SMT reranked NMT

Terumasa EHARA

Ehara NLP Research Laboratory
Seijo, Setagaya, Tokyo, JAPAN

eharate@gmail.com

Abstract

System architecture, experimental settings and experimental results of the EHR team for the WAT2017 tasks are described. We participate in three tasks: JPCen-ja, JPCzh-ja and JPCko-ja. Although the basic architecture of our system is NMT, reranking technique is conducted using SMT results. One of the major drawback of NMT is under-translation and over-translation. On the other hand, SMT infrequently makes such translations. So, using reranking of n-best NMT outputs by the SMT output, discarding such translations can be expected. We can improve BLEU score from 46.03 to 47.08 by this technique in JPCzh-ja task.

1 Introduction

Rapidly progressing of NMT techniques make paradigm change in machine translation not only for the research purpose but for the practical field. Although the NMT provides high quality and fluent translations, it has several drawbacks. One of them is under- and over-translation which is infrequent in a SMT output.

We propose a reranking method for n-best NMT outputs using a SMT output. We compare n-best NMT outputs with a SMT output by the measure of IMPACT (Echizen-ya and Araki, 2007) which is one of the automatic evaluation measure of machine translation results. The NMT output which has the highest IMPACT score referring to SMT output is selected as the system output.

In the following sections, we describe system architecture and experimental settings in section 2, experimental results and discussions in section 3 and conclusion in section 4.

2 System architecture and experimental settings

2.1 Overall system architecture

Our system architecture is shown in Figure 1. An input source sentence is fed to the NMT part and also to the SMT part. NMT part outputs n-best translations (“NMT translation 1” to “NMT translation n”) and SMT part outputs another translation (“SMT translation”). Reranking part compares NMT translations with SMT translation and reranks them. The best reranked “NMT translation i” is outputted.

2.2 NMT part

We use OpenNMT (Minh-Thang Luong et al., 2015) in NMT part.

Segmentation of English sentences is sub word based. The English segmenter segments each non-alphabetical characters (characters except for A to Z and a to z) as separate words. Segmentation of Chinese sentences and Korean sentences are both word based and character based. Word segmentation policy for these languages are described in the previous paper (Ehara, 2016). Japanese segmentation is word based, sub word based and character based. For JPCzh-ja task and JPCko-ja task, word based and character based Japanese segmenters are used. The word based Japanese segmenters are described in the previous paper (Ehara, 2016). For JPCen-ja task, we use sub word based Japanese segmenter which segments each special characters (characters except for Hiragana, Kanji, Katakana and Roman characters) as separate words, in addition to Juman’s word segmentation (Kurohashi et al., 1994).

Option settings for OpenNMT are as follows:

Source sequence length (-src_seq_len): 100 (word based), 120 (sub word based), 250 (character based);
Target sequence length (-tgt_seq_len):
100 (word based), 120 (sub word based), 250 (character based); Encoder type (-encoder_type): brnn (bidirectional recurrent NN); Replace unknown word (-replace_unk): yes; Unknown word dictionary (-phrase_table): yes (see 2.3); Beam size (-beam_size): 50; N-best size (-n_best): 50.

2.3 SMT part

Our SMT system is phrase-based SMT by Moses v.3 (Koehn et al., 2003) with default option settings. For JPCen-ja task and JPCzh-ja task, preordering is applied. The preordering system is same as described in the previous papers (Ehara, 2015; Ehara, 2016).

We use unknown dictionary for NMT part. It is made from the phrase-table of Moses. For every source word, we select the target phrase which has the highest translation probability for the source word. And the unknown word dictionary is constructed as the source word and target phrase pairs.

2.4 Reranking part

For reranking of n-best outputs of NMT part, we use automatic evaluation measure IMPACT (Echizen-ya and Araki, 2007). For the preliminary study, we compared BLEU, RIBES and IMPACT with human evaluation score JPO adequacy by the WAT2016’s evaluation results (Nakazawa et al., 2016). As the results, we found IMPACT was the best correlated score with JPO adequacy. Then we use IMPACT as the reranking measure. Reranking part calculates IMPACT score for NMT’s n-best translations with SMT translation as the reference. And the best translation which has the highest IMPACT score is outputted as the system output.

3 Experimental results and discussions

The official evaluation results of our submissions are shown in Table 1 (Nakazawa et al., 2017). In the Table 1, “Original system” means the NMT without reranking and “SMT” means SMT part of our system.

For JPCen-ja task, reranking decreases BLEU, RIBES and AMFM scores and also HUMAN score. Although the overall evaluation result doesn’t show the effectiveness of the reranking, several improvements are observed. Examples are listed in Table 2. Original translation of the example 1 has under-translation. Only the first two words (The oldest) and the punctuation mark (.) are translated in the original translation. Original translations of example 2 has also under-translation. None of words (“ACT, READ, PRE”), GBSTB, GBSTT, FXb2, PUMP, FXB, FXT, SWL, and RFX” is translated. On the other hand, reranking system does not make such under-translations. Original translation of example 3 has over-translation. “異なる(differ)” occurs two times. But the reranked translation has no over-translation.

| Task       | Data ID | System    | Segment | BLEU   | RIBES   | AMFM    | HUMAN   | JPO adeq |
|------------|---------|-----------|---------|--------|---------|---------|---------|----------|
| JPCen-ja   | 1406    | Reranking | Subword | 44.44  | 0.8610  | 0.7471  | 58.250  | ---      |
|            | 1407    | Original  | Subword | 44.63  | 0.8667  | 0.7478  | 60.000  | 4.63     |
|            | 1408    | Reranking | Word    | 36.20  | 0.8128  | 0.7237  | ---      | ---      |
|            | 1415    | Original  | Word    | 46.03  | 0.8586  | 0.7559  | ---      | ---      |
|            | 1414    | Reranking | Character | 46.52  | 0.8596  | 0.7614  | 69.750  | 4.31     |
|            | 1409    | Original  | Character | 45.27  | 0.8544  | 0.7571  | ---      | ---      |
|            | 1416    | Reranking | Word    | 40.79  | 0.8270  | 0.7384  | ---      | ---      |
|            | 1418    | Original  | Word    | 71.52  | 0.9445  | 0.8661  | 6.250   | ---      |
|            | 1417    | Reranking | Character | 70.23  | 0.9432  | 0.8623  | ---      | ---      |
|            | 1419    | Original  | Character | 69.42  | 0.9364  | 0.8605  | ---      | ---      |
|            | 1409    | Original  | Character | 71.08  | 0.9440  | 0.8645  | ---      | ---      |

Table 1: Official evaluation results (Japanese segmenter is Juman)
For JPCen-ja task, comparing our submission of data ID 1407 (EHR) and another submission (OTHER), BLEU score of the EHR is 44.63 and it is less than the OTHER’s score (50.27). On the other hand, HUMAN score of the EHR is 60.00 and it is greater than the OTHER’s score (56.25). There are 20 data that the BLEU score of EHR is less than the OTHER’s score but the HUMAN score of EHR is greater than the OTHER’s score. We examine these data and find that several data have the differences between the source expression and the reference expression. Table 3 shows examples of source, reference, EHR output and OTHER output. Example 1 has the voice change (the source is passive and the reference is active). Example 2 has the topic change (the topic of the source is “valve” and the topic of the reference is “経路 (passage)”). Example 3 has the additive translation (the source “GELD” corresponds to the reference “有機 EL ディスプレイ (OELD, organic electroluminescent display)”). Example 4 has the subtractive translation (the source has complementary information and the latter means the translation including needless information).

Table 2: JPCen-ja task’s examples having effectiveness of reranking (recover of under-translation: example 1 and 2; recover of over-translation: example 3)

| source | The oldest is the capacitive divider bridge, one arm of which consists of a reference capacitance and the other of the capacitance to be measured. |
| reference | 最も古いものは容量分圧ブリッジであり、その1つのアームは基準静電容量から成り、その他の静電容量が測定される。 |
| reranking | 最も古いものは容量分圧ブリッジであり、その1つのアームは基準キャパシタンスとその他キャパシタンスから成る。 |
| original | 最も古い。 |
| SMT | 最も古いものは、容量性ダイバリブリッジ、基準キャパシタンスと測定されるべきキャパシタンスの他の1つを構成するアームである。 |

| source | FIG. 5 shows a portion of waveforms of a simulation result of the circuit in FIG. 4, and commands (ACT, READ, PRE), GBSTB, GBSTT, Fxb 2, PUMP, FXB, FXT, SWL, and RFX are shown as signals. |
| reference | 図5には、図4の回路のシミュレーション結果の一部が示されており、信号として、コマンド（ACT, READ, PRE）、GBSTB、GBSTT、Fxb 2、PUMP、FXB、FXT、SWL、RFXが示されている。 |
| reranking | 図5は、図4の回路のシミュレーション結果の一部を示すものであり、コマンド（ACT, READ, PRE）、GBSTB、GBSTT、Fxb 2、PUMP、FXB、FXT、SWL、RFXを信号として示す。 |
| original | 図4の回路のシミュレーション結果の一部と、図4の回路のシミュレーション結果の一部を示す図である。 |
| SMT | 図5は、図4の回路のシミュレーション結果の一部を示す、及びコマンド（ACT、READ、PRE）、GBSTB、GBSTT、Fxb 2、PUMP、FXB、FXT、SWLとは、信号として示されている。 |

| source | However, specifications sometimes differ from one image forming apparatus to another. |
| reference | しかし、画像形成装置ごとにスペックが異なる場合がある。 |
| reranking | しかし、場面は、時々、1つの画像形成装置とは異なる。 |
| original | しかし、時々、1つの画像形成装置とは異なる仕様が異なる。 |
| SMT | しかし、仕様は、装置を形成する1つの画像から別の画像へ時には異なる。 |

---

1. Sentence level BLEU is calculated by mteval-v13a.pl in the Moses package.
2. For the BLEU score, “less” means “less or equal -10” and “greater” means “greater or equal 10”. If the difference of BLEU is between -10 to 10, it is considered “tie”.
3. We distinguish between “additive translation” and “over-translation”. The former means the translation including complementary information and the latter means the translation including needless information.
4. We distinguish between “subtractive translation” and “under-translation”. The former means the translation omitting complementary information and the latter means the translation omitting needless information.
“given within this range” and the reference has no such expression).

Comparing JPO adequacy score of EHR and OTHER, the EHR’s score 4.63 is less than the OTHER’s score 4.75. Although HUMAN and JPO adequacy are both human evaluation, they have a contradiction. One possible reason is that the evaluators of HUMAN score do not look at the reference translations but the evaluators of JPO adequacy score can look at the reference translations.

Comparing JPO adequacy score of EHR and OTHER, the EHR’s score 4.63 is less than the OTHER’s score 4.75. Although HUMAN and JPO adequacy are both human evaluation, they have a contradiction. One possible reason is that the evaluators of HUMAN score do not look at the reference translations but the evaluators of JPO adequacy score can look at the reference translations.

Comparing JPO adequacy score of EHR and OTHER, the EHR’s score 4.63 is less than the OTHER’s score 4.75. Although HUMAN and JPO adequacy are both human evaluation, they have a contradiction. One possible reason is that the evaluators of HUMAN score do not look at the reference translations but the evaluators of JPO adequacy score can look at the reference translations.

Table 3: Different expressions between sources and references

|   | source | reference | EHR | OTHER |
|---|--------|----------|-----|-------|
| 1 | In FIG. 6, the same symbols are used to the same elements as the elements shown in FIGS. 1-5 | 図6において、図1～図5に示す要素と同じ要素には同じ符号を付してある。 | 図6では、同じ記号が、図1～図5に示される要素と同じ要素に使用される。 | 図6において、図1～図5と同じ要素には同じ符号を付している。 |
| 2 | A sixth on-off valve 202 is disposed on the fourth bypass passage 204. | 第4バイパス経路204には第6閉閉弁202が設けられている。 | 第6のオンオフ弁202は、第4バイパス通路204上に配置される。 | 第4のバイパス通路204には、第6の閉閉弁202が配置されている。 |
| 3 | The display device 4 is, for example, an LCD and an GELD. | 表示装置4は、例えば、液晶ディスプレイ（LCD、liquidcrystaldisplay）や有機ELディスプレイ（OELD、organicelctro-luminescentdisplay）等である。 | 表示装置4は、例えばLCDおよびGELDである。 | 表示装置4は、例えばLCDやOELD（OELD、organicelctro-luminescentdisplay）等である。 |
| 4 | In this case, the proportion of the additive given within this range corresponds to 3% or lower. | この場合の添加剤の添加量は3%以下である。 | この場合、この範囲内に与えられる添加剤の割合は、3%以下に相当する。 | この場合、添加剤の割合は、3%以下である。 |

Table 3: Different expressions between sources and references

For JPCzh-ja and JPCko-ja tasks, reranking increases BLEU, RIBES and AMFM scores. However, we don’t have a HUMAN scores comparing the reranking and the original for these tasks. Examples having the effectiveness of the reranking for these tasks are shown in Table 4 and Table 5. Example 1 and 2 of Table 4 have under-translation in original translation. Example 3 of Table 4 has over-translation in original translation. Example 1 and 2 of Table 5 have under-translation in original translation.
| source | reference | reranking | original | SMT |
|--------|-----------|-----------|----------|-----|
| Table 4: JPCzh-ja task’s examples having effectiveness of reranking (recover of under-translation: example 1 and 2; recover of over-translation: example 3) |

| source | reference | reranking | original | SMT |
|--------|-----------|-----------|----------|-----|
| 1      | 図 3 (A) は、間隙通過処理を2回行う高分子組成物の製造装置について上面から装置内部を透視したときの概略透視図であり、図3(B)は、図3(A)の装置のP−Q断面における概略断面図である。 | 図3(A)は、ギャップを2回行う処理を行う高分子組成物の製造装置を上面から装置内部を透視した場合の概略透視図であり、図3(B)は、図3(A)の装置のP−Q断面における概略断面図である。 | 図3(A)は、図3(A)の装置のP−Q断面モデルの概略的な断面図である。 | 図3(A)については、2次ギャップの処理によって高分子組成物の製造装置を上面から装置内部透視時の概略斜視図であり、図3(B)は、図3(A)の装置のP−Q断面における概略断面図である。 |
| 2      | 此外、根据元素的话，一个元素有可能形成不同化合价的氧化物。 | また、元素によっては、1つの元素が異なる価数の酸化物を形成することが可能である。 | また、元素によっては、異なる元素を形成することが可能である。 | また、元素とよれば、異なる原子価元素の酸化物を形成することができる。 |
| 3      | 実施例14 | 実施例14 | 実施例14 | 実施例14 |
Table 5: JPCko-ja task’s examples having effectiveness of reranking (recover of under-translation: example 1 and 2)

| No. | source | reference | reranking | original | SMT |
|-----|--------|-----------|-----------|----------|-----|
| 1   | 이어서, 용매로서 DNP(디 이소노 닐 프탈 레이트) 183 질량부를 첨가하였다. | 次いで、溶媒として DNP (ジイソノールフタレート) 183 質量部を添加した。 | 次いで、溶媒として DNP (ジメチルホルムアミド) 183 質量部を添加した。 | 次いで、溶媒として DNP (記述이소노아닐프탈레트) 183質量部を添加した。 | 次いで、溶媒として DNP (記述이소노아닐프탈레트) 183質量部を添加した。 |
| 2   | 저장 장치 (70)는 다음과 포함 할 수 있다 (도 15 및 도 16): - 하나 혹은 그 이상의 전기 배터리 (81); 또는 - 하이브리드 배터리 (82) 및, 상기 하이브리드 배터리 (82) 와 유호하게 연결된 내연기관 (83). | 貯蔵装置70は以下を備えることができる (図15及び図16): - 1台または2台以上の電気バッテリー81;または-ハイブリッドバッテリー82と、前記ハイブリッドバッテリー82に接続されている内燃機関83。 | 記憶装置70は以下を含むことができる (図15及び図16): 1つまたはそれ以上の電気バッテリー81、又はハイブリッドバッテリー82、及び、ハイブリッドバッテリー82と有効に連結された内燃機関83。 | 記憶装置70は以下を含むことができる (図15および図16)。 | 記憶装置70は以下を含むことができる (図15および図16)。 |

For JPCzh-ja and JPCko-ja tasks, the word based translations have higher BLEU, RIBES and AMFM compared with the character based translations. However, HUMAN score of the word based translations are lower than the score of the character based translations.

For JPCzh-ja task, there are 7 data that the BLEU score of the word based translation is greater than the character based translation’s score but the HUMAN score of the word based translation is less than the character based translation’s score. Examples of such translations are listed in Table 6. Example 1 has an under-translation in the word based translation (“滴度”). Example 2 also has an under-translation in the word based translation (“実行される(進行的)”). Example 3 has miss translations both in the character based translation and the word based translation.

In the character based translation, “取付面図” is used instead of “実装面図(安裝面圖)”. And in the word based translation, “分波器モジュール” is used instead of “デュプレクサモジュール(双工器模块)”. However, the latter miss translation is more significant than the former. Example 4 has another different translations. Character based translation uses “が良い(好)”, and word based translation uses “に優れる”.

For JPCko-ja task, there are 3 data that the BLEU score of the word based translation is greater than the character based translation’s score and the HUMAN score of the word based translation is less than the character based translation’s score. Examples of such translations are listed in Table 7. Example 1 has a different translation. Literal translation of “연결” is “連結” and non-literal translation is “接続”. Example 2 shows the effectiveness of the unknown word translation in the character based translation. The expression “디 펜 타 에 리 트 리 톨 펜 타 아 크 릴 레 이 트 와 디 펜 타 에 리 트 리 톨” does not be translated in the word based translation. Example 3 has different translations. Character based translation uses “プレーキ(ブレーキ)”, and word based translation uses “ブレーキ”. 

For JPCzh-ja and JPCko-ja tasks, the word based translations have higher BLEU, RIBES and AMFM compared with the character based translations. However, HUMAN score of the word based translations are lower than the score of the character based translations.

For JPCzh-ja task, there are 7 data that the BLEU score of the word based translation is greater than the character based translation’s score but the HUMAN score of the word based translation is less than the character based translation’s score. Examples of such translations are listed in Table 6. Example 1 has an under-translation in the word based translation (“滴度”). Example 2 also has an under-translation in the word based translation (“実行される(進行的)”). Example 3 has miss translations both in the character based translation and the word based translation.

In the character based translation, “取付面図” is used instead of “実装面図(安裝面圖)”. And in the word based translation, “分波器モジュール” is used instead of “デュプレクサモジュール(双工器模块)”. However, the latter miss translation is more significant than the former. Example 4 has another different translations. Character based translation uses “が良い(好)”, and word based translation uses “に優れる”.

For JPCko-ja task, there are 3 data that the BLEU score of the word based translation is greater than the character based translation’s score and the HUMAN score of the word based translation is less than the character based translation’s score. Examples of such translations are listed in Table 7. Example 1 has a different translation. Literal translation of “연결” is “連結” and non-literal translation is “接続”. Example 2 shows the effectiveness of the unknown word translation in the character based translation. The expression “디 펜 타 에 리 트 리 톨 펜 타 아 크 릴 레 이 트 와 디 펜 타 에 리 트 리 톨” does not be translated in the word based translation. Example 3 has different translations. Character based translation uses “プレーキ(ブレーキ)”, and word based translation uses “ブレーキ”.

124
## Table 6: JPCzh-ja task's examples having BLEU and HUMAN scores contradiction

| No. | Source | Reference | Char. Based | Word Based |
|-----|--------|-----------|-------------|------------|
| 1   | 自進の一歩提示高估了CAZ028单价散装液浓度。 | このことは、CAZ028一価パルクの力価が高く見積もられていることをさらに示唆する。 | これはさらに、CAZ028単価の分散液の濃度を高めることを示唆する。 | これにより、CAZ028の一価パルクをさらに示唆する。 |
| 2   | 图9示出了电台110万释放额外的无线资源所进行的处理过程900的设计。 | 図9は、余分な無線リソースを放棄するために局110によって実行される方法900の設計を示す。 | 図9は、局110が、追加の無線リソースを解放するために行われる処理プロセス900の設計を示す。 | 図9は、追加の無線リソースを解放するために局110が処理プロセス900の設計を示す。 |
| 3   | 図4(A)是双工器模块的简要等效电路图，图4(B)是双工器模块的安装面图。 | 図4(A)はデュプレクサモジュールの概略の等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 | 図4(A)はデュプレクサモジュールの概略的な等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 | 図4(A)はデュプレクサモジュールの概略的な等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 |
| 4   | 另一方面，如果大于4.2倍，则耐加水分解性好，但基材层(B)的凝集强度降低，因此不理想。 | 另一方面、如果大于4.2倍。則耐水解性好，但基材层(B)的凝集强度降低，因此不理想。 | 另一方面，如果大于4.2倍。則耐水解性好，但基材层(B)的凝集强度降低，因此不理想。 | 另一方面、如果大于4.2倍。則耐水解性好，但基材层(B)的凝集强度降低，因此不理想。 |

| No. | Source | Reference | Char. Based | Word Based |
|-----|--------|-----------|-------------|------------|
| 1   | [8]重合性化合物は、ジペンタエリスリトールペンタアクリレートとジペンタエリスリトールヘキサアクリレートとの混合物であることが好ましい。 | この場合、受信機は、S707段階において、2次装置が接続しているか否か判断する。 | このような場合、受信機は、S707段階に進行して二段装置に連絡されたか否かを判断する。 | このような場合、受信機は、S707段階に進行して二段装置に接続されたか否かを判断する。 |
| 2   | [8]重合性化合物は、ディペンタエリスリトールペンタアクリレートとディペンタエリスリトールヘキサアクリレートとの混合物であることが好ましい。 | この場合、受信機は、S707段階において、2次装置が接続しているか否か判断する。 | このような場合、受信機は、S707段階に進行して二段装置に連絡されたか否かを判断する。 | このような場合、受信機は、S707段階に進行して二段装置に接続されたか否かを判断する。 |
| 3   | [8]重合性化合物は、ディペンタエリスリトールペンタアクリレートとディペンタエリスリトールヘキサアクリレートとの混合物であることがある。 | この場合、受信機は、S707段階において、2次装置が接続しているか否か判断する。 | このような場合、受信機は、S707段階に進行して二段装置に連絡されたか否かを判断する。 | このような場合、受信機は、S707段階に進行して二段装置に接続されたか否かを判断する。 |

## Table 7: JPCko-ja task's examples having BLEU and HUMAN scores contradiction

| No. | Source | Reference | Char. Based | Word Based |
|-----|--------|-----------|-------------|------------|
| 1   | これがさらに、CAZ028単価の分散液の濃度を高めることを示唆する。 | これはさらに、CAZ028単価の分散液の濃度を高めることを示唆する。 | これにより、CAZ028の一価パルクをさらに示唆する。 | これにより、CAZ028の一価パルクをさらに示唆する。 |
| 2   | 図9示出了电台110万释放额外的无线资源所进行的处理过程900的设计。 | 図9は、余分な無線リソースを放棄するために局110によって実行される方法900の設計を示す。 | 図9は、局110が、追加の無線リソースを解放するために行われる処理プロセス900の設計を示す。 | 図9は、追加の無線リソースを解放するために局110が処理プロセス900の設計を示す。 |
| 3   | 図4(A)是双工器模块的简要等效电路图，图4(B)是双工器模块的安装面图。 | 図4(A)はデュプレクサモジュールの概略の等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 | 図4(A)はデュプレクサモジュールの概略的な等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 | 図4(A)はデュプレクサモジュールの概略的な等価回路図であり、図4(B)はデュプレクサモジュールの実装面図である。 |

Table 6: JPCzh-ja task’s examples having BLEU and HUMAN scores contradiction (under-translation in word based: example 1 and 2; different translation: example 3 and 4)

Table 7: JPCko-ja task’s examples having BLEU and HUMAN scores contradiction (different translation: example 1 and 3; un-translation in word based: example 2)
4 Conclusion

System descriptions, experimental settings and experimental results of the EHR team are described. We participate in the 3 tasks and submitted 10 systems’ outputs. We can observe our re-ranking technique is effective to remove under-translation and over-translation which are in NMT outputs sometimes.

References

Hiroshi Echizen-ya and Kenji Araki. 2007. Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum, Proceedings of the Eleventh Machine Translation Summit (MT SUMMIT XI), Page.151-158.

Terumasa Ehara. 2015. System Combination of RBMT plus SPE and Preordering plus SMT. Proceedings of the 2nd Workshop on Asian Translation (WAT2015), pages 29–34.

Terumasa Ehara. 2016. Translation systems and experimental results of the EHR group for WAT2016 tasks. Proceedings of the 3rd Workshop on Asian Translation (WAT2016), pages 111-118.

Philipp Koehn, Franz J. Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation. Proceedings of HLTNAACL 2003, pages 48-54.

Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto and Makoto Nagao. 1994. Improvements of Japanese morphological analyzer JUMAN. Proceedings of The International Workshop on Shareable Natural Language Resources, pages 22-28.

Minh-Thang Luong, Hieu Pham and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1412–1421.

Toshiaki Nakazawa, Chenchen Ding, Hideya Mino, Isao Goto, Graham Neubig and Sadao Kurohashi. 2016. Overview of the 3rd Workshop on Asian Translation. Proceedings of the 3rd Workshop on Asian Translation (WAT2016), pages 1-46.

Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, Graham Neubig, Hideto Kazawa, Yusuke Oda, Jun Harashima and Sadao Kurohashi. 2017. Overview of the 4th Workshop on Asian Translation. Proceedings of the 4th Workshop on Asian Translation (WAT2017), pages ??-??.