Literature Survey: Pivot-based Machine Translation

Rohit More
IIT Bombay
rohitmore@cse.iitb.ac.in

Current statistical machine translation systems heavily rely on the availability of parallel corpora between the language pair involved. The good quality parallel corpus is not always available. This creates a bottleneck. One solution to solve this bottleneck is to introduce third language, named *pivot* language for which there exist good quality source-pivot and pivot-target bilingual corpora.

1 Related work

There is a substantial amount of work done in the area of pivot strategies for SMT. (De Gispert and Marino, 2006) talk about translation task between Catalan and English with use of Spanish as a pivot language. Pivoting is done using two techniques- concatenation of two SMT systems and direct approach in which Catalan-English corpus is generated and trained upon. In (Utiyama and Isahara, 2007), the authors inspect the use of pivot language through - phrase translation (phrase table creation) and sentence translation. (Wu and Wang, 2007) discuss three methods for pivot strategies namely - phrase translation (*i.e.* triangulation), transfer method and synthetic method. (Nakov and Ng, 2012) try to exploit the similarity between resource-poor languages and resource-rich languages for the translation task. (Dabre et al., 2014) used multiple decoding paths (MDP) to overcome the limitation of small sized corpora.(Paul et al., 2013) debates over criteria to be considered for selection of good pivot language. Use of source-side segmentation as pre-processing technique is demonstrated by (Kunchukuttan et al., 2014). (Goldwater and McClosky, 2005) investigates several methods for incorporating morphological information to achieve better translation from Czech to English.

Pivot strategies mentioned above focus on the situation of resource-poor languages where direct translation is either very poor or not available. Our approach, on the other hand, tries to employ pivot strategy to help improve the performance of existing direct MT system. Our attempt to integrate word segmentation with pivot strategies is first of a kind.

2 Approaches to Pivot based MT

There are methods by which the resources of pivot language can be utilized as explained in (Wu and Wang, 2009) - namely

1. Sentence Translation or Transfer Method
2. Synthetic corpus synthesis
3. Phrase table construction or Triangulation Approach

These methods are explained in brief in following sections.

2.1 Sentence Translation or Transfer Method

The transfer method first translates the source language into pivot language using source-pivot translation system, and then from pivot language to target language through the pivot-target translation system. Given a source sentence S, we can translate it into n pivot language sentences $P_1, P_2, P_3, ...P_n$ using a source-pivot translation system. Each of these n sentence, P_i then can be translated into m target language sentences $T_{i1}, T_{i2}, T_{i3}, ...T_{im}$ using pivot-target translation system. Thus, in total we will have $m \times n$ target language sentences.These sentences can then be re-scored using source-pivot and pivot-target translation scores according to method described in (Utiyama and Isahara, 2007)

If we denote source-pivot system features as h^{sp} and pivot-target features as h^{pt}, the best scoring translation is calculated using equation:

$$\hat{t} = \underset{t}{\text{argmax}} \sum_{k=1}^{L} \left(\lambda_k^{sp} h_k^{sp} (s, p) + \lambda_k^{pt} h_k^{pt} (p, t) \right) \tag{1}$$

Where, L is the number of features used in SMT systems and λ^{sp}, λ^{pt} are feature weights.

2.2 Corpus Synthesis

In order to obtain source-target corpus, there are two ways. One is, we can translate pivot language sentences from source-pivot corpus into target language sentences using the pivot-target system. The other way is, translation of pivot sentences from the pivot-target corpus into source sentences using pivot-source system.

The source-target corpora created using above two methods can then be combined to produce a final synthetic corpus.

2.3 Triangulation or Phrase table induction

The method of triangulation is described in (**?**). In this method, we train source-pivot models and pivot-target models using source-pivot and pivot-target corpora respectively. Using these two models created so far, we induce a source-target model. The two important components to be induced are - 1) phrase translation probability and 2) lexical weight.

Phrase translation probability is induced on the basis of assumption- that source and target phrases are conditionally independent when conditioned on pivot phrases. It can be given as,

$$\phi \left(\vec{s} | \vec{t} \right) = \sum_{\vec{p}} \phi \left(\vec{s} | \vec{p} \right) \phi \left(\vec{p} | \vec{t} \right) \tag{2}$$

Where, $\vec{s}$, $\vec{p}$, $\vec{t}$ are phrases in the languages L_s, L_p, L_t respectively.

Lexical Weight, according to (Koehn et al., 2003), depends on - 1) word alignment information a in a phrase pair (s, t) and 2) lexical translation probability $w(s|t)$.

To calculate lexical weight, the word alignment is induced from source-pivot and pivot-target alignment. Using the information from induced word alignment, lexical probabilities are estimated. Thus, lexical weight is calculated using induced alignment and estimated lexical probabilities.

We will take a detailed look at the mathematics behind triangulation approach

3 Mathematics of Triangulation Approach

This section will introduce the *triangulation* method that performs phrase-based SMT for the language pair $L_f - L_e$ by using two bilingual corpora of $L_f - L_p$ and $L_p - L_e$. Two translation models are trained for $L_f - L_p$ and $L_p - L_e$. Based on these models, a pivot translation model is built for $L_f - L_e$, with L_p as a pivot language. The details are extracted from Wu and Wang (Wu and Wang, 2007).

According to Equation **??**, only phrase translation probability, and the lexical weight are language dependent. They are introduced as follows:

3.1 Phrase Translation Probabilities

Using $L_f - L_p$ and $L_p - L_e$ bilingual corpora, we train two phrase translation probabilities $\phi \left(\vec{f_i} | \vec{p_i} \right)$ and $\phi \left(\vec{p_i} | \vec{e_i} \right)$, where p_i is the phrase in pivot language L_p. We obtain the phrase translation probability $\phi \left(\vec{f_i} | \vec{e_i} \right)$ according to the following model,

$$\phi \left(\vec{f_i} | \vec{e_i} \right) = \sum_{\vec{p_i}} \phi \left(\vec{f_i} | \vec{p_i}, \vec{e_i} \right) \phi \left(\vec{p_i} | \vec{e_i} \right) \tag{3}$$

The phrase translation probability $\phi \left(\vec{f_i} | \vec{p_i}, \vec{e_i} \right)$ does not depend on the phrase $\vec{e_i}$ in the language L_e, since it is estimated from the $L_f - L_p$ bilingual corpus.

Thus, equation 3 can be rewritten as

$$\phi \left(\vec{f_i} | \vec{e_i} \right) = \sum_{\vec{p_i}} \phi \left(\vec{f_i} | \vec{p_i} \right) \phi \left(\vec{p_i} | \vec{e_i} \right) \tag{4}$$

Are probability calculations correct?

Let us go step by step through the formulation of phrase translation probability $\phi \left(\vec{f_i} | \vec{e_i} \right)$.

First, we marginalize,

$$\phi \left(\vec{f_i} | \vec{e_i} \right) = \sum_{\vec{p_i}} \phi \left(\vec{f_i}, \vec{p_i} | \vec{e_i} \right) \tag{5}$$

Now we will use the chain rule,

$$\phi\left(\vec{f_i}|\vec{e_i}\right) = \sum_{\vec{p_i}} \phi\left(\vec{f_i}|\vec{p_i},\vec{e_i}\right)\phi\left(\vec{p_i}|\vec{e_i}\right) \tag{6}$$

Since, we have $L_f - L_p$ corpus available with us, the calculation of first term in the above equation will not depend on p *i.e.* $\phi\left(\vec{f_i}|\vec{p_i},\vec{e_i}\right)$ will now reduce to $\phi\left(\vec{f_i}|\vec{p_i}\right)$. Thus, the final equation will be,

$$\phi\left(\vec{f_i}|\vec{e_i}\right) = \sum_{\vec{p_i}} \phi\left(\vec{f_i}|\vec{p_i}\right)\phi\left(\vec{p_i}|\vec{e_i}\right) \tag{7}$$

3.2 Lexical Weight

According to (Koehn et al., 2003), lexical weight can be estimated using following model.

$$p_w\left(\vec{f}|\vec{e},a\right) = \prod_{i=1}^{n} \frac{1}{|j|\left(i,j\right)\in a|} \sum_{\forall(i,j)\in a} w\left(f_i|e_j\right) \tag{8}$$

In order to estimate lexical weight for our model, we first need to obtain the alignment information a between two phrases $\vec{f}$ and $\vec{e}$, and then estimate the lexical translation probability $w\left(f|e\right)$ according to the alignment information.

The alignment information for the phrase pair $\left(\vec{f},\vec{e}\right)$ can be induced from the two phrase pairs, $\left(\vec{f},\vec{p}\right)$ and $\left(\vec{p},\vec{e}\right)$. Let a_1 and a_2 be the word alignment information inside phrase pairs $\left(\vec{f},\vec{p}\right)$ and $\left(\vec{p},\vec{e}\right)$ respectively.

$$a = \{\left(f,e\right)|\exists p : \left(f,p\right)\in a_1 \& \left(p,e\right)\in a_2\} \tag{9}$$

With this induced alignment information, there exists a method to estimate the probability directly from the induced phrase pairs. This is *phrase* method. If we use K to denote the number of induced phrase pairs, we estimate co-occurring frequency of the word-pair $\left(f,e\right)$ according to the following model.

$$count\left(f,e\right) = \sum_{k=1}^{K} \phi_k\left(\vec{f}|\vec{e}\right) \sum_{i=1}^{n} \delta\left(f,f_i\right)\delta\left(e,e_{a_i}\right) \tag{10}$$

Where, $\phi_k\left(\vec{f}|\vec{e}\right)$ is phrase translation probability for phrase pair k.
$\delta\left(x,y\right) = 1 \;\; if x = y;\; otherwise\; 0$

Thus, lexical translation probability can be estimated as

$$w\left(f|e\right) = \frac{count\left(f,e\right)}{\sum_{f'} count\left(f',e\right)} \tag{11}$$

$w\left(f|e\right)$ can also be calculated using *word* method as,

$$w\left(f|e\right) = \sum_{p} w\left(f|p\right)w\left(p|e\right) sim\left(f,e;p\right) \tag{12}$$

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

3.3 Interpolated Model

If we have a small $L_f - Le$ parallel corpus, training a translation model on this corpus alone will result in the poorly performing system. The reason behind the poor performance is sparse data. In order to improve this performance, we can use additional $L_f - Lp$ and $L_p - Le$ parallel corpora. Moreover, we can also use more than one pivot languages to improve the translation performance. Different pivot language may catch different language phenomenon and can improve translation quality by adding quality $L_f - L_e$ phrase pairs.

If we include n pivot languages, n pivot models can be estimated as described in section 3. In order to combine all these models with the standard model trained with the $L_f - Le$ corpus, we use linear interpolation. The phrase translation probability and the lexical weight are estimated as shown in equation 13 and 14

$$\phi\left(\vec{f}|\vec{e}\right) = \sum_{i=0}^{n} \alpha_i \phi_i\left(\vec{f}|\vec{e}\right) \tag{13}$$

$$p_w\left(\vec{f}|\vec{e},a\right) = \sum_{i=0}^{n} \beta_i p_{w,i}\left(\vec{f}|\vec{e},a\right) \tag{14}$$

where, $\sum_{n}^{i=0} \alpha_i = 1$ and $\sum_{n}^{i=0} \beta_i = 1$

$\phi_0\left(\vec{f}|\vec{e}\right)$ and $p_{w,0}\left(\vec{f}|\vec{e},a\right)$ denote the phrase translation probability and lexical weight trained with the $L_f - Le$ corpus.

$\phi_i\left(\vec{f}|\vec{e}\right)$ and $p_{w,i}\left(\vec{f}|\vec{e},a\right)$ $(i = 1, 2, ..., n)$ are the phrase translation probability and lexical weight estimated by using pivot languages. α_i and β_i are interpolation coefficients.

4 Case Studies

4.1 Improving statistical machine translation for a resource disadvantaged language using related resource-rich languages

Due to recent developments in statistical machine translation (SMT), it is possible to build a prototype system for any language pair within hours. To achieve this, a large number of parallel sentence-aligned text is required. Such high-quality bi-texts are rare except for few pairs of languages.

Nakov and Ng (Nakov and Ng, 2012) propose a language-independent approach for improving machine translation for resource disadvantaged languages exploiting their similarity to resource-rich ones. In other words, we have a resource disadvantaged language (say X1) which is closely related to resource-rich language (say X2). X1 and X2 may have similarities in word order, vocabulary, spelling, syntax, etc. We improve translation from resource disadvantaged language X1 into resource-rich language Y using bi-text containing a limited number of parallel sentences for X1-Y and large bi-text for X2-Y. The approaches to achieve the same are discussed below.

4.1.1 Method

In order to use bi-text of one language on order to improve SMT for some related language, two general strategies are used - 1) bi-text concatenation with possible repetitions of original bi-text for balance and 2) phrase table combination, where each bi-text is used to build separate phrase table, and then two phrase tables are combined. We discuss these strategies below:

1. **Concatenating Bi-texts**

   In this approach, we can simply concatenate the bi-texts for X1-Y and X2-Y into one large bi-text. It can improve alignments obtained from X1-Y bi-text. This is because, additional sentences can provide context for rare words in that bi-text.

   Concatenation can also provide more source side translation options, thereby increasing lexical coverage and reducing the number of Out-Of-Vocabulary (OOV) words. It can also introduce new non-compositional phrases on source-side to increase the fluency. It also offers new target language phrases. Inappropriate phrases from X2 that do not exist in X1 will not match the test time input.

   However, this approach of simple concatenation can be problematic. Since the size of X2-Y bi-text is much higher than X1-Y bi-text, the former will dominate during word alignment and phrase extraction. This can affect lexical and phrase translation probabilities, thus yielding poor performance. This imbalance of bi-texts can be corrected by repeating smaller X1-Y bi-text several times so that large one does not dominate.

   Original and additional training bi-texts are combined in following ways.

   (a) **cat $\times$ 1** - Simple concatenation of original and additional bi-text to form a new training bitext, which is used to train a new phrase-based SMT system.
   (b) **cat $\times$ k** - Concatenation of k copies of original bi-text and one copy of additional bi-text to form a new training bi-text. The value of k is selected such that original bi-text approximately matches the size of additional bi-text.
   (c) **cat $\times$ k:align** - We concatenate k copies of original bi-text and one copy of additional bi-text to form a new training bi-text. Word alignments are generated from this new bi-text. Then all sentence pairs and word alignments are discarded except for one copy of original bi-text.

Thus, only word alignments from original bi-text are induced using additional statistical information from additional bi-text. These alignments are then used to build a phrase table.

2. **Combining Phrase Tables**

   The alternative way to use additional training bi-text is to build separate phrase tables. These phrase tables can be used together, merged, or interpolated.

   Phrase table construction method has many advantages. The phrase pairs extracted from X1-Y bi-text are clearly distinguished from riskier ones from X2-Y bi-text. The lexical and phrase translation probabilities are combined in proper manner. On the negative side, word-alignments for sentences in X1-Y bi-text are not improved as they were in first case. Below are the three phrase table construction strategies:

   (a) **Two tables** : Two separate phrase tables are build from two bi-texts. These tables are used as alternative decoding paths.

   (b) **Interpolation** : From original and additional bi-text, two separate phrase tables, T_{orig} and T_{extra}, are built. To combine corresponding conditional probabilities, linear interpolation is used - $Pr\left(e|s\right) = \alpha Pr_{orig}\left(e|s\right) + \left(1-\alpha\right) Pr_{extra}\left(e|s\right)$.
   The value of α is optimized over a development dataset.

   (c) **Merge** : From original and additional bi-text, two separate phrase tables, T_{orig} and T_{extra}, are built. We keep all source-target phrases from T_{orig}, adding to them those source-target phrase pairs from T_{extra} that were not present in T_{orig}. For each added phrase pair, the associated lexical and phrase translation probabilities are retained.

3. **Proposed Approach**

   The approach proposed here tries to take into account advantages and disadvantages of both the schemes discussed above. Improvement in word alignments for X1-Y bi-text is achieved by biasing word alignment process by considering additional phrases from X2-Y bi-text. It also tries to increase lexical coverage by considering additional phrases from X2-Y bi-text. The process is explained in brief below.

   (a) Use X1-Y bi-text k times and X2-Y bi-text one time to create balanced bi-text B_{rep}. Create word alignments for B_{rep} and truncate them keeping only once copy for X1-Y bi-text. By using these alignments, phrase table $T_{rep-trunc}$ is created.

   (b) By using the simple concatenation of one copy of X1-Y bi-text and one copy of X2-Y bi-text, create a bi-text called B_{cat}. Create word alignments for B_{cat} and build a phrase table T_{cat}

   (c) Now by making use of $T_{rep-trunc}$ and T_{cat}, create a new phrase table by using merging. The priority is given to $T_{rep-trunc}$ during merging.

4.1.2 Experiments and Analysis

In this paper, a number of experiments were done in order to test the similarity between the original (Indonesian and Spanish) and the auxiliary languages (Malay and Portuguese). Indonesian-English SMT is improved using Malay as auxiliary languages, while Spanish-English SMT is improved using Portuguese as pivot.

Various conclusions are drawn according to results of the experiments. It is clear that relative languages can help improve SMT. There was improvement of over 3 BLEU points in Spanish-English using Portuguese and around 1.5 BLEU points improvement in case of Indonesian-English using Malay.

Method of simple concatenation helps, but it can be problematic when additional sentences are way more than original. Concatenation works well if original bi-text is repeated enough number of times to match to the size of additional bi-text.

To give additional weighting to original phrases in merging method is a good strategy. Improvement in system is due to improvement in word alignment as well as due to increased lexical coverage.

4.2 Catalan-English Statistical Machine Translation without parallel corpus: Bridging through Spanish

This paper (De Gispert and Marino, 2006) discusses about experiments done for Catalan-English Statistical Machine Translation without an English-Catalan parallel corpus. Instead they make use of English-Spanish parallel corpus and Spanish-Catalan parallel corpus. Since, Spanish and Catalan are have close language proximity, promising results are achiever using Spanish as a bridge language.

4.2.1 Choice of Spanish as a bridge

Catalan is the Roman language spoken or understood by over 12 million people.In spite of this, there are not much parallel corpora available for Catalan-English. Spanish-English on the other side has large good quality parallel corpus. Spanish and Catalan belong to the same language family showcasing morphological and grammatical similarity. Thus, it is natural to exploit the use of Spanish as bridge language.

4.2.2 Bridging Strategies

In order to carry out English-Catalan machine translation, two strategies are implemented which are as follows:

1. **Sequential Strategy**

   This method simply concatenates two statistical machine translation systems, one between Catalan and Spanish, and the other between Spanish and English.

   This is an error additive approach, as errors from one system propagate to the input of following system.

2. **Direct Strategy**

   This strategy consists of translating the whole Spanish side of English-Spanish corpus into Catalan by using Spanish-to-Catalan SMT system, which is of a more general domain. Then, an English-Catalan is trained on this automatically translated *noisy* Catalan text. With this, the expectation is that the errors related to Spanish-Catalan system will not correlate with English test and may get very low probabilities when training English-Catalan system.

4.2.3 Baseline SMT system

The SMT system used for this experiments follows the maximum entropy framework. It maximizes log-linear combination of feature functions, as described in the following equation:

$$\hat{t}_1^I = \underset{t_1^I}{\operatorname{argmax}} \sum_{m=1}^{M} \lambda_m h_m \left(S_1^J, t_1^I\right) \quad (15)$$

Where, λ_m corresponds to weighting coefficients of log-linear combination, and the feature functions $h_m\,(s,t)$ to a logarithmic scaling of the probabilities of each model. For this approach, one translation model and four additional feature models are used. In contrast to standard phrase-based models, the translation model used here is a bilingual n-gram model expressed in tuples as bilingual units. The additional features used are - a target language model, a word bonus model, a source-to-target lexicon model, a target-to-source lexicon model.

Once these models are computed, the optimal values of logarithmic coefficients are estimated using a Simplex algorithm, an in-house implementation.

4.2.4 Results

The results, in general, show that automatic evaluation measures achieved in the Catalan-English task are similar to those of Spanish-English task. Also, sentence concatenation performs worse than direct strategy in both directions *i.e.* from English to Catalan and vice versa. Thus, nearly no loss is found when Spanish is used as a bridge to obtain the Catalan-English system. For automatic evaluation, Word Error Rate (WER), Position-independent word Error Rate (PER), and BLEU scores are used.

These experiments proved that it is possible to build a large-scale statistical machine translation system between Catalan and any other language as long as there is huge parallel corpus available between Spanish and that language.

The translation strategy discussed in this paper is limited to resource disadvantaged language which are closely related to resource-rich languages.

4.3 How to choose the best pivot language for automatic translation of low resource languages

When it comes to statistical machine translation, the first requirement is to have a good-quality parallel corpus. This situation is impossible for low resource languages such as Asian languages. Thus, recent research on multilingual statistical machine translation focuses on the use of pivot language for the translation of such resource disadvantaged languages. English, by its richness in language resources, comes first as possible pivot languages. In this paper (Paul et al., 2013), the effect of various factors on the choice of pivot language are studied.

Generally, the choice of pivot language is done on two criteria - 1) availability of language resources and 2) relatedness between source and pivot language. However, the preceding criteria might not be sufficient for choosing the best pivot language, especially for Asian languages. The recent resource shows that use of non-English language as pivot improves the system performance.

4.3.1 Coupling Strategies for pivot translation

Pivot translation is a translation from a source language to target language through an intermediate pivot language. Within SMT framework, the following strategies have been investigated.

1. *Cascading of two translation systems.*

   The first MT system translated the source language input to pivot language and second MT receives pivot language output from the first system as input. It then translated it into the target language output.

2. *Pseudo Corpus Approach*

   In the first part, a "noisy" corpus between source and target language is created by translating pivot language parts of the source-pivot corpus into the target language system using pivot-target MT system. In the second part, a single SMT system is trained on the "noisy" source-target corpus(De Gispert and Marino, 2006).

3. *Phrase-Table Composition*

   The translation models of source-pivot and pivot-target MT systems is combined into source-target MT system. This is done using the creation of source-target phrase table by merging source-pivot phrase table and pivot-target phrase table entries with identical pivot language phrases and multiplying posterior probabilities(Wu and Wang, 2007).

4. *Bridging at translation time*

   The coupling is integrated into the SMT decoding process by modeling the pivot text as hidden variable and assuming independence between source and target language sentences.

5. *Multi-pivot translation*

   Intermediate translations into several pivot languages are used to generate a final target language translation by a probabilistic combination of translation models or system combination techniques.

4.3.2 Language resources and diversity

The scope of this article was limited to the investigation of choice of pivot language using cascading of two translation systems explained above. In this article, 22 Indo-European and Asian languages are covered. For this, a multilingual *Basic Travel Expressions Corpus* (BTEC) is used.

The information of languages used is shown in table 1. These languages differ largely in word-order, segmentation unit and degree of inflection. OOV in table 1 is a percentage of OOV words while Length is average sentence length.

Language	Voc	Length	OOV	Order	Unit	Inflection
Danish	26.5K	7.2	1.0	SVO	word	high
German	25.7K	7.1	1.1	mixed	word	high
English	15.4K	7.5	0.4	SVO	word	moderate
Spanish	20.8K	7.4	0.8	SVO	word	high
French	19.3K	7.6	0.7	SVO	word	high
Hindi	33.6K	7.8	3.8	SOV	word	high
Italian	23.8K	6.7	0.9	SVO	word	high
Dutch	22.3K	7.2	1.0	mixed	word	high
Polish	36.4K	6.5	1.1	SVO	word	high
Portugese	20.8K	7.0	1.0	SVO	word	high
Brazilian Portugese	20.5K	7.2	1.0	SVO	word	high
Russian	36.2K	6.4	2.3	SVO	word	high
Arabic	47.8K	6.4	2.1	VSO	word	high
Indonesian	18.6K	6.8	0.8	SVO	word	high
Japanese	17.2K	8.5	0.5	SOV	none	moderate
Koren	17.2K	8.1	0.8	SOV	phrase	moderate
Malay	19.3K	6.8	0.8	SVO	word	high
Thai	7.4K	7.8	0.4	SVO	none	light
Tagalog	28.7	7.4	0.7	CSO	word	high
Vietnamese	9.9K	9.0	0.2	SVO	phrase	light
Chinese	13.3K	6.8	0.5	SVO	word	light
Taiwanese	39.5K	5.9	0.6	SVO	word	light

Table 1: Language Resource(BTEC)

4.3.3 Observations for pivot language selection

The results of the experiment show a large variation in BLEU score for all pivot languages, indicating that there is not a single best pivot language. The quality of given translation system largely depends on the respective source and target languages. For Indo-European pivot languages, the best language combination scores are generally higher than the ones obtained for Asian pivot languages.

For languages that are closely related, such as Portuguese and Brazilian Portuguese, the related language should be chosen as pivot language while translating from or into the respective language. The results, in general, show that pivot languages closely related to source language have a larger impact on overall pivot translation quality than pivot languages related to target language.

For Indo-European-only language pairs, only Indo-European languages perform well as pivots.

4.3.4 Indicators of Pivot Translation quality

Based on the observations made in experiments, below eight factors are identified which make a language effective pivot language for given language pair.

1. *Language Family* : A binary feature indicating whether source and target language belong to the same language family.

2. *Vocabulary* : The training data vocabulary size of source and target languages, the ratio of source and target vocabulary sizes, and the overlap between source and target vocabulary.

3. *Sentence length* : The average sentence length (computed in terms of words) of source and target training sets and the ratio of source and target sentence length.

4. *Reordering* : The amount and span of word order differences (reordering) in the training data.

5. *Language Perplexity* : The perplexity of utilized language models.

6. *Translation model entropy* : The amount of uncertainty involved in choosing candidate translation phrases.

7. *Engine performance* : The BLEU scores of the respective SMT engines used for the pivot translation experiments.

8. *Monotonicity* : The BLEU score difference of a given SMT engine for decoding with and without a reordering model.

4.3.5 Summary

To summarize, the effects of using non-English pivot languages for translations between 22 Indo-European and Asian languages were compared. The source language features are preferable for heterogeneous language pairs while target language-related features are focused more in case of homogeneous language pairs.

References

Raj Dabre, Fabien Cromieres, Sadao Kurohashi, and Pushpak Bhattacharyya. 2014. Leveraging small multilingual corpora for smt using many pivot languages.

Adrià De Gispert and Jose B Marino. 2006. Catalan-english statistical machine translation without parallel corpus: bridging through spanish. In *Proc. of 5th International Conference on Language Resources and Evaluation (LREC)*, pages 65–68. Citeseer.

Sharon Goldwater and David McClosky. 2005. Improving statistical mt through morphological analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 676–683. Association for Computational Linguistics.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics.

Anoop Kunchukuttan, Ratish Pudupully, Rajen Chatterjee, Abhijit Mishra, and Pushpak Bhattacharyya. 2014. The iit bombay smt system for icon 2014 tools contest. In *NLP Tools Contest at ICON 2014*.

Preslav Nakov and Hwee Tou Ng. 2012. Improving statistical machine translation for a resource-poor language using related resource-rich languages. *Journal of Artificial Intelligence Research*, 44(1):179–222.

Michael Paul, Andrew Finch, and Eiichrio Sumita. 2013. How to choose the best pivot language for automatic translation of low-resource languages. *ACM Transactions on Asian Language Information Processing (TALIP)*, 12(4):14.

Masao Utiyama and Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based statistical machine translation. In *HLT-NAACL*, pages 484–491.

Hua Wu and Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. *Machine Translation*, 21(3):165–181.

Hua Wu and Haifeng Wang. 2009. Revisiting pivot language approach for machine translation. 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: Volume 1-Volume 1*, pages 154–162. Association for Computational Linguistics.