

Word Alignment-Based Reordering of Source Chunks in PB-SMT

Santanu Pal¹, Sudip Kumar Naskar², Sivaji Bandyopadhyay²

¹Universität des Saarlandes, Saarbrücken, Germany

²Jadavpur University, Kolkata, India

santanu.pal@uni-saarland.de, {sudip.naskar, sbandyopadhyay}@cse.jdvu.ac.in

Abstract

Reordering poses a big challenge in statistical machine translation between distant language pairs. The paper presents how reordering between distant language pairs can be handled efficiently in phrase-based statistical machine translation. The problem of reordering between distant languages has been approached with prior reordering of the source text at chunk level to simulate the target language ordering. Prior reordering of the source chunks is performed in the present work by following the target word order suggested by word alignment. The testset is reordered using monolingual MT trained on source and reordered source. This approach of prior reordering of the source chunks was compared with pre-ordering of source words based on word alignments and the traditional approach of prior source reordering based on language-pair specific reordering rules. The effects of these reordering approaches were studied on an English–Bengali translation task, a language pair with different word order. From the experimental results it was found that word alignment based reordering of the source chunks is more effective than the other reordering approaches, and it produces statistically significant improvements over the baseline system on BLEU. On manual inspection we found significant improvements in terms of word alignments.

Keywords: Machine Translation, reordering, prior reordering, chunk-level reordering, word alignment based

1. Introduction

Reordering is the one of the most difficult problems in statistical machine translation (SMT); it presents itself differently for different language-pairs. For some language pairs (English–French, Chinese–English, etc.) only local movements are sufficient for translation, while some language-pairs have significant syntactic divergences. Particularly, SMT between SVO–SOV (e.g., English–Hindi) or SVO–VSO (e.g., English–Arabic) language pairs suffer from long-distance reordering phenomena. Most of the Indian languages are relatively free phrase-order languages; they are generally verb-final, i.e., verb phrases are positioned at the end of the sentence and local movement of words within phrases also takes place.

In SMT, language models play a crucial role in positioning the target words in an acceptable order. But language models also have their limitations; it typically considers up to 5-grams, which is not sufficient enough to make decision about a good translation. If we increase the n-gram value then the reordering cost involved is much higher in terms of computational effort and requirement; besides longer n-grams in language models suffer from data sparsity.

In the phrase-based SMT (PB-SMT) framework, reordering is typically handled by two models: distortion model and lexicalized reordering model (Koehn et al., 2005; Galley and Manning, 2008). Distortion model was proposed by the IBM Models (Brown et al., 1993). IBM models 1 and 2 define the distortion parameters in accordance with the word positions in the sentence pair instead of actual words at those positions. The distortion probability also depends on the source and target sentence lengths. Models 4 and 5 limit this by replacing absolute word positions with relative word positions. However, all these models are limited to only word movements; they do not consider phrasal movement. Koehn et al. (2003) proposed relative distortion model in PB-SMT. The model works in terms of the

difference between current phrase position and the previous phrase position in the source sentence. Basic PB-SMT model considers word movements up to 6 tokens which could be increased to consider long distance reordering; however, higher distortion limits usually result in degraded performance (Koehn et al., 2007).

Lexicalized reordering model conditions reordering on the PB-SMT phrases. It consists of three types of reordering – monotone (M), swap (S), and discontinuous (D) – by considering the orientation of previous and next phrases. The orientation is called monotone if the previous source phrase is aligned with the previous target phrase. The orientation of the swap occurs when the next source phrase is aligned with the previous phrase in the target; and the orientation is termed as discontinuous if neither of the two above mentioned cases are true, i.e. neither monotone or swap. The reordering model is built by calculating the probabilities of the phrase pairs being associated with the given orientation. Notwithstanding the reordering models used in the state-of-the-art PB-SMT, the differences in word ordering between distant languages result in poor translation quality.

The remainder of this paper is organized as follows. Section 2 briefly discusses previous research relevant to the context. Section 3 focuses on the traditional approach of tree based reordering. Section 4 describes word alignment-based approach to reordering of source chunks. Section 5 reports the tools and algorithms used along with a description of the datasets used. Section 6 presents the experiments and results together with some analysis, followed by conclusions and future work.

2. Related Work

To alleviate the problem of reordering, researchers have carried out work in two directions: one which tries to directly improve the reordering model inside the SMT system, and the other by prior reordering of the source text

so that it resembles the target word order. This section presents an overview of works that deal with prior reordering of the source text to emulate the target word order.

Prior reordering of the source text affects the performance in two ways as stated in Holmqvist et al. (2012). Firstly, it lessens the burden of the reordering model since most of the long-distance reorderings are taken care of during the reordering of the source text prior to training; only minor reorderings are performed during decoding and the translation hypothesis is constructed almost monotonically. Secondly, since statistical word alignment techniques are known to perform better for language pairs with similar word order, prior source reordering essentially should lead to more accurate word alignments and hence better translation model and improved translation quality.

Most of the works in pre-ordering rely either on automatically acquired (Xu et al., 2009; Niehues and Kolss, 2009; Genzel, 2010; Gupta et al., 2007; Habash, 2007) or handcrafted reordering rules (Collins et al., 2005; Popović and Ney, 2006). Reordering rules are usually automatically learnt from parsed training corpus and/or word alignments. Holmqvist et al. (2012) presented a method where source text is reordered to replicate the target word order based on word alignment. Then word alignment is performed between reordered source and target training data; the new word alignments are transferred back to the original training data to connect words in their original order which results in the same parallel training data with potentially improved word alignments. They reported improved translation quality for English–German and English–Swedish. They also studied the effect of this preprocessing on the word alignment quality and found that this approach resulted in improved recall but degraded precision.

Andreas et al. (2011) reported improvements in an Arabic–English translation task by using two parse fuzzification techniques that allow the translation system to select among a range of possible subject–verb reorderings.

A syntax-driven approach to reordering using association rule mining was proposed by Avinesh (2010) where reordering rules are automatically learned from parsed source side and word alignment; however it resulted in a drop in BLEU score compared to baseline MOSES.

Dan et al. (2012) proposed linguistically motivated head-finalization reordering rules based on HPSG parses in a Chinese-to-Japanese translation task and reported significant improvements in translation quality. Gupta et al. (2007) proposed a POS-based prior reordering model which learns to reorder adjectives, nouns and verbs by observing the distances between the source and target phrases using target-to-source alignments. Their model was employed as an additional feature function at the rescoring stage of PB-SMT and it resulted in improved BLEU scores in Japanese–English and German–English translation tasks. Xu et al. (2009) presented a preordering approach where handcrafted precedence rules are applied on dependency trees recursively. They applied this approach on five English to SOV languages and achieved statistically significant improvements over the respective PB-SMT baselines for all the language pairs.

Niehues and Kolss (2009) proposed automatically extract-

ing POS-based discontinuous reordering rules from word-aligned parallel data to model long-range reorderings. This method improves over applying POS-based continuous reordering rules and baseline PB-SMT.

Badr et al. (2009) presented linguistically motivated reordering rules that reorder English text to look like Arabic. To automatically detect and relocate clause-initial verbs in the Arabic side of a word-aligned parallel corpus, Bisazza and Federico (2010) proposed a chunk-based reordering technique that impacts the VSO type sentences in Arabic–English machine translation. Carpuat et al. (2010) proposed a novel approach to improve the SMT quality using a noisy syntactic parser that reorders verb-subject construction to subject-verb construction in Arabic–English SMT.

In the present paper, we proposed word alignment-based pre-ordering of source chunks which is inspired by and an extension of Holmqvist et al. (2012). However there are two important distinctions between the work presented here and in Holmqvist et al. (2012). Firstly, the main objective of Holmqvist et al. (2012) was to improve word alignment, not reordering. They do not use the reordered training set to train the final system. Contrary to Holmqvist et al. (2012), in the present work we address both the issues of word alignment and reordering, and reorder the source side of all the datasets accordingly. Secondly, Holmqvist et al. (2012) reordered source words based on word alignment, whereas we suggest reordering source chunks. We also showed that chunk-level reordering is much more effective than word-level reordering.

The motivation behind this work stems from the fact that word alignment-based pre-ordering of source words requires neither any reordering rules, nor any language dependent preprocessing. But word alignment-based reordering of source words is crucially dependent on the quality of the word alignment. The objective of the present work is to reorder the source chunks such that the source and target chunk alignments become monotone. We argue that with imperfect word alignments it might not be possible to produce perfectly monotone word alignments. However, by using these word alignments we can obtain monotone chunk associations which reduces the problem of long-range reordering to only short-range, intra-chunk reordering while preserving some source language syntax. The only language-dependent processing involved is chunking in the source language. The assumption is that human translators perform translation at chunk level rather than at the word level, and given the choices of translating from word- and chunk-reordered source text, human translators would much prefer translating from the latter.

3. Tree Based Reordering

For tree-based reordering we only consider repositioning the verbs at the end of the sentence or clause. We categorize each source (i.e., English) sentences into three basic types: simple, complex and compound, and reposition the verbs accordingly. For identifying the basic sentence type we first parse the source sentences. The tree outputs are categorized into the above mentioned three types by analyzing the structure of the tree and presence of key words such as that, which, who etc. and tags like CC, WHNP, SBAR, S,

etc.

4. Word Alignment-based Reordering

In this section we discuss a simple yet effective language-independent approach to reordering based on word alignment following Holmqvist et al. (2012). This method has the advantage that it does not necessitate any reordering rules, handcrafted or automatically acquired. It also avoids any language dependent preprocessing of the target language; it only requires chunking of the source language. In this reordering approach we first run the GIZA++ word alignment tool on the original parallel corpus bidirectionally which produces 1-to-n alignments for both directions. Then a symmetrization matrix is built on these two unidirectional word alignments and the ‘grow-diagonal-final-and’ (gdfa) heuristic is applied which produces m-to-n alignments. The gdfa heuristic is believed to be the most favourable word alignment heuristic for PB-SMT, and is used in the MOSES vanilla settings. This word alignment serves as the basis for this source reordering approach. Once the word alignment has been obtained, chunks are identified in the source (i.e., English) side of the training set which are then reordered following the word alignment. The chunk reordering process is outlined in algorithm 1. For reordering a source sentence, the algorithm starts with the chunked source sentence and the word alignment for that sentence pair. Let us consider the following chunked source sentence, the target sentence and the word alignment:

$$S = (s_1, s_2, s_3, \dots, s_p) = (C_1, C_2, C_3, \dots, C_m)$$

where

$$C_i = (s_j, \dots, s_{j+n})$$

and

$$T = (t_1, t_2, t_3, \dots, t_q)$$

where S is a source sentence, T is the corresponding target sentence, and s , t and C represent source word, target word and source chunks respectively. The alignment between words in S and T is given by:

$$A = \{a_1, a_2, \dots, a_r\}, \text{ where } a_k = [s_j, t_l]$$

For the sake of simplicity the algorithm assumes 1-based index while 0-based indexing is used for the actual alignments.

The algorithm uses a list of indices, $list_pos$, for each source chunk. $list_pos_i$ stores indices of target words which are linked to the component words of the i^{th} source chunk (C_i) via word alignment, i.e.,

$$list_pos_i = \{j : t_j \in T \wedge \exists k : s_k \in C_i \wedge [s_k - t_j] \in A\}$$

In an ideal scenario, all tokens in a source sentence, or at least some tokens in every source chunk should be aligned to some tokens in the corresponding target sentence; but that is not always the case. If no correspondence can be found with the target via word alignment for any of the tokens belonging to C_i , the source chunk position of C_i is added to $list_pos_i$. Finally, the entries in each $list_pos$ are

sorted in ascending order, and the chunks are arranged according to the first entry in the corresponding $list_pos$.

The pseudo-code of the algorithm used to reorder source chunks according to word alignment information is illustrated in Algorithm 1.

```

for  $i = 1$  to  $m$  chunks in the source do
   $new\_pos_i = \text{NULL}$ ;
  for  $j = 1$  to  $n$  source words in  $chunk_i$  do
    for  $k = 1$  to  $r$  alignments in  $A$  do
      if  $a_k = [s_j, t_l]$  then
        add  $l$  to list  $new\_pos_i$ ;
      end
    end
  end
  if  $new\_pos_i$  is  $\text{NULL}$  then
    add  $i$  to  $new\_pos_i$ ;
  end
end
for  $i = 1$  to  $m$   $new\_pos$  lists do
  Sort the items in  $new\_pos_i$  in ascending order;
end
 $reordered\_sen = \text{NULL}$ ;
while not(all  $new\_pos$  lists are empty) do
   $i =$  index of the first  $new\_pos$  list containing the
  smallest first entry;
   $j =$  first entry in  $new\_pos_i$ ;
  Append  $chunk_j$  to  $reordered\_sen$ ;
   $new\_pos_i = \text{NULL}$ ;
end

```

Algorithm 1: Word alignment-based source chunk reordering

5. Tools and Resources Used

The English–Bengali parallel dataset used for the experiments presented in this paper are from the tourism and travel domain and were taken from the EILMT¹ project. For identification of chunks, English training set and testset sentences are first POS-tagged using Stanford POS tagger². Chunks are then identified from the POS-tagged sentences using a CRF chunker³. The source side of the datasets were parsed using Stanford Parser⁴.

The MT experiments were carried out using the standard log-linear phrase-based SMT toolkit MOSES (Koehn et al., 2007). GIZA++ (Och and Ney, 2003) implementation of IBM word alignment model 4 with the grow-diagonal-final-and heuristic was used for performing word alignment. Phrase extraction was performed following Koehn et al. (2003). The feature weights were tuned using minimum error rate training (Och, 2003) on a held-out development set in terms of BLEU. For language modelling purpose we

¹The EILMT project is funded by the Department of Information Technology (DIT), Ministry of Communications and Information Technology (MCIT), Government of India.

²<http://nlp.stanford.edu/software/tagger.shtml>

³<http://crfchunker.sourceforge.net/>

⁴<http://nlp.stanford.edu/software/lex-parser.shtml>

used the SRILM toolkit (Stolcke, 2002) with Kneser-Ney smoothing (Kneser and Ney., 1995).

6. Experiments and Results

The initial parallel corpus was filtered with maximum allowable sentence length of 100 words and a maximum sentence length ratio of 1:2. 500 sentences were then randomly taken from the filtered dataset for both the development set and the test set and the rest (22,176 sentences) was treated as the training corpus. The target language model was built on the target side of the parallel corpus along with a monolingual Bengali corpus containing 488,026 words from the tourism domain. To reduce the data sparseness problem, English text in all the datasets was lowercased. We carried out all the experiments with a 4-gram static language model and maximum phrase length of 7 as they produced the best results for the baseline PB-SMT system. Table 1 presents the experimental results. We carried out experiments on tree-based reordering and word alignment-based source reordering. To compare the effect of word alignment-based reordering at chunk- and word-level, we carried out experiments on both. For the sake of completeness we also carried out experiments on word-based SMT (setting the phrase length to 1) to see whether chunk-level reordering could bring any improvement over baseline word-based SMT. We also replicated the experiment of Holmqvist et al. (2012) on this dataset. Holmqvist et al. (2012) reported 1-pass reordering experiment, while we carry out both 1-pass and 2-pass experiments. In 2-pass experiment, the process of reordering the source side is simply carried out twice, i.e., reordered source side is subjected to reordering once again. We also carried out a chunk-reordering PB-SMT experiment where the chunks are reordered based on the final alignments obtained by 1-pass experiment of Holmqvist et al. (2012).

It is to be noticed that for applying any pre-reordering technique, the testset (and in case of tuning, the development set) needs to be reordered as well using the same technique. For the tree-based reordering approach we reordered the testset and the development set using the same set of rules. For the word alignment based reordering experiments, the testset is reordered using monolingual PB-SMT systems built on the original source training data and the corresponding reordered source training data. For the monolingual PB-SMT systems, we do not perform automatic word alignment since the word alignments between the source training set and the reordered training set are already known. We create two lexical translation tables where each source word has only one translation option, i.e., the same word itself in the target, with translation probability of 1.0. It is to be noted that both these lexical translation tables are exactly the same. The phrase table and the reordering table are built on these alignments using MOSES. Since the purpose of this monolingual PB-SMT is to just reorder the source sentences, we do not use a language model for this monolingual PB-SMT model. A monolingual PB-SMT system built thus essentially just reorders the source sentences. The 'TR' column in Table 1 indicates whether the testset is reordered (using monolingual MT) in the corresponding experiment.

We carried out evaluation of the MT quality using 4 automatic MT evaluation metrics: BLEU (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005), NIST (Doddington, 2002) and TER (Snover et al., 2006). For the PB-SMT experiments, tree-based reordering brings some improvements over the PB-SMT baseline. Word alignment-based reordering at word-level also provides some improvements over the PB baseline; however the improvements are less than those obtained in tree-based reordering. Word alignment-based reordering at chunk-level improves over both and provides the overall best BLEU score among all 1-pass PB-SMT experiments (exp. 4–8). A similar trend is observed for the word-based SMT experiments for which both word- and chunk- level reordering prove to be beneficial over the baseline while chunk-level reordering appears to be more effective than word-level reordering. Our approach to alignment-based chunk-reordering (exp. 7) outperforms alignment-based word-reordering (exp. 8) described in Holmqvist et al. (2012). However, tree-based reordering produced the best scores as per TER among all 1-pass PB-SMT experiments.

The 2-pass approach to alignment-based word-reordering (exp. 9) also improves over 1-pass approach (exp. 8) across all metrics, however the improvements are small. Our final experiment (exp. 10) with chunk reordering based on the final alignments obtained by exp. 8 produces the overall best scores in BLEU and TER. Statistical significance tests were carried out using bootstrap resampling method (Koehn, 2004) and the * marked scores represent statistically significant improvements on BLEU over the respective baseline systems.

Figure 6. shows the effect of prior reordering of the source on word alignment. Figure 1.a shows the initial word alignment extracted by the baseline system for a sentence pair. Figure 1.b presents the correct (i.e., manual) alignment and figure 1.c shows the final word alignment obtained by chunk-reordered (CR, exp. 7) and word-reordered (WR, exp. 6) PB-SMT systems for the sentence pair. Figure 1.b in addition shows if the source chunks could be ordered properly (which is indeed the case here) how they could minimize the number of cross links (28 down to 2 here). The correct alignments are shown as solid lines and the wrong ones as dotted lines in figures 1.a and 1.c. English chunks are shown in brackets and Bengali chunks are shown as underlined. It is to be noted that chunking in the target side is not required in alignment-based reordering; the target side has been chunk-marked in figures 1.b and 1.c just for visualization of source-target chunk associations. In the initial word alignment (cf. figure 1.a) 11 out of 14 word associations are correct (precision=0.79, recall=0.73). However, when the source sentence is (chunk-) reordered based on this initial word alignment, the association between the source and target chunks become monotone (cf. figure 1.c). In the final word alignment between the word-reordered source and the target sentence, 12 out of 15 alignments are correct (precision=0.8, recall=0.8), an improvement over the baseline, while 13 out of 14 alignments are correct (precision=0.93, recall=0.87) between the chunk-reordered source and the target sentence. Thus, word-alignment based source reordering im-

Table 1: Evaluation results obtained on the reordering experiments.

Experiments	Prior reordering	Level	TR	Exp	BLEU	NIST	METEOR	TER
Word-based SMT	none (baseline)		no	1	8.87	3.61	0.3028	86.95
	alignment-based	word	yes	2	8.97	3.54	0.2985	88.86
		chunk	yes	3	9.94*	3.71	0.3107	86.64
Phrase-based SMT	none (baseline)		no	4	10.68	4.13	0.3035	73.37
	tree-based		yes	5	11.53*	4.22	0.3126	72.75
	alignment-based	word	yes	6	11.11	4.08	0.3073	75.34
		chunk	yes	7	12.65*	4.29	0.3144	73.00
	alignment-based reordering, 1-pass	word	no	8	11.25	4.09	0.3129	75.25
	alignment-based reordering, 2-pass	word	no	9	11.47	4.12	0.3141	75.14
	alignment-based reordering, 2-pass	chunk	yes	10	13.17*	4.28	0.3161	72.66

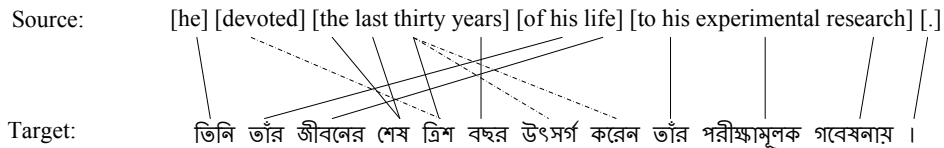


FIGURE 1.a – Initial word alignment

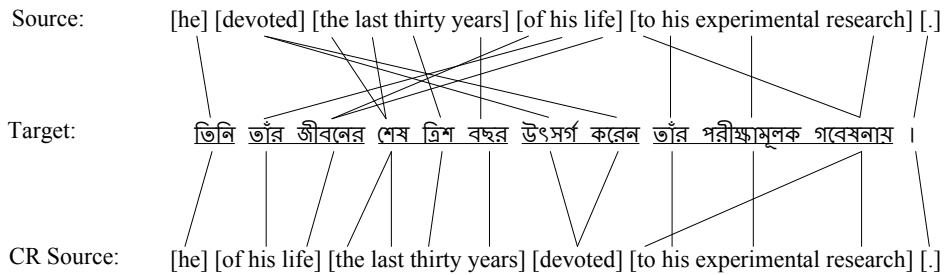


FIGURE 1.b – Correct word alignment

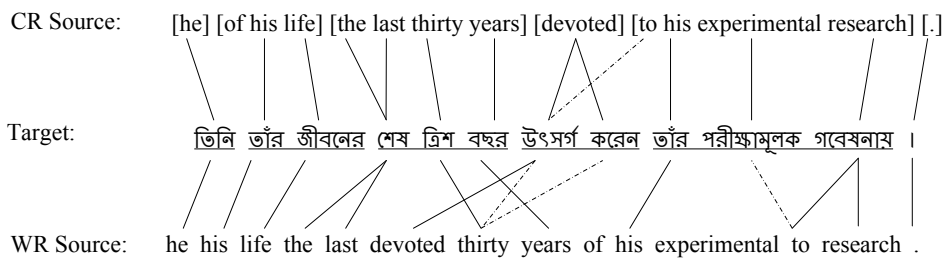


FIGURE 1.c – Final word alignment

Figure 1: Word alignments with unordered and reordered source.

proves both word alignment precision and recall. This example illustrates two important improvements: firstly, word-alignment based chunk reordering of the source results in less cross-chunk alignments, in this case zero (cf. figure 1.c), and secondly and more importantly, it improves the accuracy of the word alignment. From this example it is also evident that word-alignment based chunk reordering is more effective than word-alignment based word reordering in PB-SMT which is quite intuitive. This approach to re-

ordering can be considered as a bootstrapping approach to word alignment since this is based on word alignment and the purpose of it is to essentially improve the word alignment quality. Word alignments produced by statistical word aligners are never perfect even for sizable amount of data; if they were perfect it would have defeated the purpose of reordering. In this real world scenario it makes more sense to reorder at chunk level than at word level since both rely on imperfect word alignments while chunk-level reordering

preserves some source language syntax and is less affected by noisy word alignments.

Due to the unavailability of the gold-standard word alignment, improvement in terms of word alignment quality could not be measured empirically; however this example clearly demonstrates the usefulness of word-alignment based source chunk reordering in improving word alignment quality. Although this approach calls for the testset to be reordered (as opposed to Holmqvist et al. (2012)) and is sensitive to errors in chunking, it was still able to produce significant improvements over the baseline systems. We inspected the lexfile and phrase table sizes for the PB-SMT experiments and found that lexfile and phrase table sizes were inversely proportional to the BLEU scores obtained on them, which essentially suggests that prior reordering also reduces the data sparsity problem.

7. Conclusions

In this paper we have presented a method of source chunk pre-ordering based on word alignment. Source chunks are reordered based on their associations with the target words and the target word order. The testset is reordered using monolingual PB-SMT built on the original source training data and the reordered source training data. Our experiments showed that word alignment based source chunk pre-ordering is more effective than word alignment based source word pre-ordering and tree-based reordering; and it produced statistically significant improvements on both. On manual inspection we found significant improvements in terms of word alignments. This method also reduces the data sparsity problem. The method presented in the paper has the advantage that it does not require any language specific tools like parsers excepting a chunker for the source language.

8. Acknowledgements

The research leading to these results has received funding from the EU project EXPERT – the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme FP7/2007-2013 under REA grant agreement no. [317471] and from Science Foundation Ireland as part of the CNGL (grant 07/CE/I1142).

9. References

Andreas, J., Habash, N., and Rambow, O. (2011). Fuzzy syntactic reordering for phrase-based statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT ’11, pages 227–236, Stroudsburg, PA, USA. Association for Computational Linguistics.

Avinesh, P. V. S. (2010). A data mining approach to learn reorder rules for smt. In *Proceedings of the NAACL HLT 2010 Student Research Workshop*, HLT-SRWS ’10, pages 52–57, Stroudsburg, PA, USA. Association for Computational Linguistics.

Badr, I., Zbib, R., and Glass, J. (2009). Syntactic phrase reordering for english-to-arabic statistical machine translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational*

Linguistics, EACL ’09, pages 86–93, Stroudsburg, PA, USA. Association for Computational Linguistics.

Banerjee, S. and Lavie, A. (2005). Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72.

Bisazza, A. and Federico, M. (2010). Chunk-based verb reordering in vso sentences for arabic-english statistical machine translation. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, WMT ’10, pages 235–243, Stroudsburg, PA, USA. Association for Computational Linguistics.

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311, June.

Carpuat, M., Marton, Y., and Habash, N. (2010). Improving arabic-to-english statistical machine translation by reordering post-verbal subjects for alignment. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort ’10, pages 178–183, Stroudsburg, PA, USA. Association for Computational Linguistics.

Collins, M., Koehn, P., and Kučerová, I. (2005). Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL ’05, pages 531–540, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dan, H., Sudoh, K., Wu, X., Duh, K., Tsukada, H., and Nagata, M. (2012). Head finalization reordering for chinese-to-japanese machine translation. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, SSST-6 ’12, pages 57–66, Stroudsburg, PA, USA. Association for Computational Linguistics.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the Second International Conference on Human Language Technology Research*, HLT ’02, pages 138–145, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Galley, M. and Manning, C. D. (2008). A simple and effective hierarchical phrase reordering model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP ’08, pages 848–856, Stroudsburg, PA, USA. Association for Computational Linguistics.

Genzel, D. (2010). Automatically learning source-side reordering rules for large scale machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING ’10, pages 376–384, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gupta, D., Cettolo, M., and Federico, M. (2007). Pos-based reordering models for statistical machine translation. In *Proceedings of the MT Summit XI*, pages 207–213.

Habash, N. (2007). Syntactic preprocessing for statistical

- machine translation. In *Proceedings of the MT Summit XI*, pages 215–222.
- Holmqvist, M., Stymne, S., Ahrenberg, L., and Merkel, M. (2012). Alignment-based reordering for smt. In Chair), N. C. C., Choukri, K., Declerck, T., Doan, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Kneser, R. and Ney, H. (1995). Improved backing-off for m-gram language modeling. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Volume I*, pages 181–184.
- Koehn, P., Och, F. J., and Marcu, D. (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*, NAACL '03, pages 48–54, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Koehn, P., Axelrod, A., Mayne, R. B., Callison-burch, C., Osborne, M., and Talbot, D. (2005). Edinburgh system description for the 2005 iwslt speech translation evaluation. In *In Proc. International Workshop on Spoken Language Translation (IWSLT)*.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (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*, ACL '07, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Niehues, J. and Kolss, M. (2009). A pos-based model for long-range reorderings in smt. In *Proceedings of the Fourth Workshop on Statistical Machine Translation, StatMT '09*, pages 206–214, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, March.
- Och, F. J. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 160–167, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Popović, M. and Ney, H. (2006). Pos-based reorderings for statistical machine translation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, LREC-2006, pages 1278–1283.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *In Proceedings of Association for Machine Translation in the Americas*, pages 223–231.
- Stolcke, A. (2002). Srilm - an extensible language modeling toolkit. pages 901–904.
- Xu, P., Kang, J., Ringgaard, M., and Och, F. (2009). Using a dependency parser to improve smt for subject-object-verb languages. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 245–253, Stroudsburg, PA, USA. Association for Computational Linguistics.