Post-ordering in Statistical Machine Translation

Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada, Masaaki Nagata
NTT Communication Science Labs.
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
sudoh.katsuhito@lab.ntt.co.jp

Abstract

In the field of staistical machine translation (SMT), pre-ordering is a recently attractive approach that reorders source language words into the target language order prior to SMT decoding. It is effective for long-distance reordering in SMT, especially between languages with distant word ordering like English and Japanese. Its key idea is to decompose the SMT problem into two subproblems of translation and reordering and to solve them independently. However, most pre-ordering approaches employ syntactic parsing in the source language and reordering rules depending on a certain target language. This paper focuses on the translation in the opposite direction and proposes *post-ordering*; foreign sentences are first translated into *foreign-ordered* English, and then reordered into correctly-ordered English. The experiments on Japanese-to-English patent translation show the significant advantage of post-ordering over baseline phrase-based, hierarchical phrase-based, and syntax-based translation methods by 1.56, 0.76, and 2.77 points in BLEU, respectively.

1 Introduction

Statistical Machine Translation (SMT) consists of two major problems, translation of words or phrases and their reordering. Recent research efforts developed novel technologies such as phrase-based SMT with phrase reordering models (Koehn et al., 2003; Tillmann, 2004), and tree-based (or syntax-based) SMT (Yamada and Knight, 2001; Galley et al., 2004; Chiang, 2007). The reordering problem have been studied in many research works among various language pairs. However, it is still an important challenge in language pairs with very different word ordering such as English (Subject-Verb-Object) and Japanese (Subject-Object-Verb). That is mainly due to the computational costs both in time and space.

A recently attractive approach for this challenge is called *pre-ordering*, which reorders source language sentences into the target language word order prior to SMT decoding (Xia and McCord, 2004; Collins et al., 2005; Costa-jussà and Fonollosa, 2006; Li et al., 2007; Wang et al., 2007; Tromble and Eisner, 2009; Isozaki et al., 2010; Genzel, 2010). The pre-ordering approach is able to reorder source language words in long distance by some reordering rules or models. This effectively solves the complex reordering problem and achieves good translation performance especially in language pairs with very different word ordering. A crucial issue on the pre-ordering is to develop good reordering methods in the *source* language.

In contrast, what can we do in the translation in the opposite direction? This is a non-trivial problem because the pre-ordering techniques are usually language dependent. Even if we have a good pre-ordering technique in A-to-B translation such as reordering rules for syntactic parse trees, it cannot be used directly in B-to-A translation. Developing B-to-A pre-ordering is a different problem from A-to-B, which may require a syntactic parser and/or linguistic insights. For example in Japanese-to-English translation, pre-ordering of Japanese parse trees into English word order is not a trivial problem, while that of English parse trees into Japanese word order can be implemented by a simple rule of moving syntactic heads (Isozaki et al., 2010). This implies another question; can we utilize a pre-reordering technique in the opposite direction? If it is feasible, we can easily reverse the translation direction in which the pre-ordering approach successfully works.

This paper proposes a novel approach for this problem, which we call *post-ordering*. As the name implies, it first translates source language sentences into *source-ordered* target language sentences and then reorders these words to *correct* target language word order. Figure 1 illustrates the differences among: (a) a standard translation direction from the source language to the target language, (b) the pre-ordering approach, and (c) the post-ordering approach. As shown in Figure 1, the pre- and post-ordering are very similar methodologies. Their monotone translation parts are almost identical, except their targeting word order. The post-ordering process is the problem of reordering differently ordered sentence in the correct word order. This is the inverse problem of the pre-ordering that changes word order of correctly ordered sentences. We solve this inverse problem as a SMT problem. Once we have bitexts of source and target language and the pre-ordering rules or models for target-to-source translation, we can easily generate source-ordered target language sentences. We then use these source-ordered target sentences and the target language portion of the bitexts to train the SMT models for "source-ordered target" to "correct target" translation. This post-ordering approach has an advantage on saving the effort to develop new pre-ordering rules and models by utilizing *good* pre-ordering in the opposite direction. Note that it can be used in any language pairs in which we have pre-ordering techniques in the opposite direction.

We focus on Japanese-to-English translation in this paper, since this remains a challenging translation pair for SMT. In case of opposite direction, English-to-Japanese, Isozaki et al. (2010) proposed simple but effective English reordering rules called *Head-Finalization* that moves syntactic heads toward the end of their siblings. This Head-Finalization rule matches the head-final word ordering of Japanese and works quite well in English-to-Japanese translation. However, pre-ordering in the opposite direction by placing Japanese syntactic heads in the middle is not a trivial problem. We utilize the Head-Finalization rules to generate intermediate head-finalized English sentences called Head-Final English (HFE) and decompose Japanese-to-English translation into 1) Japanese-to-HFE translation and 2) HFE-to-English post-ordering. We achieved significant improvements from baseline (phrase-based, hierarchical phrase-based, and string-to-tree) translation methods by 1.56, 0.76, and 2.77 points in BLEU, respectively, in the experiment of patent translation.

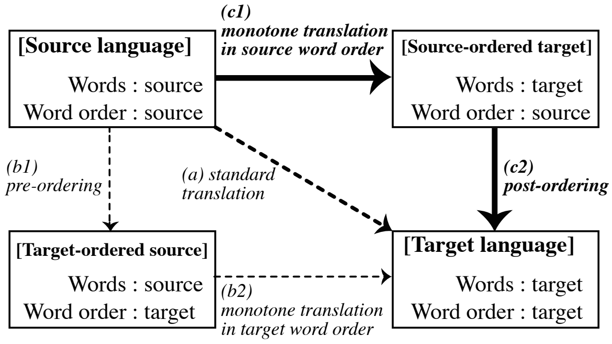

Figure 1: A typical work flow in a standard translation, pre-ordering and post-ordering approach.

The remainder of this paper is organized as follows. Section 2 briefly reviews related studies on the reordering problem and another related technology called post-editing. Section 3 presents the proposed method in detail taking Japanese-to-English translation as a test case. Section 4 reports our experiments and discusses the results. Section 5 concludes this paper with our prospects for future work.

2 Related Work

Reordering is a both theoretically and practically challenging problem in SMT. In the early period of SMT studies, reordering is modeled by distance-based constraints in translation model (Brown et al., 1993; Koehn et al., 2003). This reordering model is easy to compute and also works well in relatively similar language pair like French-to-English. The distance-based reordering constraint is not reasonable in some language pair such as English-to-Japanese, because they have very different word ordering and appropriate reordering distances of words and phrases highly depend on their syntactic roles and contexts. Tillmann (2004) proposed a lexicalized reordering model that models orientation of phrases by monotone, swap, and discontinuous. This can directly model reordering of adjacent phrases but may not work for long distance reordering, because *discontinuous* supplies few constraints for reordering. On the other hand, syntax-based SMT (Yamada and Knight, 2001; Galley et al., 2004; Graehl and Knight, 2004) is a theoretically good solution. Reordering in syntax-based SMT is modeled in a similar manner as reordering of tree nodes in the same level (siblings), regardless of their reordering distance. Although this approach have some shortcomings with parse errors and its too strong constraints, syntactic information is expected to be effective in some language pairs. Another syntactic approach, originally proposed by Wu (1997), uses formally-syntactic structure between source and target language sentences. This framework was extended as the hierarchical phrase-based SMT by Chiang (2007) and is convincing alternative in recent SMT research. The reordering models mentioned above are applied in SMT decoding and solved simultaneously with phrase translation. Xiong et al. extended the hierarchical SMT by lexicalized reordering (Xiong et al., 2006; Xiong et al., 2008). However, the integrated search requires a large computational cost both in time and space. To keep the search tractable, we constrain reordering search by its reordering distance, as so-called *distortion limit* (or *maximum span* in tree-based decoder). It effectively reduces the computational cost but it also give up long distance reordering exceeding the specified distortion limit.

A novel alternative to the reordering problem, called pre-ordering, has been studied over recent years (Xia and McCord, 2004; Collins et al., 2005; Li et al., 2007; Genzel, 2010). Xia and McCord (2004) proposed automatic reordering rule extraction for English-to-French translation; Collins et al. (2005) used linguistically-motivated hand-written rules for German-to-English translation; Li et al. (2007) presented discriminative syntactic reordering model for Chinese-to-English translation; Genzel (2010) explored English reordering rules by automatic word alignment and monotone translation for several SOV languages including Japanese. These methods apply reordering to input sentences, prior to the translation decoding. As a result, the translation decoding in the next step becomes nearly monotone and can search more phrase translation options efficiently. The pre-ordering is based on syntactic parse and can be regarded as a sub-problem of tree-to-string translation. On the other hand, there are several studies on pre-ordering without syntactic parsing. Costa-jussà and Fonollosa (2006) tackled the pre-ordering problem as SMT, using reordering tables derived from phrase tables. Tromble and Eisner (2009) applied linear ordering models to pre-ordering. Their techniques can be applied to any language pairs but rely on noisy automatic word alignment results as the reference of the reordering model training. Dyer and Resnik (2010) advanced such a pre-ordering-based translation to a novel unified approach of long-distance pre-ordering and decoding, with discriminative context-free reordering and finite-state phrase translation.

In this paper, we *reverse* the pre-ordering SMT framework for Japanese-to-English translation using English reordering rules on syntactic parse trees. There are a lot of pre-ordering studies, but this is the first work of *post-ordering* to our knowledge. The problem can be regarded as a variant of string-to-tree SMT, from Japanese sentences to English trees. We divide the string-to-tree problem into two simplified problems, which can be solved efficiently with less computational cost than a string-to-tree SMT.

Post-ordering is also highly related to post-editing technologies, which aim to correct errors in a rule-based translation (Simard et al., 2007; Dugast et al., 2007; Ehara, 2007) or a different type of SMT (Aikawa and Ruopp, 2009). There is a major difference of the post-ordering from such an post-editing framework; in the post-editing framework, the preceding translation process is a complete source-to-target translation, and the post-editing itself works as an additional process to fix errors. In contrast, the post-ordering framework divides the whole translation process into two sub-processes focusing on translation and reordering. It has an advantage that the sub-processes are simplified and easy to solve compared to a complete translation process in the post-editing.

3 Proposed method

This section presents the proposed post-ordering approach in the case of Japanese-to-English translation.

3.1 Basic framework

The post-ordering translation consists of the following two translation processes.

1) Translating input Japanese sentences into HFE sentence using the Japanese-to-HFE translation models.

2) Translating HFE sentences into English sentences using the HFE-to-English translation models.

Translation models used in these processes are trained using English and Japanese bitext and corresponding HFE sentences. The HFE sentences are generated by applying the Head-Finalization pre-ordering rules (Isozaki et al., 2010) to the English part of the bitext. This framework is illustrated in Figure 2. This is a inverse problem of English-to-Japanese translation with English pre-ordering; The HFE-to-English translation is a way to solve the inverse problem of pre-ordering by the SMT framework.

An important aspect of the post-ordering translation is that the divided two translation processes are simplified both in training and decoding. The first English-to-HFE translation only concentrate on phrase translation, so the training and decoding can be (almost) free from complex reordering problems. The second HFE-to-English translation only have to reorder English word in a appropriate ordering. It is also worth noting that the second translation model can be easily trained using word alignments between HFE and English, which are quite obvious from the reordering process.

3.2 Reordering rules for English

Here we briefly review the Head-Finalization rules for generating HFE sentences as the intermediate language in the proposed method. Figure 3 shows an example parse tree for a English sentence "The oil pressure control unit 30 operates based on a command from the CVT controller 20." In Figure 3, the nodes with bold eclipse represent syntactic heads for each tree node. The sentence is reordered by the following Head Finalization rules, as shown in the bottom part in Figure 3.

- Move each syntactic head toward the end of its siblings (except for coordination).
- Rewrite plural nouns (POS:`NNS`) with singular ones.
- Eliminate determiners "a", "an", and "the".
- Insert *pseudo*-particles for verb arguments[1]: `_va0` (arg1 of the sentence head verb), `_va1` (arg1 of other verbs), `_va2` (arg2 of verbs)

The rules depend on the corresponding Japanese characteristics: heads are usually located in the end of phrases; plural forms and determiners are not used; several kinds of particles are used as case markers. The HFE sentence has almost monotone word alignments with the corresponding Japanese sentence as shown in Figure 4. Due to the modification on the English words, the HFE-to-English translation process is not a simple reordering problem, but we believe that it can easily be solved by the SMT framework.

3.3 Two-stage translation

Now we can set up two groups of translation models (phrase tables, language models, and so on) for Japanese-to-HFE and HFE-to-English translation. The training itself can be done by a standard SMT manner using the Japanese-HFE and HFE-English bitexts.

Figure 5 shows an example of the post-ordering translation. In the first stage, the Japanese sentence is translated into HFE, with a small-distance or even no reordering in decoding. The resulting HFE sentence is head-final; the verb phrase "is composed of" is reversed and located in the end of the sentence. It also does not have determiners "a" and "the", but have a special word `_va0`. In the second stage, the intermediate HFE sentence is translated into English, with a few edits on the words themselves but with a long-distance reordering. The final English translation result seems quite good.

[1] arg1 and arg2 are swapped for passive verbs.

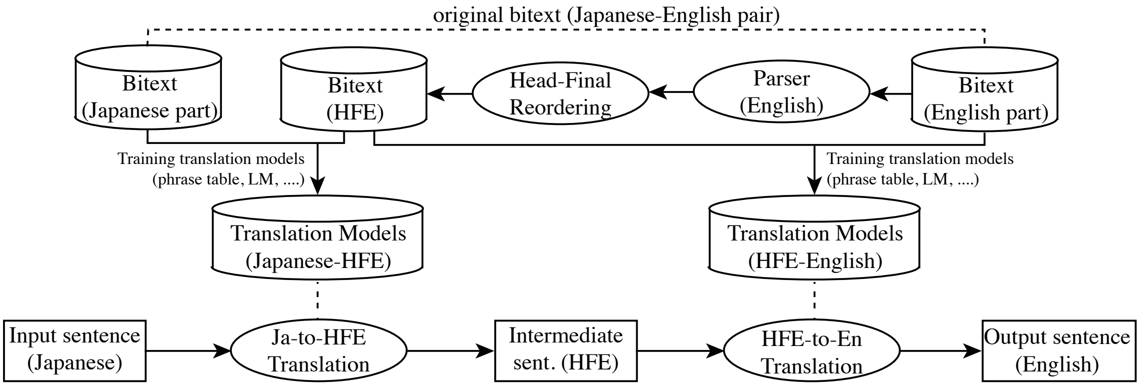

Figure 2: Post-ordering translation framework in Japanese-to-English translation.

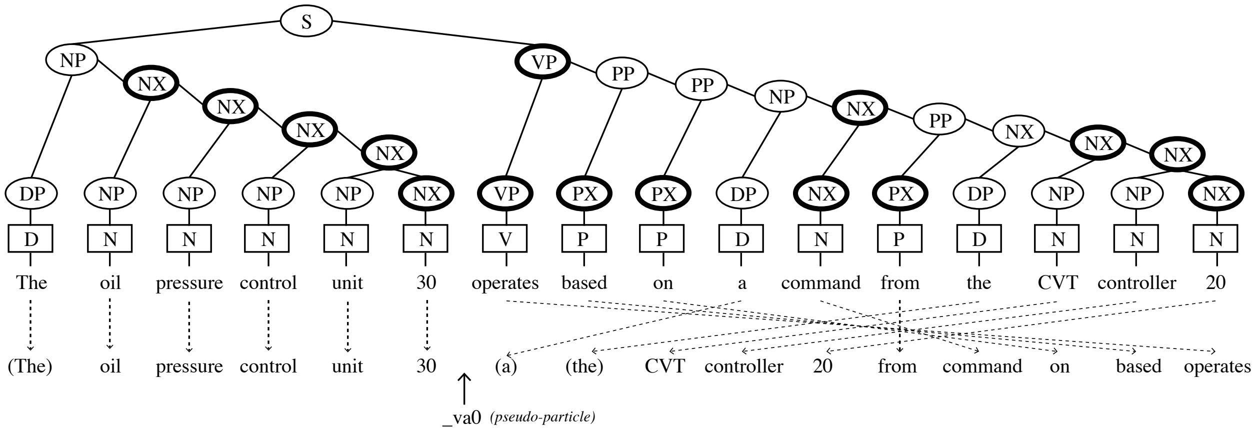

Figure 3: An example parse tree and the Head-Final reordering. Nodes with bold eclipse represent syntactic heads for each tree node. The determiners "the" and "a" are eliminated by the rules, and a *pseudo*-particle "_va0" is inserted after the subject.

4 Experiment

We investigated the advantage of our post-ordering method by the following Japanese-to-English translation experiment with the post-ordering and baseline SMTs.

4.1 Setup

We used NTCIR-9 PatentMT (NTCIR-9, 2011) English and Japanese dataset for this experiment. Some statistics of this dataset are shown in Table 1. We preprocessed the dataset by the following softwares:

- English syntactic (HPSG) parser: Enju[2](Miyao and Tsujii, 2008)
- English tokenizer: stepp (included in Enju package)
- Japanese tokenizer: Mecab[3] (with ipadic-2.7.0)

Word alignment was automatically estimated using MGIZA++[4] using bitexts of 64 or less words in the training set to avoid a problematic underflow. Language models are word 5-gram models of English and HFE, trained with SRILM[5].

4.2 Compared methods

We compared the proposed post-ordering with three baseline SMTs: a standard phrase-based SMT (PBMT) with lexicalized reordering, a hierarchical phrase-based SMT (HPBMT), and a string-to-tree syntax-based SMT (SBMT), included in Moses[6].

[2] http://www-tsujii.is.s.u-tokyo.ac.jp/enju/index.html

[3] http://mecab.sourceforge.net/

[4] http://sourceforge.net/projects/mgizapp/

[5] http://www-speech.sri.com/projects/srilm/

[6] http://sourceforge.net/projects/mosesdecoder/

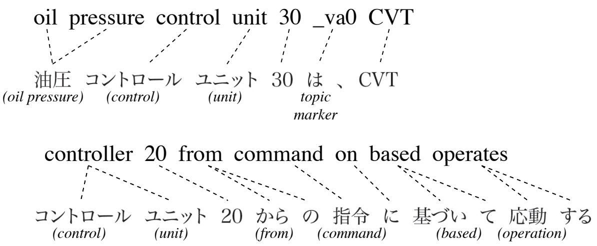

Figure 4: Word alignments between HFE and Japanese.

[Input Japanese]
第 1 f θ レンズ 17 は 、 第 1 レンズ 22 と
第 2 レンズ 23 と から 構成 さ れる 。

Japanese-to-HFE ↓ *(monotone or small distortion)*

[Intermediate Head-Final English]
first f θ lens 17 _va0 first lens 22 and
second lens 23 of composed is .

HFE-to-English ↓ *(a little edit on words, large distorion)*

[Output English]
the first f θ lens 17 is composed of
a first lens 22 and the second lens 23 .

[Reference]
the first f θ lens 17 is constituted of
a first lens 22 and a second lens 23 .

Figure 5: An example of the post-ordering translation.

English parse trees used in SBMT were identical to the ones used for generating HFE sentences in the post-ordering. The post-ordering used two Moses phrase-based decoders, one for Japanese-to-HFE and the other for HFE-to-English.

The models for these decoders were trained in the standard manner with Moses, `grow-diag-final-and` heuristics for symmetric word alignment, `msd-bidirectional-fe` lexicalized reordering (in PBMT and the post-ordering). The parameter values are optimized by minimum error rate training (MERT) (Och, 2003) with `mert-moses.pl`. One difference among configurations of the decoders was distortion limit. The Japanese-to-HFE decoder did not require long distance reordering, so we compared two conditions with the values of 0 (monotone) and 6. The HFE-to-English and PBMT decoders had to drastically reorder phrases so we used the values of 12 and 20. In the HPBMT and SBMT decoders, we used 15 for its max-chart-span option. The other decoder configurations (e.g., beam widths) were the default values of Moses.

Table 1: Data statistics. HFE stands for Head-Final English (Isozaki et al., 2010).

	Training	Dev.	Test
#sentence	3,189,025	1,000	1,000
#word (Ja)	115,877,483	37,066	35,921
#word (En)	105,966,236	33,096	33,376
#word (HFE)	100,169,813	31,228	31,331

4.3 Results

Table 2 shows the results in BLEU (Papineni et al., 2002) in case-insensitive evaluation and average decoding times[7] (on a Xeon 7460 2.66GHz computer) with the compared methods. The proposed post-ordering translation (with monotone Japanese-to-HFE translation) achieved 0.2963 in BLEU, better than the best HPBMT baseline (0.2887) by 0.76 points and the standard PBMT baseline (0.2806) by 1.57 points. The differences were statistically significant according to the bootstrap sampling test ($p < 0.05$ with HPBMT and $p < 0.01$ with PBMT, 1,000 samples) (Zhang et al., 2004), and it was consistent among all post-ordering conditions.

In the Japanese-to-HFE translation, the monotone configuration was slightly better than the reordering with the distortion limit of 6 but the difference was not significant. In the HFE-to-English translation, the difference in the distortion limit did not affect the final results. Among the baseline methods, HPBMT was better than other baselines by 0.5 points.

4.4 Discussion

The proposed post-ordering method was consistently better than the baseline methods in the experiment. To investigate the results in detail, we analyzed the Japanese-to-HFE and HFE-to-English translation individually. Tables 3 and 4 show the individual evaluation results of Japanese-to-HFE and HFE-to-English, respectively.

The Japanese-to-HFE results was 0.3195 in monotone translation. This was much better than the baseline Japanese-to-English ones by 5 points in

[7]Decoding times were obtained from Moses verbose outputs.

Table 2: Evaluation results in BLEU (case insensitive) and average decoding time. Reordering Limit stands for distortion limit in PBMT and Post-ordering and maximum chart span in HPBMT and SBMT.

Methods	Reordering Limit		BLEU	Decoding Time (sec. / sentence)		
	Ja-to-HFE	HFE-to-En		Ja-to-HFE	HFE-to-En	Total
Baseline (PBMT)	12		0.2806	3.532		3.532
	20		0.2780	7.675		7.675
Baseline (HPBMT)	15		0.2887	7.693		7.693
Baseline (SBMT)	15		0.2686	12.975		12.975
Post-ordering	0	12	0.2960	1.809	2.487	4.296
		20	**0.2963**		3.653	5.462
	6	12	0.2944	3.542	2.020	5.562
		20	0.2941		2.906	6.448

BLEU. Although the scores should not be compared directly, this implies the Japanese-to-HFE translation is a much easier problem than Japanese-to-English translation. We argue this is mainly because small distance or even no reordering was required in Japanese-to-HFE translation. Such good intermediate HFE results may be spoiled by the second HFE-to-English translation, but that was not the case in this experiment; BLEU decreased only by about 2 points in the final results shown in Table 2. The HFE-to-English translation itself achieved 0.6305 in BLEU in translating oracle HFE sentences (Head-Finalized reference English) with the distortion limit of 20.

Another important viewpoint is its decoding time. The post-ordering approach consists of two SMT processes and may take more decoding time than standard methods. An interesting finding in the experiments was the fastest post-ordering was not so slow as the fastest PBMT baseline (22% slower) or even a bit faster than the best HPBMT baseline (44% faster). The monotone Japanese-to-HFE translation was two times faster than the fastest PBMT baseline with the distortion limit of 12, and the HFE-to-English translation was also fast, compared to the baselines with the same distortion limits. This also suggests that the post-ordering can efficiently works by the two simplified SMT processes.

Table 3: Individual evaluation results of Japanese-to-HFE translation.

	Distortion Limit	BLEU
Ja-to-HFE	0	0.3195
	6	0.3186

Table 4: Individual evaluation results of HFE(oracle)-to-English translation. Note that HFE(oracle) sentences are from the reference English sentences.

	Distortion Limit	BLEU
HFE(oracle)-to-En	12	0.6191
	20	0.6305

5 Concluding remarks

This paper presents a novel post-ordering approach for translation between languages with distant word ordering. The proposed post-ordering translation is the inverse problem of the pre-ordering translation and can be solved by the two simplified translation processes: source-to-“source-ordered target” and “source-ordered target”-to-target translations. In Japanese-to-English translation, we first translate Japanese into HFE with no or small-distance reordering, then we translate HFE into English with long-distance reordering and a small number of edits on English words. The proposed post-ordering achieved significantly better translation performance in BLEU in the experiment compared to standard phrase-based, hierarchical phrase-based, and syntax-based baseline SMTs. We conclude the post-ordering is a promising way of translation in some language pairs where good pre-ordering methods have been developed in the opposite direction.

The current post-ordering implementation is an string-to-string approximation of string-to-tree translation (the inverse problem of tree-to-string pre-ordering). It may be beneficial to extend it by tree-based SMT for better post-reordering. Another interesting challenge in the post-ordering is the use of large-scale monolingual resource for training the post-ordering models. We also note that the post-ordering method has wide applicabliilty and it would be interesting to also examine other language pairs. Finally, it is worth comparing pre- and post-ordering approach on various language pairs to analyze their advantages in detail, because difficulty of pre-ordering may vary with language directions.

References

Takako Aikawa and Achim Ruopp. 2009. Chained system: A linear combination of different types of statistical machine translation systems. In *Proc. MT Summit XII*.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Michael Collins, Philipp Koehn, and Ivona Kuĉerová. 2005. Clause restructuring for statistical machine translation. In *Proc. ACL*, pages 531–540.

Marta R. Costa-jussà and José A. R. Fonollosa. 2006. Statistical machine reordering. In *Proc. EMNLP*, pages 70–76.

Loic Dugast, Jean Senellart, and Philipp Koehn. 2007. Statistical post-editing on SYSTRAN ' s rule-based translation system. In *Proc. WMT*.

Chris Dyer and Philip Resnik. 2010. Context-free reordering, finite-state translation. In *Proc. HLT-NAACL*, pages 858–866.

Terumasa Ehara. 2007. Rule based machine translation combined with statistical post editor for japanese to english patent translation. In *Proc. MT Summit XI Workshop on Patent Translation*.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In *Proc. NAACL*, pages 273–280.

Dmitriy Genzel. 2010. Automatically learning source-side reordering rules for large scale machine translation. In *Proc. COLING 2010*, pages 376–384, Beijing, China.

Jonathan Graehl and Kevin Knight. 2004. Training tree transducers. In *Proc. HLT-NAACL*, pages 105–112.

Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010. A simple reordering rule for sov languages. In *Proc. WMT-MetricsMATR*.

Phillip Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proc. HLT-NAACL*, pages 263–270.

Chi-Ho Li, Dongdong Zhang, Mu Li, Ming Zhou, Minghui Li, and Yi Guan. 2007. A probabilistic approach to syntax-based reordering for statistical machine translation. In *Proc. ACL*, pages 720–727.

Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature forest models for probabilistic hpsg parsing. *Computational Linguistics*, 34(1):35–80.

NTCIR-9. 2011. Patent Translation Task at NTCIR-9 (PatentMT). http://ntcir.nii.ac.jp/PatentMT/.

Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proc. ACL*, pages 160–167.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proc. ACL*, pages 311–318.

Michel Simard, Cyril Goutte, and Pierre Isabelle. 2007. Statistical phrase-based post-editing. In *Proc. NAACL-HLT*.

Christoph Tillmann. 2004. A unigram orientation model for statistical machine translation. In *Proc. HLT-NAACL*, pages 101–104.

Roy Tromble and Jason Eisner. 2009. Learning linear ordering problems for better translation. In *Proc. EMNLP*, pages 1007–1016.

Chao Wang, Michael Collins, , and Phillip Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In *Proc. of EMNLP*, pages 737–745.

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics*, 23(3):377–404.

Fei Xia and Michael McCord. 2004. Improving a statistical mt system with automatically learned rewrite patterns. In *Proc. COLING*, pages 508–514.

Deyi Xiong, Qun Liu, , and Shouxun Lin. 2006. Maximum entropy based phrase reordering model for statistical machine translation. In *Proc. COLING-ACL*, pages 521–528.

Deyi Xiong, Min Zhang, Aiti Aw, Haitao Mi, Qun Liu, and Shouxun Lin. 2008. Refinements in BTG-based statistical machine translation. In *Proc. IJCNLP*, pages 505–512.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proc. ACL*, pages 523–530.

Ying Zhang, Stephan Vogel, and Alex Weibel. 2004. Interpreting BLEU/NIST scores: How much improvement do we need to have a better system? In *Proc. LREC*, pages 2051–2054.