Overview of the 5th Workshop on Asian Translation

Toshiaki Nakazawa
The University of Tokyo
nakazawa@logos.t.u-tokyo.ac.jp

Katsuhito Sudoh
Nara Institute of Science and Technology
sudoh@is.naist.jp

Shohei Higashiyama and **Chenchen Ding** and **Raj Dabre**
National Institute of
Information and Communications Technology
{shohei.higashiyama, chenchen.ding, raj.dabre}@nict.go.jp

Hideya Mino and **Isao Goto**
NHK
{mino.h-gq, goto.i-es}@nhk.or.jp

Win Pa Pa
University of Conputer Study, Yangon
winpapa@ucsy.edu.mm

Anoop Kunchukuttan
Microsoft AI and Research
anoop.kunchukuttan@gmail.com

Sadao Kurohashi
Kyoto University
kuro@i.kyoto-u.ac.jp

Abstract

This paper presents the results of the shared tasks from the 5th workshop on Asian translation (WAT2018) including Ja↔En, Ja↔Zh scientific paper translation subtasks, Zh↔Ja, K↔Ja, En↔Ja patent translation subtasks, Hi↔En, My↔En mixed domain subtasks and Bn/Hi/Ml/Ta/Te/Ur/Si↔En Indic languages multilingual subtasks. For the WAT2018, 17 teams participated in the shared tasks. About 500 translation results were 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. Following the success of the previous workshops WAT2014-WAT2017 (Nakazawa et al., 2014; Nakazawa et al., 2015; Nakazawa et al., 2016; Nakazawa et al., 2017), WAT2018 brings together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas of machine translation. We have been working toward practical use of machine translation among all Asian countries.

For the 5th WAT, we adopted new translation subtasks with Myanmar ↔ English mixed domain corpus[1] and Bengali/Hindi/Malayalam/Tamil/Telugu/Urdu/Sinhalese ↔ English OpenSubtitles corpus[2] in addition to the subtasks at WAT2017.

WAT is the uniq workshop on Asian language transration with the following characteristics:

- Open innovation platform
  Due to the fixed and open test data, we can repeatedly evaluate translation systems on the same dataset over years. WAT receives submissions at any time; i.e., there is no submission deadline of translation results w.r.t automatic evaluation of translation quality.

- Domain and language pairs
  WAT is the world's first workshop that targets scientific paper domain, and Chinese↔Japanese and Korean↔Japanese language pairs. In the future, we will add more Asian languages such as Vietnamese, Thai and so on.

- Evaluation method
  Evaluation is done both automatically and manually. Firstly, all submitted translation results

[1] http://lotus.kuee.kyoto-u.ac.jp/WAT/my-en-data/

[2] http://lotus.kuee.kyoto-u.ac.jp/WAT/indic-multilingual/

32nd Pacific Asia Conference on Language, Information and Computation
The 5th Workshop on Asian Translation
Hong Kong, 1-3 December 2018

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

Table 1: Statistics for ASPEC

are automatically evaluated using three metrics: BLEU, RIBES and AMFM. Among them, selected translation results are assessed by two kinds of human evaluation: pairwise evaluation and JPO adequacy evaluation.

2 Dataset

2.1 ASPEC

ASPEC was constructed by the Japan Science and Technology Agency (JST) in collaboration with the National Institute of Information and Communications Technology (NICT). The corpus consists of a Japanese-English scientific paper abstract corpus (ASPEC-JE), which is used for ja↔en subtasks, and a Japanese-Chinese scientific paper excerpt corpus (ASPEC-JC), which is used for ja↔zh subtasks. The statistics for each corpus are shown in Table 1.

2.1.1 ASPEC-JE

The training data for ASPEC-JE was constructed by NICT from approximately two million Japanese-English scientific paper abstracts owned by JST. The data is a comparable corpus and sentence correspondences are found automatically using the method from (Utiyama and Isahara, 2007). Each sentence pair is accompanied by a similarity score that are calculated by the method and a field ID that indicates a scientific field. The correspondence between field IDs and field names, along with the frequency and occurrence ratios for the training data, are described in the README file of ASPEC-JE.

The development, development-test and test data were extracted from parallel sentences from the Japanese-English paper abstracts that exclude the sentences in the training data. Each dataset consists of 400 documents and contains sentences in each field at the same rate. 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 the training data except that there is no similarity score.

Lang	Train	Dev	DevTest	Test-N
zh-ja	1,000,000	2,000	2,000	5,204
ko-ja	1,000,000	2,000	2,000	5,230
en-ja	1,000,000	2,000	2,000	5,668

Lang	Test-N1	Test-N2	Test-N3	Test-EP
zh-ja	2,000	3,000	204	1,151
ko-ja	2,000	3,000	230	–
en-ja	2,000	3,000	668	–

Table 2: Statistics for JPC

2.1.2 ASPEC-JC

ASPEC-JC is a parallel corpus consisting of Japanese scientific papers, which come from the literature database and electronic journal site J-STAGE by JST, and their translation to Chinese with permission from the necessary academic associations. Abstracts and paragraph units are selected from the body text so as to 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). There are no documents sharing the same data across the training, development, development-test and test sets.

2.2 JPC

JPO Patent Corpus (JPC) for the patent tasks was constructed by the Japan Patent Office (JPO) in collaboration with NICT. The corpus consists of Chinese-Japanese, Korean-Japanese and English-Japanese patent descriptions whose International Patent Classification (IPC) sections are chemistry, electricity, mechanical engineering, and physics.

At WAT2018, the patent tasks has two subtasks: normal subtask and expression pattern subtask. Both subtasks uses common training, development and development-test data for each language pair. The normal subtask for three language pairs uses four test data with different characteristics:

- test-N: union of the following three sets;
- test-N1: patent documents from patent families published between 2011 and 2013;

Lang	Train	Dev	DevTest	Test
en-ja	200,000	2,000	2,000	2,000

Table 3: Statistics for JIJI Corpus

- test-N2: patent documents from patent families published between 2016 and 2017; and
- test-N3: patent documents published between 2016 and 2017 where target sentences are manually created by translating source sentences.

The expression pattern subtask for zh→ja pair uses test-EP data. The test-EP data consists of sentences annotated with expression pattern categories: title of invention (TIT), abstract (ABS), scope of claim (CLM) or description (DES). The corpus statistics are shown in Table 2. Note that training, development, development-test and test-N1 data are the same as those used in WAT2017.

2.3 JIJI Corpus

JIJI Corpus was constructed by Jiji Press Ltd. in collaboration with NICT. The corpus consists of news text that comes from Jiji Press news of various categories including politics, economy, nation, business, markets, sports and so on. The corpus is partitioned into training, development, development-test and test data, which consists of Japanese-English sentence pairs. The statistics for each corpus are shown in Table 3.

The sentence pairs in each data are identified in the same manner as that for ASPEC using the method from (Utiyama and Isahara, 2007).

2.4 IITB Corpus

IIT Bombay English-Hindi Corpus contains English-Hindi parallel corpus as well as monolingual Hindi corpus collected from a variety of sources and corpora. This corpus had been developed at the Center for Indian Language Technology, IIT Bombay over the years. The corpus is used for mixed domain tasks hi↔en. The statistics for the corpus are shown in Table 4.

2.5 Recipe Corpus

Recipe Corpus was constructed by Cookpad Inc. Each recipe consists of a title, ingredients, steps, a description and a history. Every text in titles, ingredients and steps consists of a parallel sentence while one in descriptions and histories is not always a parallel sentence. Although all of the texts in the training set can be used for training, only titles, ingredients and steps in the test set is used for evaluation. The statistics for each corpus are described in Table 5.

Lang	Train	Dev	Test	Mono
hi-en	1,492,827	520	2,507	–
hi-ja	152,692	1,566	2,000	–
hi	–	–	–	45,075,279

Table 4: Statistics for IITB Corpus. “Mono” indicates monolingual Hindi corpus.

Lang	TextType	Train	Dev	DevTest	Test
en-ja	Title	14,779	500	500	500
	Ingredient	127,244	4,274	4,188	3,935
	Step	108,993	3,303	3,086	2,804

Table 5: Statistics for Recipe Corpus

2.6 ALT and UCSY Corpus

The parallel data for Myanmar-English translation tasks at WAT2018 consists of two corpora, the ALT corpus and UCSY corpus.

- The ALT corpus is one part from the Asian Language Treebank (ALT) project (Riza et al., 2016), consisting of twenty thousand Myanmar-English parallel sentences from news articles.
- The UCSY corpus (Yi Mon Shwe Sin and Khin Mar Soe, 2018) is constructed by the NLP Lab, University of Computer Studies, Yangon (UCSY), Myanmar. The corpus consists of 200 thousand Myanmar-English parallel sentences collected from different domains, including news articles and textbooks.

The released Myanmar textual data have been tokenized into writing units and Romanized. The script for tokenization and recovery is also provided for participants,[3] so that they can make use of their own data and tools for further processing. The automatic

[3] http://www2.nict.go.jp/astrec-att/member/mutiyama/ALT/myan2roma.py

Corpus	Train	Dev	Test
ALT	17,965	993	1,007
UCSY	208,638	–	–
All	226,603	993	1,007

Table 6: Statistics for the data used in Myanmar-English translation tasks

Lang	Train	Dev	Test	Mono (src)
bn-en	337,428	500	1,000	453,859
hi-en	84,557	500	1,000	104,967
ml-en	359,423	500	1,000	402,761
ta-en	26,217	500	1,000	30,268
te-en	22,165	500	1,000	24,750
ur-en	26,619	500	1,000	29,086
si-en	521,726	500	1,000	705,793
en	–	–	–	2,891,079

Table 7: Statistics for Indic Languages Corpus

evaluation of Myanmar translation results is based on the tokenized writing units, and the human evaluation is based on the recovered Myanmar text.

The detailed composition of training, development, and test data of the Myanmar-English translation tasks are listed in Table 6.

2.7 Indic Languages Corpus

The Indic Languages Corpus covers 8 languages, namely: Bengali, Hindi, Malayalam, Tamil, Telugu, Sinhalese, Urdu and English. The corpus has been collected from OPUS[4] and belongs to the spoken language (OpenSubtitles) domain. This corpus is used for the pilot as well as multilingual English↔Indic Languages sub-tasks. The corpus is a collection of 7 bilingual parallel corpora of varying sizes, one for each Indic language and English. The parallel corpora are also accompanied by monolingual corpora from the same domain. The statistics of the parallel and monolingual corpora are given in Table 7.

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 for human evaluation. At WAT 2018, we adopted a neural machine translation (NMT) with attention mechanism as a baseline system except for the IITB tasks. We used a phrase-based statistical machine translation (SMT) system, which is the same system as that at WAT 2017, as the baseline system for the IITB tasks.

The NMT baseline systems consisted of publicly available software, and the procedures for building the systems and for translating using the systems were published on the WAT web page.[5] We used OpenNMT (Klein et al., 2017) as the implementation of the baseline NMT systems. In addition to the NMT baseline systems, we have SMT baseline systems for the tasks that started at last year or before last year. The baseline systems are shown in Tables 8, 9, and 10.

SMT baseline systems are described in the previous WAT overview paper (Nakazawa et al., 2017). 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. Because our objective is not to compare commercial RBMT systems or online translation systems from companies that did not themselves participate, the system IDs of these systems are anonymous in this paper.

[4] http://opus.nlpl.eu

[5] http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2018/baseline/baselineSystems.html

System ID	System	Type	ASPEC				JPC					
			ja-en	en-ja	ja-zh	zh-ja	ja-en	en-ja	ja-zh	zh-ja	ja-ko	ko-ja
NMT	OpenNMT's attention-based NMT	NMT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SMT Phrase	Moses' Phrase-based SMT	SMT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SMT Hiero	Moses' Hierarchical Phrase-based SMT	SMT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SMT S2T	Moses' String-to-Tree Syntax-based SMT and Berkeley parser	SMT	✓		✓		✓		✓			
SMT T2S	Moses' Tree-to-String Syntax-based SMT and Berkeley parser	SMT		✓		✓		✓		✓		
RBMT X	The Honyaku V15 (Commercial system)	RBMT	✓	✓			✓	✓				
RBMT X	ATLAS V14 (Commercial system)	RBMT	✓	✓			✓	✓				
RBMT X	PAT-Transer 2009 (Commercial system)	RBMT	✓	✓			✓	✓				
RBMT X	PC-Transer V13 (Commercial system)	RBMT										
RBMT X	J-Beijing 7 (Commercial system)	RBMT			✓	✓			✓	✓		
RBMT X	Hohrai 2011 (Commercial system)	RBMT			✓	✓				✓		
RBMT X	J Soul 9 (Commercial system)	RBMT									✓	✓
RBMT X	Korai 2011 (Commercial system)	RBMT									✓	✓
Online X	Google translate	Other	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Online X	Bing translator	Other	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AIAYN	Google's implementation of "Attention Is All You Need"	NMT	✓	✓								

Table 8: Baseline Systems I

System ID	System	Type	JIJI		IITB				Recipe		ALT	
			ja-en	en-ja	hi-en	en-hi	hi-ja	ja-hi	ja-en	en-ja	my-en	en-my
NMT	OpenNMT's NMT with attention	NMT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SMT Phrase	Moses' Phrase-based SMT	SMT	✓	✓	✓	✓	✓	✓	✓	✓		
SMT Hiero	Moses' Hierarchical Phrase-based SMT	SMT	✓	✓								
SMT S2T	Moses' String-to-Tree Syntax-based SMT and Berkeley parser	SMT	✓									
SMT T2S	Moses' Tree-to-String Syntax-based SMT and Berkeley parser	SMT		✓								
RBMT X	The Honyaku V15 (Commercial system)	RBMT	✓	✓					✓	✓		
RBMT X	PC-Transer V13 (Commercial system)	RBMT	✓	✓					✓	✓		
Online X	Google translate	Other	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Online X	Bing translator	Other	✓	✓	✓	✓	✓	✓	✓	✓		

Table 9: Baseline Systems II

System ID	System	Type	Indic	
			{bn,hi,ml,ta,te,ur,si}-en	en-{bn,hi,ml,ta,te,ur,si}
NMT	OpenNMT's NMT with attention	NMT	✓	✓
NMT M2O	OpenNMT's NMT with attention and multilingual tags (many to one)	NMT	✓	
NMT O2M	OpenNMT's NMT with attention and multilingual tags (one to many)	NMT		✓
NMT M2M	OpenNMT's NMT with attention and multilingual tags (many to many)	NMT	✓	✓

Table 10: Baseline Systems III

3.1 Training Data

We used the following data for training the NMT baseline systems.

- All of the training data for each task were used for training except for the ASPEC Japanese–English task. For the ASPEC Japanese–English task, we only used train-1.txt, which consists of one million parallel sentence pairs with high similarity scores.
- All of the development data for each task was used for validation.

3.2 Tokenization

We used the following tools for tokenization.

- Juman version 7.0[6] for Japanese segmentation.
- Stanford Word Segmenter version 2014-01-04[7] (Chinese Penn Treebank (CTB) model) for Chinese segmentation.
- The Moses toolkit for English and Indonesian tokenization.
- Mecab-ko[8] for Korean segmentation.
- Indic NLP Library[9] for Indic language segmentation.
- subword-nmt[10] for all languages.

When we built BPE-codes, we merged source and target sentences and we used 100,000 for -s option. We used 10 for vocabulary-threshold when subword-nmt applied BPE.

3.3 NMT with attention

We used the following OpenNMT configuration for the NMT with attention system.

- encoder_type = brnn
- brnn_merge = concat
- src_seq_length = 150
- tgt_seq_length = 150
- src_vocab_size = 100000
- tgt_vocab_size = 100000
- src_words_min_frequency = 1
- tgt_words_min_frequency = 1

The default values were used for the other system parameters.

For many to one, one to many, and many to many multilingual NMT (Johnson et al., 2017), we add <2XX> tags, which indicate the target language (XX is replaced by the language code), to the head of the source language sentences.

4 Automatic Evaluation

4.1 Procedure for Calculating Automatic Evaluation Score

We evaluated translation results by three metrics: BLEU (Papineni et al., 2002), RIBES (Isozaki et al., 2010) and AMFM (Banchs et al., 2015). BLEU scores were calculated using `multi-bleu.perl` in the Moses toolkit (Koehn et al., 2007). RIBES scores were calculated using `RIBES.py` version 1.02.4.[11] AMFM scores were calculated using scripts created by the technical collaborators listed in the WAT2018 web page.[12] All scores for each task were calculated using the corresponding reference translations.

Before the calculation of the automatic evaluation scores, the translation results were tokenized or segmented with tokenization/segmentation tools for each language. For Japanese segmentation, we used three different tools: Juman version 7.0 (Kurohashi et al., 1994), KyTea 0.4.6 (Neubig et al., 2011) with full SVM model[13] and MeCab 0.996 (Kudo, 2005) with IPA dictionary 2.7.0.[14] For Chinese segmentation, we used two different tools: KyTea 0.4.6 with full SVM Model in MSR model and Stanford Word Segmenter (Tseng, 2005) version 2014-06-16 with Chinese Penn Treebank (CTB) and Peking University (PKU) model.[15] For Korean segmentation, we

[6] http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN

[7] http://nlp.stanford.edu/software/segmenter.shtml

[8] https://bitbucket.org/eunjeon/mecab-ko/

[9] https://bitbucket.org/anoopk/indic_nlp_library

[10] https://github.com/rsennrich/subword-nmt

[11] http://www.kecl.ntt.co.jp/icl/lirg/ribes/index.html

[12] lotus.kuee.kyoto-u.ac.jp/WAT/WAT2018/

[13] http://www.phontron.com/kytea/model.html

[14] http://code.google.com/p/mecab/downloads/detail?name=mecab-ipadic-2.7.0-20070801.tar.gz

[15] http://nlp.stanford.edu/software/segmenter.shtml

used mecab-ko.[16] For English tokenization, we used `tokenizer.perl`[17] in the Moses toolkit. For Hindi, Bengali, Malayalam, Tamil, Telugu, Urdu and Sinhalese tokenization, we used Indic NLP Library.[18] The detailed procedures for the automatic evaluation are shown on the WAT2018 evaluation web page.[19]

4.2 Automatic Evaluation System

The automatic evaluation system receives translation results by participants and automatically gives evaluation scores to the uploaded results. As shown in Figure 1, the system requires participants to provide the following information for each submission:

- Human Evaluation: whether or not they submit the results for human evaluation;
- Publish the results of the evaluation: whether or not they permit to publish automatic evaluation scores on the WAT2018 web page.
- Task: the task you submit the results for;
- Used Other Resources: whether or not they used additional resources; and
- Method: the type of the method including SMT, RBMT, SMT and RBMT, EBMT, NMT and Other.

Evaluation scores of translation results that participants permit to be published are disclosed via the WAT2018 evaluation web page.[20] Participants can also submit the results for human evaluation using the same web interface.

This automatic evaluation system will remain available even after WAT2018. Anybody can register an account for the system by the procedures described in the registration web page. [21]

[16] https://bitbucket.org/eunjeon/mecab-ko/

[17] https://github.com/moses-smt/mosesdecoder/tree/RELEASE-2.1.1/scripts/tokenizer/tokenizer.perl

[18] https://bitbucket.org/anoopk/indic_nlp_library

[19] http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

[20] lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html

[21] http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2018/registration/index.html

5 Human Evaluation

In WAT2018, we conducted two kinds of human evaluations: *pairwise evaluation* and *JPO adequacy evaluation*.

5.1 Pairwise Evaluation

We conducted pairwise evaluation for participants' systems submitted for human evaluation. The submitted translations were evaluated by a professional translation company and *Pairwise* scores were given to the submissions by comparing with baseline translations (described in section 3).

5.1.1 Sentence Selection and Evaluation

For the pairwise evaluation, we randomly selected 400 sentences from the test set of each task. We used the same sentences as the last year for the continuous subtasks. Baseline and submitted translations were shown to annotators in random order with the input source sentence. The annotators were asked to judge which of the translations is better, or whether they are on par.

5.1.2 Voting

To guarantee the quality of the evaluations, each sentence is evaluated by 5 different annotators and the final decision is made depending on the 5 judgements. We define each judgement $j_i (i = 1, \cdots, 5)$ as:

$$j_i = \begin{cases} 1 & \text{if better than the baseline} \\ -1 & \text{if worse than the baseline} \\ 0 & \text{if the quality is the same} \end{cases}$$

The final decision D is defined as follows using $S = \sum j_i$:

$$D = \begin{cases} win & (S \geq 2) \\ loss & (S \leq -2) \\ tie & (otherwise) \end{cases}$$

5.1.3 Pairwise Score Calculation

Suppose that W is the number of *wins* compared to the baseline, L is the number of *losses* and T is the number of *ties*. The Pairwise score can be calculated by the following formula:

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

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

WAT
The Workshop on Asian Translation
Submission

SUBMISSION

Logged in as: ORGANIZER

Logout

Submission:

Human Evaluation:	☐ human evaluation
Publish the results of the evaluation:	☑ publish
Team Name:	ORGANIZER
Task:	en-ja
Submission File:	ファイルを選択 選択されていません
Used Other Resources:	☐ used other resources such as parallel corpora, monolingual corpora and parallel dictionaries in addition to official corpora
Method:	SMT
System Description (public):	100 characters or less
System Description (private):	100 characters or less

Submit

Guidelines for submission:

- System requirements:
  - The latest versions of Chrome, Firefox, Internet Explorer and Safari are supported for this site.
  - Before you submit files, you need to enable JavaScript in your browser.
- File format:
  - Submitted files should NOT be tokenized/segmented. Please check the automatic evaluation procedures.
  - Submitted files should be encoded in UTF-8 format.
  - Translated sentences in submitted files should have one sentence per line, corresponding to each test sentence. The number of lines in the submitted file and that of the corresponding test file should be the same.
- Tasks:
  - en-ja, ja-en, zh-ja, ja-zh indicate the scientific paper tasks with ASPEC.
  - HINDENen-hi, HINDENhi-en, HINDENja-hi, and HINDENhi-ja indicate the mixed domain tasks with IITB Corpus.
  - JIJIen-ja and JIJIja-en are the newswire tasks with JIJI Corpus.
  - RECIPE{ALL,TTL,STE,ING}en-ja and RECIPE{ALL,TTL,STE,ING}ja-en indicate the recipe tasks with Recipe Corpus.
  - ALTen-my and ALTmy-en indicate the mixed domain tasks with UCSY and ALT Corpus.
  - INDICen-{bn,hi,ml,ta,te,ur,si} and INDIC{bn,hi,ml,ta,te,ur,si}-en indicate the Indic languages multilingual tasks with Indic Languages Multilingual Parallel Corpus.
  - JPC{N,N1,N2,N3,EP}zh-ja ,JPC{N,N1,N2,N3}ja-zh, JPC{N,N1,N2,N3}ko-ja, JPC{N,N1,N2,N3}ja-ko, JPC{N,N1,N2,N3}en-ja, and JPC{N,N1,N2,N3}ja-en indicate the patent tasks with JPO Patent Corpus. JPCN1{zh-ja,ja-zh,ko-ja,ja-ko,en-ja,ja-en} are the same tasks as JPC{zh-ja,ja-zh,ko-ja,ja-ko,en-ja,ja-en} in WAT2015-WAT2017. AMFM is not calculated for JPC{N,N2,N3} tasks.
- Human evaluation:
  - If you want to submit the file for human evaluation, check the box "Human Evaluation". Once you upload a file with checking "Human Evaluation" you cannot change the file used for human evaluation.
  - When you submit the translation results for human evaluation, please check the checkbox of "Publish" too.
  - You can submit two files for human evaluation per task.
  - One of the files for human evaluation is recommended not to use other resources, but it is not compulsory.
- Other:
  - Team Name, Task, Used Other Resources, Method, System Description (public) , Date and Time(JST), BLEU, RIBES and AMFM will be disclosed on the Evaluation Site when you upload a file checking "Publish the results of the evaluation".
  - You can modify some fields of submitted data. Read "Guidelines for submitted data" at the bottom of this page.

Back to top

Figure 1: The interface for translation results submission

5.1.4 Confidence Interval Estimation

There are several ways to estimate a confidence interval. We chose to use bootstrap resampling (Koehn, 2004) to estimate the 95% confidence interval. The procedure is as follows:

1. randomly select 300 sentences from the 400 human evaluation sentences, and calculate the Pairwise score of the selected sentences

2. iterate the previous step 1000 times and get 1000 Pairwise scores

3. sort the 1000 scores and estimate the 95% confidence interval by discarding the top 25 scores and the bottom 25 scores

5.2 JPO Adequacy Evaluation

We conducted JPO adequacy evaluation for the top two or three participants' systems of pairwise evalution for each subtask.[22] The evaluation was carried out by translation experts based on the JPO adequacy evaluation criterion, which is originally defined by JPO to assess the quality of translated patent documents.

5.2.1 Sentence Selection and Evaluation

For the JPO adequacy evaluation, the 200 test sentences were randomly selected from the 400 test sentences used for the pairwise evaluation. For each test sentence, input source sentence, translation by participants' system, and reference translation were shown to the annotators. To guarantee the quality of the evaluation, each sentence was evaluated by two annotators. Note that the selected sentences are the same as those used in the previous workshops except for the new subtasks at WAT2018.

5.2.2 Evaluation Criterion

Table 11 shows the JPO adequacy criterion from 5 to 1. The evaluation is performed subjectively. "Important information" represents the technical factors and their relationships. The degree of importance of each element is also considered to evaluate. The percentages in each grade are rough indications for the transmission degree of the source sentence meanings. The detailed criterion is described in the JPO document (in Japanese). [23]

[22]The number of systems varies depending on the subtasks.

5	All important information is transmitted correctly. (100%)
4	Almost all important information is transmitted correctly. (80%–)
3	More than half of important information is transmitted correctly. (50%–)
2	Some of important information is transmitted correctly. (20%–)
1	Almost all important information is NOT transmitted correctly. (–20%)

Table 11: The JPO adequacy criterion

6 Participants

Table 12 shows the participants in WAT2018. The table lists 17 organizations from various countries, including Japan, China, India, Myanmar, Czech and Ireland.

More than 500 translation results by 17 teams were submitted for automatic evaluation and about 70 translation results by 16 teams were submitted for pairwise evaluation. We selected about 40 translation results for JPO adequacy evaluation according to the pairwise evaluation scores. Table 13 shows tasks for which each team submitted results by the submission deadline. Unfortunately, there were no submissions to Recipe and JIJI tasks this year.

7 Evaluation Results

In this section, the evaluation results for WAT2018 are reported from several perspectives. Some of the results for both automatic and human evaluations are also accessible at the WAT2018 website.[24]

7.1 Official Evaluation Results

Figures 2, 3, 4 and 5 show the official evaluation results of ASPEC subtasks, Figures 6, 7, 8, 9, 10, 11, 12 and 13 show those of JPC subtasks, Figures 14 and 15 show those of IITB subtasks, Figures 16 and 17 show those of ALT subtasks and Figures 18,

[23]http://www.jpo.go.jp/shiryou/toushin/chousa/tokkyohonyaku_hyouka.htm

[24]http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/

19, 20 and 21 show those of INDIC subtasks. Each figure contains automatic evaluation results (BLEU, RIBES, AM-FM), the pairwise evaluation results with confidence intervals, correlation between automatic evaluations and the pairwise evaluation, the JPO adequacy evaluation result and evaluation summary of top systems. Some of the figures for some subtasks are omitted because the pairwise evaluation was not conducted or none of the human evaluation was conducted.

The detailed automatic evaluation results are shown in Appendix A. The detailed JPO adequacy evaluation results for the selected submissions are shown in Table 14. The weights for the weighted κ (Cohen, 1968) is defined as $|Evaluation1 - Evaluation2|/4$.

7.2 Statistical Significance Testing of Pairwise Evaluation between Submissions

Tables 15 and 16 show the results of statistical significance testing of ASPEC subtasks, Table 17 shows that of IITB subtasks, Table 18 shows that of ALT subtasks and Tables 19 and 20 show those of INDIC subtasks. $\ggg$, $\gg$ and $>$ mean that the system in the row is *better* than the system in the column at a significance level of $p < 0.01$, 0.05 and 0.1 respectively. Testing is also done by the bootstrap resampling as follows:

1. randomly select 300 sentences from the 400 pairwise evaluation sentences, and calculate the Pairwise 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 workers, we calculated the Fleiss' κ (Fleiss and others, 1971) values. The results are shown in Table 21. We can see that the κ values are larger for X $\rightarrow$ J translations than for J $\rightarrow$ X translations. This may be because the majority of the workers for these language pairs are Japanese, and the evaluation of one's mother tongue is much easier than for other languages in general. The κ values for Hindi languages are relatively higt. This might be because the overall translation quality of the Hindi languages are low, and the evaluators can easily distinguish better translations from worse ones.

8 Conclusion and Future Perspective

This paper summarizes the shared tasks of WAT2018. We had 17 participants worldwide, and collected a large number of useful submissions for improving the current machine translation systems by analyzing the submissions and identifying the issues.

For the next WAT workshop, we plan to conduct documen-level evaluation using the new dataset with context for some translation subtasks and we would like to consider how to realize context-aware evaluation in WAT. Also, we are planning to do extrinsic evaluation of the translations.

Appendix A Submissions

Tables 23 to 37 summarize translation results submitted for WAT2018 human evaluation. Type, RSRC, Pair, and Adeq columns indicate type of method, use of other resources, pairwise evaluation score, and JPO adequacy evaluation score, respectively.

The tables also include results by the organizers' baselines, which are listed in Table 10. For ALT tasks, we also evaluated outputs of Online-A system and its post-processed version where the western comma (,) is replaced into Myanmar native comma (0x104a). We conducted the post-processing because Myanmar native punctuation marks are consistently used in the WAT 2018 dataset.

Team ID	Organization	Country
srcb (Li et al., 2018)	RICOH Software Research Center Beijing Co.,Ltd	China
Osaka-U (Kawara et al., 2018)	Osaka University	Japan
RGNLP (Ojha et al., 2018)	Jawaharlal Nehru University / Dublin City University	India, Ireland
TMU (Zhang et al., 2018), (Matsumura et al., 2018)	Tokyo Metropolitan University	Japan
EHR (Ehara, 2018)	Ehara NLP Research Laboratory	Japan
NICT (Wang et al., 2018b)	NICT	Japan
NICT-4 (Marie et al., 2018)	NICT	Japan
NICT-5 (Dabre et al., 2018)	NICT	Japan
XMUNLP (Wang et al., 2018a)	Xiamen University	China
UCSYNLP (Mo et al., 2018)	University of Computer Studies, Yangon	Myanmar
UCSMNLP (Thida et al., 2018)	University of Computer Studies, Mandalay	Myanmar
kmust88	Kunming University of Science and Technology	China
USTC	University of Science and Technology of China	China
CUNI (Kocmi et al., 2018)	Charles University, Prague	Czech
Anuvaad (Banerjee et al., 2018)	IIT Bombay / Microsft AI and Research, India	India
IITP-MT (Sen et al., 2018)	Indian Institute of Technology Patna	India
cvit-mt (Philip et al., 2018)	International Institute of Information Technology, Hyderabad	India

Table 12: List of participants in WAT2018

	ASPEC				JPC (N/N1/N2/N3)				JPC (EP)	IITB		ALT	
Team ID	EJ	JE	CJ	JC	EJ	CJ	JC	KJ	CJ	EH	HE	E-My	My-E
srcb	✓	✓		✓									
Osaka-U	✓	✓										✓	✓
TMU	✓	✓	✓										
EHR	✓				✓	✓		✓	✓				
NICT												✓	✓
NICT-4												✓	✓
NICT-5	✓	✓	✓	✓									✓
XMUNLP												✓	✓
UCSYNLP												✓	✓
UCSMNLP												✓	✓
kmust88												✓	
USTC						✓	✓						
CUNI										✓	✓		
cvit-mt										✓	✓		

	Indic													
Team ID	EB	BE	EH	HE	E-Ml	Ml-E	E-Ta	Ta-E	E-Te	Te-E	EU	UE	ES	SE
RGNLP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NICT-5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Anuvaad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IITP-MT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 13: Submissions for each task by each team. E, J, C, K, H, B, U, and S denote English, Japanese, Chinese, Korean, Hindi, Bengali, Urdu, and Sinhalese language, respectively.

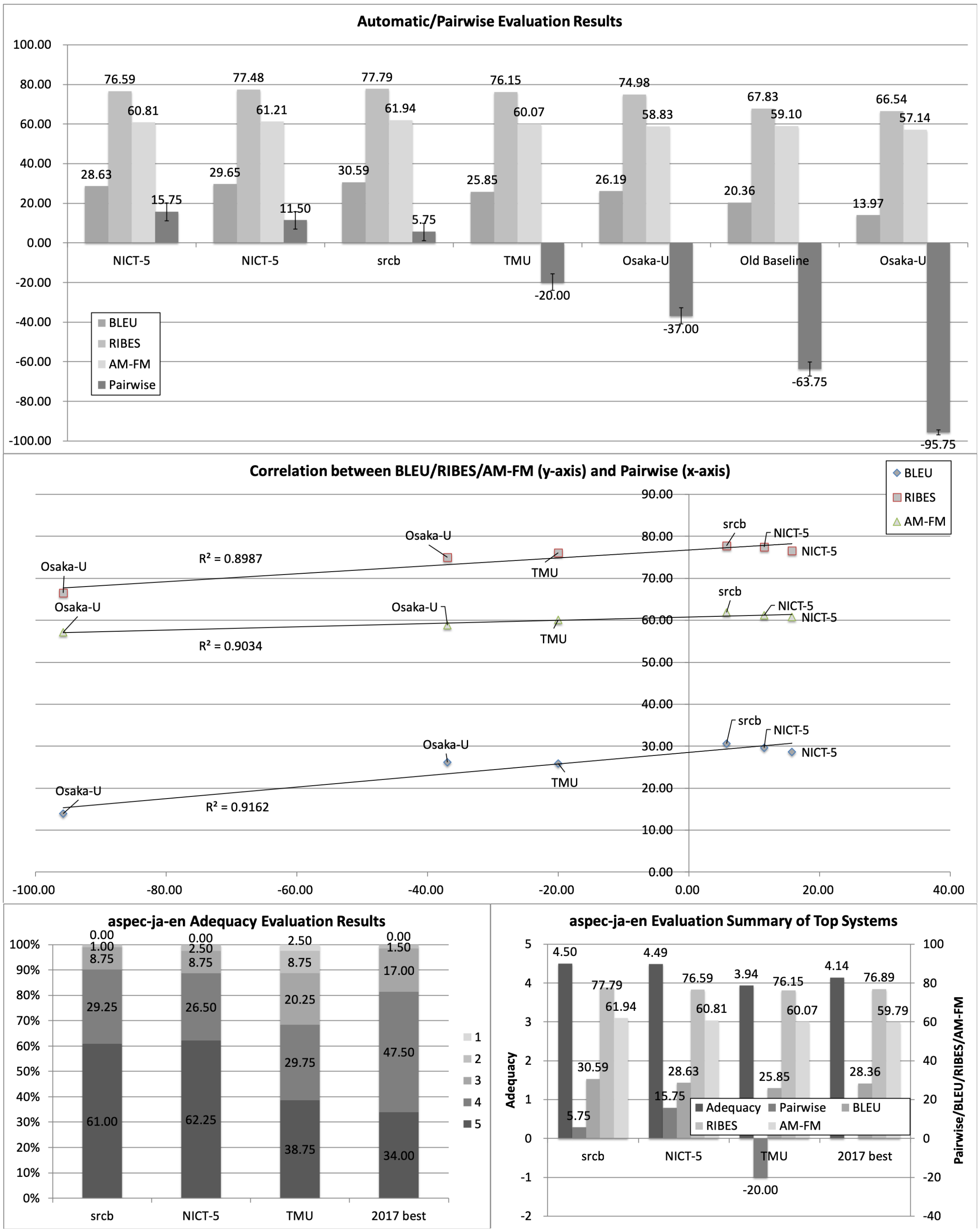


Figure 2: Official evaluation results of aspec-ja-en.

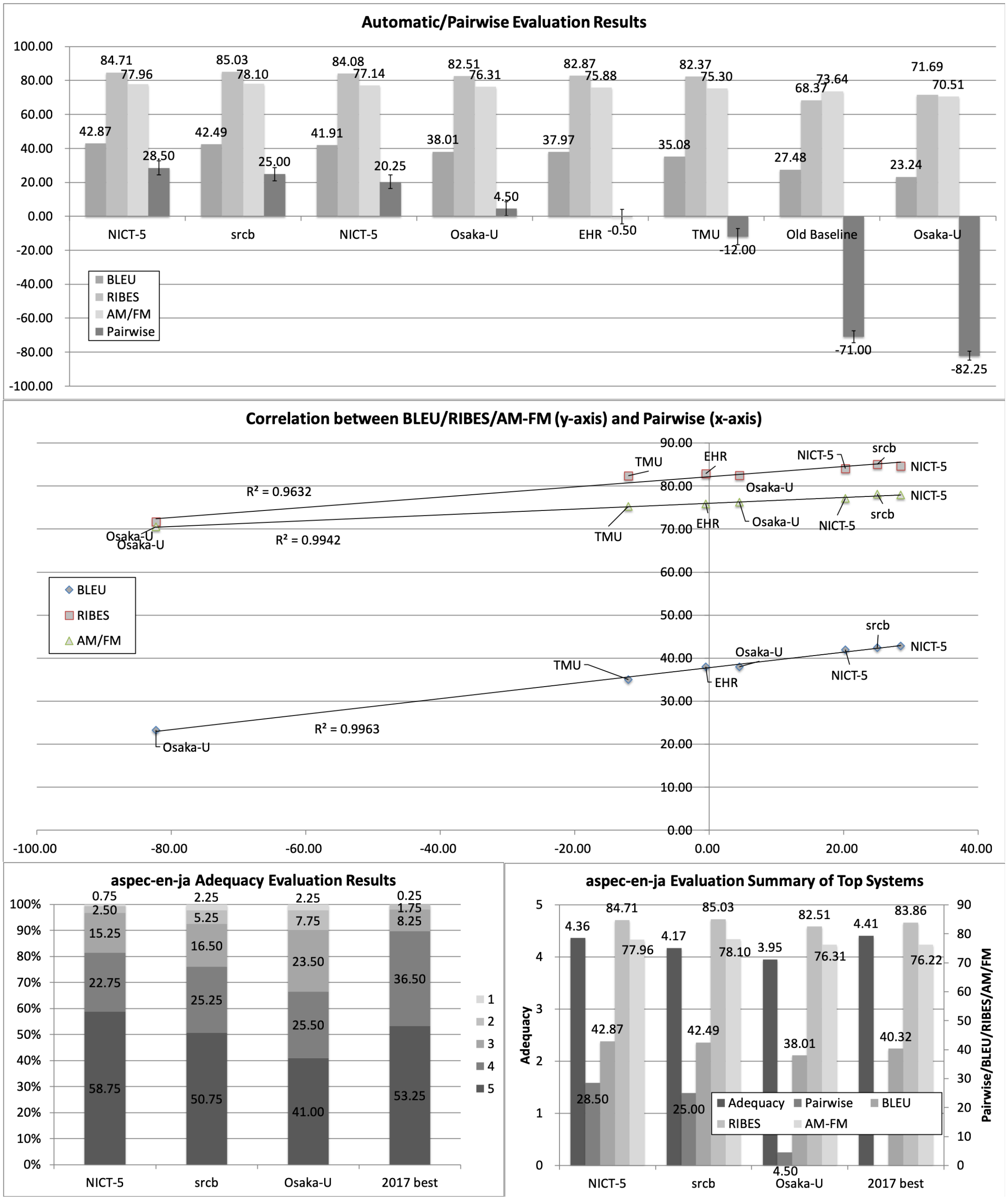


Figure 3: Official evaluation results of aspec-en-ja.

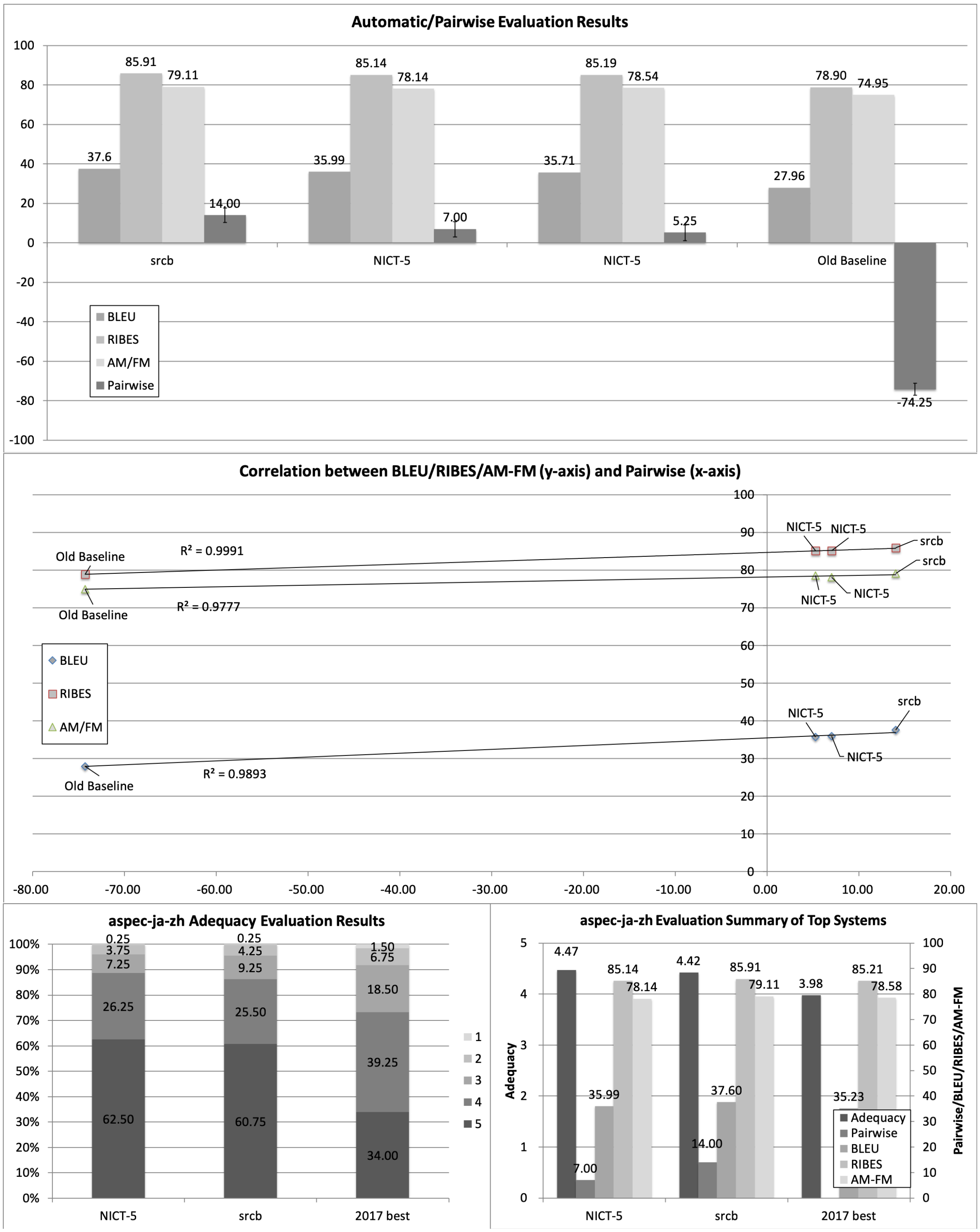


Figure 4: Official evaluation results of aspec-ja-zh.

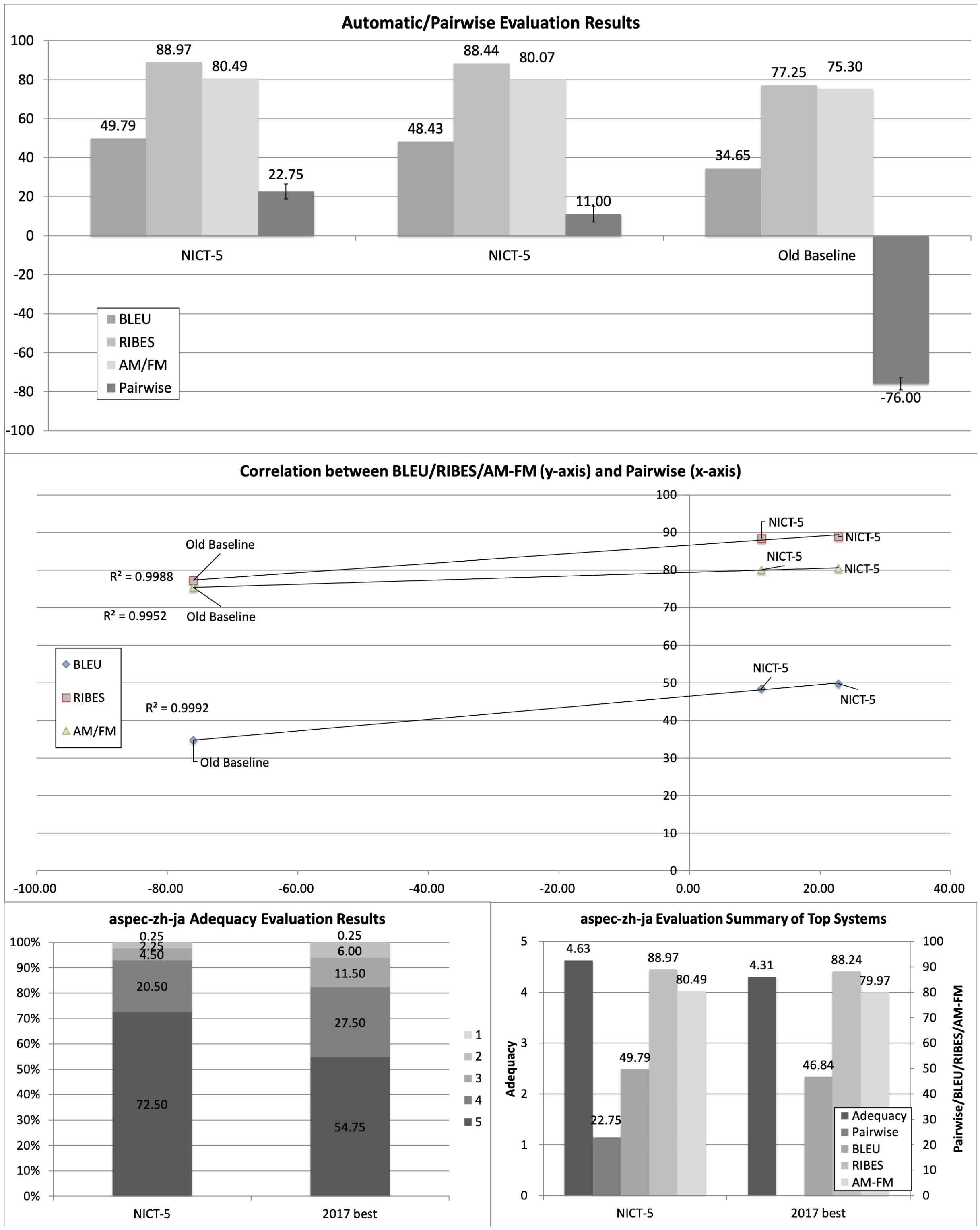


Figure 5: Official evaluation results of aspec-zh-ja.

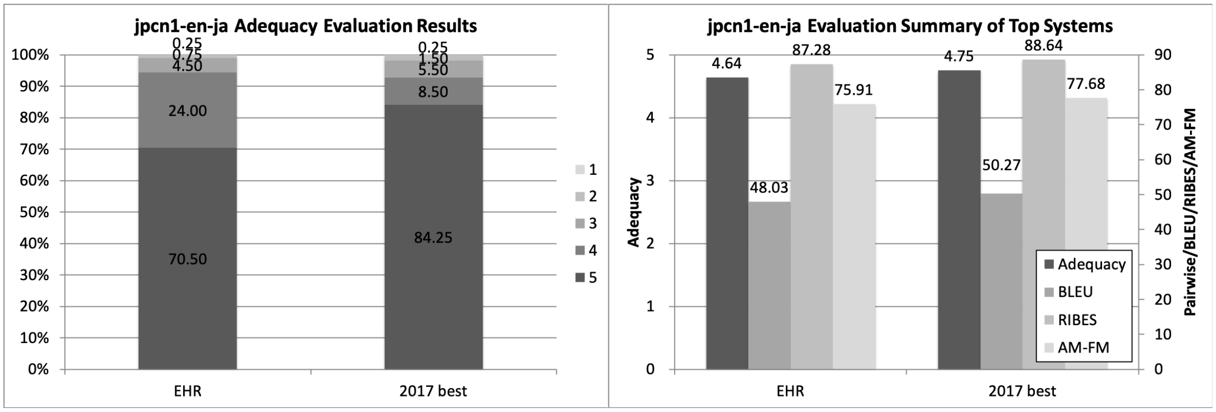


Figure 6: Official evaluation results of jpcn1-en-ja.

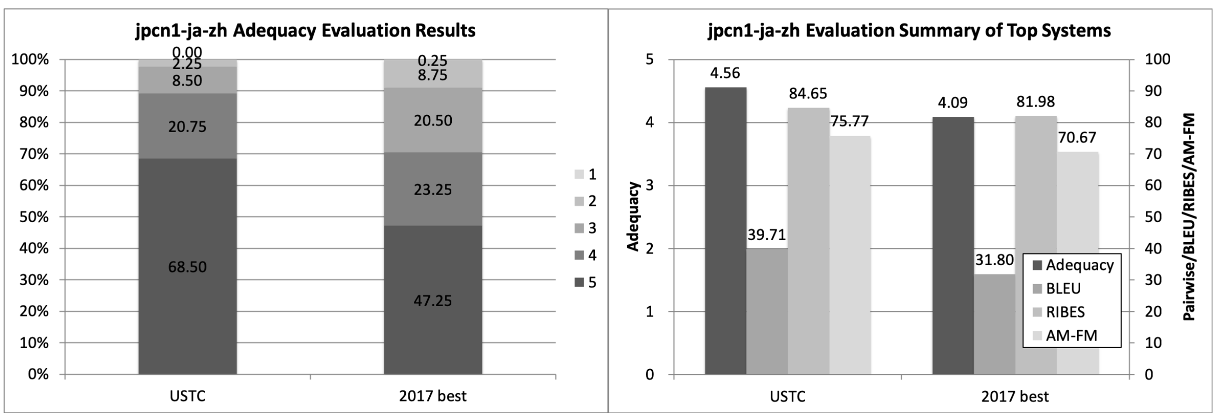


Figure 7: Official evaluation results of jpcn1-ja-zh.

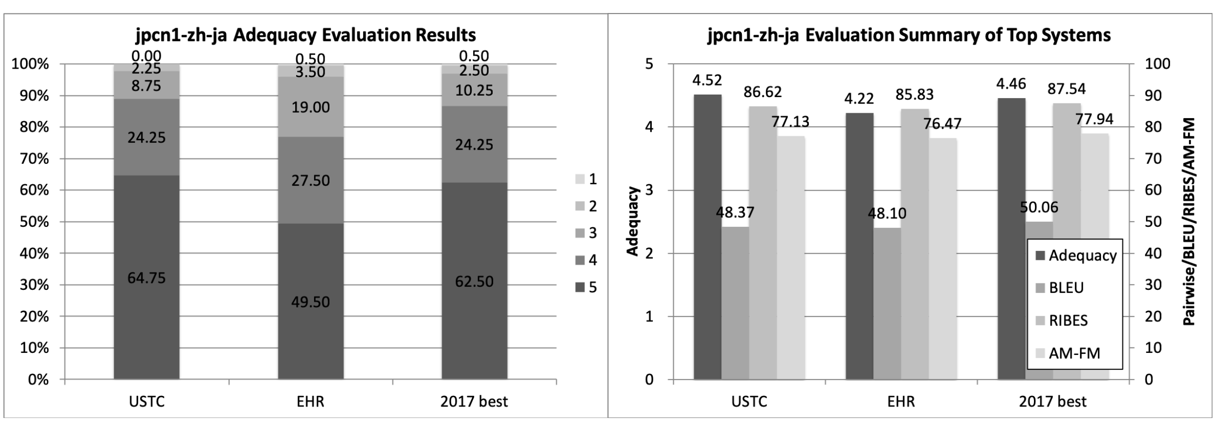


Figure 8: Official evaluation results of jpcn1-zh-ja.

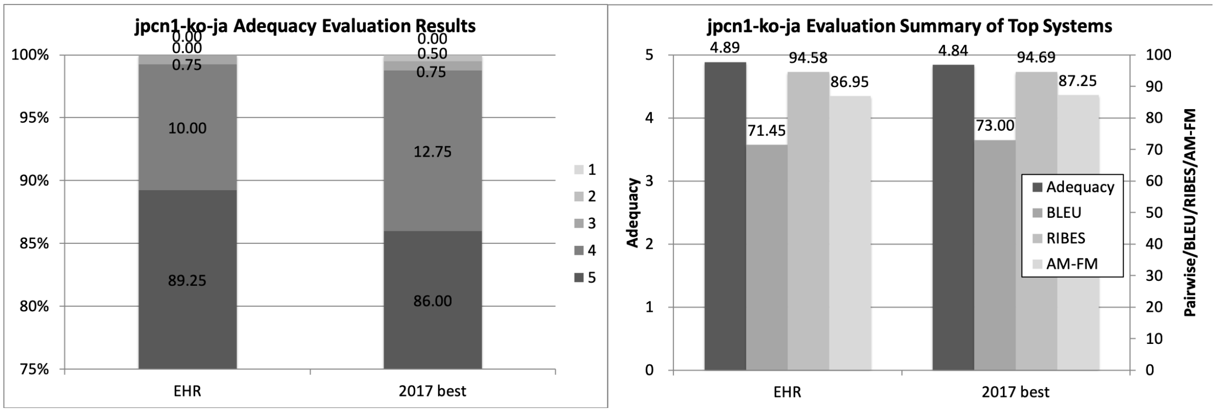


Figure 9: Official evaluation results of jpcn1-ko-ja.

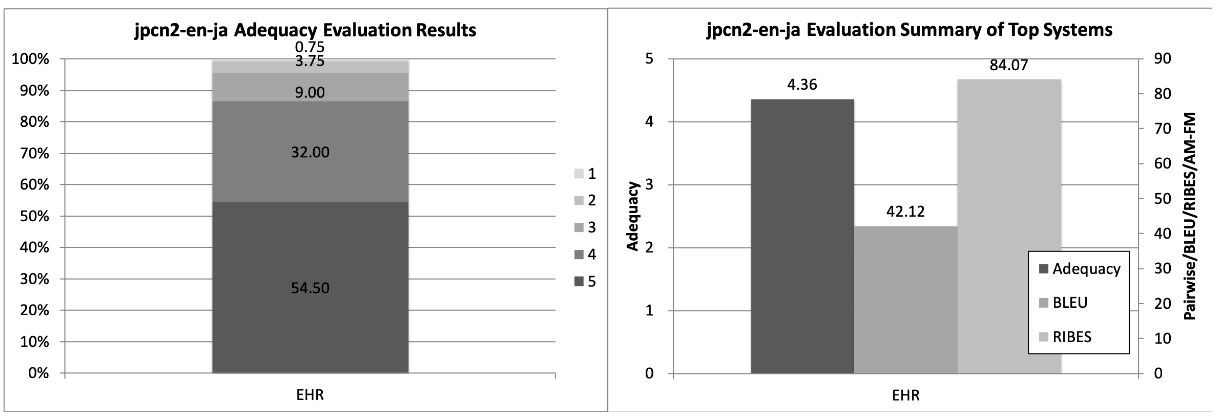


Figure 10: Official evaluation results of jpcn2-en-ja.

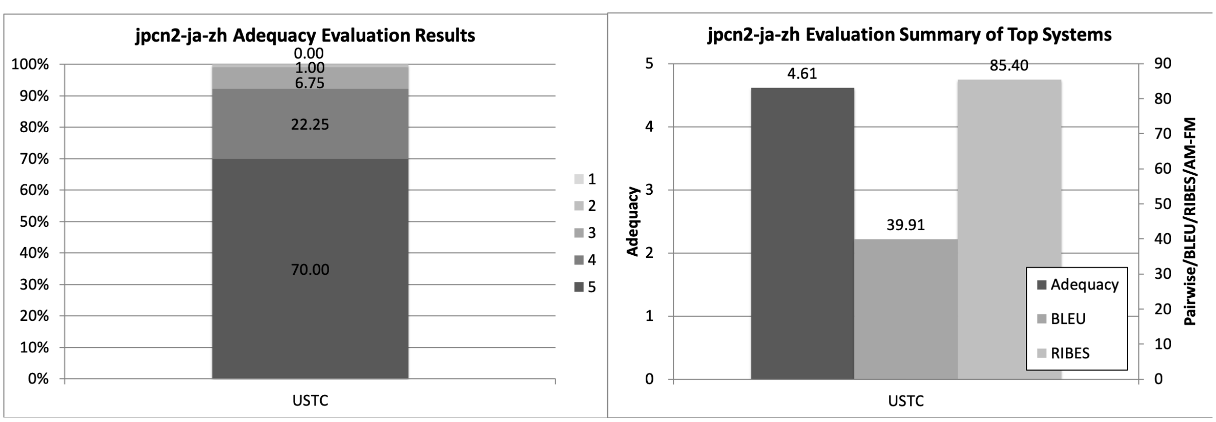
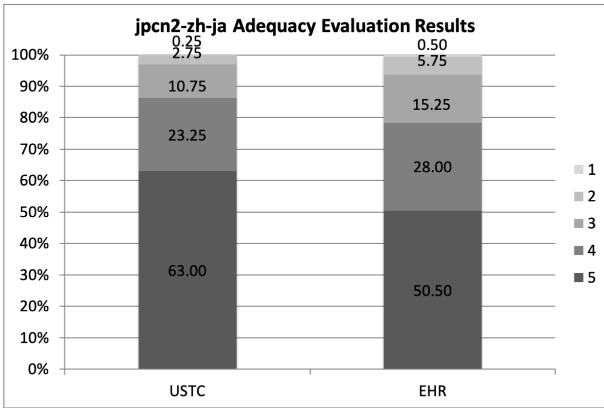


Figure 11: Official evaluation results of jpcn2-ja-zh.



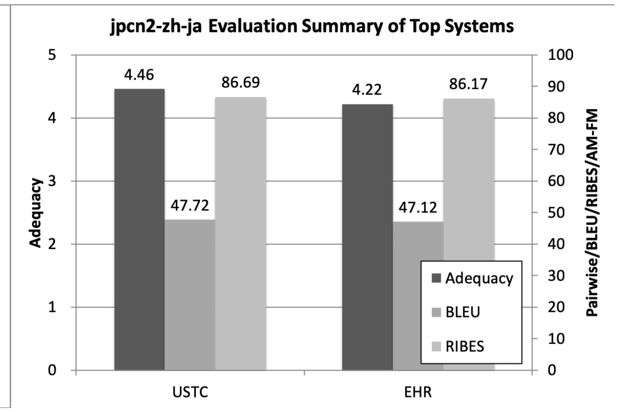
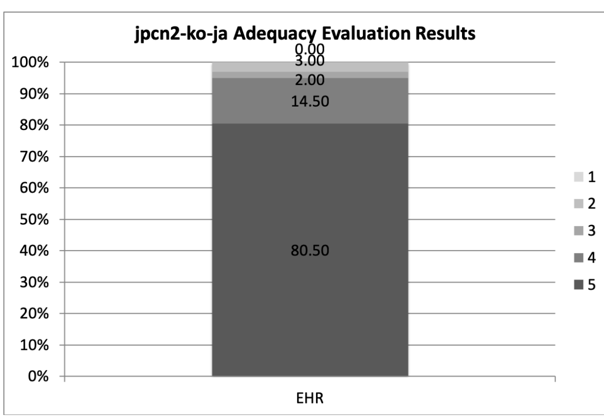


Figure 12: Official evaluation results of jpcn2-zh-ja.



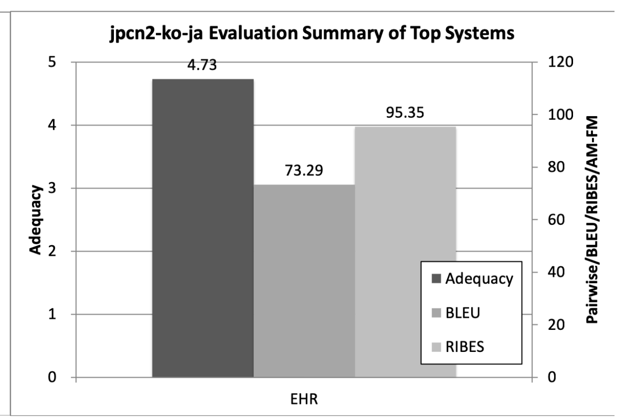


Figure 13: Official evaluation results of jpcn2-ko-ja.

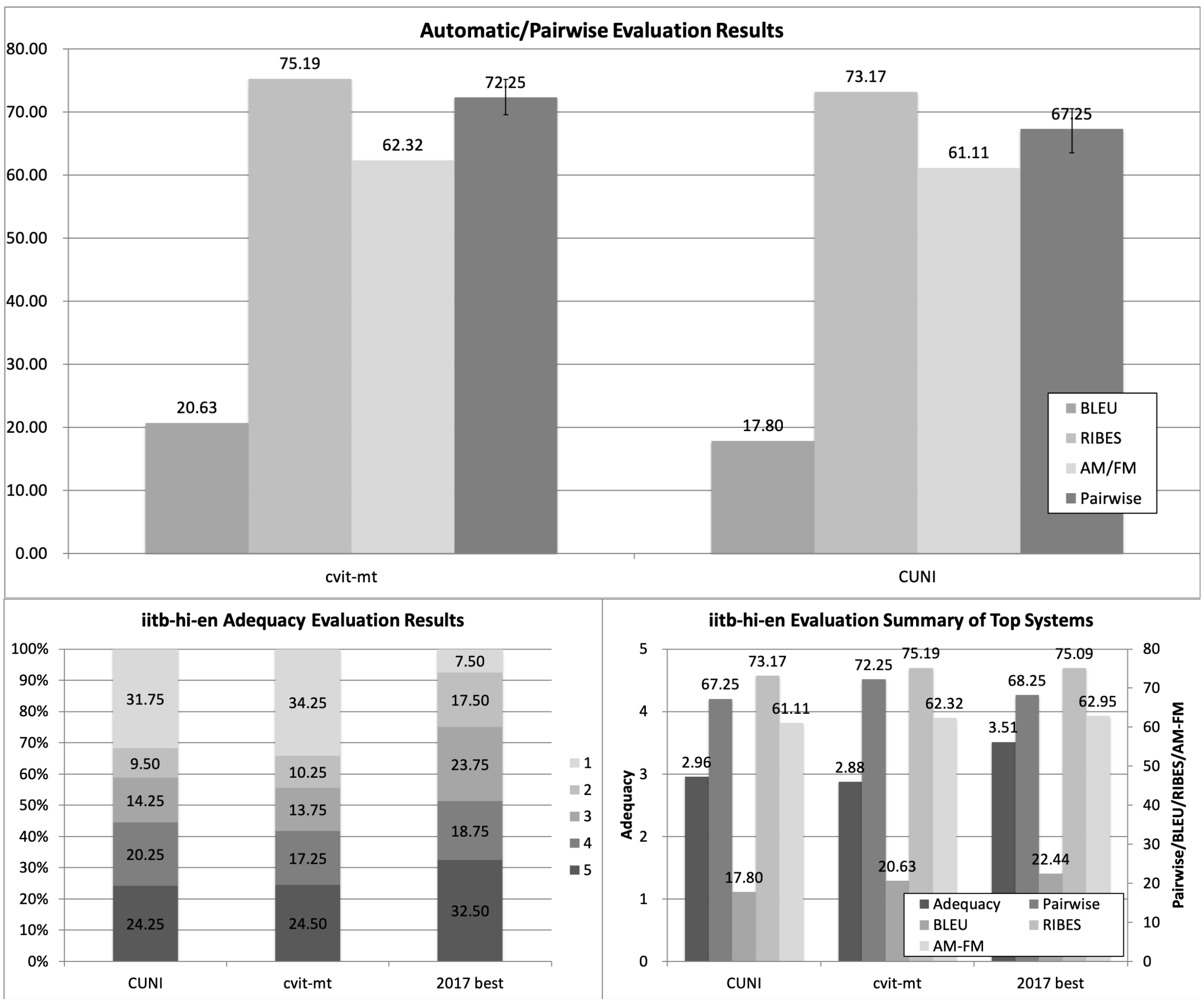


Figure 14: Official evaluation results of iitb-hi-en.

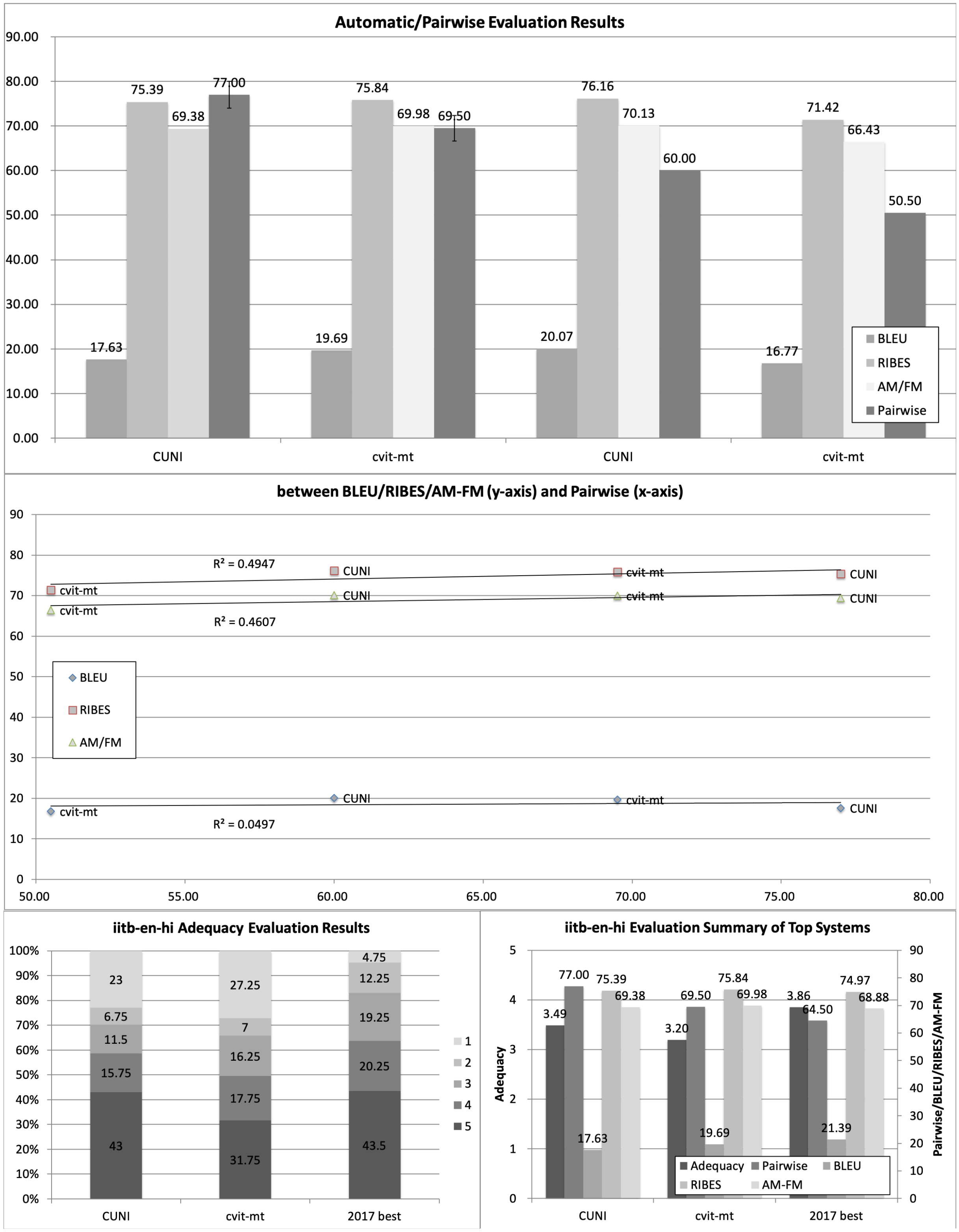


Figure 15: Official evaluation results of iitb-en-hi.

32nd Pacific Asia Conference on Language, Information and Computation
The 5th Workshop on Asian Translation
Hong Kong, 1-3 December 2018

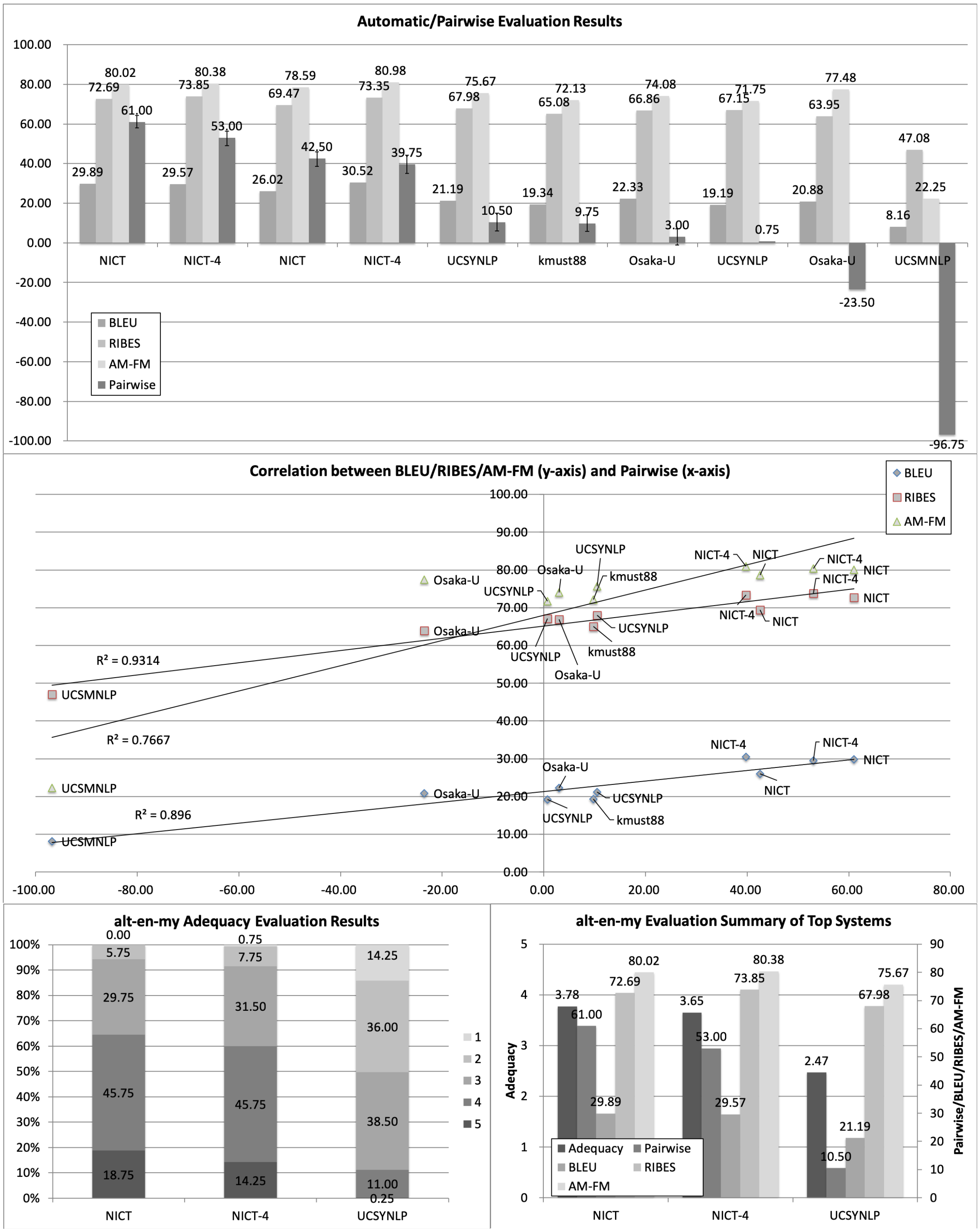


Figure 16: Official evaluation results of alt-en-my.

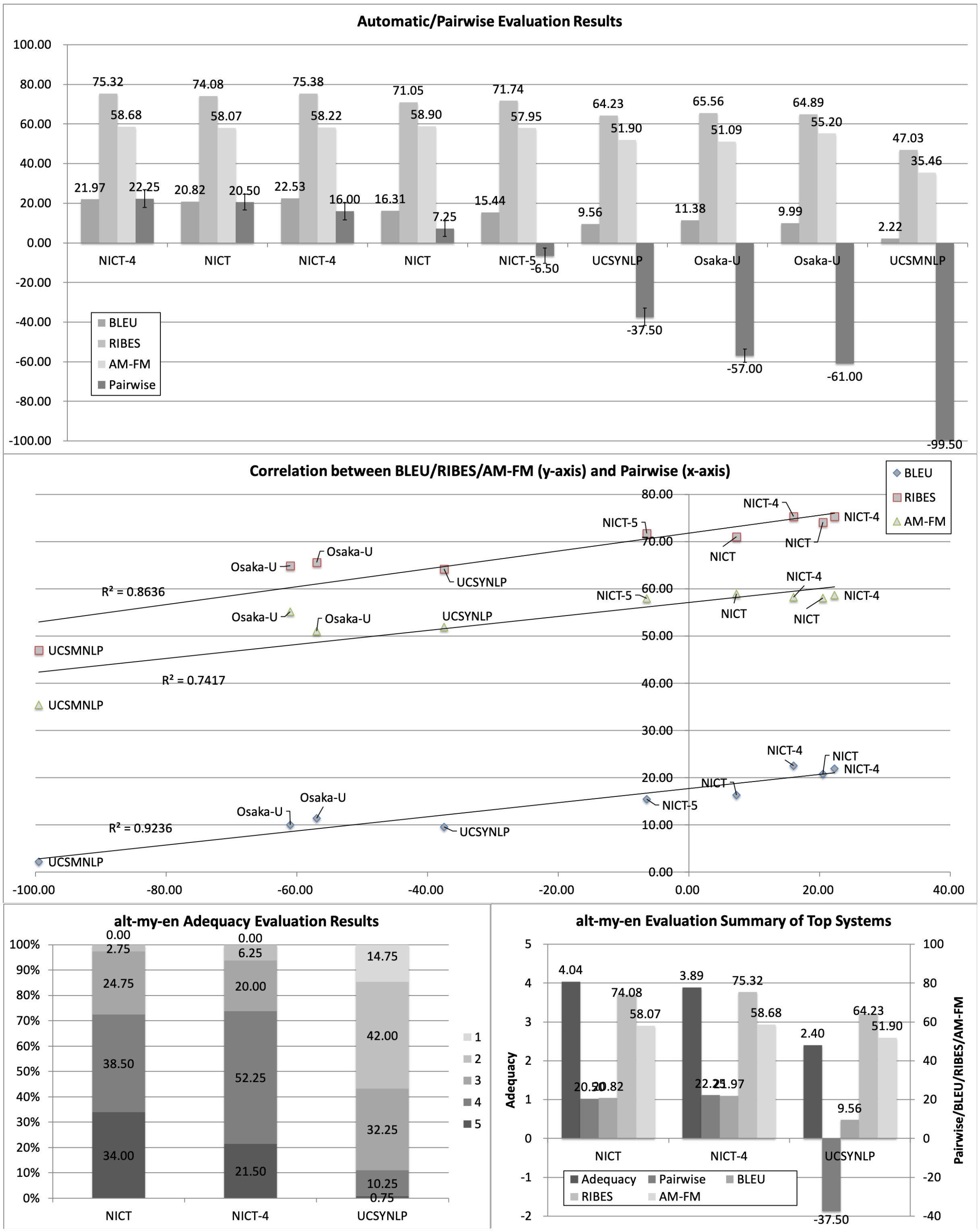


Figure 17: Official evaluation results of alt-my-en.

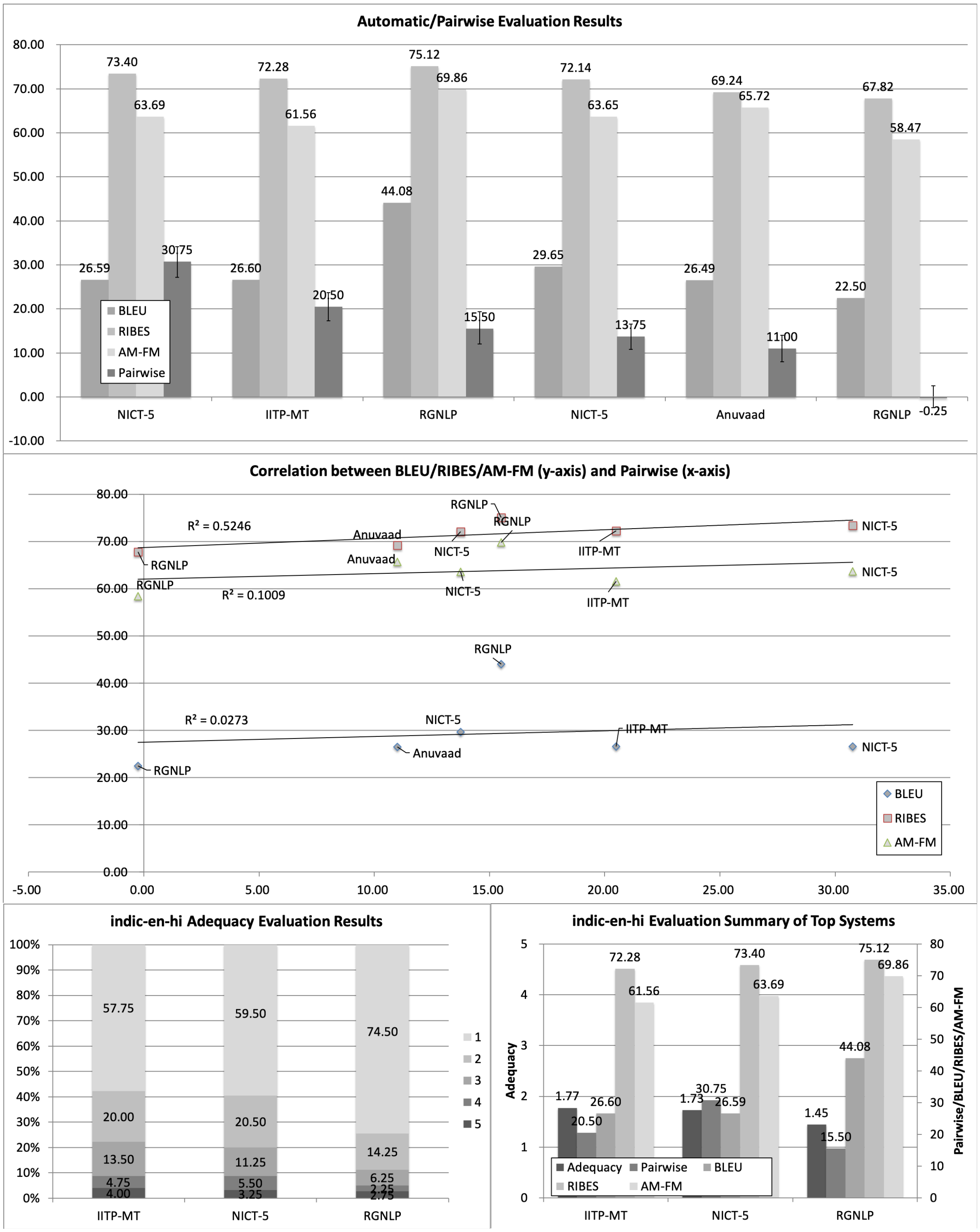


Figure 18: Official evaluation results of indic-en-hi.

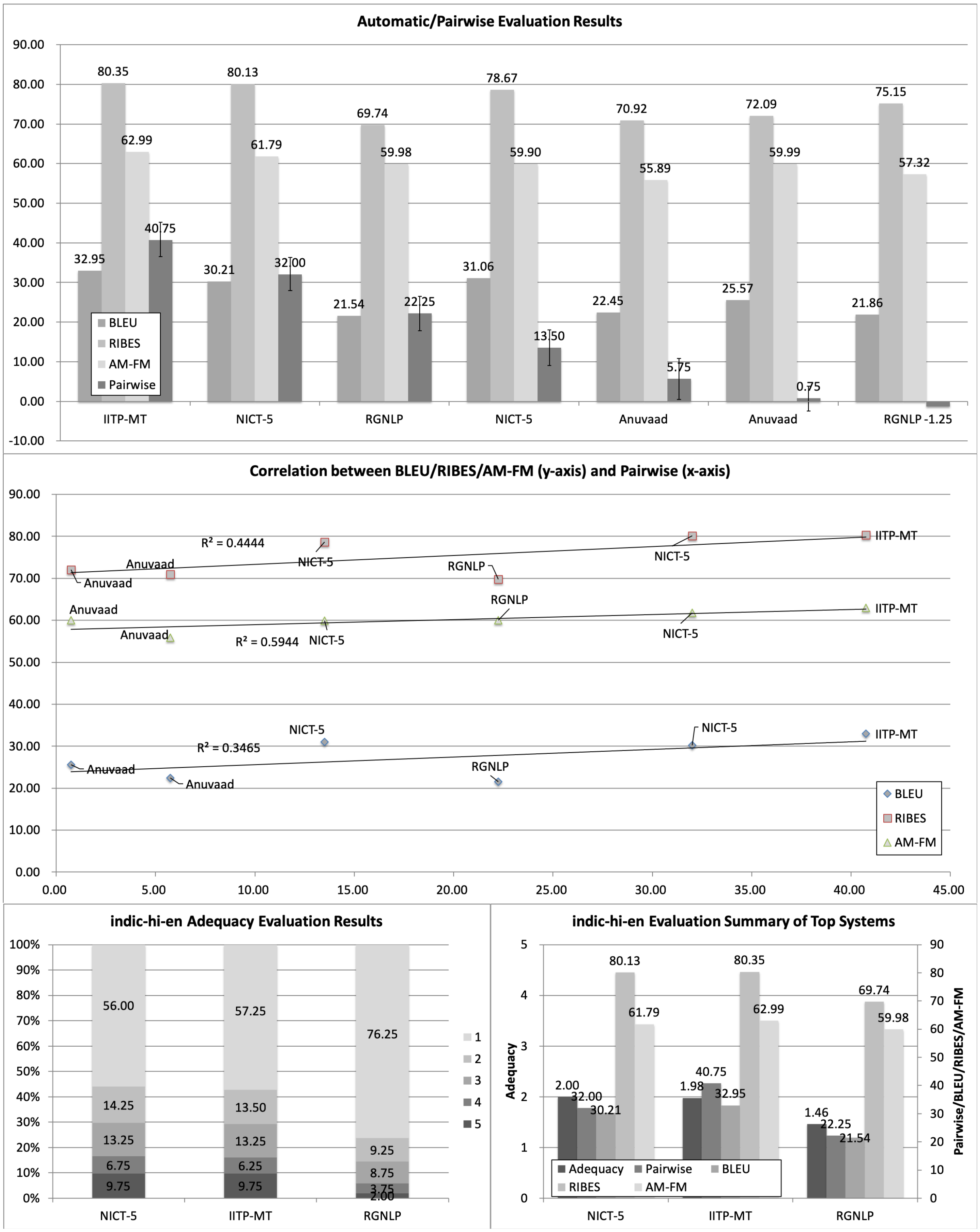


Figure 19: Official evaluation results of indic-hi-en.

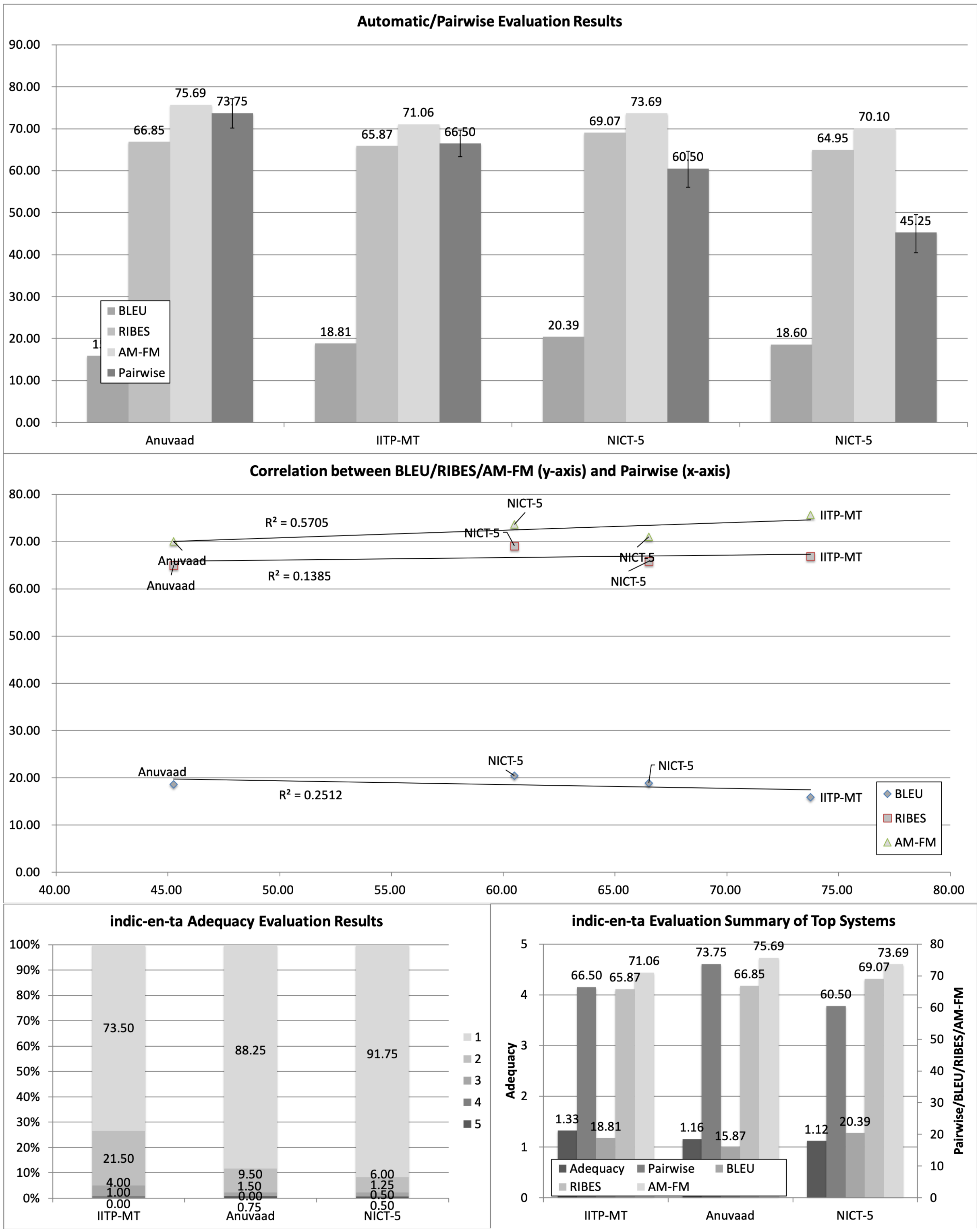


Figure 20: Official evaluation results of indic-en-ta.

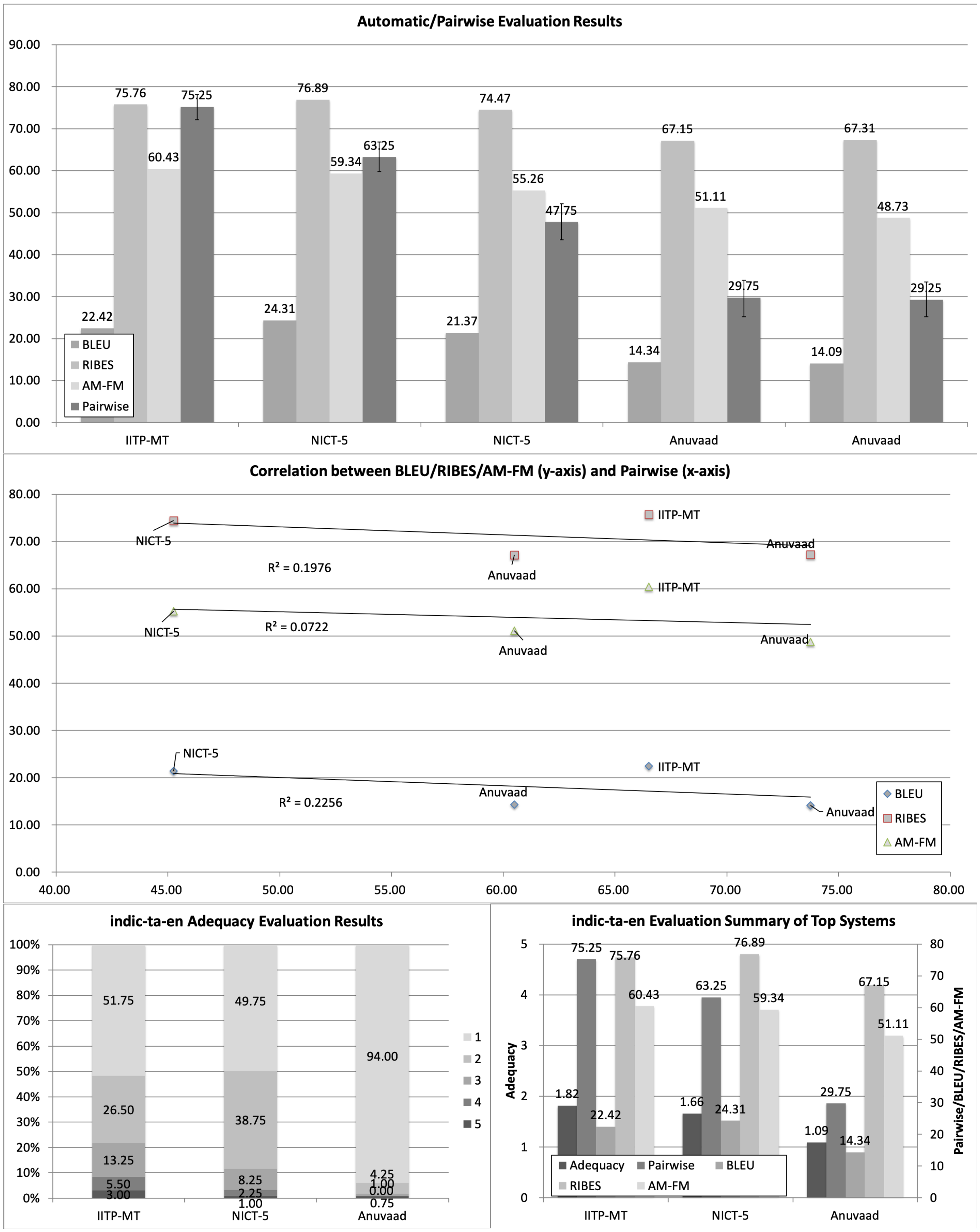


Figure 21: Official evaluation results of indic-ta-en.

Subtask	SYSTEM ID	DATA ID	Annotator A average	Annotator A variance	Annotator B average	Annotator B variance	all average	κ	weighted κ
aspec-ja-en	srcb	2474	4.37	0.49	4.63	0.44	4.50	0.15	0.25
	NICT-5	2174	4.37	0.61	4.60	0.51	4.49	0.26	0.32
	TMU	2464	3.94	0.91	3.92	1.41	3.94	0.34	0.48
	2017 best	1681	4.15	0.58	4.13	0.52	4.14	0.29	0.41
aspec-en-ja	NICT-5	2219	4.16	0.90	4.57	0.57	4.36	0.17	0.30
	srcb	2479	4.04	1.07	4.30	1.00	4.17	0.22	0.38
	Osaka-U	2439	3.74	1.34	4.17	0.88	3.95	0.25	0.42
	2017 best	1729	4.54	0.56	4.28	0.49	4.41	0.33	0.43
aspec-ja-zh	NICT-5	2266	4.67	0.32	4.27	0.90	4.47	0.28	0.36
	srcb	2473	4.69	0.30	4.16	0.98	4.42	0.19	0.24
	2017 best	1483	4.25	0.73	3.71	0.98	3.98	0.10	0.18
aspec-zh-ja	NICT-5	2267	4.78	0.26	4.48	0.67	4.63	0.31	0.33
	2017 best	1481	4.63	0.47	3.99	0.98	4.31	0.17	0.23
jpcn1-en-ja	EHR	2476	4.66	0.35	4.62	0.45	4.64	0.36	0.44
	2017 best	1454	4.74	0.45	4.76	0.38	4.75	0.32	0.48
jpcn1-ja-zh	USTC	2202	4.66	0.44	4.46	0.65	4.55	0.38	0.48
	2017 best	1465	3.99	1.12	4.19	0.94	4.09	0.22	0.32
jpcn1-zh-ja	USTC	2206	4.60	0.43	4.43	0.68	4.51	0.34	0.43
	EHR	2210	4.29	0.71	4.14	0.92	4.22	0.46	0.57
	2017 best	1484	4.41	0.68	4.51	0.64	4.46	0.26	0.34
jpcn1-ko-ja	EHR	2215	4.88	0.13	4.89	0.11	4.88	0.53	0.56
	2017 best	1448	4.82	0.24	4.87	0.11	4.84	0.55	0.55
jpcn2-en-ja	EHR	2477	4.32	0.72	4.40	0.73	4.36	0.35	0.50
jpcn2-ja-zh	USTC	2203	4.71	0.33	4.52	0.52	4.61	0.38	0.45
jpcn2-zh-ja	USTC	2207	4.54	0.48	4.38	0.82	4.46	0.42	0.56
	EHR	2211	4.37	0.65	4.08	1.07	4.22	0.33	0.46
jpcn2-ko-ja	EHR	2216	4.77	0.36	4.68	0.48	4.72	0.62	0.72
iitb-hi-en	CUNI	2381	2.96	2.55	2.96	2.52	2.96	0.48	0.76
	cvit-mt	2331	2.87	2.54	2.88	2.68	2.88	0.53	0.76
	2017 best	1511	3.43	1.64	3.60	1.74	3.51	0.22	0.45
iitb-en-hi	CUNI	2362	3.58	2.71	3.40	2.52	3.49	0.52	0.74
	cvit-mt	2254	3.21	2.58	3.18	2.56	3.20	0.64	0.81
	2017 best	1576	3.95	1.18	3.76	1.85	3.86	0.17	0.36
alt-en-my	NICT	2345	3.51	0.65	4.04	0.54	3.77	0.03	0.08
	NICT-4	2087	3.69	0.70	3.62	0.73	3.65	0.09	0.12
	UCSYNLP	2339	2.04	0.60	2.90	0.56	2.47	0.01	0.07
alt-my-en	NICT	2329	4.19	0.61	3.88	0.73	4.04	0.13	0.23
	NICT-4	2069	4.13	0.35	3.65	0.84	3.89	-0.00	0.07
	UCSYNLP	2332	2.09	0.66	2.71	0.71	2.40	0.06	0.18
indic-en-hi	IITP-MT	2354	1.92	1.35	1.63	1.03	1.77	0.27	0.48
	NICT-5	2128	1.88	1.23	1.57	1.02	1.73	0.29	0.57
	RGNLP	2417	1.46	0.82	1.44	0.86	1.45	0.48	0.67
indic-hi-en	NICT-5	2129	2.12	1.86	1.89	1.79	2.00	0.43	0.71
	IITP-MT	2347	2.02	1.91	1.94	1.75	1.98	0.43	0.67
	RGNLP	2367	1.46	0.92	1.47	0.86	1.46	0.54	0.75
indic-en-ta	IITP-MT	2356	1.36	0.42	1.28	0.29	1.32	0.33	0.43
	Anuvaad	2443	1.20	0.27	1.11	0.23	1.16	0.21	0.37
	NICT-5	2132	1.10	0.17	1.14	0.27	1.12	0.39	0.51
indic-ta-en	IITP-MT	2349	1.84	1.32	1.78	0.89	1.81	0.23	0.39
	NICT-5	2133	1.74	0.72	1.58	0.55	1.66	0.16	0.23
	Anuvaad	2400	1.15	0.33	1.03	0.05	1.09	0.10	0.14

Table 14: JPO adequacy evaluation results in detail.

	NICT-5 (2273)	srcb (2474)	TMU (2464)	Osaka-U (2440)	ORGANIZER (0006)	Osaka-U (2472)
NICT-5 (2174)	⋙	⋙	⋙	⋙	⋙	⋙
NICT-5 (2273)		⋙	⋙	⋙	⋙	⋙
srcb (2474)			⋙	⋙	⋙	⋙
TMU (2464)				⋙	⋙	⋙
Osaka-U (2440)					⋙	⋙
ORGANIZER (0006)						⋙

	srcb (2479)	NICT-5 (2048)	Osaka-U (2439)	EHR (2245)	TMU (2469)	ORGANIZER (0005)	Osaka-U (2470)
NICT-5 (2219)	>	⋙	⋙	⋙	⋙	⋙	⋙
srcb (2479)		⋙	⋙	⋙	⋙	⋙	⋙
NICT-5 (2048)			⋙	⋙	⋙	⋙	⋙
Osaka-U (2439)				>	⋙	⋙	⋙
EHR (2245)					⋙	⋙	⋙
TMU (2469)						⋙	⋙
ORGANIZER (0005)							⋙

Table 15: Statistical significance testing of the aspec-ja-en (left) and aspec-en-ja (right) Pairwise scores.

	NICT-5 (2266)	NICT-5 (2175)	ORGANIZER (0007)
srcb (2473)	⋙	⋙	⋙
NICT-5 (2266)		-	⋙
NICT-5 (2175)			⋙

	NICT-5 (2052)	ORGANIZER (0008)
NICT-5 (2267)	⋙	⋙
NICT-5 (2052)		⋙

Table 16: Statistical significance testing of the aspec-ja-zh (left) and aspec-zh-ja (right) Pairwise scores.

	cvit-mt (2254)	CUNI (2365)	cvit-mt (2251)
CUNI (2362)	⋙	⋙	⋙
cvit-mt (2254)		⋙	⋙
CUNI (2365)			⋙

	CUNI (2381)
cvit-mt (2331)	⋙

Table 17: Statistical significance testing of the iitb-en-hi (left) and iitb-hi-en (right) Pairwise scores.

	NICT-4 (2087)	NICT (2282)	NICT-4 (2287)	UCSYNLP (2339)	kmust88 (2360)	Osaka-U (2437)	UCSYNLP (2340)	Osaka-U (2471)	UCSMNLP (2337)
NICT (2345)	⋙	⋙	⋙	⋙	⋙	⋙	⋙	⋙	⋙
NICT-4 (2087)		⋙	⋙	⋙	⋙	⋙	⋙	⋙	⋙
NICT (2282)			-	⋙	⋙	⋙	⋙	⋙	⋙
NICT-4 (2287)				⋙	⋙	⋙	⋙	⋙	⋙
UCSYNLP (2339)					-	⋙	⋙	⋙	⋙
kmust88 (2360)						⋙	⋙	⋙	⋙
Osaka-U (2437)							-	⋙	⋙
UCSYNLP (2340)								⋙	⋙
Osaka-U (2471)									⋙

	NICT (2329)	NICT-4 (2290)	NICT (2281)	NICT-5 (2056)	UCSYNLP (2332)	Osaka-U (2438)	Osaka-U (2463)	UCSMNLP (2338)
NICT-4 (2069)	-	≫	⋙	⋙	⋙	⋙	⋙	⋙
NICT (2329)		>	⋙	⋙	⋙	⋙	⋙	⋙
NICT-4 (2290)			⋙	⋙	⋙	⋙	⋙	⋙
NICT (2281)				⋙	⋙	⋙	⋙	⋙
NICT-5 (2056)					⋙	⋙	⋙	⋙
UCSYNLP (2332)						⋙	⋙	⋙
Osaka-U (2438)							≫	⋙
Osaka-U (2463)								⋙

Table 18: Statistical significance testing of the alt-en-my (left) and alt-my-en (right) Pairwise scores.

	IITP-MT (2354)	RGNLP (2417)	NICT-5 (2067)	Anuvaad (2445)	RGNLP (2422)
NICT-5 (2128)	⋙	⋙	⋙	⋙	⋙
IITP-MT (2354)		≫	⋙	⋙	⋙
RGNLP (2417)			-	≫	⋙
NICT-5 (2067)				>	⋙
Anuvaad (2445)					⋙

	NICT-5 (2129)	RGNLP (2367)	NICT-5 (2066)	Anuvaad (2406)	Anuvaad (2403)	RGNLP (2383)
IITP-MT (2347)	⋙	⋙	⋙	⋙	⋙	⋙
NICT-5 (2129)		⋙	⋙	⋙	⋙	⋙
RGNLP (2367)			⋙	⋙	⋙	⋙
NICT-5 (2066)				⋙	⋙	⋙
Anuvaad (2406)					≫	⋙
Anuvaad (2403)						-

Table 19: Statistical significance testing of the indic-en-hi (left) and indic-hi-en (right) Pairwise scores.

	IITP-MT (2356)	NICT-5 (2132)	NICT-5 (2109)
Anuvaad (2443)	⋙	⋙	⋙
IITP-MT (2356)		⋙	⋙
NICT-5 (2132)			⋙

	NICT-5 (2133)	NICT-5 (2111)	Anuvaad (2400)	Anuvaad (2408)
IITP-MT (2349)	⋙	⋙	⋙	⋙
NICT-5 (2133)		⋙	⋙	⋙
NICT-5 (2111)			⋙	⋙
Anuvaad (2400)				-

Table 20: Statistical significance testing of the indic-en-ta (left) and indic-ta-en (right) Pairwise scores.

aspec-ja-en		
SYSTEM	DATA	κ
ORGANIZER	0006	0.216
TMU	2464	0.201
srcb	2474	0.183
Osaka-U	2440	0.128
Osaka-U	2472	0.130
NICT-5	2174	0.182
NICT-5	2273	0.145
ave.		0.169

aspec-en-ja		
SYSTEM	DATA	κ
ORGANIZER	0005	0.394
TMU	2469	0.450
EHR	2245	0.314
srcb	2479	0.325
Osaka-U	2439	0.302
Osaka-U	2470	0.305
NICT-5	2048	0.324
NICT-5	2219	0.256
ave.		0.334

aspec-ja-zh		
SYSTEM	DATA	κ
ORGANIZER	0007	0.254
srcb	2473	0.150
NICT-5	2175	0.162
NICT-5	2266	0.174
ave.		0.185

aspec-zh-ja		
SYSTEM	DATA	κ
ORGANIZER	0008	0.389
NICT-5	2052	0.266
NICT-5	2267	0.282
ave.		0.312

iitb-en-hi		
SYSTEM	DATA	κ
CUNI	2362	0.358
CUNI	2365	0.454
cvit-mt	2251	0.447
cvit-mt	2254	0.356
ave.		0.404

iitb-hi-en		
SYSTEM	DATA	κ
CUNI	2381	0.404
cvit-mt	2331	0.381
ave.		0.393

alt-en-my		
SYSTEM	DATA	κ
NICT	2282	0.181
NICT	2345	0.091
Osaka-U	2437	0.061
Osaka-U	2471	0.187
NICT-4	2087	0.205
NICT-4	2287	0.262
UCSYNLP	2339	0.268
UCSYNLP	2340	0.303
UCSMNLP	2337	0.212
kmust88	2360	0.275
ave.		0.205

alt-my-en		
SYSTEM	DATA	κ
NICT	2281	0.107
NICT	2329	0.202
Osaka-U	2438	0.153
Osaka-U	2463	0.161
NICT-4	2069	0.284
NICT-4	2290	0.122
NICT-5	2056	0.072
UCSYNLP	2332	0.068
UCSMNLP	2338	0.087
ave.		0.140

indic-en-hi		
SYSTEM	DATA	κ
IITP-MT	2354	0.330
RGNLP	2417	0.386
RGNLP	2422	0.417
NICT-5	2067	0.447
NICT-5	2128	0.341
Anuvaad	2445	0.437
ave.		0.393

indic-hi-en		
SYSTEM	DATA	κ
IITP-MT	2347	0.204
RGNLP	2367	0.252
RGNLP	2383	0.411
NICT-5	2066	0.327
NICT-5	2129	0.263
Anuvaad	2403	0.441
Anuvaad	2406	0.281
ave.		0.311

indic-en-ta		
SYSTEM	DATA	κ
IITP-MT	2356	0.209
NICT-5	2109	0.373
NICT-5	2132	0.443
Anuvaad	2443	0.308
ave.		0.333

indic-ta-en		
SYSTEM	DATA	κ
IITP-MT	2349	0.299
NICT-5	2111	0.256
NICT-5	2133	0.185
Anuvaad	2400	0.159
Anuvaad	2408	0.139
ave.		0.208

Table 21: The Fleiss' kappa values for the pairwise evaluation results.

System	ID	Type	RSRC	BLEU			RIBES			AMFM			Pair	Adeq
				juman	kytea	mecab	juman	kytea	mecab	juman	kytea	mecab		
NMT	1900	NMT	NO	36.37	38.48	37.15	0.824985	0.831183	0.833207	0.759910	0.759910	0.759910	–	–
NICT-5 (1)	2219	NMT	NO	42.87	44.42	43.49	0.847134	0.849399	0.853634	0.779560	0.779560	0.779560	+28.50	4.36
srcb	2479	NMT	NO	42.49	44.11	43.20	0.850318	0.852209	0.857017	0.781000	0.781000	0.781000	+25.00	4.17
NICT-5 (2)	2048	NMT	NO	41.91	43.50	42.60	0.840776	0.845042	0.849326	0.771400	0.771400	0.771400	+20.25	–
Osaka-U (1)	2439	NMT	YES	38.01	40.00	39.10	0.825061	0.829328	0.833200	0.763140	0.763140	0.763140	+4.50	3.95
EHR	2245	NMT	NO	37.97	40.00	38.66	0.828746	0.833333	0.837806	0.758750	0.758750	0.758750	-0.50	–
TMU	2469	NMT	NO	35.08	37.69	36.14	0.823653	0.829156	0.831219	0.753040	0.753040	0.753040	-12.00	–
Osaka-U (2)	2470	SMT	NO	23.24	25.50	24.26	0.716889	0.726469	0.729323	0.705050	0.705050	0.705050	-82.25	–

Table 22: ASPEC en-ja submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	1901	NMT	NO	26.91	0.764968	0.595370	–	–
NICT-5 (1)	2174	NMT	NO	28.63	0.765933	0.608070	+15.75	4.49
NICT-5 (2)	2273	NMT	NO	29.65	0.774788	0.612060	+11.50	–
srcb	2474	NMT	NO	30.59	0.777896	0.619390	+5.75	4.50
TMU	2464	NMT	NO	25.85	0.761450	0.600730	-20.00	3.94
Osaka-U (1)	2440	NMT	YES	26.19	0.749825	0.588290	-37.00	–
Osaka-U (2)	2472	SMT	NO	13.97	0.665391	0.571400	-95.75	–

Table 23: ASPEC ja-en submissions

System	ID	Type	RSRC	BLEU			RIBES			AMFM			Pair	Adeq
				juman	kytea	mecab	juman	kytea	mecab	juman	kytea	mecab		
NMT	1902	NMT	NO	43.31	43.53	43.34	0.870734	0.866281	0.870886	0.782100	0.782100	0.782100	–	–
NICT-5 (1)	2267	NMT	NO	49.79	50.66	49.89	0.889674	0.886490	0.889853	0.804920	0.804920	0.804920	+22.75	4.63
NICT-5 (2)	2052	NMT	NO	48.43	48.78	48.52	0.884426	0.879456	0.884782	0.800670	0.800670	0.800670	+11.00	–

Table 24: ASPEC zh-ja submissions

System	ID	Type	RSRC	BLEU			RIBES			AMFM			Pair	Adeq
				kytea	stanford (ctb)	stanford (pku)	kytea	stanford (ctb)	stanford (pku)	kytea	stanford (ctb)	stanford (pku)		
NMT	1903	NMT	NO	33.26	33.33	33.14	0.844322	0.844572	0.844959	0.777600	0.777600	0.777600	–	–
srcb	2473	NMT	NO	37.60	37.34	37.35	0.859132	0.858042	0.858162	0.791120	0.791120	0.791120	+14.00	4.42
NICT-5 (1)	2266	NMT	NO	35.99	35.89	35.87	0.851382	0.851416	0.850944	0.781410	0.781410	0.781410	+7.00	4.47
NICT-5 (2)	2175	NMT	NO	35.71	35.67	35.55	0.851890	0.850699	0.850580	0.785440	0.785440	0.785440	+5.25	–

Table 25: ASPEC ja-zh submissions

Task	System	ID	Type	RSRC	BLEU			RIBES			AMFM			Adeq
					juman	kytea	mecab	juman	kytea	mecab	juman	kytea	mecab	
N1	NMT	1964	NMT	NO	43.84	45.28	43.70	0.860702	0.857422	0.859818	0.744270	0.744270	0.744270	–
	EHR	2476	NMT	YES	48.03	49.24	47.86	0.872828	0.870332	0.872442	0.759120	0.759120	0.759120	4.76
N2	NMT	1936	NMT	NO	38.51	40.60	38.47	0.825565	0.824420	0.824770	–	–	–	–
	EHR	2477	NMT	YES	42.12	43.76	42.06	0.840713	0.839052	0.841133	–	–	–	4.36

Table 26: JPC N1/N2 en-ja submissions

Task	System	ID	Type	RSRC	BLEU			RIBES			AMFM			Adeq
					juman	kytea	mecab	juman	kytea	mecab	juman	kytea	mecab	
N1	NMT	1966	NMT	NO	71.42	72.21	71.68	0.945593	0.944644	0.945126	0.868760	0.868760	0.868760	–
	EHR	2215	NMT	NO	71.45	72.35	71.74	0.945771	0.944284	0.945497	0.869540	0.869540	0.869540	4.89
N2	NMT	1947	NMT	NO	70.65	71.46	70.94	0.942943	0.942543	0.943101	–	–	–	–
	EHR	2216	NMT	NO	73.29	74.25	73.61	0.953524	0.952935	0.953640	–	–	–	4.73

Table 27: JPC N1/N2 ko-ja submissions

Task	System	ID	Type	RSRC	BLEU			RIBES			AMFM			Adeq
					juman	kytea	mecab	juman	kytea	mecab	juman	kytea	mecab	
N1	NMT	1963	NMT	NO	46.32	46.73	46.11	0.857318	0.855085	0.856442	0.761820	0.761820	0.761820	
	USTC	2206	NMT	NO	48.37	49.78	48.57	0.866232	0.864284	0.865423	0.771310	0.771310	0.771310	4.52
	EHR	2210	NMT	NO	48.10	48.51	47.96	0.858259	0.855649	0.858142	0.764670	0.764670	0.764670	4.22
N2	NMT	1941	NMT	NO	45.33	46.05	45.49	0.857120	0.854593	0.857052	–	–	–	–
	USTC	2207	NMT	NO	47.72	49.45	48.24	0.866873	0.865270	0.866705	–	–	–	4.46
	EHR	2211	NMT	NO	47.12	47.71	47.14	0.861697	0.859213	0.861437	–	–	–	4.22

Table 28: JPC N1/N2 zh-ja submissions

Task	System	ID	Type	RSRC	BLEU			RIBES			AMFM			Adeq
					kytea	stanford (ctb)	stanford (pku)	kytea	stanford (ctb)	stanford (pku)	kytea	stanford (ctb)	stanford (pku)	
N1	NMT	1960	NMT	NO	39.07	40.32	39.75	0.847112	0.850851	0.850913	0.752360	0.752360	0.752360	–
	USTC	2202	NMT	NO	39.71	40.54	40.05	0.846472	0.850750	0.849818	0.757690	0.757690	0.757690	4.56
N2	NMT	1961	NMT	NO	39.14	40.28	39.77	0.847486	0.852161	0.850707	–	–	–	–
	USTC	2203	NMT	NO	39.91	40.53	40.32	0.853978	0.859330	0.857456	–	–	–	4.61

Table 29: JPC N1/N2 ja-zh submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2566	NMT	NO	13.76	0.710210	0.644860	–	–
CUNI (1)	2362	NMT	NO	17.63	0.753895	0.693830	+77.00	3.49
cvit-mt (1)	2254	NMT	YES	19.69	0.758365	0.699810	+69.50	3.20
CUNI (2)	2365	NMT	NO	20.07	0.761582	0.701300	+60.00	–
cvit-mt (2)	2251	NMT	NO	16.77	0.714197	0.664330	+50.50	–

Table 30: IITB en-hi submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2567	NMT	NO	15.44	0.718751	0.586360	–	–
cvit-mt	2331	NMT	YES	20.63	0.751883	0.623240	+72.25	2.88
CUNI	2381	NMT	NO	17.80	0.731727	0.611090	+67.25	2.96

Table 31: IITB hi-en submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
Online-A	2142	Other	YES	20.31	0.678360	0.587120	–	–
Online-A (comma→0x104a)	2143	Other	YES	20.83	0.679968	0.594230	–	–
NMT	2227	NMT	NO	22.42	0.667437	0.745550	–	–
NICT (1)	2345	NMT	NO	29.89	0.726922	0.800230	+61.00	3.78
NICT-4 (1)	2087	NMT	NO	29.57	0.738538	0.803810	+53.00	3.65
NICT (2)	2282	NMT	NO	26.02	0.694652	0.785920	+42.50	–
NICT-4 (2)	2287	Other	NO	30.52	0.733501	0.809750	+39.75	–
UCSYNLP (1)	2339	NMT	NO	21.19	0.679800	0.756710	+10.50	2.47
kmust88	2360	NMT	NO	19.34	0.650796	0.721280	+9.75	–
Osaka-U (1)	2437	NMT	YES	22.33	0.668596	0.740760	+3.00	–
UCSYNLP (2)	2340	NMT	NO	19.19	0.671461	0.717480	+0.75	–
Osaka-U (2)	2471	SMT	NO	20.88	0.639517	0.774750	-23.50	–
UCSMNLP	2337	SMT	NO	8.16	0.470758	0.222510	-96.75	–

Table 32: ALT en-my submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
Online A	2141	Other	YES	14.24	0.598345	0.576780	–	–
NMT	2228	NMT	NO	14.44	0.696861	0.525950	–	–
NICT-4 (1)	2069	NMT	NO	21.97	0.753209	0.586770	+22.25	3.89
NICT (1)	2329	NMT	NO	20.82	0.740819	0.580690	+20.50	4.04
NICT-4 (2)	2290	Other	NO	22.53	0.753767	0.582230	+16.00	–
NICT (2)	2281	NMT	NO	16.31	0.710528	0.589020	+7.25	–
NICT-5	2056	NMT	NO	15.44	0.717430	0.579520	-6.50	–
UCSYNLP	2332	NMT	NO	9.56	0.642309	0.518990	-37.50	2.40
Osaka-U (1)	2438	NMT	YES	11.38	0.655643	0.510900	-57.00	–
Osaka-U (2)	2463	NMT	NO	9.99	0.648923	0.552040	-61.00	–
UCSMNLP	2338	SMT	NO	2.22	0.470280	0.354550	-99.50	–

Table 33: ALT my-en submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2003	NMT	NO	20.78	0.682944	0.574960	–	–
NMT O2M	2146	NMT	NO	23.27	0.709586	0.601800	–	–
NMT M2M	2188	NMT	NO	22.24	0.705747	0.604040	–	–
NICT-5 (1)	2128	NMT	NO	26.59	0.734027	0.636900	+30.75	1.73
IITP-MT	2354	NMT	NO	26.60	0.722756	0.615620	+20.50	1.77
RGNLP (1)	2417	SMT	NO	44.08	0.751187	0.698550	+15.50	1.45
NICT-5 (2)	2067	NMT	NO	29.65	0.721379	0.636500	+13.75	–
Anuvaad	2445	SMT	NO	26.49	0.692385	0.657180	+11.00	–
RGNLP (2)	2422	NMT	NO	22.50	0.678207	0.584720	-0.25	–

Table 34: Indic en-hi submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2004	NMT	NO	21.15	0.752783	0.559420	–	–
NMT M2O	2099	NMT	NO	26.71	0.787645	0.586760	–	–
NMT M2M	2189	NMT	NO	26.55	0.784968	0.577010	–	–
IITP-MT	2347	NMT	NO	32.95	0.803497	0.629890	+40.75	1.98
NICT-5 (1)	2129	NMT	NO	30.21	0.801291	0.617860	+32.00	2.00
RGNLP (1)	2367	SMT	NO	21.54	0.697379	0.599760	+22.25	1.46
NICT-5 (2)	2066	NMT	NO	31.06	0.786655	0.599000	+13.50	–
Anuvaad (1)	2406	SMT	NO	22.45	0.709235	0.558850	+5.75	–
Anuvaad (2)	2403	SMT	NO	25.57	0.720866	0.599880	+0.75	–
RGNLP (2)	2383	NMT	NO	21.86	0.751517	0.573160	-1.25	–

Table 35: Indic hi-en submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2007	NMT	NO	7.12	0.457948	0.545370	–	–
NMT O2M	2148	NMT	NO	16.05	0.651935	0.706760	–	–
NMT M2M	2192	NMT	NO	15.41	0.664354	0.711080	–	–
Anuvaad	2443	SMT	NO	15.87	0.668548	0.756890	+73.75	1.16
IITP-MT	2356	NMT	NO	18.81	0.658740	0.710610	+66.50	1.33
NICT-5 (1)	2132	NMT	NO	20.39	0.690652	0.736930	+60.50	1.12
NICT-5 (2)	2109	NMT	NO	18.60	0.649454	0.700960	+45.25	–

Table 36: Indic en-ta submissions

System	ID	Type	RSRC	BLEU	RIBES	AMFM	Pair	Adeq
NMT	2008	NMT	NO	9.14	0.649417	0.488060	–	–
NMT M2O	2101	NMT	NO	19.71	0.751277	0.568020	–	–
NMT M2M	2193	NMT	NO	18.59	0.744884	0.561550	–	–
IITP-MT	2349	NMT	NO	22.42	0.757610	0.604300	+75.25	1.82
NICT-5 (1)	2133	NMT	NO	24.31	0.768865	0.593410	+63.25	1.66
NICT-5 (2)	2111	SMT	NO	21.37	0.744744	0.552630	+47.75	–
Anuvaad (1)	2400	SMT	NO	14.34	0.671535	0.511130	+29.75	1.09
Anuvaad (2)	2408	SMT	NO	14.09	0.673058	0.487250	+29.25	–

Table 37: Indic ta-en submissions

References

Rafael E. Banchs, Luis F. D'Haro, and Haizhou Li. 2015. Adequacy-fluency metrics: Evaluating mt in the continuous space model framework. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 23(3):472–482, March.

Tamali Banerjee, Anoop Kunchukuttan, and Pushpak Bhattacharyya. 2018. Multilingual Indian Language Translation System at WAT 2018: Many-to-one Phrase-based SMT. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Jacob Cohen. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213 – 220.

Raj Dabre, Anoop Kunchukuttan, Atsushi Fujita, and Eiichiro Sumita. 2018. NICT's Participation in WAT 2018: Approaches Using Multilingualism and Recurrently. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Terumasa Ehara. 2018. SMT reranked NMT (2). In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

J.L. Fleiss et al. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

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.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Yuki Kawara, Yuto Takebayashi, Chenhui Chu, and Yuki Arase. 2018. Osaka University MT Systems for WAT 2018: Rewarding, Preordering, and Domain Adaptation. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72. Association for Computational Linguistics.

Tom Kocmi, Shantipriya Parida, and Ondrej Bojar. 2018. CUNI NMT System for WAT 2018 Translation Tasks. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

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.

Yihan Li, Boyan Liu, Yixuan Tong, Shanshan Jiang, and Bin Dong. 2018. SRCB Neural Machine Translation Systems in WAT 2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Benjamin Marie, Atsushi Fujita, and Eiichiro Sumita. 2018. Combination of Statistical and Neural Machine Translation for Myanmar-English. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Yukio Matsumura, Satoru Katsumata, and Mamoru Komachi. 2018. TMU Japanese-English Neural Machine Translation System using Generative Adversarial Network for WAT 2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Hsu Myat Mo, Yi Mon Shwe Sin, Thazin Myint Oo, Win Pa Pa, Khin Mar Soe, and Ye Kyaw Thu. 2018. UCSYNLP-Lab Machine Translation Systems for WAT 2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Toshiaki Nakazawa, Hideya Mino, Isao Goto, Sadao Kurohashi, and Eiichiro Sumita. 2014. Overview of the 1st Workshop on Asian Translation. In *Proceedings of the 1st Workshop on Asian Translation (WAT2014)*, pages 1–19, Tokyo, Japan, October.

Toshiaki Nakazawa, Hideya Mino, Isao Goto, Graham Neubig, Sadao Kurohashi, and Eiichiro Sumita. 2015.

Overview of the 2nd Workshop on Asian Translation. In *Proceedings of the 2nd Workshop on Asian Translation (WAT2015)*, pages 1–28, Kyoto, Japan, October.

Toshiaki Nakazawa, Chenchen Ding, Hideya MINO, Isao Goto, Graham Neubig, and Sadao Kurohashi. 2016. Overview of the 3rd workshop on asian translation. In *Proceedings of the 3rd Workshop on Asian Translation (WAT2016)*, pages 1–46, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, Hideto Kazawa, Yusuke Oda, Graham Neubig, and Sadao Kurohashi. 2017. Overview of the 4th workshop on asian translation. In *Proceedings of the 4th Workshop on Asian Translation (WAT2017)*, pages 1–54. Asian Federation of Natural Language Processing.

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.

Atul Kr. Ojha, Koel Dutta Chowdhury, Chao-Hong Liu, and Karan Saxena. 2018. The RGNLP Machine Translation Systems for WAT 2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

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.

Jerin Philip, Vinay P. Namboodiri, and C V Jawahar. 2018. CVIT-MT Systems for WAT-2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Hammam Riza, Michael Purwoadi, Teduh Uliniansyah, Aw Ai Ti, Sharifah Mahani Aljunied, Luong Chi Mai, Vu Tat Thang, Nguyen Phuong Thai, Vichet Chea, Sethserey Sam, Sopheap Seng, Khin Mar Soe, Khin Thandar Nwet, Masao Utiyama, and Chenchen Ding. 2016. Introduction of the asian language treebank. In *In Proc. of O-COCOSDA*, pages 1–6.

Sukanta Sen, Kamal Kumar Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2018. IITP-MT at WAT2018: Transformer-based Multilingual Indic-English Neural Machine Translation System. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Aye Thida, Nway Nway Han, and Sheinn Thawtar Oo. 2018. Statistical Machine Translation Using 5-grams Word Segmentation in Decoding. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

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.

Boli Wang, Jinming Hu, Yidong Chen, and Xiaodong Shi. 2018a. XMU Neural Machine Translation Systems for WAT2018 Myanmar-English Translation Task. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Rui Wang, Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2018b. English-Myanmar NMT and SMT with Pre-ordering: NICT's machine translation systems at WAT-2018. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.

Yi Mon Shwe Sin and Khin Mar Soe. 2018. Syllable-based myanmar-english neural machine translation. In *In Proc. of ICCA*, pages 228–233.

Longtu Zhang, Yuting Zhao, and Mamoru Komachi. 2018. TMU Japanese-Chinese Unsupervised NMT System for WAT 2018 Translation Task. In *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*, Hong Kong, China, December.