Post-ordering by Parsing for Japanese-English Statistical Machine Translation

Isao Goto   **Masao Utiyama**   **Eiichiro Sumita**

Multilingual Translation Laboratory, MASTAR Project

National Institute of Information and Communications Technology

3-5 Hikaridai, Keihanna Science City, Kyoto, 619-0289, Japan

`{igoto, mutiyama, eiichiro.sumita}@nict.go.jp`

Abstract

Reordering is a difficult task in translating between widely different languages such as Japanese and English. We employ the post-ordering framework proposed by (Sudoh et al., 2011b) for Japanese to English translation and improve upon the reordering method. The existing post-ordering method reorders a sequence of target language words in a source language word order via SMT, while our method reorders the sequence by: 1) parsing the sequence to obtain syntax structures similar to a source language structure, and 2) transferring the obtained syntax structures into the syntax structures of the target language.

1 Introduction

The word reordering problem is a challenging one when translating between languages with widely different word orders such as Japanese and English. Many reordering methods have been proposed in statistical machine translation (SMT) research. Those methods can be classified into the following three types:

Type-1: Conducting the target word selection and reordering jointly. These include phrase-based SMT (Koehn et al., 2003), hierarchical phrase-based SMT (Chiang, 2007), and syntax-based SMT (Galley et al., 2004; Ding and Palmer, 2005; Liu et al., 2006; Liu et al., 2009).

Type-2: Pre-ordering (Xia and McCord, 2004; Collins et al., 2005; Tromble and Eisner, 2009; Ge, 2010; Isozaki et al., 2010b; DeNero and Uszkoreit, 2011; Wu et al., 2011). First, these methods reorder the source language sentence into the target language word order. Then, they translate the reordered source word sequence using SMT methods.

Type-3: Post-ordering (Sudoh et al., 2011b; Matusov et al., 2005). First, these methods translate the source sentence almost monotonously into a sequence of the target language words. Then, they reorder the translated word sequence into the target language word order.

This paper employs the post-ordering framework for Japanese-English translation based on the discussions given in Section 2, and improves upon the reordering method. Our method uses syntactic structures, which are essential for improving the target word order in translating long sentences between Japanese (a Subject-Object-Verb (SOV) language) and English (an SVO language).

Before explaining our method, we explain the pre-ordering method for English to Japanese used in the post-ordering framework.

In English-Japanese translation, Isozaki et al. (2010b) proposed a simple pre-ordering method that achieved the best quality in human evaluations, which were conducted for the NTCIR-9 patent machine translation task (Sudoh et al., 2011a; Goto et al., 2011). The method, which is called *head finalization*, simply moves syntactic heads to the end of corresponding syntactic constituents (e.g., phrases and clauses). This method first changes the English word order into a word order similar to Japanese word order using the head finalization rule. Then, it translates (almost monotonously) the pre-ordered

311

Figure 1: Post-ordering framework.

Figure 2: Example of post-ordering by parsing.

English words into Japanese.

There are two key reasons why this pre-ordering method works for estimating Japanese word order. The first reason is that Japanese is a typical head-final language. That is, a syntactic head word comes after nonhead (dependent) words. Second, input English sentences are parsed by a high-quality parser, Enju (Miyao and Tsujii, 2008), which outputs syntactic heads. Consequently, the parsed English input sentences can be pre-ordered into a Japanese-like word order using the head finalization rule.

Pre-ordering using the head finalization rule naturally cannot be applied to Japanese-English translation, because English is not a head-final language. If we want to pre-order Japanese sentences into an English-like word order, we therefore have to build complex rules (Sudoh et al., 2011b).

2 Post-ordering for Japanese to English

Sudoh et al. (2011b) proposed a post-ordering method for Japanese-English translation. The translation flow for the post-ordering method is shown in Figure 1, where "HFE" is an abbreviation of "Head Final English". An HFE sentence consists of English words in a Japanese-like structure. It can be constructed by applying the head-finalization rule (Isozaki et al., 2010b) to an English sentence parsed by Enju. Therefore, if good rules are applied to this HFE sentence, the underlying English sentence can be recovered. This is the key observation of the post-ordering method.

The process of post-ordering translation consists of two steps. First, the Japanese input sentence is translated into HFE almost monotonously. Then, the word order of HFE is changed into an English word order.

Training for the post-ordering method is conducted by first converting the English sentences in a Japanese-English parallel corpus into HFE sentences using the head-finalization rule. Next, a monotone phrase-based Japanese-HFE SMT model is built using the Japanese-HFE parallel corpus
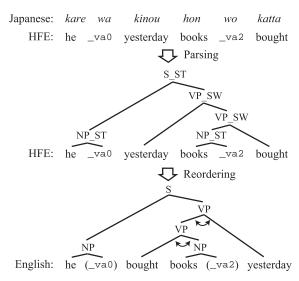
whose HFE was converted from English. Finally, an HFE-to-English word reordering model is built using the HFE-English parallel corpus.

3 Post-ordering Models

3.1 SMT Model

Sudoh et al. (2011b) have proposed using phrase-based SMT for converting HFE sentences into English sentences. The advantage of their method is that they can use off-the-shelf SMT techniques for post-ordering.

3.2 Parsing Model

Our proposed model is called the *parsing model*. The translation process for the parsing model is shown in Figure 2. In this method, we first parse the HFE sentence into a binary tree. We then swap the nodes annotated with "_SW" suffixes in this binary tree in order to produce an English sentence.

The structures of the HFE sentences, which are used for training our parsing model, can be obtained from the corresponding English sentences as follows.[1] First, each English sentence in the training Japanese-English parallel corpus is parsed into a binary tree by applying Enju. Then, for each node in this English binary tree, the two children of each node are swapped if its first child is the head node (See (Isozaki et al., 2010b) for details of the head

---

[1] The explanations of pseudo-particles (_va0 and _va2) and other details of the HFE is given in Section 4.2.

312

final rules). At the same time, these swapped nodes are annotated with "_SW". When the two nodes are not swapped, they are annotated with "_ST" (indicating "Straight"). A node with only one child is not annotated with either "_ST" or "_SW". The result is an HFE sentence in a binary tree annotated with "_SW" and "_ST" suffixes.

Observe that the HFE sentences can be regarded as binary trees annotated with syntax tags augmented with swap/straight suffixes. Therefore, the structures of these binary trees can be learnable by using an off-the-shelf grammar learning algorithm. The learned parsing model can be regarded as an ITG model (Wu, 1997) between the HFE and English sentences. [2]

In this paper, we used the Berkeley Parser (Petrov and Klein, 2007) for learning these structures. The HFE sentences can be parsed by using the learned parsing model. Then the parsed structures can be converted into their corresponding English structures by swapping the "_SW" nodes. Note that this parsing model jointly learns how to parse and swap the HFE sentences.

4 Detailed Explanation of Our Method

This section explains the proposed method, which is based on the post-ordering framework using the parsing model.

4.1 Translation Method

First, we produce N-best HFE sentences using Japanese-to-HFE monotone phrase-based SMT. Next, we produce K-best parse trees for each HFE sentence by parsing, and produce English sentences by swapping any nodes annotated with "_SW". Then we score the English sentences and select the English sentence with the highest score.

For the score of an English sentence, we use the sum of the log-linear SMT model score for Japanese-to-HFE and the logarithm of the language model probability of the English sentence.

---

[2]There are works using the ITG model in SMT: ITG was used for training pre-ordering models (DeNero and Uszkoreit, 2011); hierarchical phrase-based SMT (Chiang, 2007), which is an extension of ITG; and reordering models using ITG (Chen et al., 2009; He et al., 2010). These methods are not post-ordering methods.

4.2 HFE and Articles

This section describes the details of HFE sentences. In HFE sentences: 1) Heads are final except for coordination. 2) Pseudo-particles are inserted after verb arguments: _va0 (subject of sentence head), _va1 (subject of verb), and _va2 (object of verb). 3) Articles (a, an, the) are dropped.

In our method of HFE construction, unlike that used by (Sudoh et al., 2011b), plural nouns are left as-is instead of converted to the singular.

Applying our parsing model to an HFE sentence produces an English sentence that does not have articles, but does have pseudo-particles. We removed the pseudo-particles from the reordered sentences before calculating the probabilities used for the scores of the reordered sentences. A reordered sentence without pseudo-particles is represented by E. A language model $P(E)$ was trained from English sentences whose articles were dropped.

In order to output a genuine English sentence E' from E, articles must be inserted into E. A language model trained using genuine English sentences is used for this purpose. We try to insert one of the articles {a, an, the} or no article for each word in E. Then we calculate the maximum probability word sequence through dynamic programming for obtaining E'.

5 Experiment

5.1 Setup

We used patent sentence data for the Japanese to English translation subtask from the NTCIR-9 and 8 (Goto et al., 2011; Fujii et al., 2010). There were 2,000 test sentences for NTCIR-9 and 1,251 for NTCIR-8. XML entities included in the data were decoded to UTF-8 characters before use.

We used Enju (Miyao and Tsujii, 2008) v2.4.2 for parsing the English side of the training data. Mecab [3] v0.98 was used for the Japanese morphological analysis. The translation model was trained using sentences of 64 words or less from the training corpus as (Sudoh et al., 2011b). We used 5-gram language models using SRILM (Stolcke et al., 2011).

We used the Berkeley parser (Petrov and Klein, 2007) to train the parsing model for HFE and to

---

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

parse HFE. The parsing model was trained using 0.5 million sentences randomly selected from training sentences of 40 words or less. We used the phrase-based SMT system Moses (Koehn et al., 2007) to calculate the SMT score and to produce HFE sentences. The distortion limit was set to 0. We used 10-best Moses outputs and 10-best parsing results of Berkeley parser.

5.2 Compared Methods

We used the following 5 comparison methods: Phrase-based SMT (PBMT), Hierarchical phrase-based SMT (HPBMT), String-to-tree syntax-based SMT (SBMT), Post-ordering based on phrase-based SMT (PO-PBMT) (Sudoh et al., 2011b), and Post-ordering based on hierarchical phrase-based SMT (PO-HPBMT).

We used Moses for these 5 systems. For PO-PBMT, a distortion limit 0 was used for the Japanese-to-HFE translation and a distortion limit 20 was used for the HFE-to-English translation. The PO-HPBMT method changes the post-ordering method of PO-PBMT from a phrase-based SMT to a hierarchical phrase-based SMT. We used a max-chart-span 15 for the hierarchical phrase-based SMT. We used distortion limits of 12 or 20 for PBMT and a max-chart-span 15 for HPBMT.

The parameters for SMT were tuned by MERT using the first half of the development data with HFE converted from English.

5.3 Results and Discussion

We evaluated translation quality based on the case-insensitive automatic evaluation scores of RIBES v1.1 (Isozaki et al., 2010a) and BLEU-4. The results are shown in Table 1.

Ja-to-En	NTCIR-9		NTCIR-8	
	RIBES	BLEU	RIBES	BLEU
Proposed	**72.57**	**31.75**	**73.48**	**32.80**
PBMT (limit 12)	68.44	29.64	69.18	30.72
PBMT (limit 20)	68.86	30.13	69.63	31.22
HPBMT	69.92	30.15	70.18	30.94
SBMT	69.22	29.53	69.87	30.37
PO-PBMT	68.81	30.39	69.80	31.71
PO-HPBMT	70.47	27.49	71.34	28.78

Table 1: Evaluation results (case insensitive).

From the results, the proposed method achieved the best scores for both RIBES and BLEU for NTCIR-9 and NTCIR-8 test data. Since RIBES is sensitive to global word order and BLEU is sensitive to local word order, the effectiveness of the proposed method for both global and local reordering can be demonstrated through these comparisons.

In order to investigate the effects of our post-ordering method in detail, we conducted an "HFE-to-English reordering" experiment, which shows the main contribution of our post-ordering method in the framework of post-ordering SMT as compared with (Sudoh et al., 2011b). In this experiment, we changed the word order of the oracle-HFE sentences made from reference sentences into English, this is the same way as Table 4 in (Sudoh et al., 2011b). The results are shown in Table 2.

This results show that our post-ordering method is more effective than PO-PBMT and PO-HPBMT. Since RIBES is based on the rank order correlation coefficient, these results show that the proposed method correctly recovered the word order of the English sentences. These high scores also indicate that the parsing results for high quality HFE are fairly trustworthy.

oracle-HFE-to-En	NTCIR-9		NTCIR-8	
	RIBES	BLEU	RIBES	BLEU
Proposed	**94.66**	**80.02**	**94.93**	**79.99**
PO-PBMT	77.34	62.24	78.14	63.14
PO-HPBMT	77.99	53.62	80.85	58.34

Table 2: Evaluation resutls focusing on post-ordering.

In these experiments, we did not compare our method to pre-ordering methods. However, some groups used pre-ordering methods in the NTCIR-9 Japanese to English translation subtask. The NTT-UT (Sudoh et al., 2011a) and NAIST (Kondo et al., 2011) groups used pre-ordering methods, but could not produce RIBES and BLEU scores that both were better than those of the baseline results. In contrast, our method was able to do so.

6 Conclusion

This paper has described a new post-ordering method. The proposed method parses sentences that consist of target language words in a source language word order, and does reordering by transferring the syntactic structures similar to the source language syntactic structures into the target language syntactic structures.

References

Han-Bin Chen, Jian-Cheng Wu, and Jason S. Chang. 2009. Learning Bilingual Linguistic Reordering Model for Statistical Machine Translation. In *Proceedings of Human Language Technologies: The 2009 NAACL*, pages 254–262, Boulder, Colorado, June. Association for Computational Linguistics.

David Chiang. 2007. Hierarchical Phrase-Based Translation. *Computational Linguistics*, 33(2):201–228.

Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause Restructuring for Statistical Machine Translation. In *Proceedings of the 43rd ACL*, pages 531–540, Ann Arbor, Michigan, June. Association for Computational Linguistics.

John DeNero and Jakob Uszkoreit. 2011. Inducing Sentence Structure from Parallel Corpora for Reordering. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 193–203, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Yuan Ding and Martha Palmer. 2005. Machine Translation Using Probabilistic Synchronous Dependency Insertion Grammars. In *Proceedings of the 43rd ACL*, pages 541–548, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro, Terumasa Ehara, Hiroshi Echizen-ya, and Sayori Shimohata. 2010. Overview of the Patent Translation Task at the NTCIR-8 Workshop. In *Proceedings of NTCIR-8*, pages 371–376.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In Daniel Marcu Susan Dumais and Salim Roukos, editors, *HLT-NAACL 2004: Main Proceedings*, pages 273–280, Boston, Massachusetts, USA, May 2 - May 7. Association for Computational Linguistics.

Niyu Ge. 2010. A Direct Syntax-Driven Reordering Model for Phrase-Based Machine Translation. In *Proceedings of NAACL-HLT*, pages 849–857, Los Angeles, California, June. Association for Computational Linguistics.

Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita, and Benjamin K. Tsou. 2011. Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop. In *Proceedings of NTCIR-9*, pages 559–578.

Yanqing He, Yu Zhou, Chengqing Zong, and Huilin Wang. 2010. A Novel Reordering Model Based on Multi-layer Phrase for Statistical Machine Translation. In *Proceedings of the 23rd Coling*, pages 447–455, Beijing, China, August. Coling 2010 Organizing Committee.

Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010a. Automatic Evaluation of Translation Quality for Distant Language Pairs. In *Proceedings of the 2010 EMNLP*, pages 944–952.

Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010b. Head Finalization: A Simple Reordering Rule for SOV Languages. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 244–251, Uppsala, Sweden, July. Association for Computational Linguistics.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of the 2003 HLT-NAACL*, pages 48–54.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th ACL*, pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

Shuhei Kondo, Mamoru Komachi, Yuji Matsumoto, Katsuhito Sudoh, Kevin Duh, and Hajime Tsukada. 2011. Learning of Linear Ordering Problems and its Application to J-E Patent Translation in NTCIR-9 PatentMT. In *Proceedings of NTCIR-9*, pages 641–645.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. In *Proceedings of the 21st ACL*, pages 609–616, Sydney, Australia, July. Association for Computational Linguistics.

Yang Liu, Yajuan Lü, and Qun Liu. 2009. Improving Tree-to-Tree Translation with Packed Forests. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 558–566, Suntec, Singapore, August. Association for Computational Linguistics.

E. Matusov, S. Kanthak, and Hermann Ney. 2005. On the Integration of Speech Recognition and Statistical Machine Translation. In *Proceedings of Interspeech*, pages 3177–3180.

Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature Forest Models for Probabilistic HPSG Parsing. In *Computational Linguistics, Volume 34, Number 1*, pages 81–88.

Slav Petrov and Dan Klein. 2007. Improved Inference for Unlexicalized Parsing. In *NAACL-HLT*, pages 404–411, Rochester, New York, April. Association for Computational Linguistics.

Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at Sixteen: Update and Outlook. In *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop*.

Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, Masaaki Nagata, Xianchao Wu, Takuya Matsuzaki, and Jun'ichi Tsujii. 2011a. NTT-UT Statistical Machine Translation in NTCIR-9 PatentMT. In *Proceedings of NTCIR-9*, pages 585–592.

Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. 2011b. Post-ordering in Statistical Machine Translation. In *Proceedings of the 13th Machine Translation Summit*, pages 316–323.

Roy Tromble and Jason Eisner. 2009. Learning Linear Ordering Problems for Better Translation. In *Proceedings of the 2009 EMNLP*, pages 1007–1016, Singapore, August. Association for Computational Linguistics.

Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. 2011. Extracting Pre-ordering Rules from Chunk-based Dependency Trees for Japanese-to-English Translation. In *Proceedings of the 13th Machine Translation Summit*, pages 300–307.

Dekai Wu. 1997. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377–403.

Fei Xia and Michael McCord. 2004. Improving a Statistical MT System with Automatically Learned Rewrite Patterns. In *Proceedings of Coling*, pages 508–514, Geneva, Switzerland, Aug 23–Aug 27. COLING.