

Improving Chinese-to-Japanese Patent Translation

Using English as Pivot Language

Xianhua Li Yao Meng Hao Yu

Fujitsu R&D Centre CO., LTD, Beijing, China

{lixianhua, mengyao, yu}@cn.fujitsu.com

Abstract

This paper implements and compares three different strategies to use English as pivot language for Chinese-Japanese patent translation: corpus enrichment, sentence pivot translation and phrase pivot translation. Our results show that both corpus enrichment and phrase pivot translation strategy outperform the baseline system, while the sentence pivot translation strategy failed to improve the system. We apply the strategies on large data set and figure out approaches to improve efficiency. Finally, we perform Minimum Bayes Risk system combination on the different results of direct translation system and pivot translation systems, which significantly outperforms the direct translation system by 4.25 BLEU scores.

1 Introduction

Statistical machine translation (SMT) has made rapid progress in recent years with the support of large quantities of parallel corpora. It's quite common that we use millions of bilingual parallel sentences to train a statistical machine translation system. Unfortunately, large parallel corpora are not always available for some language pairs, or for some specific domains. For example, there are few available bilingual corpora for Chinese-to-Japanese patent translation. Many research labs and companies face data bottleneck when they do research on scarce-resourced language pairs or domains.

Much work has been done to overcome the data bottleneck problem. For example, Lu et al. (2009) exploited the existence of bilingual patent corpora and constructed a Chinese-English patent parallel corpus. Resnik and Smith (2003) took the web as a parallel corpus and mined parallel data from it. Munteanu and Marcu (2005) trained a maximum entropy classifier to extract parallel corpus from large non-parallel newspaper corpora. Our work differs in that we make use of the currently available bilingual corpora, without exploiting extra bilingual data to improve machine translation quality. In other words, we employ pivot translation strategies to improve the performance of SMT systems.

- How to apply pivot translation strategies to help scarce-resourced language translation?
- How to take advantages of different pivot translation strategies to further improve machine translation quality?

In this paper, we introduce and implement three pivot translation strategies for SMT. The first is *corpus enrichment strategy*. It translates the pivot side of source-pivot corpus and pivot-target corpus into target and source language respectively to construct source-target language pairs. With these sentence pairs, it builds up a new SMT system so as to outperform the basic system. As the corpora we employ are quite large, we select sentence pairs according to their sentence value and do experiments on different size of parallel corpus. The second is *sentence pivot translation strategy*.

It builds two SMT systems on source-pivot and pivot-target corpus respectively. When translating a source sentence into target language, it first translates it into pivot language with the source-pivot system. Then the generated sentence is translated into target language with the pivot-target system. Here, we can keep N-best for each source sentence and see the influence of different N. The third is *phrase pivot translation strategy*. It trains two phrase tables on source-pivot corpus and pivot-target corpus respectively. Then, it uses the rules with the same pivot side to induce a new rule. To limit rule table size, we only keep top M best rules, so as to reduce computational cost.

Our main contributions are as follows. Firstly, we are the first to apply pivot translation strategies on Chinese-Japanese patent SMT translation. Though similar strategies have been implemented, most of them are applied on language pairs which are from the same nature. As far as we know, no one has applied pivot translation strategies on Chinese-Japanese patent translation. Secondly, we make use of three patent corpora which are independent of each other, due to the fact that multilingual corpora are usually not easy to exploit, while others usually use corpora in which the sentences are aligned to each other across all languages, such as Europarl (Koehn, 2005). Besides, as we use large Chinese-English and English-Japanese corpora to help Chinese-Japanese SMT translation, we figure out approaches to make these pivot translation strategies practicable on such big data set. Finally, we implement three pivot translation strategies and apply minimum bayes risk (MBR) system combination on the translation results to further improve translation quality, which achieves an absolute improvement of 4.25 BLEU4 (Papineni et al., 2002) points over baseline system.

The rest of this paper is organized as follows. We describe related work making use of pivot languages (Section 2), and introduce direct SMT system and three kinds of pivot translation strategies, as well as minimum bayes risk system combination (Section3). Then, we present our experimental data and pivot translation strategy results (Section 4). Discussion on our work is in Section 5. The last section draws our conclusion and future work.

2 Related work

Pivot languages have been used for different purposes. Gollins and Sanderson (2001) used multiple pivot languages to improve cross language information retrieval. Ramirez et al. (2008) makes use of existing English resources as a pivot language to create a trilingual Japanese-Spanish-English thesaurus. Wang et al. (2006) improved word alignment for scarce-resourced languages pairs using bilingual corpora of pivot languages. Zhao et al. (2008) extracted paraphrase patterns from bilingual parallel corpora with a pivot approach.

Concerning the contribution of pivot languages to SMT, researchers have done a lot of work on it. Al-Hunaity et al. (2010) used English as pivot language to enhance Danish-Arabic SMT. Babych et al. (2007) compared the direct translation method with pivot translation strategy and confirmed that better translation quality could be achieved with pivot translation strategy. Bertoldi et al. (2008) provided theoretical formulation of SMT with pivot languages and introduced new methods for training alignment models through pivot languages. Costa-jussa et al. (2011) implemented two pivot translation strategies (the cascade system and the pseudo corpus) and performed a combination of these strategies to outperform the direct translation system. Habash and Hu (2009) compared two pivot translation strategies and gave an error analysis on their best system to show improvement. Utiyama and Isahara (2007) implemented two pivot strategies (phrase translation and sentence translation) and did experiments on the Europarl corpus to evaluate system performance. Wu and Wang (2009) revisited three pivot translation strategies and employed a hybrid method to combine RBMT and SMT systems, which significantly improved translation quality. Paul and Sumita (2011) exploited eight factors that affect the quality of pivot language and investigated the impact of these factors on pivot translation performance.

To the best of our knowledge, we are the first to apply pivot translation strategies on Chinese-Japanese patent translation. We implement three pivot translation strategies and perform a sentence level system combination on different translation results to further improve translation quality.

3 Direct phrase-based SMT and pivot translation strategies

3.1 Direct phrase-based SMT

Moses¹ is a freely available statistical machine translation system, which is also the most popular open-source platform for researchers working on SMT. Currently, Moses offers two types of translation models: phrase-based translation model (Koehn et al., 2003) and tree-based translation model. We use phrase-based Moses to build up our direct phrase-based SMT system.

In phrase-based SMT model, there are mainly three kinds of translation resources: translation rule table, language model and reordering table. Both translation rule table and reordering table are learnt from segmented sentence aligned bilingual corpus. Language model is learnt from target monolingual corpus. We employ the phrase-based Moses which uses different feature functions, such as direct phrase translation probability, inverse phrase translation probability, direct lexical weighting, inverse lexical weighting, phrase penalty, language model, distance penalty, word penalty, distortion weights et al. Feature weights are tuned on development set by Minimum Error Rate Training (MERT) (Och, 2003), using BLEU as the objective function.

When translating a source sentence f into target sentence e , the source sentence f is firstly segmented into phrases. Each phrase can be translated into different target language phrases. Phrases can be reordered. The system chooses the output \hat{e} which satisfies

$$\begin{aligned} \hat{e} &= \arg \max_e \Pr(e | f) \\ &= \arg \max_e \sum_{m=1}^N \lambda_m h_m(e, f) \end{aligned} \quad (1)$$

where λ_m denotes feature weights and $h_m(e, f)$ denotes feature functions used in phrase-based Moses.

3.2 Corpus enrichment strategy

A straightforward strategy to improve translation quality is to enrich the training corpus of the direct

translation system. However, it is not always convenient for us to collect such bilingual parallel data. Instead, we can generate source-target corpus by either translating the pivot side of source-pivot corpus into target language, or translating the pivot side of pivot-target corpus into source language, given the translation systems built on already available source-pivot corpus and pivot-target corpus respectively. For corpus translation, we can also make use of publicly available statistical machine translation systems such as Google translator et al.

In this paper, we employ Google translator API to translate the pivot side of source-pivot corpus and pivot-target corpus. One problem is that the translation process may take a long time due to our corpus size and disturbance from Google translator. Meanwhile, too many sentence pairs constructed by machine translation are not always promising because of the not-that-good translation quality of SMT systems. We should take in a reasonable size of qualified corpus to keep a balance of efficiency and effect.

We can select an amount of sentences according to sentence value which distinguishes different sentences. After that, we translate the selected sentences and add the translated parallel corpus into original training data in direct translation system. Then, we train a new system with the enriched corpus.

The sentence value is measured by sentence similarity shown in Equation (2).

$$\begin{aligned} &sentSimi(sent1, sent2) \\ &= \left(\left(\frac{count}{len(sent1)} \right)^{-1} + \left(\frac{count}{len(sent2)} \right)^{-1} \right)^{-1} \\ &= \frac{count}{len(sent1) + len(sent2)} \end{aligned} \quad (2)$$

where *count* denotes the number of shared words in the two sentences, $len(sent1)$ and $len(sent2)$ denote the length of the two sentences respectively.

We can take in sentence pairs part by part to see the influence of corpus size on machine translation quality. We believe corpus enrichment strategy can improve SMT system performance as it makes use of more translation resources.

¹ <http://www.statmt.org/moses/>

3.3 Sentence pivot translation strategy

In sentence pivot translation strategy, there must be available source-pivot and pivot-target translation systems. A source sentence s is firstly translated into n pivot sentences $p_i (i=1,2\dots n)$. Then, all pivot sentences are translated into $n \times m$ target sentences $t_{ij} (i=1,2\dots n; j=1,2\dots m)$. We choose the best translation among the $n \times m$ candidates for source sentence by employing the method described in (Utiyama and Isahara, 2007). The process is shown in Figure 1.

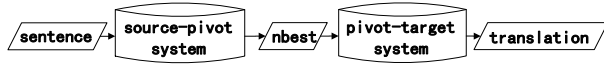


Figure 1. sentence pivot translation strategy

Suppose we use M and N features in source-pivot and pivot-target SMT systems which are $h_i^{sp} (i=1,2\dots M)$ and $h_j^{pt} (j=1,2\dots N)$ respectively, the score of target translation t_{ij} is defined as

$$S(t_{ij}) = \sum_{k=1}^M (\lambda_k^{sp} h_k^{sp}(s, p)) + \sum_{k=1}^N (\lambda_k^{pt} h_k^{pt}(p, t)) \quad (3)$$

where λ_k^{sp} and λ_k^{pt} are feature weights tuned on development set by MERT.

The best translation is that with the highest score

$$\hat{t} = \arg \max_t (S(t_{ij})) \quad (4)$$

3.4 Phrase pivot translation strategy

In phrase pivot translation strategy, a new phrase table T_{st} is generated from two existing phrase tables: one is source-to-pivot phrase table T_{sp} , the other is pivot-to-target phrase table T_{pt} . If the pivot side of two translation rules in these two tables are the same, these two rules can generate a new rule, in which the source side is the source side of the source-pivot rule and the target side is the target side of the pivot-target rule.

According to (Utiyama 2007), we estimate phrase and lexical translation probabilities for each rule as follows.

$$p(s | t) = \sum_{p \in T_{sp} \cap T_{pt}} p(s | p) p(p | t) \quad (5)$$

$$p(t | s) = \sum_{p \in T_{sp} \cap T_{pt}} p(t | p) p(p | s) \quad (6)$$

$$\phi(s | t) = \sum_{p \in T_{sp} \cap T_{pt}} \phi(s | p) \phi(p | t) \quad (7)$$

$$\phi(t | s) = \sum_{p \in T_{sp} \cap T_{pt}} \phi(t | p) \phi(p | s) \quad (8)$$

Here, $p(s | t)$ and $p(t | s)$ are phrase translation probabilities. $\phi(s | t)$ and $\phi(t | s)$ are lexical translation probabilities. $p \in T_{sp} \cap T_{pt}$ means pivot phrase p is included in T_{sp} as target side, and in T_{pt} as source side.

In phrase pivot translation strategy, the size of generated new rule table depends on the number of common phrases in target-side of T_{sp} and source-side of T_{pt} . If the number of phrase p in target side of T_{sp} is N , and in source side of T_{pt} is M , we may get $N * M$ rules maximally. The frequencies of the top 15 commonest rules in T_{sp} and T_{pt} are shown in table 1.

target-side of T_{sp}	frequency	source-side of T_{pt}	frequency
the	446189	the	848951
,	390232	,	471986
and	357239	a	309369
of	277004	of	251167
.	263823	and	250847
a	200072	to	231264
to	186682	is	191362
is	179179	in	179264
for	147076	.	145182
-	127692	, the	103469
in	123632	an	86151
with	90840	of the	82243
which	70257	by	82019
are	69505	-	77824
by	62827	, and	77554

Table 1: frequency of top 15 commonest rules in T_{sp} and T_{pt}

Corpus	Sentence pairs		Words	
			Source	Target
Chinese-Japanese (CJ)	Training set	105615	879953	1010620
	Tuning set	500	4674	5969
	Test set	1000	18552	18348/ 19122
Chinese-English (CE)	Training set	6174088	110116118	121837549
	Tuning set	1000	15963	17486
	Test set	1000	19465	17337/ 18456/ 17429
English-Japanese (EJ)	Training set	3159152	107601189	123917909
	Tuning set	1000	34171	40338
	Test set	1000	34342	38866

Table 2: Corpus details. For CJ, CE and EJ test set, we have two/three/one reference respectively

Here, we can limit the size of rule table by setting up a number limit K to filter low quality rules. We only keep the top K rules for the new rule table. The quality of the rules in the new rule table is measured by summarizing its translation and lexical probabilities.

$$Q(rule) = p(s|t) + p(t|s) + \phi(s|t) + \phi(t|s) \quad (9)$$

3.5 System combination

We use sentence level system combination to further improve the translation quality. Sentence level combination selects the best translation out from an N-best list and does not generate new translations.

With the 1-best translation results generated by direct translation system and different pivot systems, we can construct an N-best list for the source corpus. We employ MBR as a post-process to calculate the final translation.

$$E_{mbr} = \arg \min_{E'} \sum_E P(E|F)L(E, E') \quad (10)$$

where $P(E|F)$ is the posterior probability of candidate translation E , and $L(E|E')$ is the loss function. Here, we consider all the candidate translations equal, so $P(E|F)$ is a constant and can be omitted. We use $1-BLEU$ as the loss function. Thus, Equation 10 can be rewritten as

$$E_{mbr} = \arg \min_{E'} \sum_E (1-BLEU(E, E')) \quad (11)$$

$BLEU(E, E')$ is sentence level BLEU score.

4 Experiments

4.1 Datasets

We performed experiments on Chinese-Japanese (CJ), Chinese-English (CE), and English-Japanese (EJ) corpora. Corpus details are described in table 2. The training and tuning set of CJ corpus were collected from patent title and abstracts, so the sentences are quite short, while the 1000 sentence pairs of test data were extracted from patent contents, which are nearly twice as long as the ones in training and tuning set. For the CE corpus, training set consists of an in house corpus, and 1 million sentence pairs from NTCIR2011. We extracted the tuning set and test set from the training set. The EJ corpus is from NTCIR2011.

Beside these standard corpora, we also employed Google translator to translate the English side of the EJ corpus into Chinese, so as to construct a flawed CJ corpus. This flawed CJ corpus was used to enrich the original CJ corpus.

We used ICTCLAS (Zhang et al., 2003) to segment all Chinese corpora and standard Moses tokenizer to tokenize all English corpora. Mecab (Kudo 2006) was used to segment all Japanese corpora. We used GIZA++ to generate word alignment and training scripts in Moses to extract phrase pairs with maximum length 7. We employed Moses decoder to do translation with its default settings. We used Minimum Error Rate Training to tune the feature weights. SRILM (Stolcke, 2002) was employed to train a 5-gram language models with all Japanese corpus in CJ corpus and EJ corpus. Case insensitive BLEU4 was used to measure system quality.

4.2 Direct translation

We built a phrase-based Chinese-Japanese patent translation system on Chinese-Japanese corpus with Moses. As the training corpus only contained 105615 sentence pairs and most of them were rather short, the translation quality of the system was quite low, as shown in table 3.

	BLEU4
Direct translation	10.05

Table 3: BLEU of direct translation system

The direct translation system had a low quality because of the lack of training data, as well as the data quality problem as the training sentences were extracted from patent title and abstract, which were quite short and contained limited words, while the test data was from main context of patent documents.

We compared system performance with this baseline system in terms of BLEU4 scores. The percentages in later tables are relative to the BLEU4 score of this direct translation system.

4.3 Corpus enrichment

We used Google translator to translate the English side of the English-Japanese corpus into Chinese, so that to construct a Chinese-Japanese corpus, to enrich training data in 4.1. The reason why we translated English side in EJ corpus into Chinese, but not English side in CE corpus into Japanese was that we believed translation quality was much better for E-C translation than E-J translation, so the corpus we got by translating English into Chinese would be of better quality. After filtering the corpus, we got 2846799 sentence pairs.

We added the new corpus into training data in 4.1 and trained another translation system. The translation quality of this new system was measured by BLEU4 as follows.

	BLEU4	
Corpus Enrichment-All	9.22	-8.26%

Table 4: BLEU of corpus enrichment strategy

To our disappointment, adding the entire corpus into the original training corpus did not improve system performance. Contrarily, BLEU4 decreased

by 0.83. Still, this result was acceptable after we looked into the new corpus. Due to SMT system limit, the new corpus introduced in more noise than knowledge.

We ranked the sentences according to sentence value and added corpus step by step into original training corpus. Then we retrained the Moses system. The results are shown in table 5.

Corpus size added	BLEU4	
+100K	10.17	+1.19%
+200K	10.24	+1.89%
+300K	10.36	+3.08%
+400K	11.11	+10.55%
+500K	12.86	+27.96%
+600K	9.91	-1.39%
+700K	9.09	-9.55%

Table 5: BLEU of corpus enrichment strategy

As we added in more data, BLEU score improved slowly until it reached a peak point where we added in 500K sentence pairs. Then BLEU score decreased. Since we had ranked the sentences according to sentence value, we didn't test the rest sentences. We took this as the best result for corpus enrichment strategy.

4.4 sentence pivot translation strategy

We built two SMT systems for Chinese-English and English-Japanese translation with CE and EJ corpus respectively. Translation quality of these two systems was measured in terms of BLEU4 as shown in table 6.

	BLEU4
Chinese-to-English	27.84
English-to-Japanese	31.85

Table 6: BLEU of CE and EJ SMT system

For Chinese-Japanese translation, we first used Chinese-English system to translate Chinese into English. Then we used English-Japanese system to translate English into Japanese. According to Utiyama and Isahara (2007), the improvement of sentence pivot translation strategy with $n = 15$ is not significant compared to that with $n = 1$, so we kept 1 best translation for each sentence. The results are shown in table 7.

BLEU4	
9.91	-1.39%

Table 7: BLEU of sentence pivot translation strategy

As we can see from table 7, due to error accumulation, translation quality decreased a lot from BLEU4 10.05 to BLEU4 9.91. So sentence pivot translation strategy failed to improve translation quality in our experiments.

4.5 phrase pivot translation strategy

We trained two rule tables respectively on CE and EJ corpus. For each CE rule, we found the rule with the same English side in EJ rule table, and generated a new rule with C side of CE rule and J side of EJ rule. Each probability of the CJ rule was computed by minus the corresponding probabilities in CE rule and EJ rule, assuming these probabilities are independent. We kept 20 Japanese candidates for each Chinese phrase at most, and obtained a CJ rule table with 433276 rules.

We added these rules into the original rule table in direct translation system and returned the system. The results are shown in table 8.

	BLEU4	
phrase pivot	13.65	+35.82%

Table 8: BLEU of sentence pivot translation strategy

As we can see from table 8, introducing in more rules could obviously improve translation quality.

4.6 system combination

For each sentence in test set, we could get four different translation results from direct translation system and three pivot systems. We used sentence level system combination to get the final best translation. After system combination, the results are shown in table 9.

	BLEU4	
System combination	14.30	+42.29%

Table 9: BLEU of system combination

As we can see in table 9, system combination could improve translation quality significantly by 4.25 BLEU4 points compared to baseline 10.05. This is also the best result we could ever obtain.

5 Discussions and Analysis

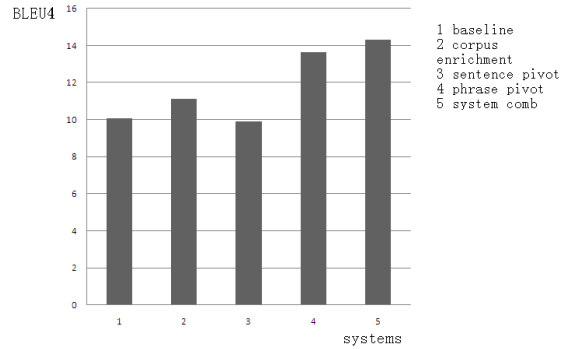


Figure 2. main results of different systems

Figure 2 shows the best machine translation performance of five different systems: baseline system, corpus enrichment system, sentence pivot translation system, phrase pivot translation system and a combined system. As we can see from Figure 2, baseline system performs better than sentence pivot translation system, while corpus enrichment system surpasses baseline system. Phrase pivot translation system obtained better BLEU score than corpus enrichment system. The combined system beat all other systems and achieved the best result. Thus, Figure 2 indicates that

systemcomb > phrase pivot > corpus enrichment > baseline > sentencepivot

where > means the system at the left hand side of it performs better than the one at the right hand side.

The reason why corpus enrichment system and phrase pivot translation system surpassed baseline system was mainly because they introduced in more translation resources into baseline system. As phrase pivot translation system introduced in selected translation rules from all pivot corpora, while corpus enrichment system only introduced in limited selected sentences, phrase pivot translation system achieved a better result. Sentence pivot translation system failed to improve translation quality, as it didn't make use of the original CJ training data, but translated the sentences only with the CE and EJ data. Its performance was also influenced by accumulative error during translation. System combination overtook all other systems as it selected the best translation from these systems for each sentence.

Source sentence	深水区域水底筑堤(坝)施工技术
English reference	Embankment (dam) construction technology at the bottom of deepwater area
Reference	深海地域の水中堤防(ダム)建設技術
Baseline result	深い水でのエリア()施工技術
System comb	深い水での 地域の海底堤防(ダム) 施工技術
Source sentence	画三条斜线处为透明或半透明材料。
English reference	Transparent or semitransparent materials are signed with three oblique lines.
Reference	斜線を描くところは透明あるいは半透明材料である。
Baseline result	絵のために三条で透明あるいは半透明の材料
System comb	画面三条 斜線 が透明あるいは半透明の材料。
Source sentence	气相制备芳族聚异氰酸酯化合物的方法
English reference	Preparation of aromatic polyisocyanate compounds in gaseous phase
Reference	ガスで芳香族化合物を作り出す方法
Baseline result	気相調製族「 聚ヘエステル 化合物の方法
System comb	気相調製 芳香族ポリイソシアネート 化合物の方法
Source sentence	过滤装置由合成树脂制成,具有重量轻和机械强度高特点。
English reference	Filtration unit is made of synthetic resin, with the characteristics of light weight and high mechanical strength.
Reference	フィルタは合成樹脂から作られ、軽量と高い機械強度の特徴がある。
Baseline result	フィルタリング装置ルーティング持つで作成した、 軽重量と機械强度高 の
System comb	フィルタリング装置ルーティングする 合成樹脂 と、は重量が 軽く と機械强度高の 正常特性 。
Source sentence	本发明涉及相当纯的粉状甘露糖醇,其在试验1中具有适中的、不过分的脆性,为40-80%
English reference	The invention relates to a very pure powder mannitol, with a modest brittleness of 40-80% in experiment 1
Reference	本発明は純の粉上マンノース糖に関し、実験1の中にあるところあいのもの、40-80%である。
Baseline result	ブツクの純粋なにかかわる粉末状のアルコールそのが試験1の中の、 不以上持つ の脆性のために
System comb	ブツクの 発明 ほぼ純粋な粉末状かかわるマンニトールそのが試験1の中、 適度 の、不以上 40~80 の脆性

Figure 3. Examples of Chinese-Japanese translation results. The differences between baseline result and our best result are highlighted in **bold**. English references are given to ease readability.

Figure 3 shows some translation examples of baseline system and system combination. As we can see from the examples, the results of system combination recognized more lexicons and achieved better translation quality.

6 Conclusions and Future Work

In this paper, we implemented three strategies (corpus enrichment, sentence pivot translation, phrase pivot translation) to make use of pivot

languages to help statistical machine translation. We also introduced approaches to make these strategies practicable on large data set. MBR sentence level system combination was employed to further improve translation quality. We applied these strategies on Chinese to Japanese patent translation using English as a pivot language. The results showed that corpus enrichment and phrase pivot translation strategies both could improve SMT quality, while sentence pivot translation

failed. After employing MBR sentence level system combination, we achieved significant improvement of SMT quality by 4.25 points in terms of BLEU. This is an absolute improvement over baseline.

Our future work would focus on exploiting pivot strategies on more advanced models (such as HPB model) to further improve Chinese-Japanese patent translation quality. Also, we would like to enhance our pivot strategies. We believe that phrase pivot translation strategy is quite promising and we would obtain more useful translation rules through phrase pivot strategy. Besides, we plan to collect more Chinese-Japanese patent corpus as the currently available corpus size is still too small. The corpus obtained would enrich the training data so as to help the learning process. We aim at high quality in Chinese-Japanese patent translation.

Acknowledgments

We would like to thank to Zhongguang Zheng, Naisheng Ge and Yiwen Fu for their helpful discussions. We also thank the anonymous reviewers for their insightful comments.

References

- Mossab Al-Hunaity, Bente Maegaard, Dorte Hansen. 2010. Using English as a pivot language to enhance Danish-Arabic statistical machine translation. In Proc. of LREC 2010: Workshop on Language Resources and Human Language Technology for Semitic Languages, pages 108-113.
- Bogdan Babych, Anthony Hartley, Serge Sharoff. 2007. Translating from under-resourced languages: comparing direct transfer against pivot translation. In Proc. of MT Summit XI, pages 29-35
- Nicola Bertoldi, Madalina Barbaiani, Marcello Federico, Roldano Cattoni. 2008. Phrase-based statistical machine translation with pivot languages. In Proc. of IWSLT 2008: Proceedings of the International Workshop on Spoken Language Translation, pages 143-149
- Mauro Cettolo, Nicola Bertoldi, Marcello Federico. 2011. Bootstrapping Arabic-Italian SMT through comparable texts and pivot translation. In Proc. of the 15th conference of the European Association for Machine Translation, pages 249-256.
- Marta R. Costa-jussà, Carlos Hernández, Rafael E. Banchs. 2011. Enhancing scarce-resource language translation through pivot combinations. In Proc. of the 5th International Joint Conference on Natural Language Processing, pages 1361-1365.
- Tim Gollins, Mark Sanderson. 2001. Improving cross language retrieval with triangulated translation. In Proc. of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pages 90-95
- Nizar Habash, Jun Hu. 2009. Improving Arabic-Chinese statistical machine translation using English as pivot language. In Proc. of the Fourth Workshop on Statistical Machine Translation, pages 173-181.
- Philipp Koehn. 2005. Europarl: a parallel corpus for statistical machine translation. MT Summit X, pages 79-86.
- Phillip Koehn, Franz Josef Och, Daniel Marcu. 2003. Statistical phrase-based translation. In Proc. of NAACL03.
- Taku Kudo. 2006. Mecab: yet another part of speech and morphological analyzer. <https://code.google.com/p/mecab/>
- Gregor Leusch, Aurélien Max, Josep Maria Crego, Hermann Ney. 2010. Multi-pivot translation by system combination. In Proc. of the 7th International Workshop on Spoken Language Translation, pages 299-306.
- Wen Li, Lei Chen, Wudaba, Miao Li. 2010. Chained machine translation using morphemes as pivot language. In Proc. of the 8th Workshop on Asian Language Resources, pages 169-177.
- Bin Lu, Benjamin K. Tsou, Jingbo Zhu, Tao Jiang, & Oi Yee Kwong. 2009. The construction of a Chinese-English patent parallel corpus. MT Summit XII: Third Workshop on Patent Translation, pages 17-24
- Dragos Stefan Munteanu, Daniel Marcu. 2005. Improving machine translation performance by exploiting non-parallel corpora. Computational Linguistics, 31(4), pages 477-504
- Franz Josef Och. 2003. Minimum Error Rate Training for Statistical Machine Translation. In Proc. of the 41st Annual Meeting of the ACL, pages 160-167.
- Franz Josef Och, Hermann Ney. 2003. A systematic comparison of various statistical alignment models. Computational Linguistics, 29(1).
- Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In Proc. of the 40th Annual Meeting on Association for Computational Linguistics, pages 311-318.

- Michael Paul, Andrew Finch, Paul R. Dixon, Eiichiro Sumita. 2011. Dialect translation: integrating Bayesian co-segmentation models with pivot-based SMT. In Proc. of DIALECTS2011: Proceedings of the First Workshop on Algorithms and Resources for Modeling of Dialects and Language Varieties, pages 1-9.
- Michael Paul, Eiichiro Sumita. 2011. Translation quality indicators for pivot-based statistical MT. In Proc. of the 5th International Joint Conference on Natural Language Processing, pages 811-818.
- Jessica Ramirez, Masayuki Asahara, Yuji Matsumoto. 2008. Japanese-Spanish thesaurus construction using English as a pivot. In Proc. of IJCNLP 2008: Third International Joint Conference on Natural Language Processing, pages 473-480.
- Philip Resnik, Noah A. Smith. 2003. The web as a parallel corpus. *Computational Linguistics*, 29(3), pages 349-380
- Andreas Stolcke. 2002. Srilm-an Extensible Language Modeling Toolkit. In Proc. of the International Conference on Spoken Language Processing, pages 901-904.
- Rie Tanaka, Yohei Murakami, Toru Ishida. 2009. Context-based approach for pivot translation services. In Proc. of IJCAI-09: Twenty-first International Joint conference on Artificial Intelligence, pages 1555-1561.
- Masao Utiyama, Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based statistical machine translation. In Proc. of Human Language Technology: the conference of the North American Chapter of the Association for Computational Linguistics, pages 484-491
- Haifeng Wang, Hua Wu, Zhanyi Liu. 2006. Word alignment for languages with scarce resources using bilingual corpora of other language pairs. In Proc. of COLING/ACL on main conference poster sessions, pages 874-881
- Hua Wu, Haifeng Wang. 2009. Revisiting pivot language approach for machine translation. In Proc. of the 47th Annual Meeting of the ACL and the 4th IJCNLP, pages 154-162.
- Hua Wu, Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. In Proc. of the 45th Annual Meeting of the Association for Computational Linguistics, pages 856-863
- Huaping Zhang, Hongkui Yu, Deyi Xiong, Qun Liu. 2003. HHMM-based Chinese lexical analyzer ICTCLAS. In Proc. of the second SIGHAN workshop on Chinese language processing, pages 184-187.
- Shiqi Zhao, Haifeng Wang, Ting Liu, Sheng Li. 2008. Pivot approach for extracting paraphrase patterns from bilingual corpora. In Proc. of HLT 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 780-788.