

# Two-Stage Pre-ordering for Japanese-to-English Statistical Machine Translation

Sho Hoshino<sup>1</sup> and Yusuke Miyao<sup>1,2</sup> Katsuhito Sudoh<sup>3</sup> and Masaaki Nagata<sup>3</sup>

<sup>1</sup>The Graduate University for Advanced Studies <sup>2</sup>National Institute of Informatics

{hoshino, yusuke}@nii.ac.jp

<sup>3</sup>NTT Communication Science Laboratories, NTT Corporation

{sudoh.katsuhito, nagata.masaaki}@lab.ntt.co.jp

## Abstract

We propose a new rule-based pre-ordering method for Japanese-to-English statistical machine translation that employs heuristic rules in two-stages. This two-stage framework contributes to experimental results that our method outperforms conventional rule-based methods in BLEU and RIBES.

## 1 Introduction

Reordering is an important strategy in statistical machine translation (SMT) to achieve high quality translation. While many reordering methods often fail in long distance reordering due to computational complexity, a promising technology called *pre-ordering* (Xia and McCord, 2004; Collins et al., 2005) has been successful for distant English-to-Japanese translation (Isozaki et al., 2010b). However, this strong effectiveness has not been shown for Japanese-to-English translation.

In this paper, we propose a novel rule-based pre-ordering method for the Japanese-to-English translation. The method utilizes simple heuristic rules in two-stages<sup>1</sup>: the inter-chunk and intra-chunk levels. Thus the method can achieve more accurate reorderings in Japanese. The translation experiments in patent domain showed that our method outperformed conventional rule-based methods, especially on the word reorderings. Our claims in this paper are summarized as follows:

1. The inter-chunk pre-ordering that relies on PAS analysis contributes to improvements in translation quality.
2. The intra-chunk pre-ordering which converts postpositional phrases into prepositional phrases further improves translation quality.
3. Thus, our two-stage framework is more effective than other pre-ordering methods.

<sup>1</sup>In this paper, we refer to Japanese *bunsetsu* as a “chunk”, a grammatical and phonological unit consists of noun, verb, or adverb followed by dependents such as particles.

## 2 Related Work

Japanese-to-English is challenging because the grammatical forms of the two languages are totally dissimilar. For instance, English is a head-initial language, and utilizes subject-verb-object (SVO) word orders, while Japanese is a *pure* head-final language, and utilizes subject-object-verb (SOV).

Komachi et al. (2006) proposed a rule-based pre-ordering method to convert SOV into SVO via a PAS analyzer. This method pre-orders inter-chunk level word orders in a single-stage, via the PAS analyzer which produced dependency trees and tagged each S, O, and V label. Then SOV sequences are converted into SVO. However, since the non-labeled words are left untouched, the effectiveness of this method is limited to simple SOV labeled matrix sentences without multiple clauses.

Katz-Brown and Collins (2008) proposed a two-stage rule-based pre-ordering method. In the first stage, SOV sequences are converted into SVO via the dependency analyzer. In the second stage, each chunk word order is naively reversed<sup>2</sup>.

Neubig et al. (2012) proposed a statistical model that was capable of learning how to pre-order word sequences from human annotated or automatically generated alignment data. However, this method has very large computational complexity to model long distance reordering.

## 3 Two-stage Pre-ordering Method

Here, we describe a new pre-ordering method which employs heuristic rules in two-stages. In the first stage, we reorganize and extend the rules described in (Komachi et al., 2006; Katz-Brown and Collins, 2008). In the second stage, we propose a new rule to consider chunk internal word orders. More precisely, **we apply four rules**: three rules for the first stage (Rule 1-1, 1-2, 1-3) and one

<sup>2</sup>Since the rule for S has not been described in detail, we provide a definition: S is a chunk followed by a topic-marker.

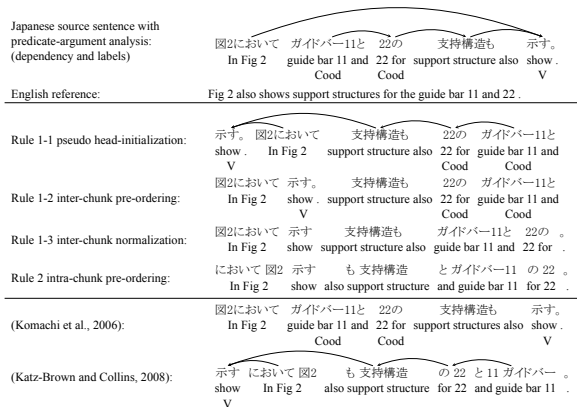


Figure 1: Pre-ordering Examples.

rule for the second stage (Rule 2). Each rule corresponds to the different linguistic nature between English and Japanese.

### 3.1 Rule 1-1 pseudo head-initialization

To output Japanese in head-initial sequences, this rule modifies Japanese dependency trees to the order that a head chunk comes first and its dependent children follow, by default. In the example, the sentence verbal head “示す show(s)” is moved to the leftmost position.

### 3.2 Rule 1-2 inter-chunk pre-ordering

This rule converts SOV into SVO. The rule can also handle a sentence which has no subject and object, due to a parsing error or a pro-drop that frequently occurs in Japanese. If there is V (a verbal head), then we apply this rule after Rule 1-1.

First, we move V instead of S or O, as we already have VSO sequences generated in Rule 1-1. Therefore, we placed V after the subject ( $V x^* S y^* \rightarrow x^* S V y^*$ ) or before the object in case a subject is not found ( $V x^* O y^* \rightarrow x^* V O y^*$ ).

Second, in the case where we only have V (without S and O), we move V before the rightmost chunk ( $V x^* y \rightarrow x^* V y$ ) to avoid head-initial outputs. In the example, since there is no subject and object, the verbal head “示す show(s)” is moved to immediately before its second dependent. On the other hand, V has been incorrectly placed in the rightmost in Komachi et al. (2006) and the leftmost in Katz-Brown and Collins (2008).

### 3.3 Rule 1-3 inter-chunk normalization

If there are coordinate clauses or punctuations, then we apply this rule after Rule 1-2. Basically,

we keep coordinate clauses and punctuations unchanged from the original word orders, by placing the coordinate clauses to the leftmost position and the punctuations to the rightmost position ( $x^* Punc y^* Cood z^* \rightarrow Cood x^* y^* z^* Punc$ ). In addition, in order to avoid comma-period sequences, we remove all commas immediately before a period.

In the example, the period “。” is moved to the rightmost position, unlike Komachi et al. (2006). And the coordinate clause “ガイドバー11と22の the guide bar 11 and 22” is restored to the original position in the source, by moving the clause to the rightmost position. While Katz-Brown and Collins (2008) does not have such a rule to restore the coordinate clause, Komachi et al. (2006) can keep it unchanged because that method does not move non-labeled words.

### 3.4 Rule 2 intra-chunk pre-ordering

For every chunk, we swap function and content words to organize pseudo prepositional phrases (Content Function  $\rightarrow$  Function Content). In the example, the chunk “ガイドバー11と the guide bar 11 AND” has three words: the two content words “ガイドバー11 the guide bar 11” and the function word “と AND”. Thus the chunk is reversed as “とガイドバー11 AND the guide bar 11”.

### 3.5 Differences to Conventional Rules

Komachi et al. (2006) did not employ Rule 1-1 and only employed Rule 1-2. In this example, the head “support structures” should be followed by the dependents “guide bar 11 and 22”, but these words are left untouched. Katz-Brown and Collins (2008) employed Rule 1-1, the most of Rule 1-2, and a rule to keep punctuations as it partially treated by Rule 1-3. However, they did not have the exceptional rule to move verb from the first position for non-subject sentences, as described in Rule 1-2. In the example, this method misplaced the sentence verbal head “show” to the first position, and the coordination clause “guide bar 11 and 22” has also been mixed.

	BLEU	RIBES
Baseline	29.19	68.48
(Katz-Brown and Collins, 2008)	27.74	66.15
(Komachi et al., 2006)	29.58	69.10
(Neubig et al., 2012)	29.93	70.15
Proposed method	<b>30.65</b>	<b>72.26</b>

Table 1: Experimental Results.

Rule 1-2	Rule 1-3	Rule 2	BLEU	RIBES
			29.19	68.48
✓			29.76	71.00
	✓		27.71	69.50
		✓	28.29	65.61
	✓	✓	28.84	70.40
✓		✓	30.41	71.74
✓	✓		<b>30.94</b>	71.34
✓	✓	✓	30.65	<b>72.26</b>

Table 2: Ablation Tests.

	BLEU	RIBES
Proposed within KNP	<b>30.65</b>	72.26
Proposed within CaboCha+SynCha	30.01	<b>72.35</b>

Table 3: Differences in Parser Configurations.

## 4 Experiments

### 4.1 Experimental Setup

In order to compare pre-ordering methods, we conducted Japanese-to-English translation experiments on a fixed data set and SMT system.

For the common data set, we used the NTCIR-9 PatentMT Test Collection Japanese-to-English Machine Translation Data<sup>3</sup> package that contains approximately 3.2 million sentence pairs for training, 500 sentence pairs for development, and 2,000 sentence pairs for testing. The Japanese sentences are tokenized by MeCab 0.994<sup>45</sup>. In addition, we employed two parser configurations for Japanese parsing: (1) the KNP configuration used KNP 4.01<sup>6</sup> (Sasano and Kurohashi, 2011) for both dependency and PAS analyzer; (2) the CaboCha+SynCha configuration used CaboCha 0.65<sup>7</sup> (Kudo and Matsumoto, 2002) for dependency analysis and SynCha 0.3<sup>8</sup> (Iida and Poesio, 2011) for PAS analysis.

For the common SMT system, we used SRILM

<sup>3</sup><http://research.nii.ac.jp/ntcir/permission/ntcir-9/perm-en-PatentMT.html>

<sup>4</sup><http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

<sup>5</sup>Since (unlike English) the Japanese language does not utilize spaces to delineate word boundaries, MeCab was used to perform the required Japanese tokenization.

<sup>6</sup><http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KNP>

<sup>7</sup><http://code.google.com/p/cabocho/>

<sup>8</sup><http://www.cl.cs.titech.ac.jp/~ryu-i/syncha/>

	BLEU	RIBES
Baseline	15.03	62.71
Proposed	<b>16.12</b>	<b>69.30</b>

Table 4: Results within a News Domain.

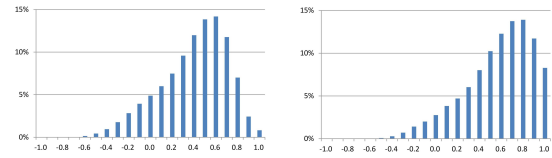


Figure 2: Kendall's  $\tau$  in Baseline and Proposed.

1.7.0 (Stolcke et al., 2011), MGIZA++ 0.7.3 (Gao and Vogel, 2008), Moses 0.91 (Koehn et al., 2007)<sup>9</sup>, and two popular evaluation metrics for Japanese-to-English: BLEU (Papineni et al., 2002) and RIBES (Isozaki et al., 2010a).

For the pre-ordering methods, we implemented rule-based methods proposed by (Komachi et al., 2006) and (Katz-Brown and Collins, 2008). In addition, an implementation<sup>10</sup> of statistical method proposed by Neubig et al. (2012) are used<sup>11</sup>. We did not use any pre-ordering in the baseline.

### 4.2 Experimental Results

Table 1 shows the experimental results for Japanese-to-English patent document translations that compare the following pre-ordering methods: the baseline (no pre-ordering), Komachi et al. (2006), Katz-Brown and Collins (2008), Neubig et al. (2012), our proposed method. We also conducted ablation tests, which consisted of a comparison of all Rule 1-2, 1-3, and 2, shown in Table 2.

The experimental results show that our proposed method outperformed all the other pre-ordering methods in terms of the BLEU and RIBES, which scored 30.65 and 72.26 points, respectively. This indicates that our two-stage pre-ordering method is better than conventional rule-based pre-ordering methods in the following aspects we found:

1. Our method and Komachi et al. (2006), both of which rely on PAS, were better than Katz-Brown and Collins (2008) which utilizes deterministic rules to obtain SOV labels.
2. Our intra-chunk pre-ordering gained a further improvement in translation accuracy as shown in Table 2. Nevertheless, we observed a 0.3 point drop in BLEU and a 0.9 point

<sup>9</sup>the following configurations are used in the system: 6-gram for language modeling, msd-bidirectional-fe for re-ordering, and MERT (Och, 2003) for tuning. After reviewing our preliminary findings, distortion limits were set to 20 for the baseline and (Komachi et al., 2006), and 10 for others.

<sup>10</sup><http://www.phontron.com/lader/>

<sup>11</sup>Only 10,000 sampled lines were used for training due to its computational complexity: During the training process, it consumed 120 GB of memory space for almost entire month.

improvement in RIBES by adding Rule 2 to Rule 1-2 and Rule 1-3, even though this combination yields better translations for native speakers. This phenomenon can be explained by the characteristic difference between BLEU and RIBES. While RIBES has a good correlation to human judgments, BLEU is said to have an uncorrelated, erratic behavior for Japanese-to-English translation (Isozaki et al., 2010a).

3. Our heuristic rules can cover more pre-ordering issues (as shown in Figure 1), and achieved further improvement in Rule 1-2 and Rule 1-3 as shown in Table 2.

In addition, as shown in Table 3, there was a 0.6 point statistically significant difference between two parser configurations (KNP and CaboCha+SynCha) in BLEU for our method. We suppose that one possible explanatory factor is the coordination structural accuracy to utilize Rule 1-3, because KNP tends to output more accurate coordination structures than CaboCha. However, it will be necessary to analyze our results in further detail to produce more definite conclusions. In any case, since such differences have been achieved simply by switching parsers, we believe that a better parsing method can be expected to produce better translation results in the future.

### 4.3 Pre-ordering Evaluation

We employed Kendall’s  $\tau$  rank correlation coefficient and its distribution as our pre-ordering criteria as described in Isozaki et al. (2010b). As shown in Figure 2, our proposed method produced much better correlation distribution than the baseline<sup>12</sup>.

Table 4 shows experimental results conducted on a news document that contains 150,000 sentence pairs created by (Utiyama and Isahara, 2003). Similar to the results shown in Table 1, our method outperformed the baseline.

## 5 Error Analysis and Discussion

Figure 3 shows an example within the proposed method. The intra-chunk rule Rule 2 moved the postposition “in” before the noun “Fig.7” and thus it makes the prepositional phrase “in Fig.7”. Also

<sup>12</sup>The average value of  $\tau$  in the proposed method is 0.575 against 0.391 in the baseline, and the percentages of sentences which have value of 0.8 or higher were 33.9% in our proposed method and 10.2% in the baseline. This 20% difference represents great reduction of word order differences.

Source: ここで、表1及び図7に示す各記号は、次のものを表している。 Here, Table 1 and Fig.7 in show each symbol, following things represent.
Reference: Here, symbols shown in Table 1 and Fig.7 represent the following items.
Proposed: でここは各記号示す及び表1に図7、いる表してをもの次の。 Here each symbol show and Table 1 in Fig.7, represent the things following.
Translated Result: In this case, the respective symbols shown in Table 1 and Fig.7 represents the followings.

Figure 3: Successful Pre-ordering Example

the coordination structure “Table 1 and Fig.7” is kept. There is still a minor verb agreement error which the verb “represent” is translated as “represents”. However, the most of errors are given via parsing process. For instance, of the first 30 sentences in the test data, we found 21 SOV tagging errors and 9 critical dependency errors, despite CaboCha is reported to have 89.8% accuracy for overall dependencies. Other methods could not translate this example correctly.

Besides, we also found that our deterministic rules cannot handle some difficult Japanese constructions. As a result, incorrect reordering had been conducted. For example, many Japanese sentences have a *topic* with a *topic-object-verb* construction, instead of subject-object-verb, because Japanese is a *topic-prominent* language. In the same 30 sentences, 18 sentences formed the topic-object-verb construction, and 4 sentences have been found as the topic-subject-object-verb construction.

## 6 Conclusion

In this paper, we proposed a new rule-based pre-ordering method for Japanese-to-English statistical machine translation, and we showed that our two-stage pre-ordering scheme was capable of solving more complex pre-ordering problems than conventional methods. From the experimental results, we found that our proposed method outperformed existing rule-based pre-ordering methods in terms of standard evaluation metrics.

## Acknowledgments

We would like to thank Ryu Iida, Mamoru Komachi, Taku Kudo, Graham Neubig, Ryohei Sasano, Wu Xianchao, and anonymous reviewers.

## References

- Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 531–540.
- Qin Gao and Stephan Vogel. 2008. Parallel implementations of word alignment tool. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 49–57, June.
- Ryu Iida and Massimo Poesio. 2011. A cross-lingual ILP solution to zero anaphora resolution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 804–813.
- 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*, pages 944–952, October.
- 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.
- Jason Katz-Brown and Michael Collins. 2008. Syntactic reordering in preprocessing for Japanese→English translation: MIT system description for NTCIR-7 patent translation task. In *Proceedings of the NTCIR-7 Workshop Meeting*.
- 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 Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180.
- Mamoru Komachi, Yuji Matsumoto, and Masaaki Nagata. 2006. Phrase reordering for statistical machine translation based on predicate-argument structure. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 77–82.
- Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*, pages 63–69.
- Graham Neubig, Taro Watanabe, and Shinsuke Mori. 2012. Inducing a discriminative parser to optimize machine translation reordering. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 843–853.
- Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 160–167.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.
- Ryohei Sasano and Sadao Kurohashi. 2011. A discriminative approach to japanese zero anaphora resolution with large-scale lexicalized case frames. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 758–766.
- Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at sixteen: Update and outlook. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*.
- Masao Utiyama and Hitoshi Isahara. 2003. Reliable measures for aligning japanese-english news articles and sentences. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 72–79.
- Fei Xia and Michael McCord. 2004. Improving a statistical MT system with automatically learned rewrite patterns. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 508–514.