Lexicalized Reordering Models for English-Myanmar Phrase-based Statistical Machine Translation

May Kyi Nyein, Khin Mar Soe
Natural Language Processing Lab., University of Computer Studies, Yangon, Myanmar
maykyinyein@ucsy.edu.mm, khinmarsoe@ucsy.edu.mm

Abstract

Reordering is extremely desirable for translation accuracy when translating between high disparity language pairs in word order. The aim of this paper is the comparative study of lexicalized reordering models (LRM) by Moses to investigate the translation performance for English-Myanmar statistical machine translation (SMT) system. The studied methods are word-based, phrase-based and hierarchical phrase-based LRM by using various orientations and distortion limits. This reordering model calculates reordering probability conditioned on the word of each phrase pair. We applied Moses phrase-based SMT (PBSMT) system to make experiments for the variants of LRM and evaluated the BLEU and RIBES scores to measure the performance of machine translation. According to this experiments, hierarchical phrase-based reordering model in MSD orientation gives the highest scores in English-Myanmar SMT system.

Keywords*: statistical machine translation (SMT), Lexicalized reordering model (LRM), orientations, reordering probabilities, Moses.*

1. Introduction

Phrase-based statistical machine translation systems have emerged state-of-the-art performance on standard translation tasks. Machine translation system has broadly focused on two main objectives, improving word translation and word order in translation output. But different syntactic structure has challenged to generate a syntactically and semantically correct word order sequence.

Especially, the word orders of English and Myanmar are very different, such as Myanmar is SOV structure and English is SVO structure. When an English sentence is translated into Myanmar sentence, the verb in the English has moved to the end of Myanmar. Myanmar word order diverges from English mostly within the noun phrase, prepositional phrase and verb phrase. So, machine translation needs to search for a good reordering that it can generate a fluent translation output.

The reordering model can help to reduce the differences of word order and are a part of machine translation system. Various methods have been proposed for SMT. In the earliest, the distance-based distortion reordering model is applied to model the phrase movements in translation. But it is unable to use linguistic context to score reordering.

Some statistical approaches are LRM that propose a reordering probability conditioned on the word of each phrase pair. Phrase pair is defined orientations with its previous phrase pair. They often categorize three orientations with previous phrase pair: monotone, swap and discontinuous. According to the orientation, LRM can further be classified into word-based reordering, phrase-based reordering and hierarchical-based reordering. The goal is to capture syntactic phenomena occurring in the foreign language.

This paper is intended to make lexicalized reordering model experiments by using various orientations and distortion limits to investigate the translation performance in English-Myanmar SMT system. The structure of this paper is designed as follows. Some representative workings on reordering models are illustrated in Section 2. Section 3 presents about the reordering issues of English-Myanmar translation. Section 4 describes phrase-based statistical machine translation system and Section 5 explains various reordering models. Section 6 presents overview of the LRM. And then, Section 7 presents the detailed experiments and results. Finally, in Section 8, we describe our conclusion of the paper.

2. Related Work

Many ideas have been proposed to handle the reordering problems.

Early phrase-based models have relied on a linear distortion feature [1]. This model works by penalizing long-ranged reordering based on the distance skipped, and it is difficult to capture syntactic word orders differences between source and target languages. The movement distance is measured on the foreign side. Without consideration of the contents in the phrases, phrase movements have not been well solved in the phrase-based translation.

Lexicalized reordering model has addressed this issue by introducing reordering probabilities on current phrase pairs. C. Tillmann [2] proposed that the block swapping is controlled by the unigram orientation model to handle swapping of predecessors. They collect block unigram counts with orientation and count how often a block occurs to the right or to the left of its predecessor block. The orientation model improves translation performance over block unigram model without reordering is used and swapping is controlled by a language model.

Koehn et al. [3] include lexicalized reordering model with three orientation types based on the actual phrases. The model tries to predict the orientation of a phrase whether the next phrase is to the left or to the right of the next phrase. This phrase orientation probability is conditioned on the current source and target phrase and relative frequencies are used to estimate the probabilities. They also optimized word alignment method, lexicalized reordering method and reordering distance limit for five different language pairs.

M. Galley et al. [4] proposed a lexicalized orientation model for hierarchical phrase-based reordering. They handle the ability to perform long-distance reordering with syntax-based systems. The reordering model relies on a hierarchical structure and enables movements of phrases that are more complex than swaps between adjacent phrases. The model maintains all the effectiveness of statistical phrase-based systems, while being able to take some key linguistic phenomena to develop parsing-based approaches.

D. Xiong et al. [5] defined two types of orientations: inverted and straight between two sequential blocks. To classify the sequential blocks, a maximum entropy classifier is used. They used features extracted from the blocks. They used lexical features and provides hierarchical phrase reordering based on features automatically learned from actual bitext. This is able to determine the reordering types that are not observed in the training data.

3. Some Reordering Issues for English-Myanmar Translation

Myanmar language is a morphological rich and verb final language. English is a verb second language (e.g., in English-to-Myanmar translation, a verb should move to the end of the clause). There are multiple word orders in Myanmar sentence for one English sentence because Myanmar is free word order language. English uses prepositions while Myanmar is postpositionally inflected with various grammatical features. Moreover, the order of noun phrases and prepositional phrases is also swapped in Myanmar as compare with English. The placement of verbs can often lead to movements over long distances.

Without reordering between these languages, words in English sentence are directly translated and there is no structural order in Myanmar language and translation cannot be meaningful. So, word reordering is one of the problems for SMT between English and Myanmar.

4. Phrase-Based SMT System

PBSMT [1] is considered as three-phase process: (1) the source sentence is segmented into phrases (blocks) which may be any sequences of adjacent of words (2) each phrase is mapped into the target language with a phrase translation table, and (3) phrases may be reordered as shown in figure 1.

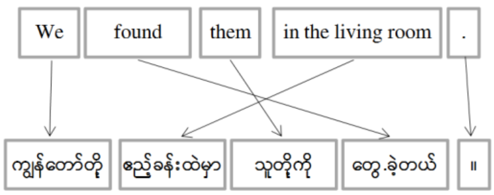


Figure 1. Phrase-based Statistical Machine Translation

The standard model of machine translation employs a log-linear approach. We search for the most possible sentence *e* given some foreign sentence *f* by maximizing the sum over a set of feature functions $h_i(e,f)$.

$$\hat{e} = \underset{e}{\operatorname{argmax}}\, p(e \mid f) \tag{1}$$

$$\underset{e}{\operatorname{argmax}}\, p(e \mid f) = \exp\left(\sum_{i=1}^{I} \lambda_i h_i(e,f)\right) \tag{2}$$

That uses a number of feature $h_i(\ )$ with weights λ_i. The feature functions are typically a language model, lexicalized reordering model, word penalty, and various translation models (lexical translation probability, phrase translation probability, etc.) [15].

5. Reordering Models

There are several reordering models that have been proposed for PBSMT systems. They can be classified into following categories [5]:

In-ordering is performed during decoding such as distance-based reordering and lexicalized reordering. It works appropriately for similar word order language pairs but is not sufficient for distant language pairs. LRM introduced lexical constraints of the phrase reordering.

Pre-ordering is made as pre-processing before decoding. These methods are subdivided into two types: (a) Rule-based methods parse input sentences and reorder the words by using rules. (b) Discriminative models learn whether children of each node should be reordered by using the parser's derivation tree as a latent variable.

Post-ordering is performed as post-processing after decoding. They have the advantage of using syntax-based features and need to use a correct parser on the target language.

6. Lexicalized Reordering Models

Lexicalized reordering model [3,2,4] have become the standard in phrase-based system. This reordering type is conditioned on the actual phrases or words. To deal with data sparsity, movement is measured in terms of orientation types, instead of exact movement distance.

Formally, given a sequence of source phrases $f = \{f_1,\ldots\ldots,f_n\}$, target phrases $e = \{e_1,\ldots\ldots e_n\}$, and a phrase alignment $a=\{a_1,\ldots\ldots a_n\}$ that expresses a source phrase f_{ai} and translated target phrase e_i, the model estimates the conditional probability of a sequence of orientations $o = \{o_1,\ldots\ldots o_n\}$. The probability is conditioned using a_{i-1} and a_i to make sure that the orientation o_i is consistent with the phrase alignment in equation 3:

$$P(o/e,f,a) \approx \prod_{i=1}^{n} P(o_i/e_i, f_{a_i}, a_{i-1}, a_i) \qquad (3)$$

There are generally three types of lexicalized models that are based on word-based LRM (Koehn et al., 2007), phrase-based LRM (Tillman, 2004), and hierarchical phrase-based LRM (Galley and Manning, 2008).

Word-based reordering model: The model determines the orientation of current phrase with respect to previous adjacent word alignment at training time, and phrase alignments at decoding time. If $(s-1, u-1)$ contains a word alignment and $(s-1, v+1)$ does not contain word alignment, orientation is set to M. Otherwise, it is set to S if $(s-1, u-1)$ does not contain word alignment and $(s-1, v+1)$ contains a word alignment shown in Fig. 2(a). In other cases, it is set to D.

Phrase-based reordering model: The model examines adjacent bilingual phrases rather than word alignments both at training and decoding time to determine orientations. If an adjacent phrase pair situated at $(s-1, u-1)$ in the phrase alignment, orientation is set to M. If an adjacent phrase pair is $(s-1, v+1)$, it is set to S *as* illustrated in Fig. 2(b) and otherwise, it is set to D.

Hierarchical phrase-based reordering model: The model allows that translated adjacent phrases are combined to form longer phrases around the current phrase. Orientation is set to M if the phrase extraction extracts a bilingual phrase pair at $(s-1, u-1)$ without limiting on maximum phrase length. If a phrase pair is extracted at position $(s-1, v+1)$, orientation is set to S *as* shown in Fig. 2(c) and otherwise, orientation is D.

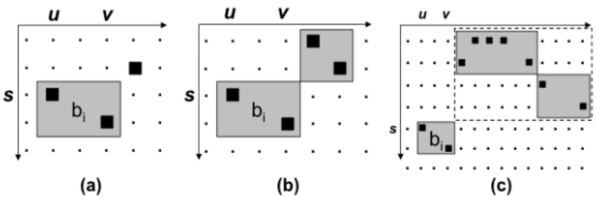


Figure 2: Swap orientations with three lexicalized reordering models: word-based, phrase-based, and hierarchical based.

In figure 2, [4] black squares are word alignments, gray squares denote phrases, [s] indicates the target-side phrases and [u,v] denotes the source-side phrases.

6.1. Reordering Orientations

Movement is measured in LRM in term of orientation types rather exact move distance. Three kinds of orientation sets are described in the following.

(1) Left/Right orientations learn whether the given source phrase is on the left of previous adjacent source phrase or not.

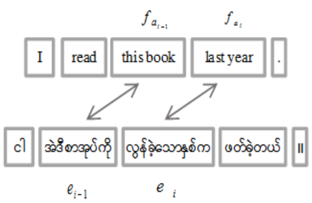
(2) MSD-based orientations that have monotone(M), swap(S), discontinuous (D). The two discontinuous labels of the MSLR are combined into one label is shown in equation 4.

$$o_i = \begin{cases} M & (a_i - a_{i-1} = 1) \\ S & (a_i - a_{i-1} = -1) \\ D & (|a_i - a_{i-1}| \neq 1) \end{cases} \quad (4)$$

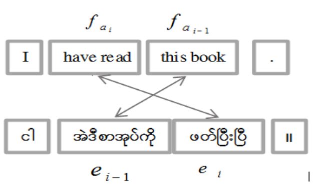
(3) MSLR-based orientations that consider four classes: monotone(M), swap(S), discontinuous-left (L), and discontinuous-right (R) are shown in equation 5.

$$o_i = \begin{cases} M & (a_i - a_{i-1} = 1) \\ S & (a_i - a_{i-1} = -1) \\ L & (a_i - a_{i-1} > 1) \\ R & (a_i - a_{i-1} < -1) \end{cases} \quad (5)$$

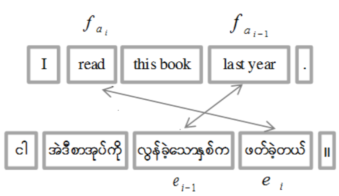
The most widely used orientation set is MSD: We will take MSD orientation type for explanation, while other orientation types (LR, MSLR) can be induced similarly.



(a) Monotone



(b) Swap



(c) Discontinuous

Figure 3. MSD orientation with respect to adjacent phrase.

Monotone means that the source phrases f_{ai}, f_{ai-1} are contiguous with respect to the target phrases e_i and e_{i-1}. Swap means f_{ai}, f_{ai-1} are adjoining and swapping. Discontinuous are not connecting to each other in the source sentence.

6.2. Reordering Probability Estimation

Most reordering models estimate a probability distribution $P(o_i \mid p_i, a_1, \ldots\ldots, a_i)$ for the *i-th* phrase pair p_i and the alignments $a_1, \ldots\ldots a_i$ of the previous target phrases. The orientation model uses maximum likelihood estimation (MLE). To compute the probability of each of the orientation types, statistically by:

$$P(o \mid f,e) = \frac{count(o,f,e)}{\sum_{o'} count(o',f,e)} \quad (6)$$

where o is the orientation type (e.g. {M, S, D}) and count ( ) returns frequency counts from the reordering file. By using additive (Laplace) smoothing with a factor σ, the estimation can be smoothed:

$$P(o \mid f,e) = \frac{\sigma + count(o,f,e)}{\sum_{o'} \sigma + count(o',f,e)} \quad (7)$$

7. Experiment and Results

7.1. Data Preparation

In this experiment, English-Myanmar bilingual sentences from ASEAN IVO Project "Open Collaboration for Developing and using Asian Language Tree-bank" (ALT) [21], ASEAN-MT [22] project and Myanmar News that consists of the various sentences from Text Books, Speaking and Local News covering difference domains are used to construct English-Myanmar parallel corpus.

The corpus consists of 143,413 parallel sentences for general domain. This is arbitrarily divided into 140,040 sentences for training, 1915 sentences for development and 1408 sentences for testing to run PBSMT Moses as shown in table 1.

Table 1. Corpus Statistics

Text Types	Parallel sentences	Training	Development	Testing
ALT	20,050	19,200	650	400
ASEAN-MT	22,768	21,733	400	435
Myanmar News	100,545	99,107	865	573
Total	143,413	140,040	1915	1408

Myanmar sentences are written as continuous sequences of syllables with no delimiting characters so word segmentation is an essential step for Myanmar language. Although ALT data and ASEAN-

MT data are already segmented, we have done manual check where necessary.

Myanmar News data are segmented by Myanmar Word Segmentation using a Combined Model [9] and checked manually. The English part of parallel data is tokenized, lowercased and cleaned using scripts in Moses tool.

When the source sentence length has one, the translation will be constantly monotonic and the reordering model does not need to learn. This kind of sentences are deleted.

7.2. Experimental Setting

For PBSMT model, we used the open source PBSMT Moses toolkit [10] for training of translation model and reordering model. The default Moses setting sets the distortion limit to 6 for LRM. If the number of words skipped is greater than 6, the translation will be pruned [17]. This makes the model fewer suitable for more syntactical different languages like English and Myanmar etc. Thus, we have set the distortion limits to 6, 9, 12 and studied the results for different LRM.

The word alignment was made by GIZA++ [13] with grow-diag-final-and it produces the possible bidirectional word alignments between source and target languages. We used KenLM [14] to train a 5-gram language model with interpolated modified Kneser-Nay discounting. Minimum error rate training (MERT) was used to tune the decoder's parameters.

All experiments were made on an Intel(R) Core (TM) i7-2600 CPU @ 3.40GHz processor with 3.8Gb of RAM and 64 bits Ubuntu 16.04.5 LTS.

7.3. Evaluation Metrics

Automated evaluation is comparing the output of machine translation system to the good reference. BLEU and RIBES scores are used as two performance statistics to measure the adequacy of translation and penalize the wrong word order of translation output.

7.3.1. BLEU

Bilingual Evaluation Understudy (BLEU) is to compare n-grams of the candidate translation with multiple reference translations in length, word choice, word order and then count the number of matches. The more the matches, the better the translation result is. [7]

7.3.2. RIBES

Rank-based Intuitive Bilingual Evaluation Measure (RIBES), which is more sensitive to reordering, is based on rank correlation coefficients modified with precision. It is used to compare the word ranks in the hypothesis with those in the reference. [8]

7.4. Results and Discussion

For evaluation the translation performance, 1,408 test sentences (not the same sentences in the training set) are used. The following tables give translation performance for systems in various LRM against the orientation types (MSD, LR and MSLR) and distortion limits (6, 9 and 12). In our experiments, we tried to find the best lexicalized reordering method and the best reordering distance limit for English to Myanmar SMT.

Table 2. Comparison of distance-based and LRM with different orientations and distortion limits in terms of BLEU score.

Model	Orientation	Distortion Limit		
		DL=6	DL=9	DL=12
distance	-	16.05	16.21	16.26
word	MSD	16.19	17.01	17.48
	LR	16.11	17.24	17.67
	MSLR	16.13	17.42	17.46
phrase	MSD	16.31	17.27	17.50
	LR	16.21	17.36	17.75
	MSLR	16.22	17.44	17.59
hierarchical	MSD	16.58	17.28	**17.78**
	LR	16.17	17.45	17.76
	MSLR	16.98	17.48	17.67

Table 3. Comparison of distance-based and LRM with different orientations and distortion limits in terms of RIBES score.

Model	Orientation	Distortion Limit		
		DL=6	DL=9	DL=12
distance	-	0.7006	0.7082	0.7139
word	MSD	0.7021	0.7139	0.7160
	LR	0.7101	0.7193	0.7231
	MSLR	0.7010	0.7108	0.7187
phrase	MSD	0.7023	0.7091	0.7172

	LR	0.7056	0.7198	0.7269
	MSLR	0.7023	0.7165	0.7197
hierarchical	MSD	0.7131	0.7149	**0.7297**
	LR	0.7042	0.7207	0.7245
	MSLR	0.7097	0.7194	0.7137

In Table 2 and 3, distance denotes the distance-based reordering model, word denotes word-based reordering model, phrase denotes phrase-based reordering model and hierarchical represents hierarchical phrase-based reordering model. DL stands for distortion limits. We used bold to indicate the highest result score.

We find that LRM obtains significantly improvements over the distance-based reordering model because reordering probabilities is conditioned on the actual phrases.

The hierarchical reordering model for all orientation types and distortion limits in this experiment achieve highest BLEU and RIBES scores than all non-hierarchical models. We also observe that phrase-based reordering models produce better results than word-based reordering models.

The highest is 17.78 BLEU score and 0.7297 RIBES score in MSD orientation trained on hierarchical reordering model and DL to 12, which suggests that this is applicable for the adequacy of the translation and the word order in English-Myanmar translation system.

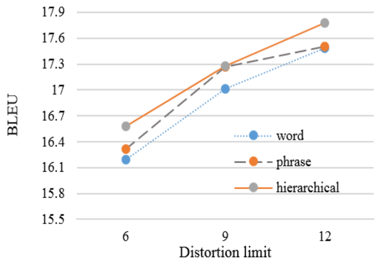


Figure 4. BLEU scores with increasing distortion limits for MSD orientation.

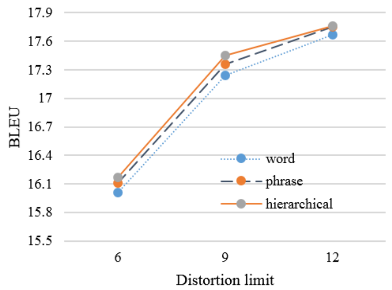


Figure 5. BLEU scores with increasing distortion limits for left/right orientation.

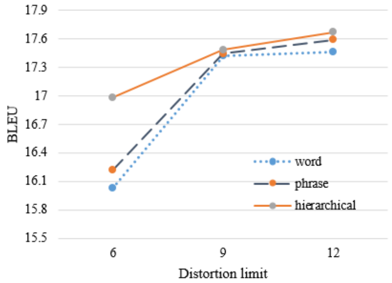


Figure 6. BLEU scores with increasing distortion limits for MSLR orientation.

The BLEU results in Table 2 are plotted as shown in figure 4, 5 and 6. These figures show the evaluation of three lexicalized reordering models based on MSD, LR and MSLR orientation types with increasing distortion limits. Since the experiments study the effect of various LRM, it is interesting to examine how the distortion limit affects translation performance. The results show that DL of 12 obtains maximum improvement (BLEU and RIBES score) for all LRM with each orientation in our experiments.

8. Conclusion

Our motivation for this paper is to investigate various lexicalized reordering models on English-Myanmar phrase-based SMT. We made experiments over word-based, phrase-based and hierarchical phrase-based LRM using different orientations and different distortion limits. According to the experiments, our results indicated that highest BLEU and RIBES scores are achieved by hierarchical phrase-based LRM with MSD orientation at DL of 12. However, the language model used is KenLM only with 5-grams. We will do experiments on SRILM and neural language models by doing more preprocessing steps such as word segmentation and POS information to test PBSMT in the future and investigate the reordering performance.

9. Acknowledgements

This work is partly supported by the Asian-MT "Network-based ASEAN Languages Translation Public Service" and the ASEAN IVO Project "Open Collaboration for Developing and Using Asian Language Treebank".

References

[1] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation", Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Santa Monica, July 22-23, 2002, pp. 48-54.

[2] C. Tillman, "A unigram orientation model for statistical machine translation", in Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, 2004, Short Papers, pages 101–104.

[3] P. Koehn, A. Axelrod, A. Birch Mayne, C. Callison Burch, M. Osborne and D. Talbot, "Edinburgh system description for the 2005 IWSLT speech translation evaluation," the International Workshop on Spoken Language Translation., 2005.

[4] M. Galley and C. D. Manning. 2008, "A simple and effective hierarchical phrase reordering model" in Proceedings of the 2008 Conference on Empirical Methods in Natural Language, Honolulu, October 2008, pages 848–856.

[5] D. Xiong, Q. Liu, S. Lin, "Maximum entropy based phrase reordering model for statistical machine translation.", In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, Sydney, July 2006, pp. 521–528.

[6] S. kanouchi, K. Sudoh, M. Komach, "Neural Reordering Model Considering Phrase Translation and Word Alignment for Phrase-based Translation", Proceedings of the 3rd Workshop on Asia Translation, pages 94-103, Osaka, Japan, December 11-17 2016.

[7] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu (2002), "BLEU: A Method for Automatic Evaluation of Machine Translation", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pages 311-318.

[8] H.Isozaki, T.Hirao, K. Duh, K. Sudoh, H. Tsukada, "Automatic Evaluation of Translation Quality for Distant Language Pairs", Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, USA, 9-11 October 2010, pp. 944-952.

[9] W. P. Pa, N. L. Thein, "Myanmar Word Segmentation using a Combined Model", e-Case2009, Singapore, January, 2009.

[10] Moses: Open source toolkit for statistical machine translation. http://www.statmt.org/moses.

[11] P. Koehn, A. Axelrod, A. B. Mayne, C. Callison-Burch, M. Osborne, and D. Talbot, "Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation," Proceedings of the International Workshop on Spoken Language Translation, pp. 68–75, 2005.

[12] P. Koehn, "Moses, Statistical Machine Translation System, User Manual and Code Guide", June 22, 2016.

[13] F. J. Och and H. Ney, "A systematic comparison of various statistical alignment models", Computational linguistics, 29(1):19–51, 2003.

[14] Heafield, Kenneath, "KenLM: Faster and Smaller Language Model Queries", Proceedings of the Sixth Workshop on Statistical Machine Translation; WMT 11, 2011, Association for Computational Linguistics, Edinburgh, Scotland, pp 187-197 ISBN- 978-1-937284-12-1.

[15] P. Koehn, A. Axelrod, A. B. Mayne, C. Callison-Burch, M. Osborne, and D. Talbot, "Edinburgh system description for the 2005 IWSLT speech translation evaluation." In IWSLT 2005.

[16] V. H. Tran, H. T. Vu, T. H. Pham, V. V. Nguyen, M. L. Nguyen, "A Reordering Model for Vietnamese-English Statistical Machine Translation Using Dependency Information.", the 2016 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future.

[17] R. Gupta, R. N. Patel, R. Shah "Learning Improved Reordering Models for Urdu, Farsi and Italian using SMT", Proceedings of the Workshop on Reordering for Statistical Machine Translation, COLING 2012, Mumbai, December 2012, pages 37–46.

[18] C. Tillmann and T. Zhang, "A Localized Prediction Model for Statistical Machine Translation," in Proc. of ACL, Ann Arbor, Michigan, 2005, pp. 557–564.

[19] J. Su, Y. Liu, Y. Liu, H. Mi, Q. Liu, "Learning Lexicalized Reordering Models from Reordering Graphs",

[20] R. Zens and H. Ney, "Discriminative reordering models for statistical machine translation.", In Proc. of Workshop on Statistical Machine Translation 2006, pages 521–528.

[21] http://www2.nict.go.jp/astrec-att/member/mutiyama/ALT

[22] http://www.aseanmt.org/index.php