Pivot Language Approach for Phrase-Based Statistical Machine Translation

Hua Wu and Haifeng Wang
Toshiba (China) Research and Development Center
5/F., Tower W2, Oriental Plaza, No.1, East Chang An Ave., Dong Cheng District
Beijing, 100738, China
{wuhua,wanghaifeng}@rdc.toshiba.com.cn

Abstract

This paper proposes a novel method for phrase-based statistical machine translation by using pivot language. To conduct translation between languages L_f and L_e with a small bilingual corpus, we bring in a third language L_p, which is named the *pivot* language. For L_f-L_p and L_p-L_e, there exist large bilingual corpora. Using only L_f-L_p and L_p-L_e bilingual corpora, we can build a translation model for L_f-L_e. The advantage of this method lies in that we can perform translation between L_f and L_e even if there is no bilingual corpus available for this language pair. Using BLEU as a metric, our pivot language method achieves an absolute improvement of 0.06 (22.13% relative) as compared with the model directly trained with 5,000 L_f-L_e sentence pairs for French-Spanish translation. Moreover, with a small L_f-L_e bilingual corpus available, our method can further improve the translation quality by using the additional L_f-L_p and L_p-L_e bilingual corpora.

1 Introduction

For statistical machine translation (SMT), phrase-based methods (Koehn et al., 2003; Och and Ney, 2004) and syntax-based methods (Wu, 1997; Alshawi et al. 2000; Yamada and Knignt, 2001; Melamed, 2004; Chiang, 2005; Quick et al., 2005; Mellebeek et al., 2006) outperform word-based methods (Brown et al., 1993). These methods need large bilingual corpora. However, for some languages pairs, only a small bilingual corpus is available, which will degrade the performance of statistical translation systems.

To solve this problem, this paper proposes a novel method for phrase-based SMT by using a pivot language. To perform translation between languages L_f and L_e, we bring in a pivot language L_p, for which there exist large bilingual corpora for language pairs L_f-L_p and L_p-L_e. With the L_f-L_p and L_p-L_e bilingual corpora, we can build a translation model for L_f-L_e by using L_p as the pivot language. We name the translation model *pivot* model. The advantage of this method lies in that we can conduct translation between L_f and L_e even if there is no bilingual corpus available for this language pair. Moreover, if a small corpus is available for L_f-L_e, we build another translation model, which is named *standard* model. Then, we build an *interpolated* model by performing linear interpolation on the standard model and the pivot model. Thus, the interpolated model can employ both the small L_f-L_e corpus and the large L_f-L_p and L_p-L_e corpora.

We perform experiments on the Europarl corpus (Koehn, 2005). Using BLEU (Papineni et al., 2002) as a metric, our method achieves an absolute improvement of 0.06 (22.13% relative) as compared with the standard model trained with 5,000 L_f-L_e sentence pairs for French-Spanish translation. The translation quality is comparable with that of the model trained with a bilingual corpus of 30,000 L_f-L_e sentence pairs. Moreover, translation quality is further boosted by using both the small L_f-L_e bilingual corpus and the large L_f-L_p and L_p-L_e corpora.

Experimental results on Chinese-Japanese translation also indicate that our method achieves satisfactory results using English as the pivot language.

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 856–863,
Prague, Czech Republic, June 2007.

The remainder of this paper is organized as follows. In section 2, we describe the related work. Section 3 briefly introduces phrase-based SMT. Section 4 and Section 5 describes our method for phrase-based SMT using pivot language. We describe the experimental results in sections 6 and 7. Lastly, we conclude in section 8.

2 Related Work

Our method is mainly related to two kinds of methods: those using pivot language and those using a small bilingual corpus or scarce resources.

For the first kind, pivot languages are employed to translate queries in cross-language information retrieval (CLIR) (Gollins and Sanderson, 2001; Kishida and Kando, 2003). These methods only used the available dictionaries to perform word by word translation. In addition, NTCIR 4 workshop organized a shared task for CLIR using pivot language. Machine translation systems are used to translate queries into pivot language sentences, and then into target sentences (Sakai et al., 2004).

Callison-Burch et al. (2006) used pivot languages for paraphrase extraction to handle the unseen phrases for phrase-based SMT. Borin (2000) and Wang et al. (2006) used pivot languages to improve word alignment. Borin (2000) used multilingual corpora to increase alignment coverage. Wang et al. (2006) induced alignment models by using two additional bilingual corpora to improve word alignment quality. Pivot Language methods were also used for translation dictionary induction (Schafer and Yarowsky, 2002), word sense disambiguation (Diab and Resnik, 2002), and so on.

For the second kind, Niessen and Ney (2004) used morpho-syntactic information for translation between language pairs with scarce resources. Vandeghinste et al. (2006) used translation dictionaries and shallow analysis tools for translation between the language pair with low resources. A shared task on word alignment was organized as part of the ACL 2005 Workshop on Building and Using Parallel Texts (Martin et al., 2005). This task focused on languages with scarce resources. For the subtask of unlimited resources, some researchers (Aswani and Gaizauskas, 2005; Lopez and Resnik, 2005; Tufis et al., 2005) used language-dependent resources such as dictionary, thesaurus, and dependency parser to improve word alignment results.

In this paper, we address the translation problem for language pairs with scarce resources by bringing in a pivot language, via which we can make use of large bilingual corpora. Our method does not need language-dependent resources or deep linguistic processing. Thus, the method is easy to be adapted to any language pair where a pivot language and corresponding large bilingual corpora are available.

3 Phrase-Based SMT

According to the translation model presented in (Koehn et al., 2003), given a source sentence $\mathbf{f}$, the best target translation $\mathbf{e}_{\text{best}}$ can be obtained according to the following model

$$\begin{aligned}\mathbf{e}_{\text{best}} &= \arg\max_{\mathbf{e}} p(\mathbf{e} \mid \mathbf{f}) \\ &= \arg\max_{\mathbf{e}} p(\mathbf{f} \mid \mathbf{e}) p_{\text{LM}}(\mathbf{e}) \omega^{\text{length}(\mathbf{e})}\end{aligned} \tag{1}$$

Where the translation model $p(\mathbf{f} \mid \mathbf{e})$ can be decomposed into

$$\begin{aligned}&p(\bar{f}_1^I \mid \bar{e}_1^I) \\ &= \prod_{i=1}^{I} \phi(\bar{f}_i \mid \bar{e}_i) d(a_i - b_{i-1}) p_{\text{w}}(\bar{f}_i \mid \bar{e}_i, a)^{\lambda}\end{aligned} \tag{2}$$

Where $\phi(\bar{f}_i \mid \bar{e}_i)$ and $d(a_i - b_{i-1})$ denote phrase translation probability and distortion probability, respectively. $p_{\text{w}}(\bar{f}_i \mid \bar{e}_i, a)$ is the lexical weight, and λ is the strength of the lexical weight.

4 Phrase-Based SMT Via Pivot Language

This section will introduce the method that performs phrase-based SMT for the language pair L_f-L_e by using the two bilingual corpora of L_f-L_p and L_p-L_e. With the two additional bilingual corpora, we train two translation models for L_f-L_p and L_p-L_e, respectively. Based on these two models, we build a pivot translation model for L_f-L_e, with L_p as a pivot language.

According to equation (2), the phrase translation probability and the lexical weight are language dependent. We will introduce them in sections 4.1 and 4.2, respectively.

4.1 Phrase Translation Probability

Using the L_f-L_p and L_p-L_e bilingual corpora, we train two phrase translation probabilities

$\phi(\bar{f}_i \mid \bar{p}_i)$ and $\phi(\bar{p}_i \mid \bar{e}_i)$, where $\bar{p}_i$ is the phrase in the pivot language L_p. Given the phrase translation probabilities $\phi(\bar{f}_i \mid \bar{p}_i)$ and $\phi(\bar{p}_i \mid \bar{e}_i)$, we obtain the phrase translation probability $\phi(\bar{f}_i \mid \bar{e}_i)$ according to the following model.

$$\phi(\bar{f}_i \mid \bar{e}_i) = \sum_{\bar{p}_i} \phi(\bar{f}_i \mid \bar{p}_i, \bar{e}_i)\phi(\bar{p}_i \mid \bar{e}_i) \qquad (3)$$

The phrase translation probability $\phi(\bar{f}_i \mid \bar{p}_i, \bar{e}_i)$ does not depend on the phase $\bar{e}_i$ in the language L_e, since it is estimated from the L_f-L_p bilingual corpus. Thus, equation (3) can be rewritten as

$$\phi(\bar{f}_i \mid \bar{e}_i) = \sum_{\bar{p}_i} \phi(\bar{f}_i \mid \bar{p}_i)\phi(\bar{p}_i \mid \bar{e}_i) \qquad (4)$$

4.2 Lexical Weight

Given a phrase pair $(\bar{f}, \bar{e})$ and a word alignment a between the source word positions $i = 1,...,n$ and the target word positions $j = 1,...,m$, the lexical weight can be estimated according to the following method (Koehn et al., 2003).

$$p_w(\bar{f} \mid \bar{e}, a) = \prod_{i=1}^{n} \frac{1}{|j \mid (i,j) \in a|} \sum_{\forall (i,j) \in a} w(f_i \mid e_j) \qquad (5)$$

In order to estimate the lexical weight, we first need to obtain the alignment information a between the two phrases $\bar{f}$ and $\bar{e}$, and then estimate the lexical translation probability $w(f \mid e)$ according to the alignment information. The alignment information of the phrase pair $(\bar{f}, \bar{e})$ can be induced from the two phrase pairs $(\bar{f}, \bar{p})$ and $(\bar{p}, \bar{e})$.

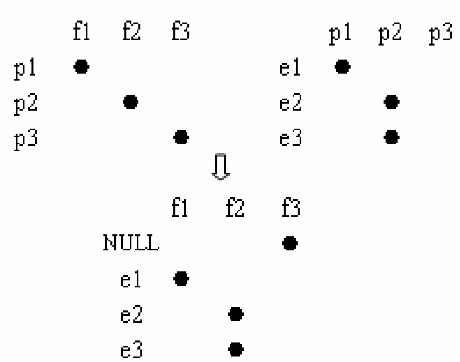

Figure 1. Alignment Information Induction

Let a_1 and a_2 represent the word alignment information inside the phrase pairs $(\bar{f}, \bar{p})$ and $(\bar{p}, \bar{e})$ respectively, then the alignment information a inside $(\bar{f}, \bar{e})$ can be obtained as shown in (6). An example is shown in Figure 1.

$$a = \{(f, e) \mid \exists p : (f, p) \in a_1 \ \& \ (p, e) \in a_2\} \qquad (6)$$

With the induced alignment information, this paper proposes a method to estimate the probability directly from the induced phrase pairs. We name this method *phrase* method. If we use K to denote the number of the induced phrase pairs, we estimate the co-occurring frequency of the word pair (f, e) according to the following model.

$$count(f, e) = \sum_{k=1}^{K} \phi_k(\bar{f} \mid \bar{e}) \sum_{i=1}^{n} \delta(f, f_i)\delta(e, e_{a_i}) \qquad (7)$$

Where $\phi_k(\bar{f} \mid \bar{e})$ is the phrase translation probability for phrase pair k. $\delta(x, y) = 1$ if $x = y$; otherwise, $\delta(x, y) = 0$. Thus, lexical translation probability can be estimated as in (8).

$$w(f \mid e) = \frac{count(f, e)}{\sum_{f'} count(f', e)} \qquad (8)$$

We also estimate the lexical translation probability $w(f \mid e)$ using the method described in (Wang et al., 2006), which is shown in (9). We named it *word* method in this paper.

$$w(f \mid e) = \sum_{p} w(f \mid p) w(p \mid e) sim(f, e; p) \qquad (9)$$

Where $w(f \mid p)$ and $w(p \mid e)$ are two lexical probabilities, and $sim(f, e; p)$ is the cross-language word similarity.

5 Interpolated Model

If we have a small L_f-L_e bilingual corpus, we can employ this corpus to estimate a translation model as described in section 3. However, this model may perform poorly due to the sparseness of the data. In order to improve its performance, we can employ the additional L_f-L_p and L_p-L_e bilingual corpora. Moreover, we can use more than one pivot language to improve the translation performance if the corresponding bilingual corpora exist. Different pivot languages may catch different linguistic phenomena, and improve translation quality for the desired language pair L_f-L_e in different ways.

If we include n pivot languages, n pivot models can be estimated using the method as described in section 4. In order to combine these n pivot models with the standard model trained with the L_f-L_e corpus, we use the linear interpolation method. The phrase translation probability and the lexical weight are estimated as shown in (10) and (11), respectively.

$$\phi(\bar{f} \mid \bar{e}) = \sum_{i=0}^{n} \alpha_i \phi_i(\bar{f} \mid \bar{e}) \tag{10}$$

$$p_{\mathrm{w}}(\bar{f} \mid \bar{e}, a) = \sum_{i=0}^{n} \beta_i p_{\mathrm{w},i}(\bar{f} \mid \bar{e}, a) \tag{11}$$

Where $\phi_0(\bar{f} \mid \bar{e})$ and $p_{\mathrm{w},0}(\bar{f} \mid \bar{e}, a)$ denote the phrase translation probability and lexical weight trained with the L_f-L_e bilingual corpus, respectively. $\phi_i(\bar{f} \mid \bar{e})$ and $p_{\mathrm{w},i}(\bar{f} \mid \bar{e}, a)$ ($i = 1,...,n$) are the phrase translation probability and lexical weight estimated by using the pivot languages. α_i and β_i are the interpolation coefficients.

6 Experiments on the Europarl Corpus

6.1 Data

A shared task to evaluate machine translation performance was organized as part of the NAACL/HLT 2006 Workshop on Statistical Machine Translation (Koehn and Monz, 2006). The shared task used the Europarl corpus (Koehn, 2005), in which four languages are involved: English, French, Spanish, and German. The shared task performed translation between English and the other three languages. In our work, we perform translation from French to the other three languages. We select French to Spanish and French to German translation that are not in the shared task because we want to use English as the pivot language. In general, for most of the languages, there exist bilingual corpora between these languages and English since English is an internationally used language.

Table 1 shows the information about the bilingual training data. In the table, "Fr", "En", "Es", and "De" denotes "French", "English", "Spanish", and "German", respectively. For the language pairs L_f-L_e not including English, the bilingual corpus is extracted from L_f-English and English-L_e since Europarl corpus is a multilingual corpus.

Language Pairs	Sentence Pairs	Source Words	Target Words
Fr-En	688,031	15,323,737	13,808,104
Fr-Es	640,661	14,148,926	13,134,411
Fr-De	639,693	14,215,058	12,155,876
Es-En	730,740	15,676,710	15,222,105
De-En	751,088	15,256,793	16,052,269
De-Es	672,813	13,246,255	14,362,615

Table 1. Training Corpus for European Languages

For the language models, we use the same data provided in the shared task. We also use the same development set and test set provided by the shared task. The in-domain test set includes 2,000 sentences and the out-of-domain test set includes 1,064 sentences for each language.

6.2 Translation System and Evaluation Method

To perform phrase-based SMT, we use Koehn's training scripts[1] and the Pharaoh decoder (Koehn, 2004). We run the decoder with its default settings and then use Koehn's implementation of minimum error rate training (Och, 2003) to tune the feature weights on the development set.

The translation quality was evaluated using a well-established automatic measure: BLEU score (Papineni et al., 2002). And we also use the tool provided in the NAACL/HLT 2006 shared task on SMT to calculate the BLEU scores.

6.3 Comparison of Different Lexical Weights

As described in section 4, we employ two methods to estimate the lexical weight in the translation model. In order to compare the two methods, we translate from French to Spanish, using English as the pivot language. We use the French-English and English-Spanish corpora described in Table 1 as training data. During training, before estimating the Spanish to French phrase translation probability, we filter those French-English and English-Spanish phrase pairs whose translation probabilities are below a fixed threshold 0.001.[2] The translation results are shown in Table 2.

[1] It is located at http://www.statmt.org/wmt06/shared-task/baseline.htm

[2] In the following experiments using pivot languages, we use the same filtering threshold for all of the language pairs.

The *phrase* method proposed in this paper performs better than the *word* method proposed in (Wang et al., 2006). This is because our method uses phrase translation probability as a confidence weight to estimate the lexical translation probability. It strengthens the frequently aligned pairs and weakens the infrequently aligned pairs. Thus, the following sections will use the *phrase* method to estimate the lexical weight.

Method	In-Domain	Out-of-Domain
Phrase	0.3212	0.2098
Word	0.2583	0.1672

Table 2. Results with Different Lexical Weights

6.4 Results of Using One Pivot Language

This section describes the translation results by using only one pivot language. For the language pair French and Spanish, we use English as the pivot language. The entire French-English and English-Spanish corpora as described in section 4 are used to train a pivot model for French-Spanish.

As described in section 5, if we have a small L_f-L_e bilingual corpus and large L_f-L_p and L_p-L_e bilingual corpora, we can obtain interpolated models.

In order to conduct the experiments, we randomly select 5K, 10K, 20K, 30K, 40K, 50K, and 100K sentence pairs from the French-Spanish corpus. Using each of these corpora, we train a standard translation model.

For each standard model, we interpolate it with the pivot model to get an interpolated model. The interpolation weights are tuned using the development set. For all the interpolated models, we set $\alpha_0 = 0.9$, $\alpha_1 = 0.1$, $\beta_0 = 0.9$, and $\beta_1 = 0.1$. We test the three kinds of models on both the in-domain and out-of-domain test sets. The results are shown in Figures 2 and 3.

The pivot model achieves BLEU scores of 0.3212 and 0.2098 on the in-domain and out-of-domain test set, respectively. It achieves an absolute improvement of 0.05 on both test sets (16.92% and 35.35% relative) over the standard model trained with 5,000 French-Spanish sentence pairs. And the performance of the pivot models are comparable with that of the standard models trained with 20,000 and 30,000 sentence pairs on the in-domain and out-of-domain test set, respectively. When the French-Spanish training corpus is increased, the standard models quickly outperform the pivot model.

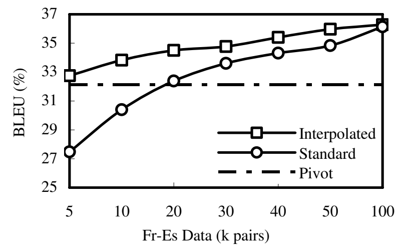

Figure 2. In-Domain French-Spanish Results

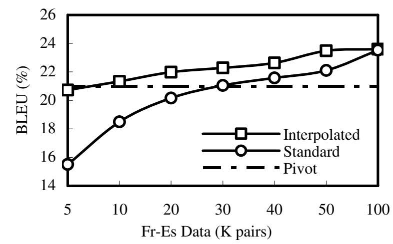

Figure 3. Out-of-Domain French-Spanish Results

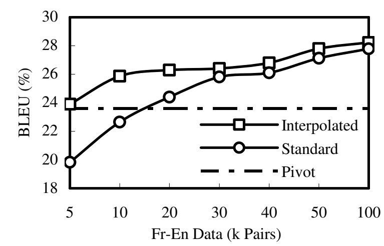

Figure 4. In-Domain French-English Results

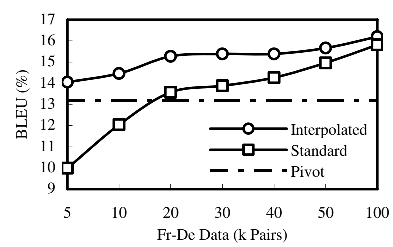

Figure 5. In-Domain French-German Results

When only a very small French-Spanish bilingual corpus is available, the interpolated method can greatly improve the translation quality. For example, when only 5,000 French-Spanish sentence pairs are available, the interpolated model outperforms the standard model by achieving a relative improvement of 17.55%, with the BLEU score improved from 0.2747 to 0.3229. With 50,000 French-Spanish sentence pairs available, the interpolated model significantly[3] improves the translation quality by achieving an absolute im-

[3] We conduct the significance test using the same method as described in (Koehn and Monz, 2006).

provement of 0.01 BLEU. When the French-Spanish training corpus increases to 100,000 sentence pairs, the interpolated model achieves almost the same result as the standard model. This indicates that our pivot language method is suitable for the language pairs with small quantities of training data available.

Besides experiments on French-Spanish translation, we also conduct translation from French to English and French to German, using German and English as the pivot language, respectively. The results on the in-domain test set[4] are shown in Figures 4 and 5. The tendency of the results is similar to that in Figure 2.

6.5 Results of Using More Than One Pivot Language

For French to Spanish translation, we also introduce German as a pivot language besides English. Using these two pivot languages, we build two different pivot models, and then perform linear interpolation on them. The interpolation weights for the English pivot model and the German pivot model are set to 0.6 and 0.4 respectively[5]. The translation results on the in-domain test set are 0.3212, 0.3077, and 0.3355 for the pivot models using English, German, and both German and English as pivot languages, respectively.

With the pivot model using both English and German as pivot languages, we interpolate it with the standard models trained with French-Spanish corpora of different sizes as described in the above section. The comparison of the translation results among the interpolated models, standard models, and the pivot model are shown in Figure 6.

It can be seen that the translation results can be further improved by using more than one pivot language. The pivot model "Pivot-En+De" using two pivot languages achieves an absolute improvement of 0.06 (22.13% relative) as compared with the standard model trained with 5,000 sentence pairs. And it achieves comparable translation result as compared with the standard model trained with 30,000 French-Spanish sentence pairs.

The results in Figure 6 also indicate the interpolated models using two pivot languages achieve the best results of all. Significance test shows that the interpolated models using two pivot languages significantly outperform those using one pivot language when less than 50,000 French-Spanish sentence pairs are available.

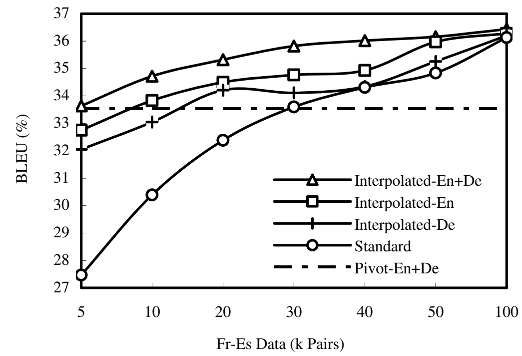

Figure 6. In-Domain French-Spanish Translation Results by Using Two Pivot Languages

6.6 Results by Using Pivot Language Related Corpora of Different Sizes

In all of the above results, the corpora used to train the pivot models are not changed. In order to examine the effect of the size of the pivot corpora, we decrease the French-English and English-French corpora. We randomly select 200,000 and 400,000 sentence pairs from both of them to train two pivot models, respectively. The translation results on the in-domain test set are 0.2376, 0.2954, and 0.3212 for the pivot models trained with 200,000, 400,000, and the entire French-English and English-Spanish corpora, respectively. The results of the interpolated models and the standard models are shown in Figure 7. The results indicate that the larger the training corpora used to train the pivot model are, the better the translation quality is.

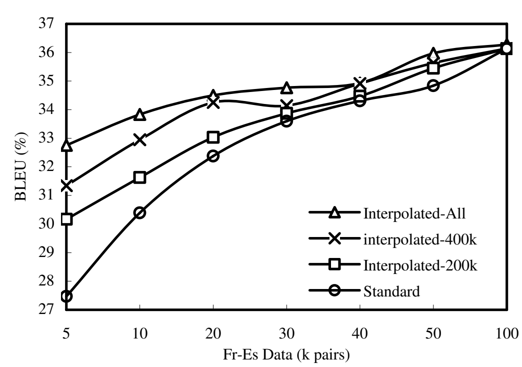

Figure 7. In-Domain French-Spanish Results by Using L_f-L_p and L_p-L_e Corpora of Different Sizes

[4] The results on the out-of-domain test set are similar to that in Figure 3. We only show the in-domain translation results in all of the following experiments because of space limit.

[5] The weights are tuned on the development set.

7 Experiments on Chinese to Japanese Translation

In section 6, translation results on the Europarl multilingual corpus indicate the effectiveness of our method. To investigate the effectiveness of our method by using independently sourced parallel corpora, we conduct Chinese-Japanese translation using English as a pivot language in this section, where the training data are not limited to a specific domain.

The data used for this experiment is the same as those used in (Wang et al., 2006). There are 21,977, 329,350, and 160,535 sentence pairs for the language pairs Chinese-Japanese, Chinese-English, and English-Japanese, respectively. The development data and testing data include 500 and 1,000 Chinese sentences respectively, with one reference for each sentence. For Japanese language model training, we use about 100M bytes Japanese corpus.

The translation result is shown in Figure 8. The pivot model only outperforms the standard model trained with 2,500 sentence pairs. This is because (1) the corpora used to train the pivot model are smaller as compared with the Europarl corpus; (2) the training data and the testing data are not limited to a specific domain; (3) The languages are not closely related.

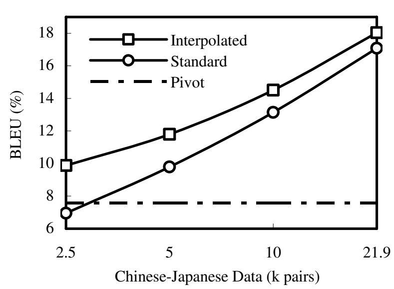

Figure 8. Chinese-Japanese Translation Results

The interpolated models significantly outperform the other models. When only 5,000 sentence pairs are available, the BLEU score increases relatively by 20.53%. With the entire (21,977 pairs) Chinese-Japanese available, the interpolated model relatively increases the BLEU score by 5.62%, from 0.1708 to 0.1804.

8 Conclusion

This paper proposed a novel method for phrase-based SMT on language pairs with a small bilingual corpus by bringing in pivot languages. To perform translation between L_f and L_e, we bring in a pivot language L_p, via which the large corpora of L_f-L_p and L_p-L_e can be used to induce a translation model for L_f-L_e. The advantage of this method is that it can perform translation between the language pair L_f-L_e even if no bilingual corpus for this pair is available. Using BLEU as a metric, our method achieves an absolute improvement of 0.06 (22.13% relative) as compared with the model directly trained with 5,000 sentence pairs for French-Spanish translation. And the translation quality is comparable with that of the model directly trained with 30,000 French-Spanish sentence pairs. The results also indicate that using more pivot languages leads to better translation quality.

With a small bilingual corpus available for L_f-L_e, we built a translation model, and interpolated it with the pivot model trained with the large L_f-L_p and L_p-L_e bilingual corpora. The results on both the Europarl corpus and Chinese-Japanese translation indicate that the interpolated models achieve the best results. Results also indicate that our pivot language approach is suitable for translation on language pairs with a small bilingual corpus. The less the L_f-L_e bilingual corpus is, the bigger the improvement is.

We also performed experiments using L_f-L_p and L_p-L_e corpora of different sizes. The results indicate that using larger training corpora to train the pivot model leads to better translation quality.

References

Hiyan Alshawi, Srinivas Bangalore, and Shona Douglas. 2000. Learning Dependency Translation Models as Collections of Finite-State Head Transducers. *Computational Linguistics*, 26(1):45-60.

Niraj Aswani and Robert Gaizauskas. 2005. Aligning Words in English-Hindi Parallel Corpora. In *Proc. of the ACL 2005 Workshop on Building and Using Parallel Texts: Data-driven Machine Translation and Beyond*, pages 115-118.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2): 263-311.

Chris Callison-Burch, Philipp Koehn, and Miles Osborne. 2006. Improved Statistical Machine Translation Using Paraphrases. In *Proc. of NAACL-2006*, pages 17-24.

Lars Borin. 2000. You'll Take the High Road and I'll Take the Low Road: Using a Third Language to Improve Bilingual Word Alignment. In *Proc. of COLING-2000*, pages 97-103.

David Chiang. 2005. A Hierarchical Phrase-Based Model for Statistical Machine Translation. In *Proc. of ACL-2005*, pages 263-270.

Mona Diab and Philip Resnik. 2002. An Unsupervised Method for Word Sense Tagging using Parallel Corpora. In *Proc. of ACL-2002*, pages 255-262.

Tim Gollins and Mark Sanderson. 2001. Improving Cross Language Information Retrieval with Triangulated Translation. In *Proc. of ACM SIGIR-2001*, pages 90-95.

Kazuaki Kishida and Noriko Kando. 2003. Two-Stage Refinement of Query Translation in a Pivot Language Approach to Cross-Lingual Information Retrieval: An Experiment at CLEF 2003. In *Proc. of CLEF-2003*. pages 253-262.

Philipp Koehn. 2004. Pharaoh: A Beam Search Decoder for Phrase-Based Statistical Machine Translation Models. In *Proc. of AMTA-2004*, pages 115-124.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proc. of MT Summit X*, pages 79-86.

Philipp Koehn and Christof Monz. 2006. Manual and Automatic Evaluation of Machine Translation between European Languages. In *Proc. of the 2006 HLT-NAACL Workshop on Statistical Machine Translation*, pages 102-121.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proc. of HLT-NAAC- 2003*, pages 127-133.

Adam Lopez and Philip Resnik. 2005. Improved HMM Alignment Models for Languages with Scarce Resources. In *Proc. of the ACL-2005 Work-shop on Building and Using Parallel Texts: Data-driven Machine Translation and Beyond*, pages 83-86.

Joel Martin, Rada Mihalcea, and Ted Pedersen. 2005. Word Alignment for Languages with Scarce Resources. In *Proc. of the ACL-2005 Workshop on Building and Using Parallel Texts: Data-driven Machine Translation and Beyond*, pages 65-74.

Dan Melamed. 2004. Statistical Machine Translation by Parsing. In *Proc. of ACL-2004*, pages 653-660.

Bart Mellebeek, Karolina Owczarzak, Declan Groves, Josef Van Genabith, and Andy Way. 2006. A Syntactic Skeleton for Statistical Machine Translation. In *Proc. of EAMT-2006*, pages 195-202.

Sonja Niessen and Hermann Ney. 2004. Statistical Machine Translation with Scarce Resources Using Morpho-Syntactic Information. *Computational linguistics*, 30(2): 181-204.

Franz Josef Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proc. of ACL-2003*, pages 160-167.

Franz Josef Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics*, 30(4):417-449.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of ACL-2002*, pages 311-318.

Chris Quirk, Arul Menezes, and Colin Cherry. 2005. Dependency Treelet Translation: Syntactically Informed Phrasal SMT. In *Proc. of ACL-2005*, pages 271-279.

Tetsuya Sakai, Makoto Koyama, Akira Kumano, and Toshihiko Manabe. 2004. Toshiba BRIDJE at NTCIR-4 CLIR: Monolingual/Bilingual IR and Flexible Feedback. In *Proc. of NTCIR 4.*

Charles Schafer and David Yarowsky. 2002. Inducing Translation Lexicons via Diverse Similarity Measures and Bridge Languages. In *Proc. of CoNLL-2002*, pages 1-7.

Haifeng Wang, Hua Wu, and Zhanyi Liu. 2006. Word Alignment for Languages with Scarce Resources Using Bilingual Corpora of Other Language Pairs. In *Proc. of COLING/ACL-2006 Main Conference Poster Sessions*, pages 874-881.

Dan Tufis, Radu Ion, Alexandru Ceausu, and Dan Stefanescu. 2005. Combined Word Alignments. In *Proc. of the ACL-2005 Workshop on Building and Using Parallel Texts: Data-driven Machine Translation and Beyond,* pages 107-110.

Vincent Vandeghinste, Ineka Schuurman, Michael Carl, Stella Markantonatou, and Toni Badia. 2006. METIS-II: Machine Translation for Low-Resource Languages. In *Proc. of LREC-2006,* pages 1284-1289.

Dekai Wu. 1997. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377-403.

Kenji Yamada and Kevin Knight. 2001. A Syntax Based Statistical Translation Model. In *Proc. of ACL-2001*, pages 523-530.