Literature Survey: Study of Reordering in Pivot Based SMT

Harshad Chavan and Deepak Patil
CFILT, IIT Bombay
{harshadpc, deepakcp}@cse.iitb.ac.in

Pivot Based SMT solves the problem of scarcity of source-target parallel corpus by introducing a third resource rich 'pivot' language. Triangulation method in Pivot Based SMT is a method that uses the pivot language to induce new phrase pairs into the phrase table, this process is known as 'Phrase Table Triangulation'. Phrase Table Triangulation has been extensively studied by many researchers. This paper surveys the past work in Pivot Based SMT, specifically in Triangulation Method. It discusses in detail, the work done in Pivot Based SMT at IIT Bombay. It also surveys the work in some of the other areas which are important for Pivot Based SMT such as, System Combination Techniques, Domain Adaptation and various Reordering Models.

1 Introduction

Machine Translation refers to the problem of mechanization or automation of translation from one natural language to another. It is a sub-field of computational linguistics. The aim of Machine Translation is, "To build translation systems that can provide a good quality translation of natural language text with minimal or no human assistance"(Hutchins and Somers, 1992). Various approaches of Machine Translation deal with this problem differently. Statistical Machine Translation(SMT) is one of these approaches and it follows a completely data driven path. It makes use of a set of sentences which are translations of each other, known as 'Parallel Corpus'. By looking at a large amount of sentences from the parallel corpus, it "learns" how to translate from one language to another.

The quality of an SMT system depends heavily on the size and the quality of the parallel corpus. Larger parallel corpora have been observed to give better results in terms of translation quality, but for some pair of languages large parallel corpus may not be easily available. This problem of low or no availability of parallel corpus is tackled by Pivot Based SMT. Pivot Based SMT uses another language other than ones in the language pair to extract some additional information and help in improving the quality of the translations produced by an SMT system. The focus of this paper is on Pivot Based SMT, its details and past work in the related areas.

2 Pivot Based SMT

It has been repetitively emphasized by researchers that the quality of a Statistical Machine Translation system depends heavily on the availability of parallel corpus. The problem of low quality translation due to unavailability of parallel corpus is tackled by using pivot based approaches. Substantial amount of work has been done in Pivot Based Strategies by many researchers. Several approaches have been proposed for use of pivot languages. Many researchers have experimented with these approaches in order to compare them. This paper gives an overview of the basics of Pivot Based SMT, the mathematical formulation that forms its basis, and several related lines of work.

2.1 Methods for Pivot Based SMT

The basic idea behind Pivot Based SMT is to use the resources available in pivot language to counter the resource scarcity of source and target languages. This can be done using one of the three methods proposed in (Wu and Wang, 2009). The difference in these methods lies in the way they make use of the pivot language and its resources. The methods are as follows:

- **Transfer Method**:

  This method makes use of two different SMT systems, one from source-pivot and other from pivot-target. These are independently trained on source-pivot and pivot-target parallel corpora respectively. This removes the need of the availability of direct corpus from source-target. When a sentence is input to the system for translation, it is first translated into pivot language using source-pivot system and then the pivot sentence is translated to target language using pivot-target system.

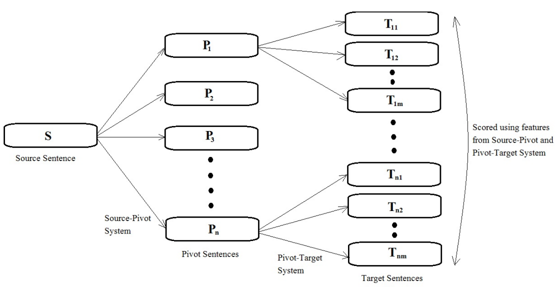

Figure 1: Transfer Method

  Now the issue here is the possibility that a sentence may get translated into more than one sentences i.e. there might be more than one translations possible for a particular sentence. For example, a gives source sentence, S is translated into n number of pivot sentences $P_1, P_2, ..P_n$ using source-pivot SMT system and each of these P_i's are translated into m target sentences $T_{i1}, T_{i2}, ...T_{im}$ using a pivot-target SMT system. In such a scenario we have total n×m candidates to choose from in order to select the best possible translation. These candidates are then scored using both source-pivot and pivot-target system scores to select the best translation using the method described in (Utiyama and Isahara, 2007).

  If f^{SP} and f^{PT} denote the features in source-pivot and pivot-target systems respectively then the best possible target translation is found using the formula

$$\hat{t} = \operatorname*{argmax}_t \sum_{k=1}^{L} (\lambda_k^{SP} f_k^{SP}(S,P) + \lambda_k^{PT} f_k^{PT}(P,T)) \tag{1}$$

  where

  L: Number of features used by SMT System
  λ_k^{SP}: weight for k^{th} feature for source-pivot system
  λ_k^{PT}: weight for k^{th} feature for pivot-target system

- **Synthetic Method or Corpus Synthesis Method**:

  If we have a source-pivot and pivot-target corpora available, we can obtain a source-target parallel corpus using two ways. We can train a source-pivot and a pivot-target system independently using the respective corpora. We then obtain translate pivot sentences from source-pivot corpus into target language sentences using the pivot-target translation system. The other option is to obtain source side translation of pivot sentences form pivot-target corpus using a pivot-source SMT system. We can then combine these two parallel corpora to get a source-target parallel corpus. Using this corpus a source-target system can be trained. Following figure explains this method.

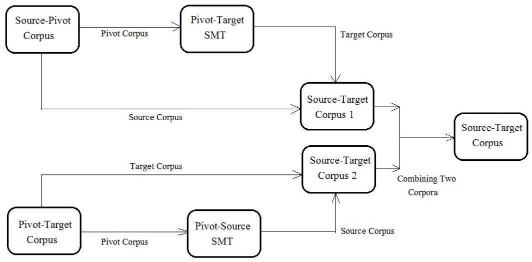

Figure 2: Synthetic Method

- **Triangulation Method**:

  Triangulation method uses the source-pivot and pivot-target translation models to actually build a source-target translation model. It refers to the triangulation of phrase tables where, phrase tables from source-pivot and pivot-target systems are triangulated to generate a source-target phrase table. The method induces some important information about source-target phrase table from the other two phrase tables.

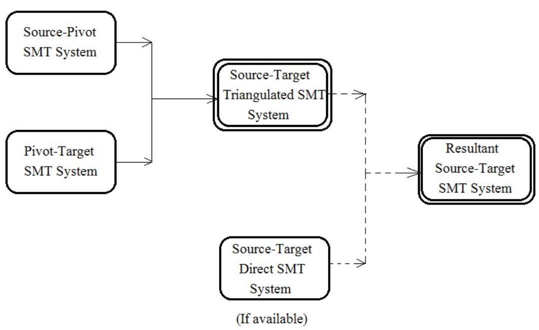

Figure 3: Triangulation Method

  Even when a small source-target parallel corpus(direct corpus) is present, the triangulation method induces some phrase pairs that are not extracted while training a source-target system on the available direct corpus. For each phrase pair that is induced by triangulation, the method needs to find out two important properties:

  - Phrase Translation Probability
  - Lexical Weights (which in turn require alignments)

Triangulation method has a mathematical formulation(Wu and Wang, 2007) through which it calculates these values and generates a source-target translation model. Since this method is going to be the focus of the discussion for the rest of the report, we look at this method closely in the next section.

3 Details Triangulation Method

As discussed in the previous section, while triangulating phrase tables, we need to induce two values for source-target phrase pairs, phrase translation probability and lexical weights. The mathematical calculations are as follows(Wu and Wang, 2007):

- **Phrase Translation Probability**:

  This value gives the probability of a particular phrase being translated into other. If S, P and T are phrases from source, pivot and target corpus respectively, and $\phi(S|P)$ and $\phi(P|T)$ are phrase translation probabilities for source-pivot and pivot-target respectively, then these two values are directly available from the source-pivot and pivot-target phrase tables. The formula for phrase translation probability of source-target is (Wu and Wang, 2007)

$$\phi(S|T) = \sum_{P} \phi(S|P).\phi(P|T) \qquad (2)$$

  The formula can be derived as follows:

  Our objective is to find out the value of $\phi(S|T)$, and for a given S and T there can be more than one P's possible. We introduce this variable through marginalization so,

$$\phi(S|T) = \sum_{P} \phi(S,P|T) \qquad (3)$$

  Applying chain rule, we get

$$\phi(S|T) = \sum_{P} \phi(S|P,T).\phi(P|T) \qquad (4)$$

  But Translation probability for source-pivot is independent of target phrase, so using

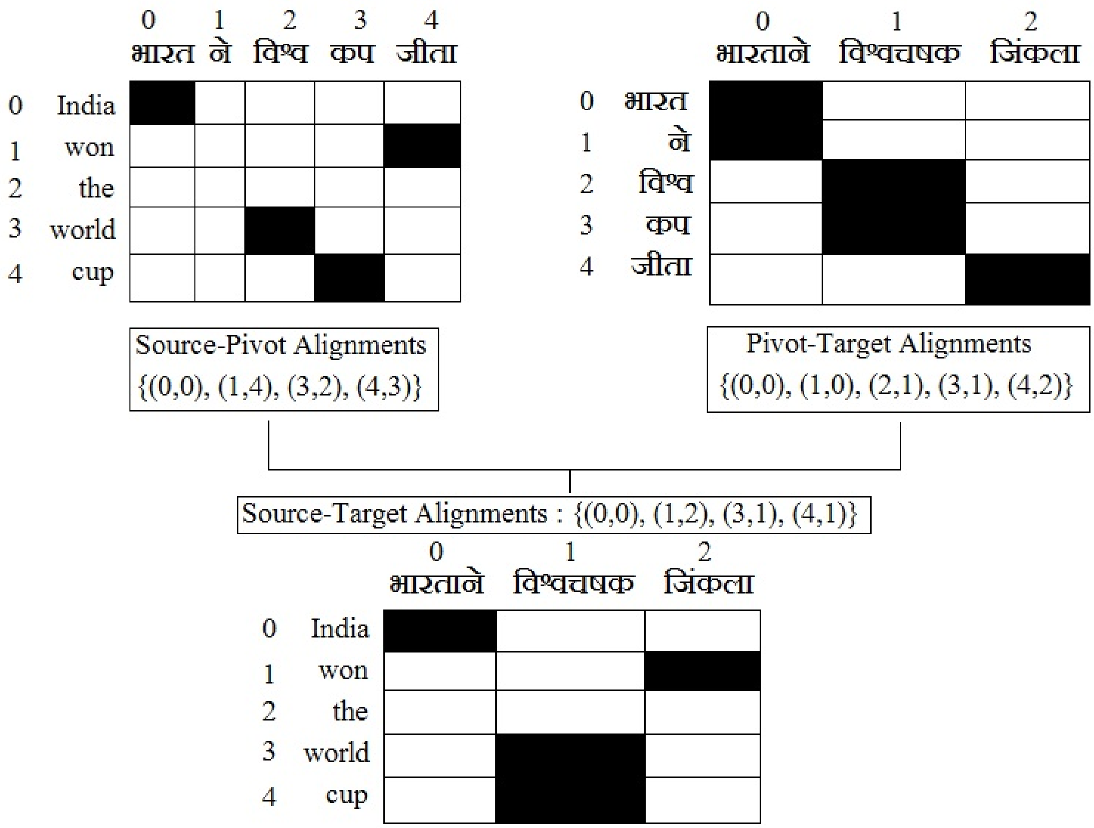

Figure 4: Obtaining Alignments during Triangulation

this independence assumption, T from first term on RHS can be ignored, therefore,

$$\phi(S|T) = \sum_{P} \phi(S|P).\phi(P|T) \qquad (5)$$

- **Lexical Weight**:

  This is an important value that helps in deciding the translation probability for a pair of phrases, when the phrase translation probability value may not be reliable. This might be the case when a phrase S from source and a phrase P from pivot are aligned to each other and have each occurred only once in the corpus. In such a case, we cannot completely rely on just the co-occurrence of these phrases for calculating the phrase translation probabilities. A good way to measure if they are translations of each other would be to look inside the phrases and map the translations at the level of words. This is incorporated using lexical weighting. If inside a phrase, the words aligned to each other have high lexical translation probability (probability that a word is translated into another), then the phrase translation probability for those pair of phrases is higher.

  Since the lexical weights depend on alignments, our first task is to obtain alignments between source phrase S and target phrase T. These can be obtained from source-pivot and pivot-target alignments as shown in the Figure 4 below. In the example shown in the Figure 4, English, Hindi and Marathi are source, pivot and target languages respectively.

  The above approach for obtaining alignments can be mathematically formulated as follows,(Wu and Wang, 2007)

$$a = \{(i,j)|\exists k : (i,k) \in a_1 \ and \ (k,j) \in a_2\} \quad (6)$$

where
a: alignment from source-target
a_1: alignment from source-pivot
a_2: alignment from pivot-target
i: index of word from source phrase
k: index of word from pivot phrase
j: index of word from target phrase

The lexical translation probability is then calculated as

$$w(s|t) = \frac{count(s,t)}{\sum_{s'} count(s',t)} \quad (7)$$

where count(s, t) is the co-occurring frequency of the word pair (s, t) in the phrase table.

Once these values are calculated, the lexical weight (L_w) for a pair of phrases (S, T) is calculated as,

$$L_w(S|T,a) = \prod_{i=1}^{n} \frac{1}{|j|(i,j) \in a|} \sum_{\forall(i,j)\in a} w(s_i|t_j) \quad (8)$$

where
s_i and t_j are words in source phrase S and target phrase T respectively.

Once we have a source-target phrase table with both phrase translation probability and lexical weights calculated for all phrase pairs, we say we have a 'Triangulated Phrase Table'. We have to use this table as a translation model when the direct corpus is not available at all. But when a small amount of direct corpus is available, it would be better if we can somehow use the knowledge from that corpus as well. Thus, we need to have some method for combining the information extracted through pivoting and the information extracted through direct corpus. We look at such methods in the next section.

4 Combination Methods

When we have a direct source-target corpus available, though small in size, we can build a phrase table from that corpus, called a 'Direct Phrase Table', as well as build a phrase table through the use of a pivot, called a 'Triangulated Phrase Table'. We need a method for using the combined knowledge of these two phrase tables while decoding a new input sentence. This can be done in one of the following ways (Dabre et al., 2014):

- **Linear Interpolation**:
  In this method, if the same phrase pair is present in both Direct and Triangulated Phrase Tables, the values from Direct Phrase Table and Triangulated phrase table are interpolated in order to get values for the final phrase table. The values that need to be interpolated are again the phrase translation probability and lexical weights. The interpolation is linear and is done using the following equations.

$$\phi(S|T) = \sum_{i=0}^{n} \alpha_i \phi_i(S|T) \quad (9)$$

$$L_w(S|T,a) = \sum_{i=0}^{n} \beta_i L_{w,i}(S|T,a) \quad (10)$$

  where $\sum_{i=0}^{n} \alpha_i = 1$ and $\sum_{i=0}^{n} \beta_i = 1$ and α_0 and β_0 are the weights for $\phi_0(S|T)$ and $L_{w,0}(S|T,a)$, phrase translation probability and lexical weight from the Direct Phrase table, respectively. Other values of i from 1 to n indicate n number of pivots and phrase translation probability lexical weight values for the triangulated phrase tables obtained by using these pivots are suffixed by i. So this model can be easily extended to more than one pivots. The weights α_i and β_i can be found using a method such as minimum error rate tuning. It has been experimentally proven that a weight of 0.9 for the direct table works well.

- **Fill-up Interpolation or Augmentation**:
In this method if a phrase pair exists in both Direct and Triangulated phrase tables, the one in Direct Phrase Table is given preference. Only the new phrase pairs from triangulated phrase table i.e. phrase pairs that were not originally present in the Direct Phrase table are added to the final phrase table and the ones present in the Direct phrase table are added as it is to the resultant phrase table. This can be viewed as "Augmenting" the direct phrase table with new entries or "filling-up" new entries in the Direct Phrase Table, hence the name.

- **Multiple Decoding Paths(MDP)**:
This is a method when all phrase tables i.e. Direct Phrase table and all triangulated phrase tables are used together during decoding as separate entities. This is a facility that Moses provides where we can provide paths for more than one translation models in the system that is built. This avoids the need of any kind of interpolation and hence avoids the noise that interpolation may bring in. This has been experimented with by Dabre et al. (Dabre et al., 2014) and has been found to be useful.

This discussion forms the basis for the use of triangulation method for pivoting. There has been a lot of work at IIT Bombay in the field of Pivot Based SMT. In the next section, we take a look at these pieces of works.

5 Pivot Based SMT at IIT Bombay

At IIT Bombay, Pivot Based SMT has been keenly studied and there have been two major paths of work that have been followed. One of them is exploring the use of more than one languages as pivots and the second is enriching Pivot Based SMT with Morphological Segmentation. In this section we discuss the details of the experiments carried out at IIT Bombay.

5.1 Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages

This was the topic of research for our senior Raj Dabre. The focus of this work was to improve Japanese-Hindi Statistical Machine Translation System using multiple pivot languages. This work was presented in NAACL 2015. (Dabre et al., 2014)

5.1.1 Motivation

Using a pivot helps because through its use, new phrase pairs can be extracted which may not be extracted when training on a Direct Corpus. For example, if we consider English to Marathi system with Hindi as a pivot, it is possible that in the small English-Marathi corpus that is available, there is no evidence that may lead to the mapping (Sun, सूर्य) being extracted. But the source-pivot table may have extracted (Sun, सूरज) and pivot-target table may have extracted (सूरज, सूर्य). From these two mappings, pivot based system can induce a new phrase pair, (Sun, सूर्य). Since pivot induces new phrase pairs, it is possible that when more than one pivots are used, each pivot may induce different phrase pairs thereby increasing the coverage of the resultant phrase table. It is also possible that each pivot captures a different phenomenon of the language and hence adds to the improvement achieved by using a single pivot.

5.1.2 Experiments

Experiments were performed for single as well as multiple pivot settings. Languages chosen were Japanese and Hindi and 7 other languages were chosen as pivots. The corpus used was a freely available multilingual Bible corpus of 29780 sentence tuples. A tuple contains several sentences, one for each language. This was divided into 29000 training tuples, 280 tuning tuples and 500 test tuples. Phrase table triangulation was performed for all 7 languages. 7 pivot languages were as follows:

- Chinese, Korean (Closer to source)
- Marathi, Kannada, Telugu (Closer to target)

- Paite (Sino Tibetian)
- Esperanto (not close to either source or pivot)

The phrase tables were combined using all combination techniques mentioned in Section 4 i.e Linear Interpolation, Fill-up Interpolation and Multiple Decoding Paths(MDP). The experiments lead to some interesting observations. Those are listed in the next subsection.

5.1.3 Observations and Conclusions

Improvement in BLEU score from 33.86 to 38.22 was observed for Japanese to Hindi SMT system(Dabre et al., 2014) when all 7 pivots were used. The reverse system i.e. Hindi to Japanese, showed a BLEU score improvement from 37.47 to 41.08(Dabre et al., 2014). Along with the quantitative improvement, some interesting observations were made during the experiments. Those are listed below:

- Pivot languages closer to either source or target language can act as a good pivot. Languages from the same family generally act as good pivots but there are several exceptions to this rule and the family of a language cannot be the sole criteria while selecting a language as a pivot.
- Morphological similarity of pivot to source and target is another factor that affects the translation quality. Languages having rich morphological features and high agglutination are always a major hurdle in improving the SMT translation quality.
- In case of single pivot systems, improvements are achieved when Interpolation methods, (linear or fill-up) are used, but these improvements are small. The reason behind this being that the corpus used was a multilingual one and thus was same in size for direct system as well as each pivot system. Interpolation method achieves greater improvements when direct corpus is small in size as compared to pivot corpus.
- In case of multiple pivot systems, interpolation improves the BLEU scores significantly, this leads to the conclusion that each pivot induces a new set of phrase pairs and all pivots together have a composite effect on translation quality.
- Interpolation methods may not perform well in some cases since they disturb the probability space either by changing probability values of phrase pairs or by adding new phrase pairs. Since log-linear combination does not modify the probability space, combining table using MDP leads to better results.

It was observed during these experiments that pivot obviously helps in extracting phrase translation information, but the question of whether it extracts good reordering information was not explored in this work. Our work discussed in the later sections of this paper tries to answer this question partially.

5.2 Augmenting Pivot Based SMT with Word Segmentation

This was the topic of research for our senior Rohit More. The focus of this work was to study the effect of morphological segmentation on the quality of Pivot Based SMT systems.

5.2.1 Motivation

One of the major reasons of poor performance of SMT systems is agglutination of languages. Languages which are morphologically rich and highly agglutinative tend to damage the quality of SMT system, since they may give rise to data sparsity. Take an example of English and Marathi. The phrase “instead of doing” which has 3 words in English, is translated as one word in Marathi, “करण्यापेक्षा”(karaNyApekShA). Examples like these show that Marathi is morphologically rich as compared to English. It has been found that morphological segmentation reduces data sparsity and assists in improving the performance of many of the NLP tasks, e.g. POS Tagging. This happens since the systems no longer work on the surface forms of the words. Use of a pivot already reduces data sparsity by extracting new phrase pairs, so it was an interesting prospect to explore how morphological segmentation would help in improving the scenario further.

5.2.2 Experiments

Experiments were performed for different scenarios. In the first set of experiments, only the source corpus was morphologically segmented. In the second set of experiments, both source and pivot corpora were morphologically segmented. For segmentation, Morfessor (Virpioja et al., 2013) and Indic-nlp-library[1] were used. During this work as well, the effect of multiple pivots was explored. If the source corpus is morphologically segmented, the system is called a "source-morphed" system. Similarly "pivot-morphed" systems are defined. All combinations of {Single Pivot, Multiple Pivots} × {Source-morphed, Pivot-morphed} systems were explored in these experiments. Only Multiple Decoding Paths (MDP) was used for combination of Direct and Pivot systems. The corpus used for these experiments was ILCI multilingual corpus from Health and Tourism Domain. It was divided into 46277 sentences for training corpus, 500 sentence for tuning corpus and 2000 sentences for testing corpus.

5.2.3 Observations and Conclusions

Several interesting observations made during these experiments are listed below:

- Systems using only the triangulated phrase tables perform poorly as compared to direct phrase table systems. Since the corpus is multilingual and triangulated phrase table has only newly extracted phrase pairs, so it cannot beat the system that has phrase pairs directly extracted from the corpus.

- In case of experiments without morphological segmentation, Dravidian languages never come up as good pivots. The reason behind this may be the highly agglutinative nature of Dravidian languages.

- English also does not surface as a good pivot. This is because English is not close to either the source or the target since source and target were always Indian languages.

- Using more number of pivots improves the BLEU scores. This vindicates the conclusion stated in previous section, that each pivot induces a new set of phrase pairs.

- Languages with simpler morphological features mostly act as good pivots since they extract better alignments and hence extract better phrase pairs. During the experiments, Hindi, Gujarati, Punjabi mostly surfaced as good pivots. All of them have simple morphological features.

- Source-morphed systems improve the BLEU score only when source is morphologically complex. This leads to the conclusion that morphological segmentation is only helpful when the language being segmented is morphologically complex.

- Only pivot-morphed systems perform poorer than only source-morphed systems and systems where source and pivot both are morphed perform better than both of them. Systems where both source and pivot are morphed and multiple pivots are used perform even better.

- From the experiments, a few distinct features that are indicative of a good pivot were listed. For a language to be considered as a good pivotm it should have
  1. Vocabulary overlap with source or target
  2. Simple Morphological features
  3. Closeness with source or target language

This was a crucial piece of work and explored a new direction of morphological segmentation in triangulation. Most of the work in Pivot Based SMT has been for the triangulation of phrase tables, but there exists a possibility that triangulation of reordering tables might increase

[1] https://github.com/anoopkunchukuttan/indic_nlp_library

the performance even further. From another point of view, both phrase table and reordering table are important parts of an SMT system. If a method for reordering table triangulation can be devised so that better reordering probability values are extracted for newly added phrase pairs, it would complete the process of triangulation and complement the phrase table triangulation process. We explore this possibility in the next section of this paper. We devise a formulation for reordering table triangulation and show improvements in translation quality.

6 Past work in Reordering Models

Lexicalized Reordering Model was first proposed by Tillmann (Tillmann, 2004). Lexicalized Reordering is different from the distortion model in the classic SMT Distortion model gives a probability distribution over relative distances between the phrases in a sentence, whereas Lexicalized reordering model works on a concept of orientation of a phrase with respect to phrases adjacent to it in a sentence. Tillman proposed orientations for "blocks" of phrases where a "block" is a set of one or more phrases from a sentence. He proposed 3 kinds of orientations, Left, Right and Neutral. A block is said to have a 'Right' orientation if it has a left predecessor block, a 'Left' orientation if it has a predecessor block on its right and a 'Neutral' orientation if the predecessor block is not adjacent to it. The following diagram will help in understanding these orientations.

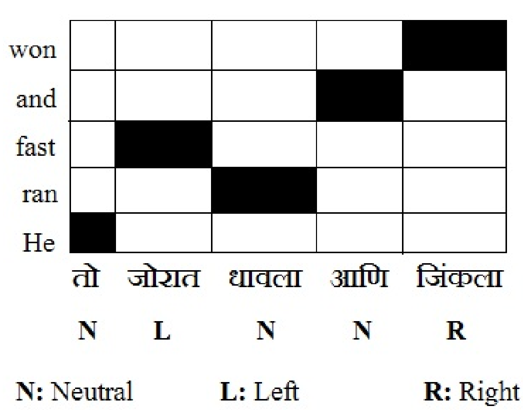

Figure 5: Orientations Proposed by Tillman

Ohashi et al. proposed a different phrase reordering model (Ohashi et al., 2005). Their reordering model gave importance to the distance between the adjacent phrases along with their positioning. They proposed four kinds of reordering based on relative distance between phrases:

- Monotone: when two source phrase are adjacent and the target phrases who are translations of those source phrases are also adjacent and source phrases are in the same order as target phrases.
- Monotone-Gap: when two source phrases are not adjacent but their order is same as that of the target phrases which are their translations
- Reverse: when two source phrases are adjacent and their order is opposite to the order of target phrases which are their translations
- Reverse-Gap: when two source phrases are not adjacent but their order is opposite to the order of target phrases which are their translations

The reordering model on which we focus in this report and around which most of the discussion in this report takes place is a Lexicalized Reordering model proposed by Koehn et al. (Koehn et al., 2005). This model is similar to the one proposed by Tillman in that it does not give importance to the relative distance between phrases. This model proposes three kinds of orientations - Monotone, Swap and Discontinuous. Moses also uses this Reordering model in its implementation (Koehn et al., 2007). These reordering orientations are explained in detail in the next section.

7 Reordering Orientations: Monotone, Swap and Discontinuous

A reordering orientation is a property of a phrase pair i.e. a source phrase and a target phrase which are translation of each other. Let S_1, S_2 be phrases in the source sentence, T_1, T_2

be their corresponding translations in the target sentence and T_1 is a phrase that immediately precedes T_2 in the target sentence. Then for the phrase pair (S_2,T_2) the orientation is (Koehn et al., 2005):

- **Monotone**, if S_1 is a phrase that immediately precedes S_2 in the source sentence.

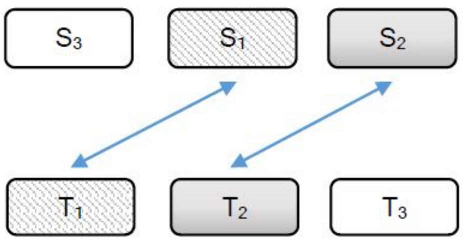

Figure 6: Monotone Orientation

- **Swap**, if S_1 is a phrase that immediately succeeds S_2 in the source sentence.

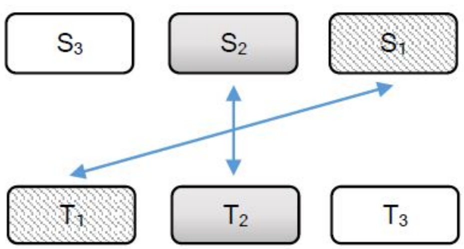

Figure 7: Swap Orientation

- **Discontinuous**, if S_1 and S_2 are not adjacent to each other in the source sentence.

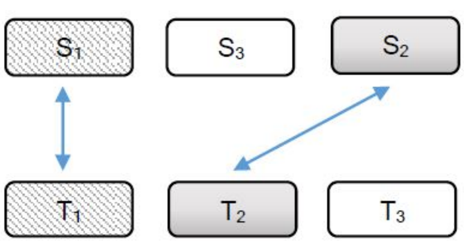

Figure 8: Discontinuous Orientation

Let us consider sentences from English-Marathi language pair shown in Figure 9.

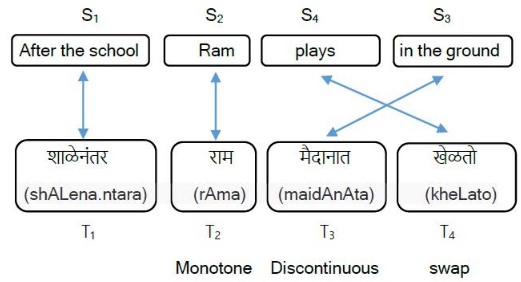

Figure 9: Example depicting all orientations

As per the definitions discussed above, the orientation exhibited by the phrase pair (Ram, राम) is a monotone since, the translation of the previous phrase of राम i.e. 'After the school' is also the previous phrase of 'Ram'. Similarly by definition, (plays, खेळतो) has a swap orientation and (in the ground, मैदानात) has a discontinuous orientation.

8 Domain Adaptation

8.1 Introduction

Domain Adaptation for Machine Translation is a well studied problem in SMT. It is an important concept that helps in closing the gap between the training and the testing situations for an SMT system from the point of view of domains. Domain Adaptation is crucial in scenarios where a very good SMT system, probably trained on a large corpus, already exists for Domain D_1 and we need to leverage it and adapt it to work for another domain, D_2 without losing all the knowledge it has acquired through domain D_1. In other words, we need to leverage the knowledge of the already existing out-of-domain SMT system while adapting to a new domain using some in-domain corpus.

The approach of training an SMT system on in-domain corpus and then incorporating it to a larger SMT system trained on out-of-domain corpus, relies on the fact that substantial amount of in-domain training corpus is available for source language to target language pair. In reality, such a corpus may not be available. Pivot based SMT has been shown to help in such a scenario. In a pivot based approach, a pivot language is chosen such that there is substantially large in-domain corpus available for

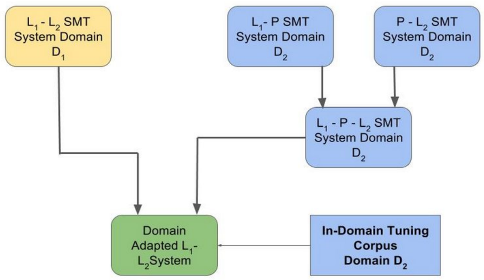

Figure 10: Description of The Method

source-pivot and pivot-target language pairs. In this paper we present a case study on English-Hindi language pair, with Marathi used as a pivot. The domains used are tourism and health and we present results for Domain Adaptation in both directions.

In this paper we also go a step further and claim that if a small amount of in-domain tuning corpus is available, then the performance of the SMT system can be improved further. This idea is interesting because if just the presence of in-domain tuning corpus without the availability of in-domain training corpus, improves the performance then it will provide a way for adapting an existing SMT system to new domain with a very little effort in collecting in-domain corpus from source to target language.

8.2 Related Work

The concept of Domain adaptation was initially introduced by Daume III and Marcu (2006) for improving the performance of statistical classifiers. Koehn and Schroeder (2007) brought this concept in the field of Statistical Machine Translation. They studied Domain Adaptation by using in-domain and out-of-domain translation model and some of its variants including in-domain language model, interpolated language model and two language models(in-domain and out-of-domain) which proved to be helpful in improving the results.

Wu et al. (2008) and Daumé III and Jagarlamudi (2011) tried to improve the results of Domain Adaptation experiments with the help of large in-domain monolingual corpus. They used this corpus to build the in-domain language model which can be integrated with the out-of-domain trained translation model to improve the translation.

Wu et al. (2008) and Daumé III and Jagarlamudi (2011) proposed that even though obtaining in-domain bilingual corpus is difficult we can easily obtain or mine in-domain bilingual dictionary. They proved that appending this in-domain bilingual dictionary with the out-of-domain bilingual corpus helps improve the translation in Domain adaptation scenario. As presence of little in-domain corpus created positive impact on the results Bertoldi and Federico (2009) took a step further to explore this concept. He proposed an approach that takes advantage of the large in-domain monolingual corpus to build better translation model. He used corpus synthesis method translate in-domain monolingual sentences in their counterpart language which then can be used to help in Domain Adaptation either by appending this corpus to the out-of-domain bilingual corpus or by building a separate translation model on this generated in-domain bilingual corpus.

Daumé III and Jagarlamudi (2011) mentioned in the paper that availability of very small in-domain bilingual corpus for tuning purpose can further improve the results by great margin. Nakov (2008) showed that along with the in-domain and out-of-domain data, sentence level syntactic paraphrases and recaser helps in improving the Domain Adaptation results.

Tiedemann (2012) showed that even if we use in-domain data in only one leg of pivoting (i.e. either in source-pivot or in pivot-target), we can still get an improvement. These experiments were performed for both Character based and Word based Statistical Machine Translation systems and in both the cases they got improvement. They also showed impact of closeness of the pivot language with either of the source language or the target language on Domain Adaptation experiments by experimenting on various domains for English-Norwegian translation

References

Nicola Bertoldi and Marcello Federico. 2009. Domain adaptation for statistical machine translation with monolingual resources. In *Proceedings of the fourth workshop on statistical machine translation*, pages 182–189. Association for Computational Linguistics.

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

Hal Daumé III and Jagadeesh Jagarlamudi. 2011. Domain adaptation for machine translation by mining unseen words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 407–412. Association for Computational Linguistics.

Hal Daume III and Daniel Marcu. 2006. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, pages 101–126.

William John Hutchins and Harold L Somers. 1992. *An introduction to machine translation*, volume 362. Academic Press London.

Philipp Koehn and Josh Schroeder. 2007. Experiments in domain adaptation for statistical machine translation. In *Proceedings of the second workshop on statistical machine translation*, pages 224–227. Association for Computational Linguistics.

Philipp Koehn, Amittai Axelrod, Alexandra Birch, Chris Callison-Burch, Miles Osborne, David Talbot, and Michael White. 2005. Edinburgh system description for the 2005 iwslt speech translation evaluation. In *IWSLT*, pages 68–75.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180. Association for Computational Linguistics.

Preslav Nakov. 2008. Improving english-spanish statistical machine translation: experiments in domain adaptation, sentence paraphrasing, tokenization, and recasing. In *Proceedings of the third workshop on statistical machine translation*, pages 147–150. Association for Computational Linguistics.

Kazuteru Ohashi, Kazuhide Yamamoto, Kuniko Saito, and Masaaki Nagata. 2005. Nut-ntt statistical machine translation system for iwslt 2005. In *IWSLT*, pages 118–123.

Jörg Tiedemann. 2012. Character-based pivot translation for under-resourced languages and domains. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 141–151. Association for Computational Linguistics.

Christoph Tillmann. 2004. A unigram orientation model for statistical machine translation. In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 101–104. Association for Computational Linguistics.

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

Sami Virpioja, Peter Smit, Stig-Arne Grönroos, Mikko Kurimo, et al. 2013. Morfessor 2.0: Python implementation and extensions for morfessor baseline.

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.

Hua Wu, Haifeng Wang, and Chengqing Zong. 2008. Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 993–1000. Association for Computational Linguistics.