

# System Combination of RBMT plus SPE and Preordering plus SMT

Terumasa EHARA

Ehara NLP Research Laboratory  
Seijo, Setagaya, Tokyo, JAPAN

eharate @ gmail.com

## Abstract

System architecture, experimental settings and evaluation results of EHR group in the en-ja, zh-ja, JPCzh-ja and JPCko-ja tasks are described. Our system concept is combination of a rule based method and a statistical method. System combination of rule-based machine translation (RBMT), RBMT plus statistical post-editing (SPE) and preordering plus statistical machine translation (SMT) is conducted. From the multiple outputs of three systems, candidate selection part selects the best output by language model score. For JPCzh-ja task devtest data translation, SPE improves BLEU score by 17.81, preordering improves BLEU score by 1.89 and system combination improves BLEU score by 0.26.

## 1 Introduction

Two main processes of machine translation are lexical transfer and structural transfer. Machine translation techniques and related techniques are classified in terms of these two processes as shown in Table 1.

Technique	Lex. Trans.	Struct. Trans.
RBMT	✓	✓
SMT	✓	✓
Monotone SPE	✓	
Preordering		✓
Monotone SMT	✓	

Table 1: Classification of MT and related techniques.

RBMT and SMT conduct lexical transfer and structural transfer simultaneously. On the other hand, monotone SPE and monotone SMT, which are technically the same process, conduct lexical transfer only. Preordering conducts structural transfer only.

We have made researches combining a rule based method and a statistical method that is RBMT plus SPE (Ehara, 2014). This year, we add preordering plus SMT part to our system for WAT2015. This new system is also the combination of rule based method (RBMT and preordering) and statistical method (SPE and SMT).

## 2 System architecture

Our basic system architecture is shown in Figure 1.

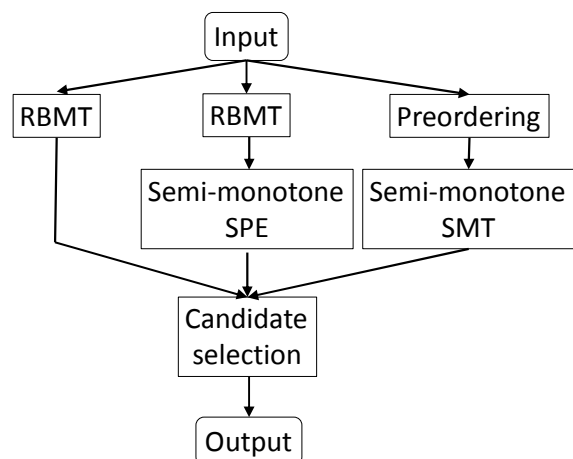


Figure 1: Basic system architecture.

Input sentence is fed to RBMT system, RBMT plus SPE system and Preordering plus SMT system<sup>1</sup>. From outputs of three systems, candidate selection part selects best output as the system output. Here, our SPE and SMT are semi-monotone,

<sup>1</sup> For JPCko-ja task, we don't use preordering part.

because distortion limit of decoder is set to 6 instead of 0. We will explain details of the each part of the system in the following subsections.

## 2.1 RBMT part

We use a commercial based RBMT system for the RBMT part. We, also, use user terminology dictionaries for zh-ja, JPCzh-ja and JPCko-ja tasks. For zh-ja and JPCzh-ja tasks, we use a part of Chinese to Japanese technical term dictionary of JPO (Japan Patent Office) (Japio, 2015)<sup>2</sup>. Original JPO’s dictionary includes 2,210,294 words (nouns and verbs). We filter out all verbs and the nouns which have identical Japanese translations with the commercial based RBMT outputs. As the result, we select 1,463,265 terms for the user dictionary for JPCzh-ja and zh-ja tasks. For JPCko-ja task, we make a user dictionary from the training corpus of the task. We get 434,334 terms for the user dictionary for the JPCko-ja task. For en-ja task, we don’t use any user dictionary.

We also use sentence pattern dictionary for JPCzh-ja task. We use only 13 sentence patterns for the task.

## 2.2 SPE part

SPE part intends to improve the translation quality of the output of the RBMT part. All target languages of the tasks are Japanese. So SPE part translates Japanese to Japanese. We use phrase based Moses (Koehn et al., 2003) with default options as the SPE engine. Word segmenter for Japanese is Juman ver.7.01 (Kurohashi et al., 1994).

Translation models (TM) of each task is built from RBMT output and reference translation of the training corpus of each task. Training corpus size (number of sentence pairs) will be listed in Table 3.

Language model (LM) is built from the following monolingual corpora. LM for en-ja task and zh-ja task is built from target side of the training corpora both of the en-ja task and zh-ja task (3.6M sentences). LM is built by Implz tool in Moses (KenLM) with order 6. LM for JPCko-ja task and JPCzh-ja task is built from target side of the training corpora both of the JPCko-ja task and JPCzh-

ja task and also Japanese side of NTCIR-10’s training corpora of PatentMT task EJ subtask (Goto et al. 2013)<sup>3</sup>. Total number of sentences for this LM training is 5M. This LM is also built by Implz with order 6.

Distortion limit (DL) of the decoder is set to 6. Usual setting of DL between Japanese and English or between Japanese and Chinese is 10 or over. So we call our SPE part semi-monotone SPE.

## 2.3 Preordering part

Preordering is one of the effective technique to improve structural transfer accuracy (Isozaki, 2010). Our preordering method uses context free parsing rules with reordering rules. Figure 2 shows examples of parsing rules with reordering rules and example of parsed phrases<sup>4</sup>. First example is English grammar rule with reordering rule. The right hand side of the parsing rule “ADVP VBN PP” is reordered to “PP ADVP VBN” by the reordering rule “2 0 1”<sup>5</sup>. Second example is Chinese grammar rule with reordering rule. Reordering rules are built by a heuristic algorithm.

<p>VP ⇒ ADVP VBN PP 2 0 1 (widely utilized in many fields)</p> <p>VP ⇒ VV NP IP 1 2 0 (使 各个 电动机 13 旋转 驱动)</p>
---

Figure 2: Example of parsing rules and reordering rules with examples

Parsing accuracy directly affects preordering accuracy. We use Stanford Chinese word segmenter (Tseng et al., 2005; Chang et al., 2008) and Berkeley parser (Petrov et al., 2006) as the parsing engine of our preordering part. Two improvements for the parsing are conducted. First one is grammar improvement for Chinese grammar. For en-ja task, we use the original English grammar of the Berkeley parser, eng\_sm6.gr. For JPCko-ja task, we don’t conduct preordering because of the similarity of word order of Korean and Japanese. For zh-ja task and JPCzh-ja task, we improve the original Chinese grammar, chn\_sm5.gr. It will be explained in section 2.3.1. Second improvement of parsing is reranking of k-best parse trees that will be explained in section 2.3.2.

<sup>2</sup> <https://alaginrc.nict.go.jp/resources/jpo-info/jpo-list.html>

<sup>3</sup> Dev, devtest and test data of JPCko-ja task and JPCzh-ja task are extracted from Japanese patent documents published in 2013. On the other hand, NTCIR-10’s training corpora is extracted from Japanese patent documents published in 1990 to 2005. They are not overlapped.

<sup>4</sup> All sample sentences and phrases in this paper are from ASPEC Corpora or JPO Patent Corpora provided by the workshop organizer.

<sup>5</sup> Reordering rule “2 0 1” means that position of right hand side tags permutes from “0 1 2” to “2 0 1”. Then, “ADVP VBN PP” is reordered to “PP ADVP VBN”.

### 2.3.1 Grammar improvement

The idea for grammar improvement is to use word alignment of JPCzh-ja bilingual training corpus. Firstly, word alignment is conducted from JPCzh-ja training corpus (1M sentence pairs) by GIZA++ (Och and Ney, 2003). For each sentence pairs, we decide sentence head word both for Japanese and Chinese using word alignment. Since Japanese is a typical head final language, head word of Japanese sentence is positioned at the end of the sentence. So it is easy to find the sentence head word for Japanese sentences. We find Chinese sentence head word as the aligned word to the Japanese sentence head word. For example, in the ja-zh sentence pair shown in Figure 3, Japanese sentence head word “発生する” is aligned to Chinese word “产生”. So “产生” is decided as the head word of this Chinese sentence. We consider it as the gold standard head word.

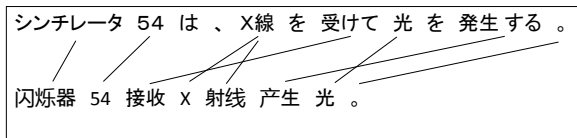
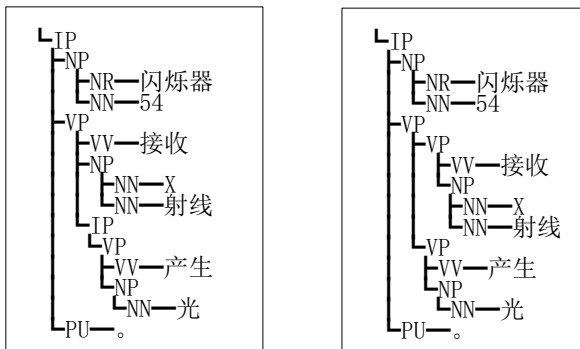


Figure 3: Example of zh-ja word alignment.



(a) top ranked tree (b) second ranked tree  
Figure 4: Parsed trees.

Next step is to make a tree bank. Chinese sentences of training corpus are parsed by the original grammar i.e. chn\_sm5.gr, and we get k-best parse trees for each sentence (k is set to 100). Then we select the best parse, in which the sentence head word is same as the gold standard head word. For example, the Chinese sentence in Figure 3 is k-best parsed shown in Figure 4. The top ranked tree has “接收” as the sentence head word and the second ranked tree has “产生” as the sentence head word, which is same as the gold standard. So we

<sup>6</sup> Devtest data is not included in the training data for the grammar improvement. Gold standard of Chinese

put the second ranked tree to our tree bank in this case.

From 1M JPCzh-ja training corpus, the number of second or lower ranked tree is selected is about 151K. Re-training Berkeley Chinese grammar using this 151K tree bank, we get new grammar named chn\_jpo.gr. Comparing original chn\_sm5.gr and chn\_jpo.gr, the agreement rate of sentence head word of top ranked parse tree and gold standard for devtest data is shown in Table 2<sup>6</sup>.

Agreement rate rises about 13%.

Grammar	Agreement rate (%)
chn_sm5.gr	50.5
chn_jpo.gr	63.0

Table 2: Agreement rate of sentence head word for the devtest data

### 2.3.2 Reranking of k-best parse trees

Another improvement of preordering part is reranking of k-best parse trees. For the training corpus, reordered source sentence is compared to gold standard reordered source sentence. Here, gold standard reordered source sentence is determined using alignment to a target sentence. For example, Chinese sentence shown in Figure 3 “闪烁器 54 接收 X 射线 产生 光。” is gold standard reordered to “闪烁器 54 X 射线 接收 光 产生。” using alignment to the target sentence. This comparison is measured by word error rate and select the parse tree which has the minimum word error rate in the k-best parse trees as the best parse tree. The parse tree shown in Figure 4 (a) is reordered to “闪烁器 54 X 射线 光 产生 接收。” and the parse tree shown in Figure 4 (b) is reordered to “闪烁器 54 X 射线 接收 光 产生。”. Then, Figure 4 (b) is selected as the best parse tree in this case.

For the dev, devtest and test sets, we use LM based reranking to select the best parse tree. Firstly, we make reordered source sentence corpus from the training corpus by the above method and build LM using this corpus. Next, we select the best parse tree which has the maximum LM score in the k-best reordered sentences in dev, devtest and test sets. Here, LM score of a sentence is a score calculated by the tool “query” in Moses divided by the number of words in the sentence.

sentence head for devtest data is determined by the method described above.

## 2.4 SMT part

SMT part uses phrase based Moses same as SPE part. For JPCko-ja task, SMT part translates source sentences to target sentences as the usual phrase based SMT. As the segmenter for Korean, we use MeCab-ko<sup>7</sup>. For JPCzh-ja task, zh-ja task and en-ja task, SMT part translates reordered source sentences to the target. Reordering is made by the method described in 2.3. Distortion limit is set to 6 both JPCko-ja task and other tasks. So, we call our SMT semi-monotone SMT. LM for SMT is same as LM for SPE. TM is trained from the training corpus provided by the workshop organizer. Training corpus sizes (number of sentence pairs) are listed in Table 3.

Task	Corpus size
JPCko-ja	994,998
JPCzh-ja	995,385
zh-ja	668,468
en-ja	2,351,575

Table 3: Training corpus size

## 2.5 Candidate selection part

The last part of our translation system is a candidate selection part. This part select the candidate which has the maximum LM score from the outputs of RBMT part, RBMT+SPE part and Preordering+SMT part. Here, LM score is calculated from the LM for SMT part by the method described in section 2.3.2.

## 2.6 Other ad-hoc processing

For JPCko-ja task, we conduct an ad-hoc preprocessing for Korean source sentences of the train, dev, devtest and test corpora and their RBMT outputs. It is deletion of brackets surrounding the number, because the use of brackets between Korean and Japanese is different shown in Figure 5. In Korean sentence, number “2” is surrounded by the brackets. However, in Japanese sentence, number “2” is not surrounded by the brackets. So we delete brackets surrounding the number in Korean side to improve alignment accuracy of brackets.

<p>또한, 산화피막(2)이 존재하는 경우에는, ...          また、酸化皮膜2が存在する場合には、...</p>
---

Figure 5: Different bracket usage in Korean and Japanese.

Another ad-hoc processing is to convert all half width characters in RBMT and SMT outputs to full width characters, because Japanese document tend to use full width characters.

## 2.7 Issues for context-aware machine translation

We have no consideration for context-awareness in our system.

# 3 Experimental results

## 3.1 Translation results

Table 4 shows the official evaluation results of the translation of the test data (Nakazawa et al., 2015). In all cases, BLEU and RIBES are calculated using Japanese segmenter Juman.

Task	System	BLEU	RIBES	HUMAN
JPCko-ja	RBMT+SPE	70.13	0.9419	6.500
	SMT	70.81	0.9430	--
	Combination	70.67	0.9430	10.250
JPCzh-ja	RBMT+SPE	40.35	0.8195	8.250
	Preordering+SMT	40.70	0.8243	--
	Combination	41.06	0.8270	22.000
zh-ja	RBMT+SPE	35.59	0.8158	--
	Preordering+SMT	39.43	0.8377	--
	Combination	37.90	0.8260	25.750
en-ja	RBMT+SPE	30.46	0.7685	--
	Preordering+SMT	29.78	0.7536	32.500
	Combination	30.88	0.7657	--

Table 4: Evaluation results of the translation

In JPCko-ja task and JPCzh-ja task, system combination using candidate selection by LM score is more accurate than RBMT+SPE system both in automatic and human evaluation. In zh-ja task, Preordering+SMT system has higher BLEU and RIBES than system combination. However, we don't have human score for preordering+SMT system for the zh-ja task.

## 3.2 Candidate selection results

Table 5 shows the candidate selection results. Most of outputs of RBMT part are not selected. Outputs of RBMT+SPE part and outputs of preordering+SMT part are selected about half and half.

<sup>7</sup> <https://bitbucket.org/eunjeon/mecab-ko/>

Task	RBMT	RBMT+SPE	Preordering+SMT	Total
JPCko-ja	25	1177	798	2000
JPCzh-ja	2	875	1123	2000
zh-ja	9	1270	828	2107
en-ja	5	658	1149	1812

Table 5: The number of each system outputs selected by the candidate selection part.

To confirm effectiveness of candidate selection process, we compare LM scores and human evaluation scores for JPCzh-ja task. Table 6 shows human evaluation score of SPE<sup>8</sup> outputs and SMT outputs when the case of LM score for SMT output exceeds LM score for SPE output.

SMT\SPE	-1	0	1
-1	6	4	2
0	25	101	26
1	8	32	18

Table 6: The number of human evaluation scores for SPE outputs and SMT outputs when the case of LM score of SMT output exceeds LM score of SPE output.

From the Table 6, we can see this candidate selection process makes human score better in 65 cases (SMT > SPE) and worse in 32 cases (SPE > SMT). The number of tie cases is 125.

To investigate worsen cases, we show several translation examples. Table 7 shows SMT output and SPE output, baseline (BASE) output, reference (REF) and source (SRC) of two translation examples. In these cases, LM score of SMT output is greater than LM score of SPE output. But human score of SMT output (-1) is less than human score of SPE output (1).

In the first example, the word “オートコリメータ” is less general than the word “コリメータ”. Actually, LM score of the former word is -5.61676 and LM score of the latter word is -4.12944. Then LM score of the former sentence is less than LM score of the latter sentence.

In the second example, “アノード410と陰極415” is less general than “アノード415とカソード410” in our LM. Actually, LM score of SMT output sentence is -66.1355 and the LM score of the sentence which is converted the term “アノード415とカソード410” of SMT output to “アノード410とカソード4

15” is -71.1855. These two examples indicate the limitation of LM score based candidate selection method.

SMT	図5は、コリメータ装置による調整動作を説明するための概略図である。
SPE	図5は、オートコリメータ装置による調整動作を説明するための概略図である。
BASE	図5は、自動コリメータ装置による調整動作を説明するための概略図である。
REF	図5は、オートコリメータ装置による調整動作を説明するための概略図である。
SRC	図5是用于说明由自动准直仪装置进行的调整操作的概略图。

SMT	図4は、名称からアノード415とカソード410との間に挟まれた発光層405のOLED400とを含む第1の例示的な実施の形態について説明する。
SPE	図4は、概念上からアノード410と陰極415との間の発光層405のOLED400の第1の例示的な実施形態を含むと説明する。
BASE	図4は、陽極410とカソード415との間の発光層405のOLED400の第1で概念的に説明は、例示的な実施形態が挟まれている。
REF	図4は、アノード410とカソード415とに挟まれた発光層405を含むOLED400の第1実施例の概念図である。
SRC	図4从概念上说明包括夹在阳极410与阴极415之间的发射层405的OLED 400的第一示例性实施方案。

Table 7: Example of candidate selection part making worse output.

### 3.3 Other results

Table 8 shows other evaluation results for JPCzh-ja task on devtest data translation.

System	Additional feature	BLEU	RIBES
RBMT		16.55	0.7192
RBMT	User dictionary	23.54	0.7510
RBMT+SPE	User dictionary	41.35	0.8203
SMT	(DL=10)	40.86	0.8071
Preordering+SMT	original grammar	40.15	0.8164
Preordering+SMT	improved grammar	41.84	0.8218
Preordering+SMT	improved grammar + reranking of parse trees	42.75	0.8237
System combination		43.01	0.8265

Table 8: Evaluation results for JPCzh-ja devtest data translation

User dictionary, SPE and preordering greatly improve RIBES score. Improving of grammar, reranking of parse trees and system combination slightly improve RIBES score. For BLEU score,

<sup>8</sup> In this section, the term “SPE” is used for RBMT+SPE and the term “SMT” is used for preordering+SMT for the simplicity.

results are almost similar as RIBES case. However, preordering with original grammar makes BLEU score worse compared with simple SMT. RBMT+SPE with user dictionary improves BLEU score by 17.81 compared with simple RBMT. Preordering+SMT with the improved grammar and reranking of parse trees improves BLEU score by 1.89 compared with simple SMT with DL 10. System combination improves BLEU score by 0.26 compared with preordering+SMT.

#### 4 Conclusion

System architecture, experimental settings and evaluation results of the EHR group in the WAT2015 tasks are described. Our system design concept is combining of rule-based method and statistical method and it gives the good effect to the translation accuracy. One of the future issues is to improve parsing accuracy both in RBMT part and preordering part.

#### Reference

- Pi-Chuan Chang, Michel Galley and Chris Manning. 2008. Optimizing Chinese Word Segmentation for Machine Translation Performance. *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 224-232.
- Terumasa Ehara. 2014. A machine translation system combining rule-based machine translation and statistical post-editing, *Proceedings of the 1st Workshop on Asian Translation (WAT2014)*, pages 50-54.
- Isao Goto, Ka Po Chow, Bin Lu, Eiichiro Sumita and Benjamin K. Tsou. 2013. Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop, *Proceedings of the 10th NTCIR Conference*, pages 260-286.
- Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada and Kevin Duh. 2010. Head finalization: A simple reordering rule for SOV languages. *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, pages 244-251.
- Japio (Japan Patent Information Organization). 2015. Investigation report of dictionary improvement and quality evaluation of machine translation for Chinese patent documents (in Japanese).
- Philipp Koehn, Franz J. Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation, *Proceedings of HLT-NAACL 2003*, pages 48-54.
- Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto and Makoto Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. *Proceedings of The International Workshop on Shareable Natural Language Resources*, pages 22-28.
- Toshiaki Nakazawa, Hideya Mino, Isao Goto, Graham Neubig, Sadao Kurohashi and Eiichiro Sumita. 2015. Overview of the 2nd Workshop on Asian Translation, *Proceedings of the 2nd Workshop on Asian Translation (WAT2015)*.
- Franz Josef Och, Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, volume 29, number 1, pages 19-51.
- Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation, *Proceedings of COLING-ACL 2006*, pages 433-440.
- Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and Christopher Manning. 2005. A Conditional Random Field Word Segmenter. In *Fourth SIGHAN Workshop on Chinese Language Processing*.