.194	8.70	0.249	0.841–0.847

Table 3: Results for captoids.

Since the judges were asked to evaluate the proposed translations out of context, we allowed them to give an answer of "not sure", as well as "correct" and "incorrect". Our accuracy scores are therefore given as a range, where the low score combines answers of "not sure" and "incorrect", and the high score combines answers of "not sure" and "correct".

The "total accuracy" column gives results at different levels of coverage over all the translation pairs generated by our basic method. For a more detailed analysis, the remaining columns provide a breakdown for single-word translations, translations involving multiwords given to us by the parser ("multiword accuracy"), and new multiwords hypothesized by our procedure ("compound accuracy"). As the table shows, our performance is quite good on single-word translations, with accuracy of around 80% even at our cut-off of 63% type coverage, which represents 99% of the tokens in the corpus.

To compare our results more directly with Melamed's published results on single-word translation, we show Table 2, where both coverage and accuracy are given for single-word translations only. According to the standard of correctness Melamed uses that is closest to ours, he reports 92% accuracy at 36% type coverage, 89% accuracy at 46% type coverage, and 87% accuracy at 90% type coverage, on a set of 300,000 aligned sentence pairs from the French-English Hansard corpus of Candian Parliament proceedings. Our accuracies at the first two of these coverage points are 88–90% and 84–87%, which is slightly lower than Melamed, but given the different corpus, different judges, and different evaluation conditions, one cannot draw any definite conclusions about which method is more accurate at these coverage levels. Our method, however, does not produce any result approaching 90% type coverage, and accuracy appears to start dropping rapidly below 56% type coverage. Nevertheless, this still represents good accuracy up to 97% token coverage.

Returning to Table 1, we see that our accuracy on multiwords is much lower than on single words, especially the multiwords hypothesized by our learning procedure. The results are much better, however, when we look at the results for our specialized method for finding translations of captoids, as shown in Table 3. Our accuracy at nearly 20% type coverage is around 84%, which is higher than our accuracy for general translation pairs (76–80%) at the same type coverage level. It is lower than our single-word translation accuracy (90–91%) at this coverage level, but it is striking how close it is, given far less data. At 20% type coverage of single words, there are 389 tokens per word type, while at 20% type coverage of captoids, there are fewer than 9 tokens per captoid type. In fact, further analysis shows that of the 2561 captoid translation pairs, 947 have only a single example of the English captoid in the training data, yet our accuracy on these is around 82%. We note, however, that our captoid learning procedure cuts off at around 20% type coverage, which is only 25% token coverage for these items.

6 Conclusions

We have evaluated our approach and found it to be comparable in accuracy on single-word translations to Melamed's results (which appear to be

the best previous results, as far as one can tell given the lack of standard test corpora) up to nearly 60% type coverage and 97% token coverage. Space does not permit a detailed comparison of Melamed's methods to ours, but we repeat that ours are far simpler to implement and much faster to run. Our approach to generating translations involving muti-word compounds performs less well in general, but the special-case modification of it to deal with captoids performs with very high accuracy for those captoids it is able to find a translation for. Based on these results, the focus of our future work will be to try to extend our region of high-accuracy single-word translation to higher levels of coverage, improve the accuracy of our general method for finding multiword translations, and extend the coverage of our method for translating captoids.

References

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2):263–311.

R. Catizone, G. Russell, and S. Warwick. 1989. Deriving translation data from bilingual texts. In *Proceedings of the First International Lexical Acquisition Workshop*, Detroit, MI.

T. Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

P Fung. 1995. A pattern matching method for finding noun and proper noun translations from noisy parallel corpora, In *Proceedings of the 33rd Annual Meeting*, pages 236–243, Boston, MA. Association for Computational Linguistics.

W. Gale and K. Church . 1991. Identifying owrd correspondences in parallel texts. In *Proceedings Speech and Natural Language Workshop*, pages 152–157, Asilomar, CA. DARPA.

A. Kumano and H. Hirakawa. 1994. Building an MT dictionary from parallel texts based on linguistic and statistical information. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 76–81, Kyoto, Japan.

I. D. Melamed. 1995. Automatic evaluation and uniform filter cascades for inducing N-best translation lexicons. In *Proceedings of the Third Workshop on Very Large Corpora*, pages 184–198, Cambridge, MA.

I. D. Melamed. 1996. Automatic construction of clean broad coverage translation lexicons. In *Proceedings of the 2nd Conference of the Association for Machine Translation in the Americas*, pages 125-134, Montreal, Canada.

I. D. Melamed. 1997. Automatic discovery of non-compositional compounds in parallel data. In *Proceedings of the 2nd Conference on Enpirical Methods in Natural Language Processing (EMNLP '97)*, Providence, RI.

I. D. Melamed. 2000. Models of Translational Equivalence. *Computational Linguistics*, 26(2):221–249.

J. Pinkham and M. Corston-Oliver. 2001. Adding Domain Specificity to an MT System. In *Proceedings of the Workshop on Data-Driven Machine Translation*, 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France.

S. Richardson, W. B. Dolan, M. Corston-Oliver, and A. Menezes. 2001. Overcoming the customization bottleneck using example-based MT. In *Proceedings of the Workshop on Data-Driven Machine Translation*, 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France.

D. Wu and X. Xia. 1994. Learning an English-Chinese lexicon from a parallel corpus. In *Proceedings of the 1st Conference of the Association for Machine Translation in the Americas*, pages 206–213, Columbia, MD.