# **Overview of the 1st Workshop on Asian Translation**

Toshiaki Nakazawa

Hideya Mino

Japan Science and Technology Agency National Institute of Information and Isao Goto NHK

\_\_\_\_\_

Communications Technology

nakazawa@pa.jst.jp hideya.mino@nict.go.jp goto.i-es@nhk.or.jp

Sadao Kurohashi Kyoto University Eiichiro Sumita National Institute of Information and Communications Technology eiichiro.sumita@nict.go.jp

kuro@i.kyoto-u.ac.jp

### Abstract

This paper presents the results of the 1st workshop on Asian translation (WMT2014) shared tasks, which included  $J\leftrightarrow E$  translation subtasks and  $J\leftrightarrow C$  translation subtasks. As the first year of WAT, 12 institutions participated to the shared tasks. More than 300 translation results have been submitted to the automatic evaluation server, and selected submissions were manually evaluated.

# 1 Introduction

The Workshop on Asian Translation (WAT) is a new open evaluation campaign focusing on Asian languages. We would like to invite a broad range of participants and conduct various forms of machine translation experiments and evaluation. Collecting and sharing our knowledge will allow us to understand the essence of machine translation and the problems to be solved. We are working toward the practical use of machine translation among all Asian countries.

For the 1st WAT, we chose scientific papers as the targeted domain, and selected the languages Japanese, Chinese and English.

What makes WAT unique:

• Open innovation platform

The test data is fixed and open, so you can repeat evaluations on the same data and confirm changes in translation accuracy over time. WAT has no deadline for the automatic translation quality evaluation (continuous evaluation), so you can submit translation results at any time.

• Domain and language pairs WAT is the world's first workshop that uses

LangPair	Train	Dev	DevTest	Test
ASPEC-JE	3,008,500	1,790	1,784	1,812
ASPEC-JC	672,315	2,090	2,148	2,107

Table 1:	Statistics	of ASPEC.
----------	------------	-----------

scientific papers as a domain and Japanese-Chinese as a language pair. In the future, we will add more Asian languages, such as Korean, Vietnamese, Indonesian, Thai, Myanmar and so on.

• Evaluation method Evaluation will be done by both automatic and human evaluation. For human evaluation, WAT will use crowdsourcing, which is low cost and allows multiple evaluations.

# 2 Dataset

WAT uses Asian Scientific Paper Excerpt Corpus  $(ASPEC)^1$  as the dataset. ASPEC is constructed by the Japan Science and Technology Agency (JST) in collaboration with the National Institute of Information and Communications Technology (NICT). It consists of a Japanese-English scientific paper abstract corpus (ASPEC-JE), which is used for J $\leftrightarrow$ E subtasks, and a Japanese-Chinese scientific paper excerpt corpus (ASPEC-JC), which is used for J $\leftrightarrow$ C subtasks. The statistics of each corpus are described in Table1.

# 2.1 ASPEC-JE

The training data of ASPEC-JE was constructed by the NICT from approximately 2 million Japanese-English scientific paper abstracts owned by the JST. Because the paper abstracts are kind

Proceedings of the 1st Workshop on Asian Translation (WAT2014), pages 1–19, Tokyo, Japan, 4th October 2014. 2014 Copyright is held by the author(s).

1

<sup>&</sup>lt;sup>1</sup>http://lotus.kuee.kyoto-u.ac.jp/ASPEC/

of comparable corpora, the sentence correspondences are automatically found using the method of (Utiyama and Isahara, 2007). Each sentence pair is accompanied with the similarity score and the field symbol. The similarity scores are calculated by the method of (Utiyama and Isahara, 2007). The field symbols are single letters A-Z and show the scientific field of each document<sup>2</sup>. The correspondance between the symbols and field names, along with the frequency and occurance ratios for the training data, are given in the README of ASPEC-JE.

The development, development-test and test data were extracted from parallel sentences from Japanese-English paper abstracts owned by JST that are not contained in the training data. Each data set contains 400 documents. Furthermore, the data has been selected to contain the same relative field coverage across each data set. The document alignment was conducted automatically and only documents with a 1-to-1 alignment are included. It is therefore possible to restore the original documents. The format is the same as for the training data except that there is no similarity score.

# 2.2 ASPEC-JC

ASPEC-JC is a parallel corpus consisting of Japanese scientific papers from the literature database and electronic journal site J-STAGE of JST that have been translated to Chinese after receiving permission from the necessary academic associations. The parts selected were abstracts and paragraph units from the body text, as these contain the highest overall vocabulary coverage.

The development, development-test and test data are extracted at random from documents containing single paragraphs across the entire corpus. Each set contains 400 paragraphs (documents). Therefore there are no documents sharing the same data across the training, development, development-test and test sets.

# **3** Baseline Systems

Human evaluations were conducted as pairwise comparisons between the translation results for a specific baseline system and translation results for each participant's system. That is, the specific baseline system was the standard of human evaluation. A phrase-based statistical machine translation (SMT) system was adopted as the specific

<sup>2</sup>http://opac.jst.go.jp/bunrui/index.html

baseline system at WAT 2014.

In addition to the results for the baseline phrasebased SMT system, we produced results for the baseline systems that consisted of a hierarchical phrase-based SMT system, a string-to-tree syntaxbased SMT system, a tree-to-string syntax-based SMT system, five commercial rule-based machine translation (RBMT) systems, and two online translation systems. The SMT baseline systems consisted of publicly available software, and the procedures for building the systems and translating using the systems were published on the WAT 2014 web page<sup>3</sup>. We used Moses (Koehn et al., 2007; Hoang et al., 2009) as the implementation of the baseline SMT systems. The Berkeley parser (Petrov et al., 2006) was used to obtain syntactic annotations. The baseline systems are shown in Table 2.

The commercial RBMT systems and the online translation systems were operated by the organizers. We note that these RBMT companies and online translation companies did not submit themselves. Since our objective is not to compare commercial RBMT systems or online translation systems from companies that did not themselves participate, the system description of these systems are anonymized in this paper.

We describe the detail of the baseline SMT systems.

# 3.1 Data for Training

We used the following data for the training of the SMT baseline systems.

- Training data for the language model: All of the target language sentences in the parallel corpus.
- Training data for the translation model: Sentences that were 40 words or less in length. (For Japanese–English training data, we only used train-1.txt, which consisted of 1 million parallel sentence pairs with high similarity scores.)
- Development data for tuning: All of the development data.

### 3.2 Common Settings for Baseline SMT

We used the following tools for tokenization.

• Juman version 7.0<sup>4</sup> for Japanese segmentation.

<sup>&</sup>lt;sup>3</sup>http://lotus.kuee.kyoto-u.ac.jp/WAT/

<sup>&</sup>lt;sup>4</sup>http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN

System ID	System	Туре	JE	EJ	JC	CJ
SMT Phrase	Moses' Phrase-based SMT	SMT	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
SMT Hiero	Moses' Hierarchical Phrase-based SMT	SMT	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
SMT S2T	Moses' String-to-Tree Syntax-based SMT and Berkeley parser	SMT	$\checkmark$		$\checkmark$	
SMT T2S	Moses' Tree-to-String Syntax-based SMT and Berkeley parser	SMT		$\checkmark$		$\checkmark$
RBMT X	The Honyaku V15 (Commercial system)	RBMT	$\checkmark$	$\checkmark$		
RBMT X	ATLAS V14 (Commercial system)	RBMT	$\checkmark$	$\checkmark$		
RBMT X	PAT-Transer 2009 (Commercial system)	RBMT	$\checkmark$	$\checkmark$		
RBMT X	J-Beijing 7 (Commercial system)	RBMT			$\checkmark$	$\checkmark$
RBMT X	Hohrai 2011 (Commercial system)	RBMT			$\checkmark$	$\checkmark$
Online X	Google translate (July, 2014)	(SMT)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Online X	Bing translator (July, 2014)	(SMT)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

 Table 2: Baseline Systems

- Stanford Word Segmenter version 2014-01-04<sup>5</sup> (Chinese Penn Treebank (CTB) model) for Chinese segmentation.
- The Moses toolkit for English tokenization.

To obtain word alignments, GIZA++ and growdiag-final-and heuristics were used. We used 5gram language models with modified Kneser-Ney smoothing, which were built using a tool in the Moses toolkit (Heafield et al., 2013).

## 3.3 Phrase-based SMT

We used the following Moses' configuration for the phrase-based SMT system.

- distortion-limit = 20
- msd-bidirectional-fe lexicalized reordering
- Phrase score option: GoodTuring

The default values were used for the other system parameters.

## 3.4 Hierarchical Phrase-based SMT

We used the following Moses' configuration for the hierarchical phrase-based SMT system.

- max-chart-span = 1000
- Phrase score option: GoodTuring

The default values were used for the other system parameters.

### 3.5 String-to-Tree Syntax-based SMT

We used Berkeley parser to obtain target language syntax. We used the following Moses' configuration for the string-to-tree syntax-based SMT system.

- max-chart-span = 1000
- Phrase score option: GoodTuring

• Phrase extraction options: MaxSpan = 1000, MinHoleSource = 1, and NonTermConsec-Source.

The default values were used for the other system parameters.

# 3.6 Tree-to-String Syntax-based SMT

We used Berkeley parser to obtain source language syntax. We used the following Moses' configuration for the baseline tree-to-string syntax-based SMT system.

- max-chart-span = 1000
- Phrase score option: GoodTuring
- Phrase extraction options: MaxSpan = 1000, MinHoleSource = 1, MinWords = 0, NonTermConsecSource, and AllowOnlyUnalignedWords.

The default values were used for the other system parameters.

# 4 Automatic Evaluation

# 4.1 Procedure of Calculating Automatic Evaluation Score

We calculated automatic evaluation scores of the translation results applying two popular metrics: BLEU (Papineni et al., 2002) and RIBES (Isozaki et al., 2010). BLEU scores were calculated with *multi-bleu.perl* distributed with the Moses toolkit (Koehn et al., 2007). RIBES scores were calculated with *RIBES.py* version 1.02.4 <sup>6</sup>. All scores of each task were calculated using one reference. Before the calculation of the automatic evaluation scores, the translation results have been tokenized with word segmentation tools on each language.

<sup>&</sup>lt;sup>5</sup>http://nlp.stanford.edu/software/segmenter.shtml

<sup>&</sup>lt;sup>6</sup>http://www.kecl.ntt.co.jp/icl/lirg/ribes/index.html

刻 科学技術振興機構 NGG WARTER (1) 成 都大学	辰興機構 Appende	NICO		》 京都大 KYOTO UNIVE																	
								L	he 1st	Workshop Submi	WAT 2014 The 1st Workshop on Asian Translation Submission Site	slation	-								
										Octob Toky	October 4, 2014 Tokyo, Japan										
SUBMISSION	Z																				
Logged in as: test-user Logout	user																				
Submission: Human Evaluation: human ( Publish the results of the evaluation:	⊡humar € ≪publisl	<ul> <li>human evaluation</li> <li>publish</li> </ul>																			
Team Name: Task	test-user en-ia \$																				
l File Other	ファイル used th	ファイルを選択 ファイルを選択 used the other resources li	\未選択 rces like paı	allel corpus, n	nonolingua	ıl corpus, pı	<ul> <li>ファイルを選択 ファイル未選択</li> <li>ファイル未選択</li> <li>Used the other resources like parallel corpus, monolingual corpus, parallel dictionary, and</li> </ul>														
Resources: 4 Method:	so on in 8 SMT	so on in addition to ASPEC SMT ‡	SPEC		)																
System Description (disclosure):								100 characters or less													
System Description (non-disclosure):								100 characters or less													
Submit								য													
Notice for submission:	11: see should b	ai hohoono o		ţ																	
<ul> <li>Submission thes should be encoded in 1.1-3 horing a sentence in each line which assigned Translated sentences in submission files should be put with a sentence in each line which assigned Team Name, Task. Josel Other Resources, Method, System Description (disclosure), Joate and Ti If you want to submit the file for human evaluation, check the box of "Human Evaluation". Once.</li> <li>You can submit the file for human evaluation twice per each task.</li> <li>You can modify some information of submitted data. Read the "Notice for submitted data" below.</li> </ul>	tes snoud itences in s Task, Usec submit the vit the file f fy some in	be encoded in ubmission filk 1 Other Resou 7 file for huma or human eva formation of s	ULT-S TOT es should be inces, Metho in evaluation function twic ubmitted da	nat. put with a sen d, System Des 1, check the bo e per each task ita. Read the "1	ntence in ca scription (d x of "Hurr k. Notice for :	ach line wh lisclosure), nan Evalual submitted d	tich assigned to the co , Date and Time(JST) tion". Once you uplot data" below.	rresponding to BLEU and F id the file with	sst sentence. T KIBES will be i checking "Hi	he number of lines in the disclosed at Evluation S liman Evaluation S liman Evaluation " you c	<ul> <li>Promusion the should be moded in 11-1-5 format.</li> <li>Translated sentences in submission file abould be puil sentence in each line which assigned to the corresponding test file should be equal.</li> <li>Translated sentences in submission file should be puil sentence. The number of lines in the submission file and the corresponding test file should be equal.</li> <li>Translated sentences in submission file should be puil sentence in each line which assigned to the corresponding test file should be equal.</li> <li>Team Name, Task, Used Other Resources, Method, System Description (disclosure), Date and Time(JST), BLEU and RIBES will be disclosed at Evitation Site when you upload the file with checking "Publish the results of the evaluation".</li> <li>If you want to submit the file for human evaluation, check the box of "Human Evaluation". Once you upload the file with checking "Human Evaluation" you can't change the file for human evaluation.</li> <li>You can submit the file for human evaluation.</li> </ul>	ponding test i tith checking " valuation.	file should b "Publish the	e equal. results of the evi	aluation".						
Submitted Data:	of Submitte	Part of Date																			Back to top
-									ð	Syst	System Description			BLEU				RIBES			
nr Withdraw	Locked	Evaluation 1	Publish	Date/Time	Team	Task	Filename	Method	d Resources	(disclosure)	(non-disclosure)	mi	kyt me	mec mos s	std- std- ctb pku	mnį	kyt	mec mos	os std- ctb	std- pku	HUMAN

Figure 1: The submission web page for participants

For Japanese segmentation we use three different tools, which are Juman version 7.0 (Kurohashi et al., 1994), KyTea 0.4.6 (Neubig et al., 2011) with Full SVM model <sup>7</sup> and MeCab 0.996 (Kudo, 2005) with IPA dictionary 2.7.0 <sup>8</sup>. For Chinese segmentation we use two different tools, which are KyTea 0.4.6 with Full SVM Model in MSR model and Stanford Word Segmenter version 2014-06-16 with Chinese Penn Treebank (CTB) and Peking University (PKU) model <sup>9</sup> (Tseng, 2005). For English segmentation we use tokenizer.perl <sup>10</sup> in the Moses toolkit.

The detailed procedures for the automatic evaluation are shown at WAT2014 evaluation web page <sup>11</sup>.

### 4.2 Automatic Evaluation System

The participants submit the translation results via an automatic evaluation system deployed at WAT2014 web page, which give them automatic evaluation scores of the results they upload. Figure 1 shows the submitting interface for participants. The system requires the participants to provide the following information when they upload the translation results:

- Subtask  $(J \leftrightarrow E, J \leftrightarrow C)$
- Method (SMT, RBMT, SMT and RBMT, EBMT, Other)
- Existence of the use of other resources in addition to ASPEC
- Permission of publishing the automatic evaluation scores on WAT2014 web page

The server of the system keeps all submitted information including translation results or scores and participants can confirm the only information they uploaded. The information of translation results which the participant permits to publish is disclosed on the web page. In addition to submitting the translation results for automatic evaluation, participants submit the results for human evaluation with the same web interface. This automatic evaluation system will be available even after WAT2014. Everybody can use the system by registering on the registration web page  $^{12}$ .

## **5** Human Evaluation

#### 5.1 Using Crowdsourcing

As all the MT researchers know, the human evaluation costs a lot of time and money. One of the solutions to reduce them is using crowdsourcing. Other machine translation evaluation campaigns such as IWSLT (2011, 2012) and WMT (2012, 2013) also used crowdsourcing for the human evaluation. Recently, there are so many crowdsourcing services in the world: Amazon Mechanical Turk<sup>13</sup>, CrowdFlower<sup>14</sup>, Yahoo Crowdsourcing<sup>15</sup>, Lancers<sup>16</sup> and so on. Among these services, we used Lancers for the human evaluation of WAT2014.

There are two reasons of choosing Lancers. One is that we can set the category of the crowdsourcing task ('Translation' for this case). We can reach the appropriate workers by setting the appropriate categories. The other reason is that we can ask the task to the identity-verified workers. This function guarantee the quality of the workers. These two advantages can keep the evaluation quality at higher level.

### 5.2 Human Evaluation Method

For the human evaluation, we randomly chose *documents* from the Test set of ASPEC data, in total 400 sentence pairs for JE and JC. We excluded the documents which contains a sentence with longer than 100 Japanese characters. Each submission is compared with the baseline translation (Phrase-based SMT, described in Section 3) and given *HUMAN* score.

# 5.2.1 Pairwise Evaluation of Sentences

We conducted pairwise evaluation of each test sentence of the 400 sentences. The input sentence and two translations (the baseline and a submission) are shown to the workers, and the workers are asked to judge which translation is better than the other, or they are of the same quality. The order of the two translations are at random. Figure 2 shows the illustration of the evaluation.

<sup>&</sup>lt;sup>7</sup>http://www.phontron.com/kytea/model.html <sup>8</sup>http://code.google.com/p/mecab/downloads/detail?

name=mecab-ipadic-2.7.0-20070801.tar.gz

<sup>&</sup>lt;sup>9</sup>http://nlp.stanford.edu/software/segmenter.shtml <sup>10</sup>https://github.com/moses-smt/mosesdecoder/tree/

RELEASE-2.1.1/scripts/tokenizer/tokenizer.perl

<sup>&</sup>lt;sup>11</sup>http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

<sup>&</sup>lt;sup>12</sup>http://lotus.kuee.kyoto-u.ac.jp/WAT/registration/index.html
<sup>13</sup>https://www.mturk.com

<sup>&</sup>lt;sup>14</sup>http://www.crowdflower.com

<sup>&</sup>lt;sup>15</sup>http://crowdsourcing.yahoo.co.jp (Japanese service)

<sup>&</sup>lt;sup>16</sup>http://www.lancers.jp (Japanese service)

# 2つの機械翻訳結果の優劣判断

```
科学技術論文の英語入力文に対する日本語の機械翻訳結果が2つ表示されています。
どちらの翻訳がより正しいかを判断してください。
優劣がつけられない場合は、同程度としてください。
```

入力文: Details of dose rate of "Fugen Power Plant" can be calculated by using DERS software.
 翻訳文1: ふげん発電所」の線量率の詳細はDERSソフトウェアを用いて計算できる。
 翻訳文2: 「ふげん発電所の線量率の詳細を用いて計算することができる「DERsソフトウェアである。
 1つ目の翻訳の方が良い 2つ目の翻訳の方が良い 同程度

Figure 2: The illustration of the crowdsourcing evaluation. The workers are asked to judge which translation is better, or the same.

Worker 1	Α	Α	Α	A	A	Α	Tie	Tie	Tie	В
Worker 2	Α	Α	Α	Tie	Tie	В	Tie	Tie	В	В
Worker 3	Α	Tie	В	Tie	В	В	Tie	В	В	В
Decision	Α	А	Α	A	Tie	В	Tie	В	В	В

Table 3: The combinations of human judgements and the final decision of each sentence pairs from system A and B.

## 5.2.2 Voting

The crowdsourcing workers are not specialists, thus the quality of the judgements are not necessarily precise. To guarantee the quality of the evaluation, each sentence is evaluated by 3 different workers and the final decision is made by the voting of the judgements. Table 3 shows all the combinations of the worker judgements and the final decision.

# 5.2.3 HUMAN Score Calculation

Suppose W to be the number of wins compared to the baseline, L to be the number of *losses* and T to be the number of *ties*, the HUMAN score, which is the official human evaluation score of WAT2014, can be calculated by the following formula:

$$HUMAN = 100 \times \frac{W - L}{W + L + T}$$

From the definition, the HUMAN score ranges between -100 and 100.

#### 5.2.4 Confidence Interval Estimation

As there are several ways to estimate the confidence interval, we chose the bootstrap resampling (Koehn, 2004) to estimated 95% confidence interval. The procedure is as follows:

- 1. randomly select 300 sentences from the 400 human evaluation sentences, and calculate the HUMAN score on the selected sentences
- 2. iterate the previous step 1000 times and get 1000 HUMAN scores
- 3. sort the 1000 scores and estimate the 95% confidence interval by discarding top and bottom 25 scores

### 5.3 Cost of Evaluation

One big benefit of using crowdsourcing is that we can reduce the cost of evaluations. In WAT2014, one judgement costs 5 JPY. The evaluation of a submission requires 3 (judgements)  $\times$  400 (sentence pairs) = 1,200 judgements and it costs 5  $\times$  1,200 = 6,000 JPY. The time for the evaluation differs depending on the translation direction. On the average, one evaluation finished in a couple of days.

# 6 Participants List

Table 4 shows the list of participants to WAT2014. There are not only the Japanese organizations, but some organizations came from outside Japan. 12 teams submitted one or more translation results to the automatic evaluation server, and 11 teams submitted one or more translation results to the human evaluation.

# 7 Evaluation Results

In this section, the evaluation results of WAT2014 are reported from several perspectives. Parts of the results of both automatic and human evaluations are also accessible at WAT2014 website<sup>17</sup>.

## 7.1 Official Automatic Evaluation Results

Figure 3 shows the official automatic evaluation results of the representative submissions and baseline systems. The automatic evaluation results of all the submissions are shown in Section Appendix A.

# 7.2 Official Human Evaluation Results

#### **HUMAN scores**

Figure 4 shows the official human evaluation results. The error bars in the figures show the 95%confidence interval (see Section 5.2.4). Note that overlapping the error bars between two submissions does not necessarily mean that there is no significant difference. If an error bar crosses the x-axis (HUMAN score = 0), it means that there is no significant difference between the submission and the baseline (SMT Phrase).

From the results, the followings can be observed:

- The best SMT system achieved better quality than RBMT system.
- The translation quality of the widely used systems was Phrase-based SMT < Hierarchical PBSMT < Syntax-based SMT (S2T and T2S).
- Forest-to-String Syntax-based SMT system (Neubig, 2014) achieved the best quality for all the translation directions.

# Statistical Significance Testing between Submissions

Tables 5, 6, 7 and 8 show the results of statistical significance testings of JE, EJ, JC and CJ translations respectively where all the pairs of submissions are tested.  $\gg$ ,  $\gg$  and > mean that the system in the row is *better* than the system in the column by p < 0.01, 0.05, 0.1 respectively. The test-

ings are also done by the bootstrap resampling as follows:

- 1. randomly select 300 sentences from the 400 human evaluation sentences, and calculate the HUMAN scores on the selected sentences for both systems
- 2. iterate the previous step 1000 times and count the number of wins (W), losses (L) and ties (T)
- 3. calculate  $p = \frac{L}{W+L}$

### **Inter-annotator Agreement**

To assess the reliability of agreement between the crowdsourcing workers, we calculated the Fleiss' kappa (Fleiss and others, 1971) values. The results are shown in Table 9. We can see that the Kappa values are larger for  $X \rightarrow J$  translations than  $J \rightarrow X$  translations. This may be because we used Japanese crowdsourcing service for the evaluation and the majority of the crowdsourcing workers are Japanese. The MT evaluation of their mother tongue is much easier than the others in general.

# Case Study: Direct Comparison and Relative Comparison

Looking at evaluation results of WEBLIO-EJ1 1 and 2 submissions (see Table 12), the automatic and human evaluations are inconsistent: the WEBLIO-EJ1 2 is consistently better than WEBLIO-EJ1 1 in the automatic evaluation, however it is much worse in the human evaluation. According to the descriptions of the two submissions, the difference of the two is whether it uses the forest input or not. It is natural that using the forest input improves the translation quality, thus we conducted the human evaluation of WEBLIO-EJ1 2 compared to WEBLIO-EJ1 1, which means we used WEBLIO-EJ1 1 as the baseline for the human evaluation.

The HUMAN score was  $2.50 \pm 4.17$  which means there is no significant difference between the two, and this result is far from the results of the official results. Actually, taking the confidence intervals into consideration, this conclusion can be derived under some probability.

The Fleiss' kappa value was 0.528 and it is much higher than the other  $E \rightarrow J$  human evaluations. This may be because the outputs of the two systems are quite similar and it is very easy for the

<sup>&</sup>lt;sup>17</sup>http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

	non-removal	removal
JE BLEU	0.46489	0.95098
JE RIBES	0.78255	0.83691
EJ BLEU	0.41524	0.84418
EJ RIBES	0.75105	0.85730
JC BLEU	0.49240	0.07937
JC RIBES	0.38695	0.10198
CJ BLEU	0.78713	0.82592
CJ RIBES	0.70081	0.83209
	•	

Table 10: The changes of correlations  $(R^2)$  before and after removing RBMT and online systems.

workers to judge which translation is better. If two translations have both better and worse parts than the other, the workers would evaluate differently from person to person.

# 7.3 Correlation between Automatic and Human Evaluations

Figure 5 shows the correlations between automatic evaluation measures (BLEU/RIBES) and the HU-MAN score. It is well known that the automatic and human evaluations do not have good correlations for RBMT and online systems. Removing these systems from the graph changes the correlation values ( $R^2$ ) like in Table 10. The correlation becomes much better after removing the RBMT and online systems for all the translation directions other than J $\rightarrow$ C.

# 8 Submitted Data

The number of published automatic evaluation results of 12 teams exceeded 100 by the day of WAT2014 workshop and 37 translation results for human evaluation was submitted by 11 teams. We will organize the all submitted data for human evaluation and make it public.

# 9 Conclusion and Future Perspective

This paper summarized the WAT2014 machine translation evaluation campaing. We had 12 participants worldwide, and collected a large number of submissions which are useful to improve the current machine translation systems by analyzing the submissions and finding the issues.

For the next WAT workshop, we are planning to conduct context-aware MT evaluations. The test data of WAT is prepared using the paragraph as a unit, while almost all other evaluation campaigns use the sentence as a unit. Therefore, it is suitable to investigate the importance of the context for the translation. Also, we are very happy to include other languages if there are available resources.

# Appendix A Submissions

Tables 11, 12, 13 and 14 summarize all the submissions listed in the automatic evaluation server at the point of WAT2014 workshop (4th, October, 2014). The OTHER RESOURCES column shows the use of other resources such as parallel corpora, monolingual corpora and parallel dictionaries in addition to ASPEC.

Team ID	Organization	JE	EJ	JC	CJ
NAIST (Neubig, 2014)	Nara Institute of Science and Technology	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
EIWA (Ehara, 2014)	Yamanashi Eiwa College	$\checkmark$			$\checkmark$
Kyoto-U (Richardson et al., 2014)	Kyoto University	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
WEBLIO-EJ1 (Zhu, 2014)	Weblio, Inc.		$\checkmark$		
TMU (Ohwada et al., 2014)	Tokyo Metropolitan University	$\checkmark$			
BJTUNLP (Cai et al., 2014)	Beijing Jiaotong University			$\checkmark$	
NII (Hoshino et al., 2014)	National Institute of Informatics	$\checkmark$			
SAS_MT (Wang et al., 2014)	SAS Research and Development Co., Ltd		$\checkmark$		$\checkmark$
Sense (Tan and Bond, 2014)	Saarland University & Nanyang Technological University	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
NICT (Ding et al., 2014)	National Institute of Information and Communication Technology			$\checkmark$	
TOSHIBA (Sonoh et al., 2014)	Toshiba Corporation	$\checkmark$		$\checkmark$	
WASUIPS (Yang and Lepage, 2014)	Waseda University			√*	√*

Table 4: The list of participants which submitted translation results to WAT2014 and their participations to each subtasks. (\*Only submitted to automatic evaluations.)

	NAIST 2	Kyoto-U 1	IT S2T	TOSHIBA 1	RBMT D	EIWA	Kyoto-U 2	TOSHIBA 2	Online D	IT Hiero	Sense	[]	[2	IU 1	IU 2
	NA	Ky	SMT	TO	RB	EIV	Ky	TO	On	SMT	Ser	IIN	IIN	TMU	TMU
NAIST 1	-														
NAIST 2		$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Kyoto-U 1			-	-	-	-	-	>	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
SMT S2T				-	-	-	>	>	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
TOSHIBA 1					-	-	-	-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
RBMT D						-	-	-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
EIWA							-	-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Kyoto-U 2								-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
TOSHIBA 2									$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Online D										$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
SMT Hiero											$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Sense												$\gg$	$\gg$	$\gg$	$\gg$
NII 1													$\gg$	$\gg$	$\gg$
NII 2														-	-
TMU 1															-

Table 5: Statistical significance testing of JE results.

					5						
	NAIST 2	WEBLIO-EJ1	Online A	Kyoto-U 1	WEBLIO-EJ1	SMT T2S	Kyoto-U 2	SMT Hiero	SAS_MT	Sense	RBMT B
NAIST 1	≫										
NAIST 2		$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
WEBLIO-EJ1 1			-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Online A				>	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Kyoto-U 1					-	>	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
WEBLIO-EJ1 2						-	-	>	$\gg$	$\gg$	$\gg$
SMT T2S							-	-	$\gg$	$\gg$	$\gg$
Kyoto-U 2								-	$\gg$	$\gg$	$\gg$
SMT Hiero									>	$\gg$	$\gg$
SAS_MT										$\gg$	$\gg$
Sense											-

Table 6: Statistical significance testing of EJ results.

















Figure 3: The official automatic evaluation results.



Figure 4: The official human evaluation results.

	SMT S2T	Sense	NICT	SMT Hiero	NAIST 2	TOSHIBA 1	Kyoto-U 1	BJTUNLP	TOSHIBA 2	Kyoto-U 2	Online D	RBMT B
NAIST 1	-	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>	>>>>
SMT S2T		-	$\gg$	$\gg$	>>>>	$\gg$	$\gg$	$\gg$	$\gg$	>>>>	$\gg$	$\gg$
Sense			-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
NICT				-	>	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
SMT Hiero					-	-	-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
NAIST 2						-	-	>	$\gg$	$\gg$	$\gg$	$\gg$
TOSHIBA 1							-	>	$\gg$	$\gg$	$\gg$	$\gg$
Kyoto-U 1								-	>	$\gg$	$\gg$	$\gg$
BJTUNLP									-	>	$\gg$	$\gg$
TOSHIBA 2										-	$\gg$	$\gg$
Kyoto-U 2											$\gg$	$\gg$
Online D												>

Table 7: Statistical significance testing of JC results.

	NAIST 2	SAS_MT	SMT T2S	EIWA	Kyoto-U 1	Kyoto-U 2	SMT Hiero	Sense	Online A	RBMT A
NAIST 1										
NAIST 2		$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
SAS_MT			$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	>>>>
SMT T2S				-	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
EIWA					$\gg$	$\gg$	$\gg$	$\gg$	$\gg$	$\gg$
Kyoto-U 1						-	-	$\gg$	$\gg$	$\gg$
Kyoto-U 2							-	$\gg$	$\gg$	$\gg$
SMT Hiero								$\gg$	$\gg$	>>>>
Sense									$\gg$	>>>>
Online A										>>>>

Table 8: Statistical significance testing of CJ results.

JE							
System ID	Kappa			JC			
NAIST 1	0.162	EJ		System ID	Kappa	CJ	
NAIST 2	0.047	System ID	Kappa	NAIST 1	0.077	System ID	Kappa
SMT S2T	0.099	NAIST 1	0.280	SMT S2T	0.069	NAIST 1	0.168
Kyoto-U 1	0.070	NAIST 2	0.250	Sense	0.009	NAIST 1 NAIST 2	0.108
TOSHIBA 1	0.098	WEBLIO-EJ1 1	0.238	NICT	0.067	SAS_MT	0.203
RBMT D	0.075	Online A	0.219	SMT Hiero	0.202	SMT T2S	0.107
EIWA	0.083	Kyoto-U 1	0.216	NAIST 2	0.093	EIWA	0.230
Kyoto-U 2	0.139	WEBLIO-EJ1 2	0.240	TOSHIBA 1	0.095	Kyoto-U 1	0.175
TOSHIBA 2	0.078	SMT T2S	0.240	Kyoto-U 1	0.089	Kyoto-U 2	0.199
Online D	0.055	Kyoto-U 2	0.229	BJTUNLP	0.198	SMT Hiero	0.130
SMT Hiero	0.119	SMT Hiero	0.277	TOSHIBA 2	0.198	Sense	0.274
Sense	0.245	SAS_MT	0.248	Kyoto-U 2	0.163	Online A	0.220
NII 1	0.119	Sense	0.395	Online D	0.035	RBMT A	0.239
NII 2	0.086	RBMT B	0.217	RBMT B	0.083	ave.	0.130
TMU 1	0.091	ave.	0.254		0.003	avc.	0.200
TMU 2	0.136			ave.	0.101		
ave.	0.106						

Table 9: The Fleiss' kappa values of human evaluation results.



Figure 5: The correlations between BLEU/RIBES and HUMAN scores.

HUMAN SYSTEM DESCRIPTION	+7.75 [Hierarchical Phrase-based SMT	- Phrase-based SMT	+25.50 String-to-Tree SMT	+13.75 Online D	- RBMT E	- RBMT F	- Online C	+23.00 RBMT D	+40.50 [Travatar-based Forest-to-String SMT System with Extra Dictionaries	+37.50 [Travatar-based Forest-to-String SMT System	<ul> <li>Travatar-based Forest-to-String SMT System (Tuned BLEU+RIBES)</li> </ul>	+22.50 Combination of RBMT and SPE(statistical post editing)	<ul> <li>Our baseline system using 3M parallel sentences</li> </ul>	+21.25 Our new baseline system after several modifications	+25.00 Our new baseline system after several modifications + 20-best parses, KN7, RNNLM reranking	-17.25 Our baseline system with preordering method	-17.20 Our baseline system with another preordering method	- Our baseline system	-5.75 Our Baseline	-14.25 Our Baseline with Preordering	+1.25 Paraphrase max10	- Baseline SMT	- Context sensitive SMT	- SMT with lexicon	- SMT with lexicon X5	+20.25 RBMT system	+23.25 RBMT with SPE(Statistical Post Editing) system	
		0.645137 —			0.663851 -	0.661387 —	0.624827 —	0.683378 +2					0.689829 -		0.698953 +2							0.640393 —	0.641377 —	0.646133 —	0.637375 —	0.687122 +2	7936	
5								15.29										15.40					18.00			15.69	20.61	
OTHER RESOURCES BLEU	NO	NO	NO	YES	YES	YES	YES	YES	YES	NO	NO	YES	NO	NO	NO	ON	NO	NO	ON	NO	ON	NO	NO	NO	NO	YES	YES	
Q	SMT	SMT	SMT	Other	Other	Other	Other	Other	SMT	SMT	SMT	MT and RBMT	EBMT	EBMT	1 262 EBMT	SMT	SMT	SMT	SMT	SMT	SMT	SMT	SMT	SMT	SMT	RBMT	241 SMT and RBMT	
Ð	2	9	6			79					125	116 S	136	256	262	300	301	307	271	272	164	185	191	205	206	240	241 S	
SYSTEM ID	SMT Hiero	SMT Phrase	SMT S2T	Online D	<b>RBMT E</b>	<b>RBMT</b> F	Online C	RBMT D	NAIST 1	NAIST 2	NAIST 3	EIWA	Kyoto-U 3	Kyoto-U 2	FKyoto-U1	TMU 2	TMU 1	TMU 3	NII 1	NII 2	Sense 1	Sense 2	Sense 3	Sense 4		<b>TOSHIBA 2</b>	TOSHIBA 1	

Table 11: JE submissions

HIMAN SYSTEM DESCRIPTION		Phrase-based SMT	+34.25 Tree-to-String SMT	+42.50 Online A	+0.75 RBMT B	. RBMT A	· Online B	RBMT C	+31.50 Hierarchical Phrase-based SMT	+56.25 Travatar-based Forest-to-String SMT System	+51.50 Travatar-based Forest-to-String SMT System (Tuned	Our baseline system using 3M parallel sentences	Using n-best parses and RNNLM	+33.75 Our new baseline system after several modifications	+38.00 Our new baseline system after several modifications + 20-best	parses, KN7, RNNLM reranking	+43.25 Weblio Pre-reordering SMT System Baseline	+36.00 Weblio Pre-reordering SMT System (with forest inputs)	+27.50 Syntactic reordering Hierarchical SMT (using part of data)	- Paraphrase max10	+3.75 Baseline SMT	Context sensitive SMT	SMT with 20x lexicon	SMT with lexicon X5	
HIIM			+34.	+42.	Ŷ				+31.	+56.	+51.			+33.	+38.		+43.	+36.	+27.						
	mecab	0.695390	0.760516	0.725848	0.682683	0.678934	0.698126	0.676018	0.747722	0.806581	0.811081	0.756545	0.766495	0.766435	0.771545		0.792616	0.795027	0.771605	0.703139	0.703049	0.697257	0.693560	0.691011	
RIBES	kytea	0.691926	0.758031	0.723486	0.680748	0.676464	0.693390	0.672645	0.746978	0.801520	0.807010	0.755744	0.765251	0.764049	0.770908		0.786902	0.790066	0.770948	0.699334	0.700583	0.694678	0.689623	0.688801	
	juman	0.683735	0.748883	0.718019	0.671958	0.670167	0.687797	0.668372	0.734705	0.796079	0.801742	0.743969	0.755629	0.752058	0.766435		0.782066	0.785015	0.759415	0.690718	0.690464	0.684467	0.681194	0.679666	
	mecab	28.27	32.10	20.17	13.48	13.16	17.36	12.14	30.94	35.81	35.67	29.59	30.84	30.46	31.73		33.26	33.40	31.47	28.72	28.66	27.15	27.81	28.16	
BLEU	kytea	29.80	33.44	21.63	14.85	14.43	18.67	13.32	32.56	37.16	37.15	31.61	32.78	32.46	33.55		34.87	35.04	33.00	30.27	30.18	28.46	29.15	29.54	
	juman	27.48	31.05	19.66	13.18	12.86	17.04	12.19	30.19	35.03	34.84	28.93	30.25	29.76	31.09		32.53	32.69	30.47	27.88	27.92	26.59	27.00	27.33	
OTHER	RESOURCES	ON	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO		ON	NO	NO	NO	NO	NO	NO	NO	
METHOD		SMT	SMT	Other	Other	Other	Other	Other	SMT	SMT	SMT	 EBMT	EBMT	EBMT	EBMT		SMT	SMT	SMT	SMT	SMT	SMT	SMT	SMT	
Ē		5	12	34	99	68	91	95	367	118	126	134	186	253	267		132	202	264	163	184	190	265	274	
CVSTEM ID		SMT Phrase	SMT T2S	Online A	RBMT B	<b>RBMT A</b>	Online B	RBMT C	SMT Hiero	NAIST 1	NAIST 2	Kyoto-U 3	Kyoto-U 4	Kyoto-U 2	Kyoto-U 1		WEBLIO-EJ1	WEBLIO-EJ12	SAS_MT	Sense 2	Sense 1	Sense 3	Sense 4	Sense 5	

Table 12: EJ submissions

SYSTEM ID	A	METHOD	OTHER		BLEU	•		RIBES		HUMAN SYSTE	HUMAN SYSTEM DESCRIPTION
			RESOURCES	kytea	stanford (ctb)	stanford (pku)	kytea	stanford (ctb)	stanford (pku)	2	
SMT Hiero	m	SMT	ON	27.71	27.70	27.35	0.809128	0.809561	0.811394	+3.75 Hierarc	+3.75 Hierarchical Phrase-based SMT
SMT Phrase	7	SMT		27.96	28.01	27.68	0.788961	0.790263	0.790937	Phrase-	Phrase-based SMT
SMT S2T	10	SMT		28.65	28.65	28.35	0.807606	0.809457	0.808417	+14.00 String-to-Tree SMT	to-Tree SMT
Online D	37	Other		9.37	8.93	8.84	0.606905	0.606328	0.604149	-14.50 Online D	D
Online C	216	Other		7.26	7.01	6.72	0.612808	0.613075	0.611563	— Online C	C
RBMT B	243	RBMT		17.86	17.75	17.49	0.744818	0.745885	0.743794	-20.00 RBMT B	B
RBMT C	244	RBMT	ON	9.62	9.96	9.59	0.642278	0.648758	0.645385	RBMT C	C
NAIST 1	122	SMT	ON	30.53	30.46	30.25	0.818040	0.819406	0.819492	+17.75 Travata	+17.75 Travatar-based Forest-to-String SMT System
NAIST 2	123	SMT	NO	29.83	29.77	29.54	0.829627	0.830839	0.830529	+1.25 Travatar-based	
										BLEU+	BLEU+RIBES)
Kyoto-U 3	18	EBMT	NO	26.69	26.48	26.30	0.796402	0.798084	0.798383	Our bas	Our baseline system
Kyoto-U 1	257	EBMT	NO	27.21	27.02	26.83	0.791270	0.792166	0.790743	-0.75 Our nev	Our new baseline system after several modifications
Kyoto-U 2	259	EBMT	NO	27.67	27.44	27.34	0.788321	0.789069	0.788206	-8.75 Our nev	-8.75 Our new baseline system after several modifications + 20-best
										parses,	parses, KN7, RNNLM reranking
BJTUNLP	224	SMT	ON	24.12	23.76	23.55	0.794834	0.796186	0.793054	-3.75 SMT	
Sense 2	175	SMT	ON	27.92	28.03	27.67	0.793876	0.796589	0.797332	SMT	
Sense 1	201	SMT	NO	23.09	22.94	23.04	0.779495	0.779502	0.780262	+10.00 Character based SMT	ter based SMT
NICT	260	SMT	ON	27.98	28.18	27.84	0.806070	0.808684	0.807809	+6.50 Pre-reo	+6.50 Pre-reordering for phrase-based SMT (dependency parsing +
1										manual rules)	rules)
<sup>op</sup> TOSHIBA 2	236	RBMT	YES	19.28	18.93	18.82	0.764491	0.765346	0.763931	-5.25 RBMT system	system
TOSHIBA 1	238	238 SMT and RBMT	YES	27.42	26.82	26.79	0.804444	0.803302	0.803980	+0.75 RBMT	RBMT with SPE(Statistical Post Editing) system
WASUIPS 1	371	SMT	ON	22.71	22.49	22.39	0.776323	0.777615	0.777327	Our bas	Our baseline system (segmentation tools: urheen and mecab.
										moses: 1.0).	1.0).
WASHIPS 2	373	SMT	YES	24 70	24 25	24.28	0.790030	0 790460	0 790898	Our ha	Our baseline system + additional quasi-parallel cormus (se $\sigma$ -
		TIME		0	07-E7	07.17	00000000	0010710	0,000,000	mentati	Jui vascuito system + autoronai quasi-paranei vorpus (seg- mentation tools: urheen and mecab, moses: 1.0).
WASHIPS 3	376	SMT	ON	25 44	25 04	24 98	0 794244	0 793945	0 794823	Our ha	Dur haseline system (seomentation tools: urheen and mecah
										moses:	moses: 2.1.1).
WASUIPS 4	377	SMT	YES	25.60	25.10	25.07	0.794716	0.795786	0.795594	Our bas	Our baseline system + additional quasi-parallel corpus segmen-
										tation to	ation tools: urheen and mecab, moses: 2.1.1).
WASUIPS 5		SMT	NO	22.01	21.81	21.61	0.767418	0.767414	0.766092	Our bas	Dur baseline system (segmentation tools: kytea, moses: 1.0).
WASUIPS 6	382	SMT	YES	22.20	22.02	21.91	0.771952	0.773341	0.772107	Our ba	Dur baseline system + additional quasi-parallel corpus (seg-
										mentati	nentation tools: kytea, moses: 1.0).
WASUIPS 7	385	SMT	NO	25.45	25.10	25.01	0.793819	0.793308	0.793029	Our bas	Dur baseline system (segmentation tools: kytea, moses: 2.1.1).
WASUIPS 8	386	SMT	YES	25.68	25.01	25.11	0.795721	0.795504	0.795129	Our ba	Jur baseline system + additional quasi-parallel corpus (seg-
										mentati	mentation tools: kytea, moses: 2.1.1).
WASUIPS 9	389	SMT	NO	25.08	24.81	24.64	0.790498	0.791430	0.790142	Our bas	Jur baseline system (segmentation tools: stanford-ctb and ju-
										man, m	man, moses: 2.1.1).
WASUIPS 10 390	390	SMT	YES	25.63	25.30	25.18	0.794646	0.795307	0.794024	— Our ba	Our baseline system + additional quasi-parallel corpus (seg-
					_					mentati	mentation tools: stanford-ctb and juman, moses: 2.1.1).
						E		•			
						, L	hle 13. IC	Table 13. IC submissions	2010		

Table 13: JC submissions

D         MELHUU         RESOLRCIES jiman         Wrea         mecab         Junan         Kytea         Mathio         Junan         Kytea         Junan         Junan         Kytea         Junan         Junan <th <="" th=""><th></th><th>E</th><th>ACTION</th><th>OTHER</th><th></th><th>BLEU</th><th></th><th></th><th>RIBES</th><th></th><th></th></th>	<th></th> <th>E</th> <th>ACTION</th> <th>OTHER</th> <th></th> <th>BLEU</th> <th></th> <th></th> <th>RIBES</th> <th></th> <th></th>		E	ACTION	OTHER		BLEU			RIBES		
<ul> <li>ero</li> <li>4 SMT</li> <li>NO</li> <li>55.4</li> <li>0.81046</li> <li>0.77349</li> <li>0.77349</li> <li>0.80755</li> <li>4.7</li> <li>0.80755</li> <li>3.80</li> <li>3.81</li> <li>3.81</li> <li>3.80</li> <li>3.81</li> <li>3.82</li> <li>3.84</li> <li>3.83</li> <li>3.845</li> <li>3.845&lt;</li></ul>		A	METHOD	RESOURCES		kytea	mecab	juman				
Interview         8         NMT         NO         34.65         35.16         34.77         0.77292         0.55344         0.771005         1.1           3         3         0.0her         YES         10.47         0.5532         0.360706         0.5533139         0.5553         1.1         0.55323         0.55323139         0.55035         1.1         0.55323         0.5607706         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.500750         0.70136         0.70136         0.70136         0.70136         0.70136 <t< td=""><td>Hiero</td><td>4</td><td>SMT</td><td>ON</td><td>35.43</td><td>35.91</td><td>35.64</td><td>0.810406</td><td></td><td>0.807665</td><td>+4.75 Hierarchical Phrase-based SMT</td></t<>	Hiero	4	SMT	ON	35.43	35.91	35.64	0.810406		0.807665	+4.75 Hierarchical Phrase-based SMT	
<ul> <li>35 [13 SMT NO 36.52 37.07 36.64 0.85592 0.580176 0.503765</li> <li>4 220 RBMT NO 8.39 870 0.666277 0.503755 0.580176 0.503756</li> <li>7 220 RBMT NO 8.39 8.70 8.30 0.664179 0.6661730 0.563139</li> <li>1 12 20 SMT NO 40.21 41.29 40.35 0.566176 0.660776 0.660776</li> <li>1 24 SMT NO 40.21 41.29 40.36 0.641183 0.565402 0.653139</li> <li>1 21 20 SMT NO 40.21 41.29 40.36 0.845486 0.838092 0.845655 + 4</li> <li>1 31 28 SMT and RBMT YES 13.55 9 33.74 33.87 0.811350 0.840566 0.800566 0.800566</li> <li>1 31 31 35 RBMT NO 40.21 41.22 40.35 33.37 0.811350 0.845486 0.787105 0.7732466 - 7</li> <li>1 31 31 35 RBMT NO 33.35 9 33.74 33.87 0.811350 0.800566 0.800566 - 0.7732466</li> <li>1 2 228 EBMT NO 34.75 35.89 34.43 0.82259 0.787105 0.778146 - 7</li> <li>1 2 228 EBMT NO 34.75 35.89 34.43 0.820529 0.787105 0.777147 - 7</li> <li>1 2 228 SMT NO 34.75 35.89 34.64 0.771197 0.78310 0.800356 - 7</li> <li>1 2 282 SMT NO 34.75 35.89 34.64 0.771197 0.78349 0.7771046 - 7</li> <li>1 2 282 SMT NO 34.56 35.08 34.64 0.771197 0.778446 - 7</li> <li>1 1 2 28 SMT NO 34.56 35.08 34.64 0.771197 0.778446 - 7</li> <li>1 1 2 28 SMT NO 34.56 35.08 34.64 0.771197 0.778446 - 7</li> <li>1 1 2 68 33.74 33.74 33.74 33.74 33.74 33.74 33.74 33.07293468 - 7</li> <li>1 2 2 3 SMT NO 34.55 35.08 34.64 0.771197 0.778446 - 7</li> <li>1 1 2 2 8 SMT NO 34.56 35.08 34.64 0.771197 0.778447 - 7</li> <li>1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</li></ul>	Phrase	×	SMT	ON	34.65	35.16	34.77	0.772498		0.771005	Phrase-based SMT	
A         36         Other         YES         11.63         13.21         11.87         0.559925         0.598172         0.598773         1.           3         215         0ther         YES         10.48         11.26         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.607760         0.633119         -         0.72386         -         0.673866         -         0.723866         -         0.723860         0.778862         0.778181 <td></td> <td>13</td> <td>SMT</td> <td>ON</td> <td>36.52</td> <td>37.07</td> <td>36.64</td> <td>0.825292</td> <td></td> <td>0.825025</td> <td>+16.00 Tree-to-String SMT</td>		13	SMT	ON	36.52	37.07	36.64	0.825292		0.825025	+16.00 Tree-to-String SMT	
3         215         Other         YES         10.48         11.26         10.47         0.660773         0.552400         0.660779         -           D         229         RBMT         NO         9.37         9.87         9.35         0.666773         0.552400         0.661739         -           D         229         RBMT         NO         40.11         41.29         9.35         0.41389         0.552400         0.661739         -           1         17         RBMT         NO         40.11         41.29         40.30         0.842477         0.8438092         0.847235         +           1         177         SBMT         NO         40.11         41.29         40.35         0.84545         0.833892         0.847255         +         +           1         173         EBMT         NO         33.57         33.33         33.45         0.741080         0.78039         0.807506         0.780393         0.780312         -         +		36	Other	YES	11.63	13.21	11.87	0.595925	0.598172	0.598573	-21.75 Online A	
A         239         RBMT         NO         9.37         9.87         9.35         0.666277         0.553402         0.661730         1.30           1         120         SMIT         NO         8.30         0.84448         0.853319         0.833319         1.4           2         124         SMIT         NO         40.11         41.28         0.845486         0.8338092         0.653319         1.5           137         RBMT         VES         18.69         18.33         3.3.74         0.845486         0.833694         4.4           137         RBMT         VES         33.26         33.3.74         33.3.75         0.811350         0.803564         4.4           137         RBMT         NO         33.26         33.3.74         33.3.70         0.8133916         0.780356         0.730386         4.45652         4.4           13         RBMT         NO         33.26         33.3.74         33.3.74         0.813695         0.730381         0.730386         4.76538         0.730381         0.730386         4.76538         0.730381         0.730386         4.76538         0.770386         0.770386         0.770386         0.770386         0.770386         0.770386         0.770		215	Other	YES	10.48	11.26	10.47	0.600733		0.600706	Online B	
D         242         RBMT         NO         8.39         8.70         8.30         0.641189         0.626400         0.633319         -           1         120         SMIT         NO         40.11         41.29         40.30         0.843486         0.838922         0.843525         +           1         120         SMIT         NO         40.21         40.82         0.841486         0.838092         0.845255         +           137         RBMT         YES         18.69         18.33         18.32         0.841186         0.732666         0.8080504         +           138         BMIT         NO         33.57         33.345         0.8111350         0.800596         0.808593         -           138         BMIT         NO         33.57         33.33         33.345         0.731656         0.736532         -           14         135         BMIT         NO         33.57         33.345         0.800599         0.780531         0.802539         0.780531         0.802539         0.780531         0.802539         0.780531         0.802539         0.780531         0.802539         0.780531         0.802539         0.780531         0.802539         0.781738         0.		239	RBMT	NO	9.37	9.87	9.35	0.666277		0.661730	-37.75 RBMT A	
I         120         SMT         NO         40.11         41.29         40.30         0.842477         0.8434824         0.842235         +           2         124         SMT         NO         40.11         41.29         40.15         0.845486         0.8338992         0.845605         +         +           137         RBMT         YES         33.36         0.311560         0.800566         0.800566         0.300569         -         -         -         +         +         +         +         +         +         +         +         +         +         -		242	RBMT	NO	8.39	8.70	8.30	0.641189		0.633319	RBMT D	
2         124         SMT         NO         40.21         40.82         40.15         0.845486         0.836902         0.845625         +           137         RBMT         YES         35.33         33.74         33.87         0.811350         0.800506         0.808504         +         +           137         RBMT         YES         33.53         33.74         33.87         0.811350         0.800506         0.808504         +         +           138         RBMT         NO         33.26         35.09         33.67         0.811350         0.800506         0.808504         +         +           13         EBMT         NO         33.26         35.39         33.45         0.811350         0.807535         0.817308         0.817368         +         +           11         268         EBMT         NO         35.58         34.45         0.771975         0.76470         0.771081         -         -           17         268         EBMT         NO         35.45         35.34         0.807535         0.817308         0.817368         +         -           17         268         SMT         NO         37.42         37.07         0.81739	ST 1	120	SMT	ON	40.11	41.29	40.30	0.842477		0.842235	-	
137         RBMT         YES         18.69         18.33         18.32         0.740183         0.720281         0.732466         +           13         I38 SNT and RBMT         YES         33.53         33.74         33.87         0.801566         0.803504         +         +           13         EBMT         NO         33.57         33.74         33.87         0.811350         0.800566         0.803504         +         +           12         258         EBMT         NO         33.57         35.43         33.457         0.811350         0.800596         0.800390         0.785310         576633         -         -         776352         -         -         776352         -         -         786336         -         786336         -         786336         -         786336         -         7736352         -         -         773425         0.781368         -         -         773438         0.7864012         -         773648         -         -         773448         0.770346         -         773448         -         773445         -         -         773442         0.771475         -         17475         -         17475         -         7734423         0.764	ST 2	124	SMT	NO	40.21	40.82	40.15	0.845486		0.845625	Travatar-based Forest-to-String	
137         RBMT         YES         18.69         18.33         0.740183         0.770281         0.779264         -           13         138         SMT and RBMT         YES         33.50         33.57         33.57         0.791560         0.7761761         0.771081         0.770181         0.770181         0.770181         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761761         0.7761776					4	0	0				BLEU+RIBES)	
138 SMT and RBMT         YES         33.57         33.74         33.87         0.800506         0.817368         1           1         2         2         3         3         3         3         3         3         3         3         3         3         3         3         3         3         3 <td>A 2</td> <td>137</td> <td>RBMT</td> <td></td> <td>18.69</td> <td>18.33</td> <td>18.32</td> <td>0.740183</td> <td></td> <td>0.732466</td> <td>— RBMT plus user dictionary</td>	A 2	137	RBMT		18.69	18.33	18.32	0.740183		0.732466	— RBMT plus user dictionary	
73         133         EBMT         NO         33.26         35.09         33.62         0.73105         0.787105         0.787105         0.78105         0.783016         0.783015         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.781032         0.771081         0.7710812         0.7710812         0.7710812         0.7710812         0.781032         <	A 1	138,	SMT and RBMT		33.53	33.74	33.87	0.811350		0.808504	+15.00 RBMT with user dictionary plus SPE(statistical post editing)	
14         135         EBMT         NO         32.68         33.30         32.45         0.7853016         0.7853016         0.7853016         0.7853016         0.786332         1           12         2288         EBMT         NO         33.57         34.43         33.45         0.800449         0.7953016         0.786335         0.817368         -           17         26.8         EBMT         NO         35.57         34.43         35.45         37.07         0.825551         0.817368         -           17         26.3         SMT         NO         35.56         37.07         0.825551         0.817368         +           174         SMT         NO         34.56         35.08         33.464         0.771975         0.774338         0.774303         0.774303         0.774303         0.7743012         -           200         SMT         NO         37.46         28.09         28.0027         0.789495         0.770846         -         -         0.774338         0.7743012         0.7844012         -         -         0.774303         0.774303         0.774303         0.774302         0.7844022         -         -         -         -         -         -         5.814	e	133		ON	33.26	35.09	33.62	0.791680		0.791269	— Using n-best parses and RNNLM	
12         258         EBMT         NO         33.57         34.43         33.45         0.800949         0.75539         0.800986           11         268         EBMT         NO         34.75         35.89         34.43         33.45         0.800535         0.800986         1           12         253         SMT         NO         36.58         36.22         36.10         0.822180         0.807535         0.817368         1           17         263         SMT         NO         34.56         33.86         34.64         0.771975         0.770846         1           200         SMT         NO         33.66         33.86         33.46         0.779183         0.771342         0.771348         1734012           PS1         369         SMT         NO         33.66         33.86         33.46         0.779183         0.776446         1           PS1         3092         SMT         NO         21.65         28.09         28.200         0.779183         0.771433         0.771434         0.771475         1           PS1         SMT         YES         30.44         30.92         28.25         32.54         0.779433         0.7774340         0.7	4	135		NO	32.68	33.30	32.45	0.786229	0.783016	0.786352	Our baseline system	
J 1         268         EBMT         NO         34.75         35.89         34.83         0.802629         0.798631         0.802535         0.807548         +         +           T 1         200         33.66         33.86         33.46         0.7711975         0.774338         0.7744012         -         784012         -         784012         -         784012         -         -         784012         -         784012         -         -         784012         -         -         784012         -         784012         -         -         784012         -         784012         -         784012         -         784012         0.778476         0.784422		258		NO	33.57	34.43	33.45	0.800949	0.795390	0.800986	+6.00 Our new baseline system after several modifications	
T         232         SMT         NO         36.53         36.10         887355         0.807535         0.817368         4-22.50           T1         263         SMT         NO         37.65         37.07         0.824170         0.825551         0.833048         4-22.50           T1         263         SMT         NO         37.65         37.07         0.834170         0.825551         0.833048         4-22.50           T14         SMT         NO         37.65         33.86         33.46         0.779183         0.771081         -1.00           PS 1         369         SMT         NO         33.66         33.86         33.46         0.779183         0.774338         0.784012         -1.00           PS 2         370         SMT         NO         27.66         28.09         28.20         0.779183         0.774338         0.784012         -1.00           PS 3         374         30.92         30.86         0.779183         0.776446            PS 3         374         SMT         NO         31.87         32.26         0.779142         0.781475            PS 3         375         SMT         YES         32.19	1	268		NO	34.75	35.89	34.83	0.802629	0.798631	0.802930	+7.50 Our new baseline system after several modifications + 20-best	
T2         232         36.10         0.822180         0.807535         0.817368            T1         263         SMT         NO         37.42         37.65         37.07         0.834170         0.837355         0.817368         +22.50           T1         263         SMT         NO         37.45         37.65         37.07         0.834170         0.87535         0.817368         +22.50           PS1         309         SMT         NO         34.66         0.779183         0.774338         0.7740181         -1.00           PS1         369         SMT         NO         27.66         28.09         28.20         0.779183         0.774338         0.7844012         -1.00           PS1         369         SMT         NO         31.87         30.92         38.26         0.794303         0.776349         0.770846            PS 3         374         SMT         YES         32.26         32.324         0.779338         0.776422            PS 3         374         SMT         YES         32.26         32.264         0.774423         0.777876         0.786422            PS 3         SMT         YES											parses, KN7, RNNLM reranking	
TI         Z63         SMT         NO         37.42         37.65         37.07         0.834170         0.825551         0.833048         +22.50           PS1         500         SMT         NO         34.56         35.08         34.64         0.771975         0.771081            PS1         369         SMT         NO         34.56         33.86         33.46         0.779183         0.771081            PS1         369         SMT         NO         23.66         33.86         33.46         0.779183         0.774338         0.7784012         -1.00           PS1         369         SMT         NO         21.66         28.09         28.20         0.779183         0.763494         0.770846            PS3         374         SMT         NO         31.87         32.26         32.26         0.774423         0.77876         0.78422            PS4         379         SMT         YES         32.19         32.25         32.24         0.774423         0.778766         0.767073            PS4         380         SMT         YES         27.43         0.774423         0.767073         0.767073 <td< td=""><td>~</td><td>232</td><td>SMT</td><td>NO</td><td>36.58</td><td>36.22</td><td>36.10</td><td>0.822180</td><td></td><td>0.817368</td><td><ul> <li>— Syntactic reordering phrase-based SMT (SAS token tool)</li> </ul></td></td<>	~	232	SMT	NO	36.58	36.22	36.10	0.822180		0.817368	<ul> <li>— Syntactic reordering phrase-based SMT (SAS token tool)</li> </ul>	
174         SMT         NO         34.56         35.08         34.64         0.771975         0.766470         0.771081            PS I         369         SMT         NO         33.46         33.346         0.779183         0.77038         0.771081            PS I         369         SMT         NO         33.466         33.346         0.779183         0.774338         0.774046            PS 2         370         SMT         YES         30.44         30.92         38.08         0.779183         0.773142         0.784012         -1.00           PS 3         374         SMT         YES         30.44         30.92         38.08         0.794303         0.773142         0.786422            PS 4         375         SMT         YES         32.26         32.54         0.79338         0.773749         0.767703            PS 4         370         SMT         YES         27.86         28.89         28.00         0.774423         0.767703            PS 5         380         SMT         YES         27.86         28.28         0.774423         0.769409         0.767073	_	263	SMT	NO	37.42	37.65	37.07	0.834170		0.833048	+22.50 Syntactic reordering Hierarchical SMT (using SAS token tool)	
200         SMT         NO         33.66         33.36         33.46         0.789495         0.774338         0.784012         -1.00           370         SMT         NO         27.66         28.09         28.20         0.779183         0.774338         0.78446            370         SMT         NO         27.66         28.09         28.20         0.779183         0.773142         0.781475            374         SMT         YES         30.44         30.92         38.26         0.794303         0.773142         0.781475            375         SMT         NO         31.87         32.26         32.54         0.794303         0.77876         0.786422            379         SMT         YES         32.19         32.55         32.54         0.795338         0.77876         0.786422            370         SMT         YES         32.19         32.54         0.795338         0.778750         0.786423            383         SMT         YES         27.86         28.89         27.43         0.776550         0.769409            384         SMT         YES         32.43		174		NO	34.56	35.08	34.64	0.771975		0.771081	SMT	
369       SMT       NO       27.66       28.09       28.20       0.779183       0.762949       0.770846          370       SMT       YES       30.44       30.92       30.86       0.789824       0.773142       0.781475          374       SMT       YES       30.44       30.92       30.86       0.789824       0.773142       0.781475          375       SMT       NO       31.87       32.26       32.24       0.794303       0.777876       0.786422          379       SMT       YES       32.19       32.55       32.54       0.795838       0.778776       0.787791          379       SMT       YES       27.43       0.774423       0.767073          380       SMT       NO       27.37       28.28       28.00       0.776550       0.767073          383       SMT       NO       32.08       23.43       0.776550       0.775168       0.767073          384       SMT       YES       32.43       0.796220       0.778075       0.780465          387       SMT       NO       32.55       32.47       0.79		200		NO	33.66	33.86	33.46	0.789495		0.784012	-1.00 Character based SMT	
370       SMT       YES       30.44       30.92       30.86       0.789824       0.773142       0.781475        0         374       SMT       NO       31.87       32.26       32.26       0.794303       0.777876       0.786422        0         375       SMT       YES       32.19       32.55       32.54       0.794303       0.777876       0.786422        0         379       SMT       YES       32.19       32.55       32.54       0.795838       0.77037       0.787591        0         379       SMT       YES       27.43       0.774423       0.776773       0.767073        0         383       SMT       YES       27.43       0.774423       0.775570       0.776721       0.767073        0         383       SMT       YES       32.08       33.09       32.18       0.795220       0.776721       0.767073        0         383       SMT       YES       32.43       0.796220       0.778075       0.778065        0         384       SMT       NO       32.52       32.69       32.47       0.796520 <td< td=""><td>UIPS 1</td><td>369</td><td>SMT</td><td>ON</td><td>27.66</td><td>28.09</td><td>28.20</td><td>0.779183</td><td></td><td>0.770846</td><td>— Our baseline system (segmentation tools: urheen and mecab,</td></td<>	UIPS 1	369	SMT	ON	27.66	28.09	28.20	0.779183		0.770846	— Our baseline system (segmentation tools: urheen and mecab,	
370       SMT       YES       30.44       30.92       30.86       0.789824       0.7713142       0.781475          374       SMT       NO       31.87       32.26       32.26       0.794303       0.777876       0.786422        0         375       SMT       YES       32.19       32.55       32.54       0.795838       0.777876       0.786422        0         379       SMT       YES       32.19       32.55       32.54       0.795838       0.776703       0.787591        0         379       SMT       YES       27.43       0.774423       0.756721       0.767073        0         383       SMT       YES       27.43       0.776550       0.7756721       0.767073        0         384       SMT       YES       32.08       32.18       0.795230       0.778075       0.787665        0         387       SMT       YES       32.43       0.796220       0.778075       0.790107        0         388       SMT       NO       32.52       32.81       32.59       0.796579       0.790107        0 <td></td> <td>moses: 1.0).</td>											moses: 1.0).	
374       SMT       NO       31.87       32.26       32.26       0.794303       0.777876       0.786422          375       SMT       YES       32.19       32.55       32.54       0.795838       0.7780627       0.786422          379       SMT       YES       32.19       32.55       32.54       0.795838       0.780027       0.787591          379       SMT       NO       27.37       28.28       28.00       0.776550       0.756721       0.767073          383       SMT       YES       27.86       28.89       28.00       0.776550       0.7767073           383       SMT       YES       27.43       0.776550       0.776721       0.769409           384       SMT       NO       32.08       33.09       32.18       0.795220       0.778075       0.789657           387       SMT       NO       32.52       32.69       32.47       0.796059       0.789657            388       SMT       NO       32.52       32.81       32.59       0.796059       0.790107	UIPS 2	370	SMT	YES	30.44	30.92	30.86	0.789824		0.781475	<ul> <li>Our baseline system + additional quasi-parallel corpus (segmentation tools: urbeen and mercal moses: 1.0)</li> </ul>	
375       SMT       YES       32.19       32.55       32.54       0.795838       0.780027       0.787591        0         379       SMT       NO       27.37       28.28       27.43       0.774423       0.767073        0         380       SMT       YES       27.86       28.28       27.43       0.774423       0.767073        0         381       SMT       YES       27.86       28.89       28.00       0.776550       0.7756721       0.769409        0         383       SMT       NO       32.08       33.09       32.18       0.793230       0.778075       0.789657        0         384       SMT       YES       32.43       33.36       32.44       0.796220       0.778075       0.789657        0         387       SMT       NO       32.52       32.69       32.47       0.796059       0.780402       0.790107        0         388       SMT       YES       32.65       32.81       32.59       0.796777       0.791219        0        0        0        0        0 <td>UIPS 3</td> <td>374</td> <td>SMT</td> <td>ON</td> <td>31.87</td> <td>32.26</td> <td>32.26</td> <td>0.794303</td> <td></td> <td>0.786422</td> <td>Our baseline system (segmentation tools: urheen and mecab,</td>	UIPS 3	374	SMT	ON	31.87	32.26	32.26	0.794303		0.786422	Our baseline system (segmentation tools: urheen and mecab,	
375       SMT       YES       32.19       32.55       32.54       0.795838       0.778027       0.787591        0         379       SMT       YES       32.19       32.55       32.54       0.75571       0.787591        0         379       SMT       YES       27.37       28.28       27.43       0.774423       0.756721       0.767073        0         380       SMT       YES       27.86       28.89       28.00       0.776550       0.756721       0.769409        0         383       SMT       YES       32.08       33.09       32.18       0.795230       0.778075       0.787665        0         384       SMT       YES       32.43       33.36       32.48       0.796220       0.778075       0.789657        0         387       SMT       NO       32.52       32.69       32.47       0.796059       0.778075       0.790107        0         388       SMT       YES       32.65       32.81       32.59       0.796777       0.791219        0        0        0        0											moses: 2.1.1).	
379       SMT       NO       27.37       28.28       27.43       0.774423       0.753749       0.767073          380       SMT       YES       27.86       28.89       28.00       0.776550       0.756721       0.769409          383       SMT       NO       32.08       33.09       32.18       0.776550       0.776565        0         383       SMT       NO       32.08       33.09       32.18       0.793230       0.775168       0.787665        0         384       SMT       YES       32.43       33.36       32.48       0.795220       0.778075       0.789657        0         387       SMT       NO       32.52       32.69       32.47       0.796059       0.780402       0.790107        0         388       SMT       YES       32.65       32.81       32.59       0.796777       0.791219        0	UIPS 4	375	SMT	YES	32.19	32.55	32.54	0.795838		0.787591	Our baseline system + additional quasi-parallel corpus (seg-	
380       SMT       YES       27.86       28.89       28.00       0.775550       0.756721       0.769409          383       SMT       NO       32.08       33.09       32.18       0.775550       0.776550       0.776565        0         383       SMT       NO       32.08       33.09       32.18       0.793230       0.775168       0.787665        0         384       SMT       YES       32.43       33.36       32.48       0.795220       0.778075       0.789657        0         387       SMT       NO       32.52       32.69       32.47       0.796059       0.778075       0.790107        0         388       SMT       YES       32.65       32.81       32.59       0.796777       0.791219        0	S SdIII		TMS	ON	77 37	28.28	27 43	0 774423		0 767073	Dur baseline system (seomentation tools: kytea moses: 10)	
383       SMT       NO       32.08       33.09       32.18       0.793230       0.775168       0.787665        0         384       SMT       YES       32.43       33.36       32.48       0.796220       0.778075       0.789657        0         387       SMT       NO       32.52       32.69       32.47       0.796059       0.780402       0.790107        0         388       SMT       YES       32.65       32.81       32.59       0.796777       0.781733       0.791219        0	UIPS 6		SMT	YES	27.86	28.89	28.00	0.776550		0.769409	Our baseline system + additional quasi-parallel corpus (seg-	
SMT         NO         32.08         33.09         32.18         0.793230         0.775168         0.787665          0           SMT         YES         32.43         33.36         32.48         0.796220         0.778075         0.789657          0           SMT         YES         32.43         33.36         32.48         0.796220         0.778075         0.789657          0           SMT         NO         32.52         32.69         32.47         0.796059         0.780402         0.790107          0           SMT         YES         32.65         32.81         32.59         0.796777         0.781733         0.791219          0											mentation tools: kytea, moses: 1.0).	
SMT         YES         32.43         33.36         32.48         0.796220         0.778075         0.789657          0           SMT         NO         32.52         32.69         32.47         0.796059         0.780402         0.790107          0           SMT         YES         32.65         32.81         32.59         0.796777         0.781733         0.791219          0	UIPS 7	383		NO	32.08	33.09	32.18	0.793230		0.787665	<ul> <li>Our baseline system (segmentation tools: kytea, moses: 2.1.1).</li> </ul>	
SMT         NO         32.52         32.69         32.47         0.796059         0.780402         0.790107          0           SMT         YES         32.65         32.81         32.59         0.796777         0.781733         0.791219          0	NIPS 8	384		YES	32.43	33.36	32.48	0.796220		0.789657	— Our baseline system + additional quasi-parallel corpus (seg-	
SMT YES 32.65 32.81 32.59 0.796777 0.781733 0.791219 0	6 SAIN	387	SMT	NO	32.52	32.69	32.47	0.796059		0.790107	Our baseline system (segmentation tools: stanford-ctb and ju-	
	UIPS 10	388	SMT	YES	32.65	32.81	32.59	2776777		0.791219	man, moses: 2.1.1). — Our baseline system + additional quasi-narallel cornus (see-	
						10.10					mentation tools: stanford-ctb and juman, moses: 2.1.1).	

Table 14: CJ submissions

### References

- Jingsheng Cai, Yujie Zhang, Hua Shan, and Jinan Xu. 2014. System Description: Dependency-based Preordering for Japanese-Chinese Machine Translation. In Proceedings of the 1st Workshop on Asian Translation (WAT2014).
- Chenchen Ding, Masao Utiyama, Eiichiro Sumita, and Mikio Yamamoto. 2014. Word Order Does NOT Differ Significantly Between Chinese and Japanese. In Proceedings of the 1st Workshop on Asian Translation (WAT2014).
- Terumasa Ehara. 2014. A machine translation system combining rule-based machine translation and statistical post-editing. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014).*
- J.L. Fleiss et al. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.
- Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H. Clark, and Philipp Koehn. 2013. Scalable modified kneser-ney language model estimation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 690–696, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Hieu Hoang, Philipp Koehn, and Adam Lopez. 2009. A unified framework for phrase-based, hierarchical, and syntax-based statistical machine translation. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 152–159.
- Sho Hoshino, Hubert Soyer, Yusuke Miyao, and Akiko Aizawa. 2014. Japanese to English Machine Translation using Preordering and Compositional Distributed Semantics. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014)*.
- Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. 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.
- 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 Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session.
- Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 388–395, Barcelona, Spain, July. Association for Computational Linguistics.

- T. Kudo. 2005. Mecab : Yet another part-of-speech and morphological analyzer. http://mecab.sourceforge.net/.
- Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto, and Makoto Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. In Proceedings of The International Workshop on Sharable Natural Language, pages 22–28.
- Graham Neubig, Yosuke Nakata, and Shinsuke Mori. 2011. Pointwise prediction for robust, adaptable japanese morphological analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2, HLT '11, pages 529– 533, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Graham Neubig. 2014. Forest-to-String SMT for Asian Language Translation: NAIST at WAT 2014. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014).*
- Kenichi Ohwada, Ryosuke Miyazaki, and Mamoru Komachi. 2014. Predicate-Argument Structurebased Preordering for Japanese-English Statistical Machine Translation of Scientific Papers. In Proceedings of the 1st Workshop on Asian Translation (WAT2014).
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL*, pages 311– 318.
- Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pages 433–440, Sydney, Australia, July. Association for Computational Linguistics.
- John Richardson, Fabien Cromières, Toshiaki Nakazawa, and Sadao Kurohashi. 2014. KyotoEBMT System Description for the 1st Workshop on Asian Translation. In Proceedings of the 1st Workshop on Asian Translation (WAT2014).
- Satoshi Sonoh, Satoshi Kinoshita, Hiroyuki Tanaka, and Satoshi Kamatani. 2014. Toshiba MT System Description for the WAT2014 Workshop. In *Proceedings of the 1st Workshop on Asian Translation* (WAT2014).
- Liling Tan and Francis Bond. 2014. Manipulating Input Data in Machine Translation. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014).*
- Huihsin Tseng. 2005. A conditional random field word segmenter. In *In Fourth SIGHAN Workshop on Chinese Language Processing*.

- Masao Utiyama and Hitoshi Isahara. 2007. A japanese-english patent parallel corpus. In *MT summit XI*, pages 475–482.
- Rui Wang, Xu Yang, and Yan Gao. 2014. The SAS Statistical Machine Translation System for WAT 2014. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014)*.
- Wei Yang and Yves Lepage. 2014. Consistent Improvement in Translation Quality of Chinese– Japanese Technical Texts by Adding Additional Quasi-parallel Training Data. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014).*
- Zhongyuan Zhu. 2014. Weblio Pre-reordering Statistical Machine Translation System. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014).*