Global Pre-ordering for Improving Sublanguage Translation

Masaru Fuji[†‡] **Masao Utiyama**[†] **Eiichiro Sumita**[†] **Yuji Matsumoto**[‡]
† National Institute of Information and Communication Technology
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan
`{fuji.masaru,mutiyama,eiichiro.sumita}@nict.go.jp`

‡ Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, Japan
`{fuji.masaru.fe1,matsu}@is.naist.jp`

Abstract

When translating formal documents, capturing the sentence structure specific to the sublanguage is extremely necessary to obtain high-quality translations. This paper proposes a novel global reordering method with particular focus on long-distance reordering for capturing the global sentence structure of a sublanguage. The proposed method learns global reordering models from a non-annotated parallel corpus and works in conjunction with conventional syntactic reordering. Experimental results on the patent abstract sublanguage show substantial gains of more than 25 points in the RIBES metric and comparable BLEU scores both for Japanese-to-English and English-to-Japanese translations.

1 Introduction

Formal documents such as legal and technical documents often form sublanguages. Previous studies have highlighted that capturing the sentence structure specific to the sublanguage is extremely necessary for obtaining high-quality translations especially between distant languages (Buchmann et al., 1984; Luckhardt, 1991; Marcu et al., 2000). Figure 1 illustrates two pairs of bilingual sentences specific to the sublanguage of patent abstracts. In both sentence pairs, the global sentence structure ABC in the source sentences must be reordered to CBA in the target sentences to produce a structurally appropriate translation. Each of the components ABC must then be syntactically reordered to complete the reordering.

Various attempts have been made along this line of research. One such method is the skeleton-based statistical machine translation (SMT) which uses a syntactic parser to extract the global sentence structure, or the *skeleton*, from syntactic trees and uses conventional SMT to train global reordering (Mellebeek et al., 2006; Xiao et al., 2014). However, the performance of this method is limited by syntactic parsing, therefore the global reordering has low accuracy where the accuracy of syntactic parsing is low. Another approach involves manually preparing synchronous context-free grammar rules for capturing the global sentence structure of the target sublanguage (Fuji et al., 2015). However, this method requires manual preparation of rules. Both methods are unsuitable for formal documents such as patent abstracts, because they fail to adapt to sentences with various expressions, for which manual preparation of rules is complex.

This paper describes a novel global reordering method for capturing sublanguage-specific global sentence structure to supplement the performance of conventional syntactic reordering. The method learns a global pre-ordering model from non-annotated corpora without using syntactic parsing and uses this model to perform global pre-ordering on newly inputted sentences. As the global pre-ordering method does not rely on syntactic parsing, it is not affected by the degradation of parsing accuracy, and is readily applicable to new sublanguages. Globally pre-ordered sentence segments are then syntactically reordered before being translated by SMT.

In this empirical study on the patent abstract sublanguage in Japanese-to-English and English-to-Japanese translations, the translation quality of the sublanguage was improved when global pre-ordering

Pair 1	Japanese	[[_A_ アンテナ資源を有効に活用して信頼性の高い通信を行うことができる][_B_ 通信装置を][_C_ 提供すること。]]
	Japanese (word-for-word translation)	[[_A_ Antenna resources effectively utilizing reliability high communication perform capable][_B_ communication apparatus][_C_ to provide.]]
	English	[[_C_ To provide][_B_ a communication apparatus][_A_ capable of performing highly reliable communication by effectively utilizing antenna resources.]]

Pair 2	Japanese	[[_A_ 高画質な画像を形成できる][_B_ 画像形成装置を][_C_ 提供する。]]
	Japanese (word-for-word translation)	[[_A_ High quality images form enable][_B_ image formation device][_C_ to provide.]]
	English	[[_C_ To provide][_B_ an image formation device][_A_ which enables high quality images to be formed.]]

Figure 1: Example of sublanguage-specific bilingual sentences requiring global reordering. A, B, C are the sentence segments constituting global sentence structures.

was combined with syntactic pre-ordering. A statistically significant improvement was observed against the syntactic pre-ordering alone, and a substantial gain of more than 25 points in RIBES score against the baseline was observed for both Japanese-to-English and English-to-Japanese translations, and the BLEU scores remained comparable.

2 Related Work

The hierarchical phrase-based method (Chiang, 2005) is one of the early attempts at reordering for SMT. In this method, reordering rules are automatically extracted from non-annotated text corpora during the training phase, and the reordering rules are applied in decoding. As the method does not require syntactic parsing and learns from raw text corpora, it is highly portable. However, this method does not specifically capture global sentence structures.

The tree-to-string and string-to-tree SMTs are the methods which employ syntactic parsing, whenever it is available, either for the source or for the target language to improve the translation of the language pair (Yamada and Knight, 2001; Ambati and Chen, 2007). However, these methods too are not specifically designed for capturing global sentence structures.

The skeleton-based SMT is a method particularly focusing on the reordering of global sentence structure (Mellebeek et al., 2006; Xiao et al., 2014). It uses a syntactic parser to extract the global sentence structure, or the *skeleton*, from syntactic trees, and uses conventional SMT to train global reordering. Another related approach is the reordering method based on predicate-argument structure (Komachi et al., 2006). However, the performance of sentence structure extraction tends to be low when the accuracy of the syntactic parsing is low.

The syntactic pre-ordering is the state-of-the-art method which has substantially improved reordering accuracy, and hence the translation quality (Isozaki et al., 2010b; Goto et al., 2015; de Gispert et al., 2015; Hoshino et al., 2015). However, the adaptation of this method to a new domain requires manually parsed corpora for the target domains. In addition, the method does not have a specific function for capturing global sentence structure. Thus, we apply here our proposed global reordering model as a preprocessor to this syntactic reordering method to ensure the capturing of global sentence structures.

3 Global Pre-ordering Method

We propose a novel global reordering method for capturing sublanguage-specific global sentence structure. On the basis of the finding that sublanguage-specific global structures can be detected using relatively shallow analysis of sentences (Buchmann et al., 1984), we extract from the training set the n-grams frequently occurring in sentences involving global reordering and use these n-grams to detect the global structure of newly inputted sentences.

信号 を 発信できる で 画像 装置 を 提供 する

To
provide
an
image
device
capable
of
sending
signals
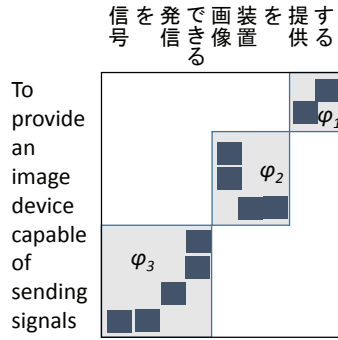
φ_1 φ_2 φ_3

Figure 2: An example of segments arranged in swap orientations for English to Japanese translation

For example, Figure 1 shows two sentence pairs in the training set that contain global reordering, where the segments ABC in the source sentence must be reordered globally to CBA in the target sentence to obtain structurally appropriate translations. With segment boundaries represented by the symbol "|", the extraction of unigrams on both sides of the two segment boundaries of sentence E1 of Figure 1 yields

$$\{provide, |, a\} \ \{apparatus, |, capable\}.$$

When we input the sentence *"To provide a heating apparatus capable of maintaining the temperature,"* this is matched against the above unigrams. Thus, the segment boundary positions are detected as *"To provide | a heating apparatus | capable of maintaining the temperature."* The detected segments are then reordered globally to yield the sentence *"Capable of maintaining the temperature | a heating apparatus | to provide,"* which has the appropriate global sentence structure for the target Japanese sentence. Each segment is then syntactically reordered before inputting to English-to-Japanese SMT.

The method consists of two steps. Step (i): we extract sentence pairs containing global reordering from the training corpus. We call this subset of the training corpus the *global reordering corpus*. Step (ii): we extract features from the source sentences of the global reordering corpus, and use these features to detect the segments of newly inputted sentences. We then reorder these detected segments globally. In step (ii), we experiment with a detection method based on heuristics, as well as a method based on machine learning. Steps (i) and (ii) are described in the following subsections.

3.1 Extraction of Sentence Pairs Containing Global Reordering

We extract sentences containing global reordering from the training corpus and store them in the global reordering corpus; they can subsequently be used for training and prediction. We consider that a sentence pair contains global reordering if the segments in the target sentence appear in swap orientation (Galley and Manning, 2008) to the source segments, when the sentences are divided into two or three segments each. Figure 2 shows an example of a sentence pair involving global reordering with the sentence divided into three segments. We take the following steps:

1. We divide each source and target sentence into two or three segments. The candidate segments start at all possible word positions in the sentence. Here, a sentence pair consisting of K segments is represented as $(\phi_1, \phi_2 \cdots \phi_K)$, where ϕ_k consists of the kth phrase of the source sentence and $\alpha_k{}^{th}$ phrase of the target sentence. These segments meet the standard phrase extraction constraint.
2. By referring to the alignment table, the source and target phrases of ϕ_k are considered to be in swap orientation if $\alpha_k = \alpha_{k+1} + 1$.
3. From the candidates produced in step1, we select all segments satisfying the conditions of step 2. If there is more than one candidate, we select the segment candidate based on the head directionality of the source sentence. For a head-initial language, such as English, we select the candidate for which ϕ_K has the largest length. For a head-final language, such as Japanese, we select the candidate for which ϕ_1 has the largest length.

ID	n-grams	len	freq
m1	*prevent, |*	1	2217
m2	*To, prevent, |*	2	1002
m3	*To, prevent, |, imperfect*	3	120
m4	*To, prevent, |, imperfect, coating*	4	18

Figure 3: Example of n-gram matching against an input sentence containing two segments. The input sentence is "*To prevent imperfect coating and painting.*"

3.2 Training and Prediction of Global Reordering

3.2.1 Heuristics-based Method

In the heuristics-based method, we extract n-grams from the source sentences of the global reordering corpus and match these n-grams against a newly inputted sentence to perform global reordering. We call this method *heuristics-based*, because automatic learning is not used for optimizing the extraction and matching processes of the n-grams, but rather, we heuristically find the optimal setting for the given training data. Below, we describe the extraction and matching processes.

N-gram extraction We extract n-grams occurring on both sides of the segment boundary between adjacent segments ϕ_k and ϕ_{k+1}. In the heuristic-based method, n can assume different values in the left- and right-hand sides of the segment boundary. Let B be the index of the first word in ϕ_{k+1}, and f be the source sentence. Then the range of n-grams extracted on the left-hand side of f is as follows where nL is the value n of the n-gram.

$$(f_{B-nL}, f_{B-nL+1} \cdots f_{B-1}) \tag{1}$$

Likewise, the range of n-grams extracted from the right-hand side of f is as follows where nR denotes the value n of the n-gram.

$$(f_B \cdots f_{B+nR-2}, f_{B+nR-1}) \tag{2}$$

Decoding The decoding process of our global reordering is based on n-gram matching. We hypothesize that the matching candidate is more reliable (i) when the length of the n-gram matching is larger and/or (ii) when the occurrence frequency of the n-grams is higher. Thus, we heuristically determine the following score where len denotes the length of n-gram matching and $freq$ denotes the occurrence frequency of the n-grams. We calculate the score for all matching candidates and select the candidate that has the highest score.

$$\log(freq) \times len \tag{3}$$

Figure 3 shows an example of the decoding process for an input sentence containing two segments, i.e., $K = 2$, with one segment boundary. $m1$ through $m4$ are the n-grams matching the input sentence "*To prevent imperfect coating and painting,*" where "|" denotes the position of the segment boundary. The matching length is indicated by len which is the sum of nL and nR on both sides of the segment boundary. For example, for $m3$, the occurrence frequency is given as 120 and len is calculated such that $len = nL + nR = 2 + 1 = 3$. A score is calculated using equation 3 for all candidates, $m1$ through $m4$, and the candidate obtaining the highest score is used to determine the segment boundary.

3.2.2 Machine Learning-based Method

As the heuristic method involves intuitive determination of settings, which makes it difficult to optimize the performance of the system, we introduce machine learning to facilitate the optimization of segment detection. We regard segment boundary prediction as a binary classification task and use support vector machine (SVM) models to perform training and prediction. We train an SVM model to predict whether each of the word positions in the input sentence is a segment boundary by providing the features relating to the word in question. We use two types of features, as described below, for SVMs, both for training and prediction.
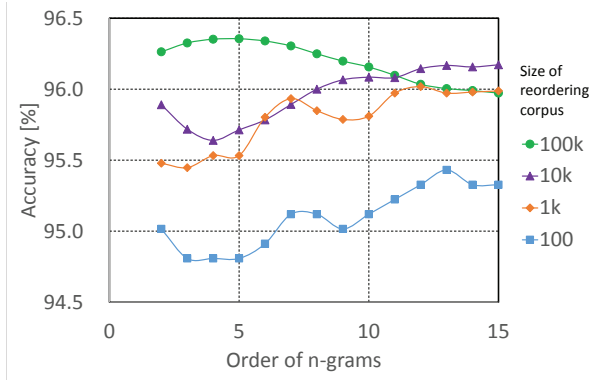
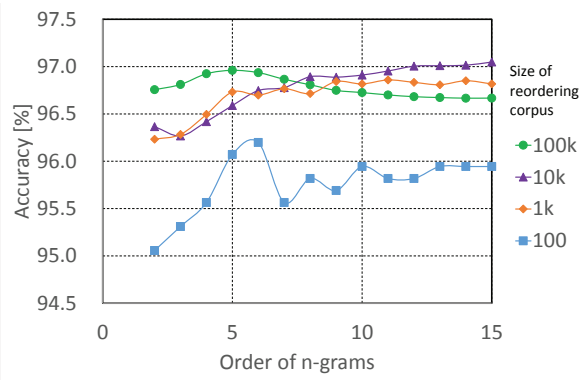Figure 4: Variation in the boundary prediction accuracy for Japanese input

Figure 5: Variation in the boundary prediction accuracy for English input

- **N-grams**: Here, n-grams are extracted from both sides of the word under training/prediction. In contrast to the heuristics-based method, for simplicity, we use here the same value of n for n-grams in the left- and right-hand sides of the examined word. The n-grams used are as follows, where f is the sentence, i is the index of the word in question, and n is the value n of n-grams.

$$(f_{i-n+1}, f_{i-n+2} \cdots f_i \cdots f_{i+n-1}, f_{i+n}) \tag{4}$$

- **Position in the sentence**: The position of the word under training/prediction is provided as a feature. This feature is introduced to differentiate multiple occurrences of identical n-grams within the same sentence. The position value is calculated as the position of the word counted from the beginning of the sentence divided by the number of words contained in the sentence. This is shown as follows, where i denotes the index of the word in question and F is the number of words contained in the sentence.

$$\frac{i}{F} \tag{5}$$

In the prediction process, we extract the features corresponding to the word position i and then input these features to the SVM model to make a prediction for i. By repeating this prediction process for every i in the sentence, we obtain a sentence with each position i marked either as a segment boundary or as *not* a segment boundary. These predicted segments are then reordered globally to produce the global sentence structure of the target language.

4 Experiments

In this section, we first describe the reordering configuration for depicting the effect of global pre-ordering. We then describe the primary preparation of global reordering, followed by a description of the settings used in our translation experiment.

4.1 Reordering Configuration

To illustrate the effect of introducing global pre-ordering, we evaluate the following four reordering configurations: **(T1)** Baseline SMT without any reordering; **(T2)** T1 with global pre-ordering only. The input sentence is globally pre-ordered, and this reordered sentence is translated and evaluated; **(T3)** T1 with conventional syntactic pre-ordering (Goto et al., 2015). The input sentence is pre-ordered using conventional syntactic pre-ordering and the reordered sentence is translated and evaluated; and **(T4)** T1 with a combination of syntactic and global pre-ordering. The input sentence is globally pre-ordered, each segment is reordered using syntactic pre-ordering and the reordered sentence is translated and evaluated.

4.2 Preparation of Global Reordering

In preparation for global pre-ordering, we calibrated the machine learning-based detection to determine the optimal feature set for detecting segments. To determine the optimal feature set, we plotted the

prediction accuracy with respect to the size of the global reordering corpus and value n of n-grams. As our support vector machines, we used liblinear 1.94 (Fan et al., 2008) for training and prediction.

Figure 4 shows the variation in the prediction accuracy with respect to the size of the global reordering corpus and the order of an n-gram for Japanese input. Figure 5 shows the same for English input. The *accuracy* is the average accuracy of a ten-fold cross-validation for the global reordering corpus. From the calibration shown in the tables, we select the settings producing the highest prediction accuracy, namely, a value of $five$ for the n of n-grams and a size of $100k$ for the global reordering corpus, for both Japanese and English inputs.

4.3 Translation Experiment Setup

Data As our experimental data, we use the Patent Abstracts of Japan (PAJ), the English translations of Japanese patent abstracts. We automatically align (Utiyama and Isahara, 2007) PAJ with the corresponding original Japanese abstracts, from which we randomly select 1,000,000 sentence pairs for training, 1,000 for development and 1,000 for testing. This training data for the translation experiment are also used for training global reordering as described in the previous subsection. Out of the 1,000 sentences in the test set, we extract the sentences that show any matching with the n-grams and use these sentences for our evaluation. In our experiments, the number of sentences actually used for evaluation is 300.

Baseline SMT The baseline system for our experiment is Moses phrase-based SMT (Koehn et al., 2007) with the default distortion limit of six. We use KenLM (Heafield et al., 2013) for training language models and SyMGIZA++ (Junczys-Dowmunt and Szal, 2010) for word alignment. The weights of the models are tuned with the n-best batch MIRA (Cherry and Foster, 2012). As variants of the baseline, we also evaluate the translation output of the Moses phrase-based SMT with a distortion limit of 20, as well as that of the Moses hierarchical phrase-based (Chiang, 2005) SMT with the default maximum chart span of ten.

Conventional syntactic pre-ordering Syntactic pre-ordering is implemented on the Berkeley Parser. The input sentences are parsed using the Berkeley Parser, and the binary nodes are swapped by the classifier (Goto et al., 2015). As a variant of conventional reordering, we also use a reordering model based on the top-down bracketing transducer grammar (TDBTG) (Nakagawa, 2015). We use the output of mkcls and SyMGIZA++ obtained during the preparation of the baseline SMT for training TDBTG-based reordering.

Global pre-ordering Global pre-ordering consists of the detection of segment boundaries and the reordering of the detected segments. Out of the 1,000,000 phrase-aligned sentence pairs in the training set for SMT, we use the first 100,000 sentence pairs for extracting the sentence pairs containing global reordering. We only use a portion of the SMT training data due to the slow execution speed of the current implementation of the software program for extracting sentence pairs containing global reordering. We evaluate both the heuristic and the machine learning-based methods for comparison.

Evaluation metrics We use the RIBES (Isozaki et al., 2010a) and the BLEU (Papineni et al., 2002) scores as evaluation metrics. We use both metrics because n-gram-based metrics such as BLEU alone cannot fully illustrate the effects of global reordering. RIBES is an evaluation metric based on rank correlation which measures long-range relationships and is reported to show much higher correlation with human evaluation than BLEU for evaluating document translations between distant languages (Isozaki and Kouchi, 2015).

5 Results

The evaluation results based on the present translation experiment are shown in Tables 1 and 2 for Japanese-to-English and English-to-Japanese translations respectively, listing the RIBES and BLEU scores computed for each of the four reordering configurations. The numbers in the brackets refer to the improvement over the baseline phrase-based SMT with a distortion limit of six.

A substantial gain of more than 25 points in the RIBES scores compared to the baseline is observed for both Japanese-to-English and English-to-Japanese translations, when global pre-ordering is used in con-

Table 1: Evaluation of Japanese-to-English translation where *glob-pre* denotes global pre-ordering and *pre* denotes conventional syntactic pre-ordering, *dl* denotes distortion limit, HPB denotes hierarchical phrase-based SMT and TDBTG denotes reordering based on top-down bracketing transduction grammar. The bold numbers indicate a statistically insignificant difference from the best system performance according to the bootstrap resampling method at $p = 0.05$ (Koehn, 2004).

	Reordering config		Settings			Results	
	glob-pre	pre	SMT	glob-pre	pre	RIBES	BLEU
T1			PB dl=6			44.9	17.9
			PB dl=20			53.7 (+8.8)	21.3 (+3.4)
			HPB			54.9 (+10.0)	23.1 (+5.2)
T2	√		PB dl=6	heuristic		61.7 (+16.8)	19.6 (+1.7)
			PB dl=6	SVM		61.0 (+16.1)	19.3 (+1.4)
T3		√	PB dl=6		TDBTG	64.6 (+19.7)	22.3 (+4.4)
			PB dl=6		syntactic	64.9 (+20.0)	**25.5 (+7.6)**
T4	√	√	PB dl=6	heuristic	syntactic	**71.3 (+26.4)**	**25.3 (+7.4)**
			PB dl=6	SVM	syntactic	**72.1 (+27.2)**	**25.6 (+7.7)**

Table 2: Evaluation of English-to-Japanese translation

	Reordering config		Settings			Results	
	glob-pre	pre	SMT	glob-pre	pre	RIBES	BLEU
T1			PB dl=6			43.2	27.9
			PB dl=20			54.4 (+11.1)	29.0 (+1.1)
			HPB			59.1 (+15.8)	32.1 (+4.2)
T2	√		PB dl=6	heuristic		59.5 (+16.2)	28.4 (+0.5)
			PB dl=6	SVM		65.3 (+22.1)	29.1 (+1.2)
T3		√	PB dl=6		TDBTG	**77.7 (+34.5)**	34.9 (+7.0)
			PB dl=6		syntactic	76.1 (+32.8)	**36.9 (+9.0)**
T4	√	√	PB dl=6	heuristic	syntactic	**77.3 (+34.1)**	**36.5 (+8.6)**
			PB dl=6	SVM	syntactic	**77.7 (+34.5)**	**36.5 (+8.6)**

junction with conventional pre-ordering. Also, the combination of global syntactic pre-ordering performs significantly better than syntactic pre-ordering alone. The BLEU score is not as sensitive to the introduction of global reordering, probably because the improvement is mainly concerned with long-distance reordering. We will further discuss the matter of evaluation metrics in the following section.

Figure 6 shows typical translations of the four reordering configurations: T1, T2, T3 and T4. Compared with the reference, the baseline (T1) fails to produce segment A and fails to output segments B and C in the correct order. In addition, the word order within each segment is not appropriate. The baseline with global pre-ordering (T2) successfully produces all three segments in the correct order, although the quality within each segment is not improved. The translation using conventional pre-ordering alone (T3) improves the local word order, while it fails to arrange the segments in the correct order. The translation with global and syntactic pre-ordering (T4) successfully produces the segments in the correct order, while at the same time improving the word order in each of the segments.

6 Analysis

To evaluate the ability of our proposed method to produce appropriate sentence structures in translated sentences, we count the number of sentences with correct global sentence structures among the translated test sentences. We think this is an important figure because the failure to produce the correct global sentence structure leads to inappropriate translation in most sublanguage-specific translations. We consider

Reference: [$_A$ To provide] [$_B$ a toner cake layer forming apparatus] [$_C$ which forms a toner cake layer having a high solid content and which can be actuated by an electrostatic printing engine.]
T1: [$_C$ Solid content of high toner cake layer for generating an electrostatic print engine operates in] [$_B$ a toner cake layer forming device.]
T2: [$_A$ To provide] [$_B$ toner cake layer forming apparatus] [$_C$ of the solid content of high toner cake layer for generating an electrostatic print engine can be operated.]
T3: [$_C$ For generating toner cake layer having a high solids content and] [$_A$ to provide] [$_B$ a toner cake layer forming device] [$_C$ which can be operated by an electrostatic printing engine.]
T4: [$_A$ To provide] [$_B$ a toner cake layer forming device] [$_C$ for generating toner cake layer having a high solid content, and operable by an electrostatic printing engine.]

Figure 6: Typical translations

Table 3: Number of correct global reordering in 100 test sentences

	Reordering config		Settings			Correct global reordering in 100 test sentences	
	glob-pre	pre	SMT	glob-pre	pre	Japanese-to English	English-to Japanese
T1			PB dl=6			4	6
			PB dl=20			12	15
			HPB			11	28
T2	√		PB dl=6	heuristic		24	23
			PB dl=6	SVM		31	26
T3		√	PB dl=6		TDBTG	21	59
			PB dl=6		syntactic	27	67
T4	√	√	PB dl=6	heuristic	syntactic	46	68
			PB dl=6	SVM	syntactic	58	63

that a translated sentence has a correct global sentence structure if it meets the following two criteria: (a) The translated sentence actually has the sentence structure CBA, where the source sentence structure ABC must be reordered to CBA in the target sentence. All the segments must be present in the correct order in the translated sentence; (b) All the words in each of the ABC segments in the source sentence must appear in an undivided segment in the target sentence. We randomly select a portion of the translated test sentences and manually counted the number of sentences meeting these criteria.

Table 3 shows the number of correct global reordering in Japanese-to-English and English-to-Japanese translations out of the 100 sentences randomly selected from the test set. The table shows that T4, which combines global and syntactic reordering, has a largely improved sentence structure compared with T1 and T2. In case of Japanese-to-English translation, the performance of T4 is much higher than T3, the state-of-the-art reordering alone. In case of an English-to-Japanese translation, the performance of the syntactic reordering is already considerably higher than the baseline and hence the performance of T4 is comparable to that of T3. The prominent improvement in BLEU scores by HPB observed in Tables 1 and 2 do not appear as prominent in Table 3, probably because HPB deals with more local reordering which is reflected well by BLEU score, but does not contribute much to global reordering.

7 Conclusion

In this paper, we proposed a global pre-ordering method that supplements conventional syntactic pre-ordering and improves translation quality for sublanguages. The proposed method learns global reordering models without syntactic parsing from a non-annotated corpus. Our experimental results on the patent abstract sublanguage show substantial gains of more than 25 points in RIBES and comparable BLEU scores for Japanese-to-English and English-to-Japanese translations.

References

Vamshi Ambati and Wei Chen. 2007. *Cross Lingual Syntax Projection for Resource-Poor Languages*. CMU.

Beat Buchmann, Susan Warwick-Armstrong, and Patrick Shane. 1984. Design of a machine translation system for a sublanguage. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics, Proceedings of COLING '84, July 2-6, 1984, Stanford University, California, USA.*, pages 334–337.

Colin Cherry and George Foster. 2012. Batch tuning strategies for statistical machine translation. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 427–436, Stroudsburg, PA, USA. Association for Computational Linguistics.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 263–270, Stroudsburg, PA, USA. Association for Computational Linguistics.

Adrià de Gispert, Gonzalo Iglesias, and William Byrne. 2015. Fast and accurate preordering for SMT using neural networks. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015)*, June.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Masaru Fuji, Atsushi Fujita, Masao Utiyama, Eiichiro Sumita, and Yuji Matsumoto. 2015. Patent claim translation based on sublanguage-specific sentence structure. In *Proceedings of the 15th Machine Translation Summit*, pages 1–16.

Michel Galley and Christopher D. Manning. 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.

Isao Goto, Masao Utiyama, Eiichiro Sumita, and Sadao Kurohashi. 2015. Preordering using a target-language parser via cross-language syntactic projection for statistical machine translation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 14(3):13:1–13:23, June.

Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H Clark, and Philipp Koehn. 2013. Scalable modified Kneser-Ney language model estimation. In *ACL (2)*, pages 690–696.

Sho Hoshino, Yusuke Miyao, Katsuhito Sudoh, Katsuhiko Hayashi, and Masaaki Nagata. 2015. Discriminative preordering meets Kendall's τ maximization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 139–144, Beijing, China, July. Association for Computational Linguistics.

Hideki Isozaki and Natsume Kouchi. 2015. Dependency analysis of scrambled references for better evaluation of Japanese translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 450–456, Lisbon, Portugal, September. Association for Computational Linguistics.

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 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 944–952, Stroudsburg, PA, USA. Association for Computational Linguistics.

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*, WMT '10, pages 244–251, Stroudsburg, PA, USA. Association for Computational Linguistics.

Marcin Junczys-Dowmunt and Arkadiusz Szal. 2010. SyMGiza++: A tool for parallel computation of symmetrized word alignment models. In *International Multiconference on Computer Science and Information Technology - IMCSIT 2010, Wisla, Poland, 18-20 October 2010, Proceedings*, pages 397–401.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 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.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing , EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, 25-26 July 2004, Barcelona, Spain*, pages 388–395.

Mamoru Komachi, Yuji Matsumoto, and Masaaki Nagata. 2006. Phrase reordering for statistical machine translation based on predicate-argument structure. In *IWSLT*, pages 77–82. Citeseer.

Heinz-Dirk Luckhardt. 1991. Sublanguages in machine translation. In *EACL 1991, 5th Conference of the European Chapter of the Association for Computational Linguistics, April 9-11, 1991, Congress Hall, Alexanderplatz, Berlin, Germany*, pages 306–308.

Daniel Marcu, Lynn Carlson, and Maki Watanabe. 2000. The automatic translation of discourse structures. In *ANLP*, pages 9–17.

Bart Mellebeek, Karolina Owczarzak, Declan Groves, Josef Van Genabith, and Andy Way. 2006. A syntactic skeleton for statistical machine translation. In *In Proceedings of the 11th Conference of the European Association for Machine Translation*, pages 195–202.

Tetsuji Nakagawa. 2015. Efficient top-down btg parsing for machine translation preordering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 208–218, Beijing, China, July. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 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.

Masao Utiyama and Hitoshi Isahara. 2007. A Japanese-English patent parallel corpus. In *Proceedings of the Eleventh Machine Translation Summit*, pages 475–482.

Tong Xiao, Jingbo Zhu, and Chunliang Zhang. 2014. A hybrid approach to skeleton-based translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, pages 563–568.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics.