

Active Semi-Supervised Learning for Improving Word Alignment

Vamshi Ambati, Stephan Vogel and Jaime Carbonell

{vamshi, vogel, jgc}@cs.cmu.edu

Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Abstract

Word alignment models form an important part of building statistical machine translation systems. Semi-supervised word alignment aims to improve the accuracy of automatic word alignment by incorporating full or partial alignments acquired from humans. Such dedicated elicitation effort is often expensive and depends on availability of bilingual speakers for the language-pair. In this paper we study active learning query strategies to carefully identify highly uncertain or most informative alignment links that are proposed under an unsupervised word alignment model. Manual correction of such informative links can then be applied to create a labeled dataset used by a semi-supervised word alignment model. Our experiments show that using active learning leads to maximal reduction of alignment error rates with reduced human effort.

1 Introduction

The success of statistical approaches to Machine Translation (MT) can be attributed to the IBM models (Brown et al., 1993) that characterize *word-level* alignments in parallel corpora. Parameters of these alignment models are learnt in an unsupervised manner using the EM algorithm over *sentence-level* aligned parallel corpora. While the ease of automatically aligning sentences at the word-level with tools like GIZA++ (Och and Ney, 2003) has enabled fast development of statistical machine translation (SMT) systems for various language pairs, the quality of alignment is typically quite low for language

pairs that diverge from the independence assumptions made by the generative models. Also, an immense amount of parallel data enables better estimation of the model parameters, but a large number of language pairs still lack parallel data.

Two directions of research have been pursued for improving generative word alignment. The first is to relax or update the independence assumptions based on more information, usually syntactic, from the language pairs (Cherry and Lin, 2006). The second is to use extra annotation, typically *word-level* human alignment for some sentence pairs, in conjunction with the parallel data to learn alignment in a semi-supervised manner. Our research is in the direction of the latter, and aims to reduce the effort involved in hand-generation of word alignments by using active learning strategies for careful selection of word pairs to seek alignment.

Active learning for MT has not yet been explored to its full potential. Much of the literature has explored one task – selecting sentences to translate and add to the training corpus (Haffari et al., 2009). In this paper we explore active learning for word alignment, where the input to the active learner is a sentence pair (s_1^J, t_1^I) , present in two different languages $\mathcal{S} = \{s^*\}$ and $\mathcal{T} = \{t^*\}$, and the annotation elicited from human is a set of links $\{(j, i) : j = 0 \cdots J; i = 0 \cdots I\}$. Unlike previous approaches, our work does not require elicitation of full alignment for the sentence pair, which could be effort-intensive. We use standard active learning query strategies to selectively elicit partial alignment information. This partial alignment information is then fed into a semi-supervised word aligner which per-

forms an improved word alignment over the entire parallel corpus.

Rest of the paper is organized as follows. We present related work in Section 2. Section 3 gives an overview of unsupervised word alignment models and its semi-supervised improvisation. Section 4 details our active learning framework with discussion of the link selection strategies in Section 5. Experiments in Section 6 have shown that our selection strategies reduce alignment error rates significantly over baseline. We conclude with discussion on future work.

2 Related Work

Semi-supervised learning is a broader area of Machine Learning, focusing on improving the learning process by usage of unlabeled data in conjunction with labeled data (Chapelle et al., 2006). Many semi-supervised learning algorithms use co-training framework, which assumes that the dataset has multiple views, and training different classifiers on a non-overlapping subset of these features provides additional labeled data (Zhu, 2005). Active query selection for training a semi-supervised learning algorithm is an interesting method that has been applied to clustering problems. Tomanek and Hahn (2009) applied active semi supervised learning to the sequence-labeling problem. Tur et al. (2005) describe active and semi-supervised learning methods for reducing labeling effort for spoken language understanding. They train supervised classification algorithms for the task of call classification and apply it to a large unlabeled dataset to select the least confident instances for human labeling.

Researchers have begun to explore semi-supervised word alignment models that use both labeled and unlabeled data. Fraser and Marcu (2006) pose the problem of alignment as a search problem in log-linear space with features coming from the IBM alignment models. The log-linear model is trained on the available labeled data to improve performance. They propose a semi-supervised training algorithm which alternates between discriminative error training on the labeled data to learn the weighting parameters and maximum-likelihood EM training on unlabeled data to estimate the parameters. Callison-Burch et

al. (2004) also improve alignment by interpolating human alignments with automatic alignments. They observe that while working with such datasets, alignments of higher quality should be given a much higher weight than the lower-quality alignments. Wu et al. (2006) learn separate models from labeled and unlabeled data using the standard EM algorithm. The two models are then interpolated as a learner in the semi-supervised AdaBoost algorithm to improve word alignment.

Active learning has been applied to various fields of Natural Language Processing like statistical parsing, entity recognition among others (Hwa, 2004; Tang et al., 2001; Shen et al., 2004). In case of MT, the potential of active learning has remained largely unexplored. For Statistical Machine Translation, application of active learning has been focused on the task of selecting the most informative sentences to train the model, in order to reduce cost of data acquisition. Recent work in this area discussed multiple query selection strategies for a Statistical Phrase Based Translation system (Haffari et al., 2009). Their framework requires source text to be translated by the system and the translated data is used in a self-training setting to train MT models. To our knowledge, we are not aware of any work that has looked at reducing human effort by selective elicitation of alignment information using active learning techniques.

3 Word Alignment

3.1 IBM models

IBM models provide a generative framework for performing word alignment of parallel corpus. Given two strings from source and target languages $s_1^J = s_1, \dots, s_j, \dots, s_J$ and $t_1^I = t_1, \dots, t_i, \dots, t_I$, an alignment \mathcal{A} is defined as a subset of the Cartesian product of the word indices as shown in Eq 1. In IBM models, since alignment is treated as a function, all the source positions must be covered exactly once (Brown et al., 1993).

$$\mathcal{A} \subseteq \{(j, i) : j = 0 \dots J; i = 0 \dots I\} \quad (1)$$

For the task of translation, we would ideally want to model $P(s_1^J | t_1^J)$, which is the probability of observing source sentence s_1^J given target sentence t_1^J . This requires a lot of parallel corpus for estimation

and so it is then factored over the word alignment A for the sentence pair, which is a hidden variable. Word alignment is therefore a by-product in the process of modeling translation. We can also represent the same under some parameterization of θ , which is the model we are interested to estimate.

$$P(s_1^J | t_1^I) = \sum_{a_1^J} Pr(s_1^J, A | t_1^I) \quad (2)$$

$$= \sum_A p_\theta(s_1^J, A | t_1^I) \quad (3)$$

Given a parallel corpus U of sentence pairs $\{(s_k, t_k) : k = 1, \dots, K\}$ the parameters can be estimated by maximizing the conditional likelihood over the data. IBM models (Brown et al., 1993) from 1 to 5 are different ways of factoring the probability model to estimate the parameter set θ . For example in the simplest of the models, IBM model 1, only the lexical translation probability is considered treating each word being translated independent of the other words.

$$\hat{\theta} = \arg \max_{\theta} \prod_{k=1}^K \sum_A p_\theta(s_k, A | t_k) \quad (4)$$

The parameters of the model above are estimated as $\hat{\theta}$, using the EM algorithm. We can also extract the *Viterbi alignment*, \hat{A} , for all the sentence pairs, which is the alignment with the highest probability under the current model parameters θ :

$$\hat{A} = \arg \max_A p_{\hat{\theta}}(s_1^J, A | t_1^I) \quad (5)$$

The alignment models are asymmetric and differ with the choice of translation direction. We can therefore perform the above after switching the direction of the language pair and obtain models and Viterbi alignments for the corpus as represented below:

$$\hat{\theta} = \arg \max_{\theta} \prod_{k=1}^K \sum_a p_\theta(t_k, a | s_k) \quad (6)$$

$$\hat{A} = \arg \max_A p_{\hat{\theta}}(t_1^I, A | s_1^J) \quad (7)$$

Given the Viterbi alignment for each sentence pair in the parallel corpus, we can also compute the word-level alignment probabilities using simple relative likelihood estimation for both the directions.

As we will discuss in Section 5, the alignments and the computed lexicons form an important part of our link selection strategies.

$$P(s_j / t_i) = \frac{\sum_s \text{count}(t_i, s_j; \hat{A})}{\sum_s \text{count}(t_i)} \quad (8)$$

$$P(t_i / s_j) = \frac{\sum_s \text{count}(t_i, s_j; \hat{A})}{\sum_s \text{count}(s_j)} \quad (9)$$

We perform all our experiments on a symmetrized alignment that combines the bidirectional alignments using heuristics as discussed in (Koehn et al., 2007). We represent this alignment as $A = \{a_{ij} : i = 0 \dots J \in s_1^J; j = 0 \dots I \in t_1^I\}$.

3.2 Semi-Supervised Word Alignment

We use an extended version of MGIZA++ (Gao and Vogel, 2008) to perform the constrained semi-supervised word alignment. To get full benefit from the manual alignments, MGIZA++ modifies all alignment models used in the standard training procedure, i.e. the IBM1, HMM, IBM3 and IBM4 models. Manual alignments are incorporated in the EM training phase of these models as constraints that restrict the summation over all possible alignment paths. Typically in the EM procedure for IBM models, the training procedure requires for each source sentence position, the summation over all positions in the target sentence. The manual alignments allow for one-to-many alignments and many-to-many alignments in both directions. For each position i in the source sentence, there can be more than one manually aligned target word. The restricted training will allow only those paths, which are consistent with the manual alignments. Therefore, the restriction of the alignment paths reduces to restricting the summation in EM.

4 Active Learning for Word Alignment

Active learning attempts to optimize performance by selecting the most informative instances to label, where ‘informativeness’ is defined as maximal expected improvement in accuracy. The objective is to select optimal instance for an external expert to label and then run the learning method on the newly-labeled and previously-labeled instances to minimize prediction or translation error, repeating until either the maximal number of external queries

is reached or a desired accuracy level is achieved. Several studies (Tong and Koller, 2002; Nguyen and Smeulders, 2004; Donmez and Carbonell, 2008) show that active learning greatly helps to reduce the labeling effort in various classification tasks.

We discuss our active learning setup for word alignment in Algorithm 1. We start with an unlabeled dataset $U = \{(S_k, T_k)\}$, indexed by k , and a seed pool of partial alignment links $A_0 = \{a_{ij}^k, \forall s_i \in S_k, t_j \in T_k\}$. Each a_{ij}^k represents an alignment link from a sentence pair k that connects source word s_i with t_j .

This is usually an empty set at iteration $t = 0$. We iterate for T iterations. We take a pool-based active learning strategy, where we have access to all the automatically aligned links and we can score the links based on our active learning query strategy. The query strategy uses the automatically trained alignment model θ_t from the current iteration t , for scoring the links. Re-training and re-tuning an SMT system for each link at a time is computationally infeasible. We therefore perform batch learning by selecting a set of N links scored high by our query strategy. We seek manual corrections for the selected links and add the alignment data to the current labeled dataset. The word-level aligned labeled dataset is then provided to our semi-supervised word alignment algorithm, which uses it to produce the alignment model θ_{t+1} for U .

Algorithm 1 AL FOR WORD ALIGNMENT

- 1: Unlabeled Data Set: $U = \{(s_k, t_k)\}$
- 2: Manual Alignment Set : $A_0 = \{a_{ij}^k, \forall s_i \in S_k, t_j \in T_k\}$
- 3: Train Semi-supervised Word Alignment using $(U, A_0) \rightarrow \theta_0$
- 4: N : batch size
- 5: **for** $t = 0$ to T **do**
- 6: $L_t = \text{LinkSelection}(U, A_t, \theta_t, N)$
- 7: Request Human Alignment for L_t
- 8: $A_{t+1} = A_t + L_t$
- 9: Re-train Semi-Supervised Word Alignment on $(U, A_{t+1}) \rightarrow \theta_{t+1}$
- 10: **end for**

We can iteratively perform the algorithm for a defined number of iterations T or until a certain desired performance is reached, which is measured by align-

ment error rate (AER) (Fraser and Marcu, 2007) in the case of word alignment. In a more typical scenario, since reducing human effort or cost of elicitation is the objective, we iterate until the available budget is exhausted.

5 Query Strategies for Link Selection

We propose multiple query selection strategies for our active learning setup. The scoring criteria is designed to select alignment links across sentence pairs that are highly uncertain under current automatic translation models. These links are difficult to align correctly by automatic alignment and will cause incorrect phrase pairs to be extracted in the translation model, in turn hurting the translation quality of the SMT system. Manual correction of such links produces the maximal benefit to the model. We would ideally like to elicit the least number of manual corrections possible in order to reduce the cost of data acquisition. In this section we discuss our link selection strategies based on the standard active learning paradigm of ‘uncertainty sampling’ (Lewis and Catlett, 1994). We use the automatically trained translation model θ_t for scoring each link for uncertainty. In particular θ_t consists of bidirectional lexicon tables computed from the bidirectional alignments as discussed in Section 3.

5.1 Uncertainty based: Bidirectional Alignment Scores

The automatic Viterbi alignment produced by the alignment models is used to obtain translation lexicons, as discussed in Section 3. These lexicons capture the conditional distributions of source-given-target $P(s/t)$ and target-given-source $P(t/s)$ probabilities at the word level where $s_i \in S$ and $t_j \in T$. We define certainty of a link as the harmonic mean of the bidirectional probabilities. The selection strategy selects the least scoring links according to the formula below which corresponds to links with maximum uncertainty:

$$\text{Score}(a_{ij}/s_1^I, t_1^J) = \frac{2 * P(t_j/s_i) * P(s_i/t_j)}{P(t_j/s_i) + P(s_i/t_j)} \quad (10)$$

5.2 Confidence Based: Posterior Alignment probabilities

Confidence estimation for MT output is an interesting area with meaningful initial exploration (Blatz

et al., 2004; Ueffing and Ney, 2007). Given a sentence pair (s_1^I, t_1^J) and its word alignment, we compute two confidence metrics at alignment link level – based on the posterior link probability and a simple IBM Model 1 as seen in Equation 13. We select the alignment links that the initial word aligner is least confident according to our metric and seek manual correction of the links. We use t_{2s} to denote computation using higher order (IBM4) target-given-source models and s_{2t} to denote source-given-target models. Targeting some of the uncertain parts of word alignment has already been shown to improve translation quality in SMT (Huang, 2009). In our current work, we use confidence metrics as an active learning sampling strategy to obtain most informative links. We also experiment with other confidence metrics as discussed in (Ueffing and Ney, 2007), especially the IBM 1 model score metric which showed some improvement as well.

$$P_{t2s}(a_{ij}, t_1^J/s_1^I) = \frac{p_{t2s}(t_j/s_i, a_{ij} \in A)}{\sum_i^M p_{t2s}(t_j/s_i)} \quad (11)$$

$$P_{s2t}(a_{ij}, s_1^I/t_1^J) = \frac{p_{s2t}(s_i/t_j, a_{ij} \in A)}{\sum_i^N p_{s2t}(s_i/t_j)} \quad (12)$$

$$Conf(a_{ij}/S, T) = \frac{2 * P_{t2s} * P_{s2t}}{P_{t2s} + P_{s2t}} \quad (13)$$

5.3 Agreement Based: Query by Committee

The generative alignments produced differ based on the choice of direction of the language pair. We use A_{s2t} to denote alignment in the source to target direction and A_{t2s} to denote the target to source direction. We consider these alignments to be two experts that have two different views of the alignment process. We formulate our query strategy to select links, where the agreement differs across these two alignments. In general query by committee is a standard sampling strategy in active learning(Freund et al., 1997), where the committee consists of any number of experts with varying opinions, in this case alignments in different directions. We formulate a query by committee sampling strategy for word alignment as shown in Equation 14. In order to break ties, we extend this approach to select the link with higher average frequency of occurrence of words involved in the link.

Language	Sentences	Words	
		Src	Tgt
Ch-En	21,863	424,683	524,882
Ar-En	29,876	630,101	821,938

Table 1: Corpus Statistics of Human Data

Alignment	Automatic Links	Manual Links
Ch-En	491,887	588,075
Ar-En	786,223	712,583

Table 2: Alignment Statistics of Human Data

$$Score(a_{ij}) = \alpha \quad \text{where} \quad (14)$$

$$\alpha = \begin{cases} 2 & a_{ij} \in A_{t2s} \cap A_{s2t} \\ 1 & a_{ij} \in A_{t2s} \cup A_{s2t} \\ 0 & \text{otherwise} \end{cases}$$

6 Experiments

6.1 Data Analysis

To run our active learning and semi-supervised word alignment experiments iteratively, we simulate the setup by using a parallel corpus for which the gold standard human alignment is already available. We experiment with two language pairs - Chinese-English and Arabic-English. Corpus-level statistics for both language pairs can be seen in Table 1 and their alignment link level statistics can be seen in Table 2. Both datasets were released by LDC as part of the GALE project.

Chinese-English dataset consists of 21,863 sentence pairs with complete manual alignment. The human alignment for this dataset is much denser than the automatic word alignment. On an average each source word is linked to more than one target word. Similarly, the Arabic-English dataset consisting of 29,876 sentence pairs also has a denser manual alignment. Automatic word alignment in both cases was computed as a symmetrized version of the bidirectional alignments obtained from using GIZA++ (Och and Ney, 2003) in each direction separately.

6.2 Word Alignment Results

We first perform an unsupervised word alignment of the parallel corpus. We then use the learned model

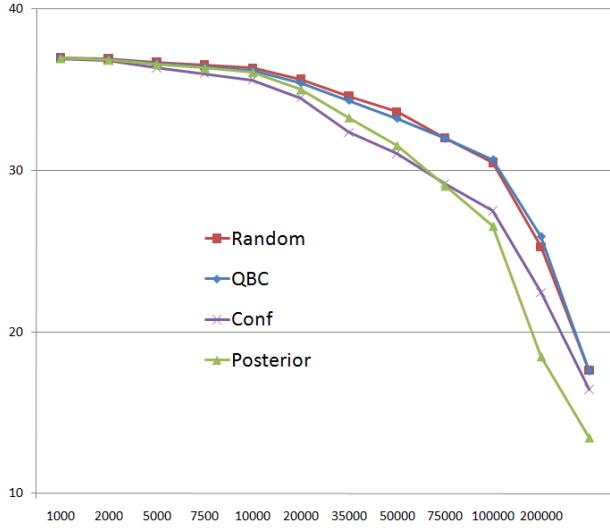


Figure 1: Chinese-English: Link Selection Results

in running our link selection algorithm over the entire alignments to determine the most uncertain links according to each active learning strategy. The links are then looked up in the gold standard human alignment database and corrected. In scenarios where an alignment link is not present in the gold standard data for the source word, we introduce a NULL alignment constraint, else we select all the links as given in the gold standard. The aim of our work is to show that active learning can help in selecting informative alignment links, which if manually labeled can reduce the overall alignment error rate of the given corpus. We, therefore measure the reduction of alignment error rate (AER) of a semi-supervised word aligner that uses this extra information to align the corpus. We plot performance curves for both Chinese-English, Figure 1 and Arabic-English, Figure 2, with number of manual links elicited on x-axis and AER on y-axis. In each iteration of the experiment, we gradually increase the number of links selected from gold standard and make them available to the semi-supervised word aligner and measure the overall reduction of AER on the corpus. We compare our link selection strategies to a baseline approach, where links are selected at random for manual correction.

All our approaches perform equally or better than the baseline for both language pairs. Query by committee (qbc) performs similar to the baseline in Chinese-English and only slightly better for Arabic-

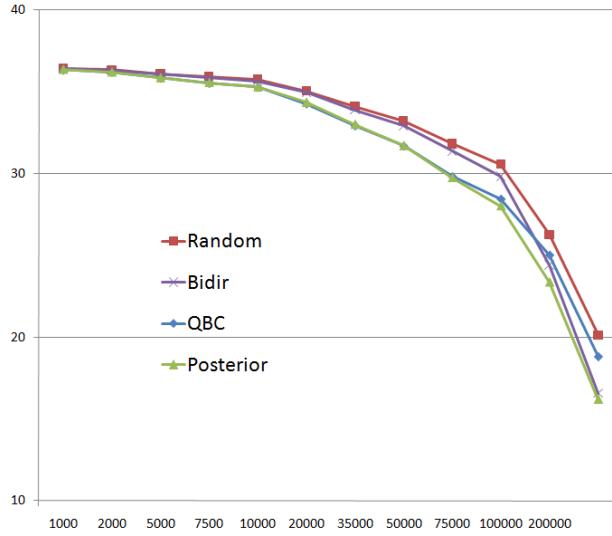


Figure 2: Arabic-English: Link Selection Results

English. This could be due to our committee consisting of two alignments that differ only in direction and so are not sufficient in deciding for uncertainty. We will be exploring alternative formulations to this strategy. Confidence based and uncertainty based metrics perform significantly better than the baseline in both language pairs. We can interpret the improvements in two ways. For the same number of manual alignments elicited, our selection strategies select links that provide higher reduction of error when compared to the baseline. An alternative interpretation is that assuming a uniform cost per link, our best selection strategy achieves similar performance to the baseline, at a much lower cost of elicitation.

6.3 Translation Results

We also perform end-to-end machine translation experiments to show that our improvement of alignment quality leads to an improvement of translation scores. For Chinese-English, we train a standard phrase-based SMT system (Koehn et al., 2007) over the available 21,863 sentences. We tune on the MT-Eval 2004 dataset and test on a subset of MT-Eval 2005 dataset consisting of 631 sentences. The language model we use is built using only the English side of the parallel corpus. We understand that this language model is not the optimal choice, but we are interested in testing the word alignment accuracy, which primarily affects the translation model.

Cn-En	BLEU	METEOR
Baseline	18.82	42.70
Human Alignment	19.96	44.22
Active Selection 20%	19.34	43.25

Table 3: Effect of Alignment on Translation Quality

We first obtain the baseline score by training in an unsupervised manner, where no manual alignment is used. We also train a configuration, where we substitute the final word alignment with gold standard manual alignment for the entire parallel corpus. This is an upper bound on the translation accuracy that can be achieved by any alignment link selection algorithm for this dataset. We now take our best link selection criteria, which is the confidence based method and re-train the MT system after eliciting manual information for only 20% of the alignment links. We observe that at this point we have reduced the AER from 37.09 to 26.57. The translation accuracy reported in Table 3, as measured by BLEU (Papineni et al., 2002) and METEOR (Lavie and Agarwal, 2007), also shows significant improvement and approaches the quality achieved using gold standard data. We did not perform MT experiments with Arabic-English dataset due to the incompatibility of tokenization schemes between the manually aligned parallel corpora and publicly available evaluation sets.

7 Conclusion

Word-Alignment is a particularly challenging problem and has been addressed in a completely unsupervised manner thus far (Brown et al., 1993). While generative alignment models have been successful, lack of sufficient data, model assumptions and local optimum during training are well known problems. Semi-supervised techniques use partial manual alignment data to address some of these issues. We have shown that active learning strategies can reduce the effort involved in eliciting human alignment data. The reduction in effort is due to careful selection of maximally uncertain links that provide the most benefit to the alignment model when used in a semi-supervised training fashion. Experiments on Chinese-English have shown considerable improvements.

8 Future Work

In future, we wish to work with word alignments for other language pairs as well as study the effect of manual alignments by varying the size of available parallel data. We also plan to obtain alignments from non-experts over online marketplaces like Amazon Mechanical Turk to further reduce the cost of annotation. We will be experimenting with obtaining full-alignment vs. partial alignment from non-experts. Our hypothesis is that, humans are good at performing tasks of smaller size and so we can extract high quality alignments in the partial alignment case. Cost of link annotation in our current work is assumed to be uniform, but this needs to be revisited. We will also experiment with active learning techniques for identifying sentence pairs with very low alignment confidence, where obtaining full-alignment is equivalent to obtaining multiple partial alignments.

Acknowledgments

This research was partially supported by DARPA under grant NBCHC080097. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the DARPA. The first author would like to thank Qin Gao for the semi-supervised word alignment software and help with running experiments.

References

- John Blatz, Erin Fitzgerald, George Foster, Simona Gandrabi, Cyril Goutte, Alex Kulesza, Alberto Sanchis, and Nicola Ueffing. 2004. Confidence estimation for machine translation. In *Proceedings of Coling 2004*, pages 315–321, Geneva, Switzerland, Aug 23–Aug 27. COLING.
- Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Chris Callison-Burch, David Talbot, and Miles Osborne. 2004. Statistical machine translation with word- and sentence-aligned parallel corpora. In *ACL 2004*, page 175, Morristown, NJ, USA. Association for Computational Linguistics.
- O. Chapelle, B. Schölkopf, and A. Zien, editors. 2006. *Semi-Supervised Learning*. MIT Press, Cambridge, MA.

Colin Cherry and Dekang Lin. 2006. Soft syntactic constraints for word alignment through discriminative training. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 105–112, Morristown, NJ, USA.

Pinar Donmez and Jaime G. Carbonell. 2008. Optimizing estimated loss reduction for active sampling in rank learning. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, pages 248–255, New York, NY, USA. ACM.

Alexander Fraser and Daniel Marcu. 2006. Semi-supervised training for statistical word alignment. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 769–776, Morristown, NJ, USA. Association for Computational Linguistics.

Alexander Fraser and Daniel Marcu. 2007. Measuring word alignment quality for statistical machine translation. *Comput. Linguist.*, 33(3):293–303.

Yoav Freund, Sebastian H. Seung, Eli Shamir, and Nafali Tishby. 1997. Selective sampling using the query by committee algorithm. *Machine. Learning.*, 28(2-3):133–168.

Qin Gao and Stephan Vogel. 2008. Parallel implementations of word alignment tool. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 49–57, Columbus, Ohio, June. Association for Computational Linguistics.

Gholamreza Haffari, Maxim Roy, and Anoop Sarkar. 2009. Active learning for statistical phrase-based machine translation. In *Proceedings of HLT NAACL 2009*, pages 415–423, Boulder, Colorado, June. Association for Computational Linguistics.

Fei Huang. 2009. Confidence measure for word alignment. In *Proceedings of the Joint ACL and IJCNLP*, pages 932–940, Suntec, Singapore, August. Association for Computational Linguistics.

Rebecca Hwa. 2004. Sample selection for statistical parsing. *Comput. Linguist.*, 30(3):253–276.

Philipp Koehn, Hieu Hoang, Alexandra Birch Mayne, Christopher Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *ACL Demonstration Session*.

Alon Lavie and Abhaya Agarwal. 2007. Meteor: an automatic metric for mt evaluation with high levels of correlation with human judgments. In *WMT 2007*, pages 228–231, Morristown, NJ, USA.

David D. Lewis and Jason Catlett. 1994. Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the Eleventh International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann.

Hieu T. Nguyen and Arnold Smeulders. 2004. Active learning using pre-clustering. In *ICML*.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, pages 19–51.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL 2002*, pages 311–318, Morristown, NJ, USA.

Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan. 2004. Multi-criteria-based active learning for named entity recognition. In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 589, Morristown, NJ, USA. Association for Computational Linguistics.

Min Tang, Xiaoqiang Luo, and Salim Roukos. 2001. Active learning for statistical natural language parsing. In *ACL '02*, pages 120–127, Morristown, NJ, USA.

Katrin Tomanek and Udo Hahn. 2009. Semi-supervised active learning for sequence labeling. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1039–1047, Suntec, Singapore, August. Association for Computational Linguistics.

Simon Tong and Daphne Koller. 2002. Support vector machine active learning with applications to text classification. *Journal of Machine Learning*, pages 45–66.

Gokhan Tur, Dilek Hakkani-Tr, and Robert E. Schapire. 2005. Combining active and semi-supervised learning for spoken language understanding. *Speech Communication*, 45(2):171 – 186.

Nicola Ueffing and Hermann Ney. 2007. Word-level confidence estimation for machine translation. *Comput. Linguist.*, 33(1):9–40.

Hua Wu, Haifeng Wang, and Zhanyi Liu. 2006. Boosting statistical word alignment using labeled and unlabeled data. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 913–920, Morristown, NJ, USA. Association for Computational Linguistics.

X. Zhu. 2005. Semi-Supervised Learning Literature Survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison. http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf.