

## SMT reranked NMT (2)

Terumasa Ehara

Ehara NLP Research Laboratory

Seijo, Setagaya, Tokyo Japan

<http://www.ne.jp/asahi/eharate/eharate/>

### Abstract

System architecture, experimental settings and experimental results of the EHR team for the WAT2018 tasks are described. System architecture is same as the WAT2017 submission. We participate four tasks this time. We can improve 1.5 to 3.0 BLEU score compared with the baseline scores. On the other hand, the pairwise evaluation score of our system to the baseline is negative in ASPEC en-ja task. Our technique also decreases the number of under-translation that is frequently appears in a NMT output.

### 1 Introduction

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 have proposed a reranking method for n-best NMT outputs using a SMT output (Ehara, 2017). Ehara (2017) shows under-translation can be reduced by the method. This time, we use the same technique. Figure 1 shows our system architecture.

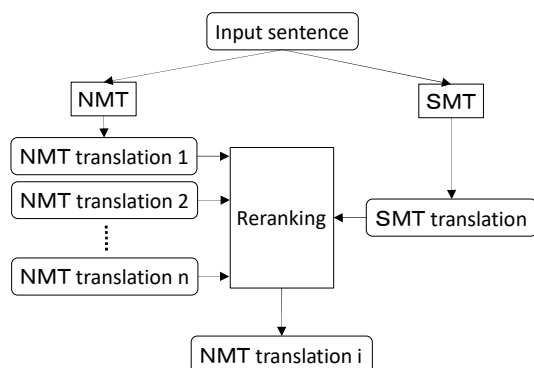


Figure 1: System architecture

An input sentence is fed to NMT part and to SMT part. N-best outputs of NMT are reranked by the SMT output as a reference. Reranking measure is BLEU, this time, instead of IMPACT in WAT2017. NMT translation  $i$  having highest BLEU score is output as a system output.

### 2 Experimental setting

In WAT2018 (Nakazawa et al., 2018), we participate four tasks: ASPEC en-ja, JPC zh-ja, JPC ko-ja and JPC en-ja tasks. We use OpenNMT (Minh-Thang Luong et al., 2015) in NMT part. Our SMT part is phrase-based SMT by Moses v.3 (Koehn et al., 2003) with default option settings. We adopt preordering in SMT for en-ja and zh-ja tasks. Option settings for NMT part is described in Ehara (2017). Segmentation policies for NMT are as follows. In JPC zh-ja task, zh part is segmented by a character and ja part is segmented by a word except for alpha-numeric and symbol character sequences. Alpha-numeric and symbol character sequences are segmented by a character. In JPC ko-ja task, ko part and ja part are both segmented by a character. In JPC en-ja task and ASPEC en-ja task, en parts and ja parts are both segmented by a sub word. For JPC en-ja task, vocabulary size of en part is 60,217 and vocabulary size of ja part is 94,542. For ASPEC en-ja task, vocabulary size of en part is 62,203 and vocabulary size of ja part is 107,145.

### 3 Experimental results

Experimental results are shown in Table 1 with training data size (number of sentences) and final epoch number of NMT training. Training data for JPC en-ja task include WAT2018’s training data and NTCIR-10’s training data (Goto et al., 2013). The differences of three data ID in four subtasks of

JPC en-ja task are epoch number of NMT training. They are 13, 18 and 20.

We can improve 1.5 to 3.0 BLEU score compared with the baseline scores.

JPO adequacy scores are similar to the last workshop’s scores of EHR team (Adeq/2017).

Task	Data size	Sub task	Data ID	Epoch	BLEU	RIBES	Pairwise	JPO Adeq.	Adeq/2017	
JPC zh-ja	1,000,000	N	N	2209	13	45.55	0.8567	---	---	
			N1	2210	13	48.10	0.8583	---	4.22	4.31
			N2	2211	13	47.12	0.8617	---	4.22	---
			N3	2212	13	39.65	0.7673	---	---	---
			EP	2213	13	34.23	0.8051	---	---	---
JPC ko-ja	1,000,000	N	N	2214	13	71.70	0.9485	---	---	
			N1	2215	13	71.45	0.9458	---	4.89	4.81
			N2	2216	13	73.29	0.9535	---	4.73	---
JPC en-ja	4,043,073	N	N	2217	13	53.83	0.9074	---	---	
			N	2283	13	44.71	0.8525	---	---	
			N	2394	18	45.27	0.8549	---	---	
			N	2475	20	45.39	0.8548	---	---	
			N	2284	13	47.57	0.8707	---	---	
			N1	2395	18	48.01	0.8734	---	---	
			N1	2476	20	48.03	0.8728	---	4.64	4.63
			N2	2285	13	41.54	0.8376	---	---	
			N2	2396	18	41.99	0.8405	---	---	
			N2	2477	20	42.12	0.8407	---	4.36	---
			N3	2286	13	50.04	0.8649	---	---	
			N3	2397	18	51.17	0.8642	---	---	
			N3	2478	20	51.42	0.8645	---	---	
ASPEC en-ja	1,502,767	---	2245	13	37.97	0.8287	-0.500	---		

Table 1: Experimental results

For ASPEC en-ja task, our result makes 37.97 BLEU score comparing that baseline result has 36.37 BLEU score. On the other hand, pairwise evaluation score of our system is minus 0.5. We calculate the difference of sentence level BLEU of our translation and baseline translation (diff BLEU)<sup>1</sup>. Figure 2 shows scattering graph of diff BLEU and pairwise evaluation score (sum of five evaluator’s scores).

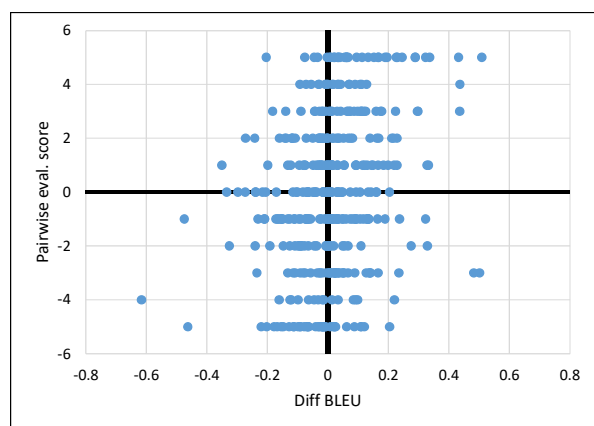


Figure 2: Scattering graph of diff BLEU and pairwise evaluation score (ASPEC en-ja task)

<sup>1</sup> Diff BLEU in Figure 2 is scaled from -1 to 1. On the other hand, BLEU in Table 1 is scaled from 0 to 100. NIST tool

Data points in first quadrant and third quadrant have no contradiction in two scores. Data points in second quadrant and fourth quadrant have contradiction in two scores. Number of data in each quadrant is listed in Table 2. Several examples of contradicted data are examined in the next section.

Diff BLEU \ Pairwise	< 0	≥ 0
< 0	94	84
≥ 0	85	137

Table 2: Contingency table of diff BLEU and pairwise evaluation score.

We examine under- and over-translations in ASPEC en-ja task’s 1,812 test data. The result is shown in Table 3. In Table 3, “reranked” means outputs with reranking, “unreranked” means outputs without reranking, “normal” means translation which includes neither under nor over translation, “under” means under-translation, “over” means over-translation and “under & over” means translation which includes both under and over translation.

		reranked				total
		normal	under	over	under & over	
unreranked	normal	1496	10	9	1	1516
	under	206	71	3	2	282
	over	4	0	6	0	10
	under & over	2	0	1	1	4
	total	1708	81	19	4	1812

Table 3: Number of under- and over-translation in unreranked translation and reranked translation for ASPEC en-ja task test data

From Table 3, we can see that in 1812 test sentences, 286 unreranked outputs include under-translation and only 85 reranked outputs include under-translation. On the other hand, in 1812 test sentences, 14 unreranked outputs include over-translation and 23 reranked outputs include over-translation. We can say our reranking technique decreases the number of under-translation, but cannot decrease the number of over-translation. Several examples of these data are described in the next section.

mteval-v13a.pl (<ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v13a.pl>) is used to calculate sentence level BLEU.

## 4 Analysis of translation results

### 4.1 Comparison of translation results with reranking and without reranking

We compare translation results with reranking and without reranking in this subsection for ASPEC en-ja task. We compute sentence level BLEU of both translation results. Table 4 shows the cases that the BLEU score of a translation with reranking is higher than the BLEU score of a translation without reranking. Table 5 shows the opposite cases.

In these tables, “snt. #” means sentence number of the test set, “src” means source sentence, “ref” means reference sentence, “smt” means translation result by the SMT part, “unreranked” means translation result without reranking, “reranked” means translation result with reranking by our method, “BLEU\_f” means BLEU score of the “unreranked” and BLEU\_r means BLEU score of the “reranked”.

From Table 4, we can see several under-translations in the unreranked are corrected in reranked (see snt. # 1269 and 44).

item	value
snt. #	815
src	Method of the DCT is explained .
ref	DCT の方法について解説した。
smt	DCT の方法について解説した。
unreranked	DCT の方法を説明した。
reranked	DCT の方法について解説した。
BLEU_f	0.2665
BLEU_r	1
remarks	BLEU_f is very low, however unreranked has similar meaning to ref.

snt. #	183
src	The dynamic equilibrium surface form is dependent on annealing condition and miss - cut direction of the substrate .
ref	動的平衡表面形態はアニーリング条件と基板のミスカット方向に依存した。
smt	動的平衡表面形態がアニーリング条件と基板のmiss-cut方向に依存した。
unreranked	動的平衡表面形状は基板のアニーリング条件とミスカット方向に依存する。
reranked	動的平衡表面形態はアニーリング条件と基板のミスカット方向に依存した。
BLEU_f	0.3921
BLEU_r	1
remarks	Dependee of "of the substrate" is " annealing condition and miss-cut direction" in unreranked that is incorrectly interpreted but is "miss-cut direction" in smt and reranked that is correctly interpreted.

snt. #	1236
src	This paper describes the application of the package software .
ref	業務パッケージソフトの適用について述べた
smt	本稿では、パッケージソフトの適用について述べた。
unreranked	パッケージソフトウェアの応用について述べた。
reranked	パッケージソフトの適用について述べた。
BLEU_f	0.2336
BLEU_r	0.8409
remarks	"Application" is translated to "適用" in ref, smt and reranked and it is translated to "応用" in unreranked.

snt. #	1269
src	Survey items are as follows : 1) Trends of engineers ( study in 1992 fiscal year and literature search ) . 2) Survey on actual conditions and consciousness of engineers ( questionnairing for each individual engineer ) .
ref	調査項目は、1) 技術者の動向(平成4年度調査と文献調査)、2) 技術者の実態と意識に関する調査(技術者個人に対するアンケート調査)である。
smt	調査項目は以下の通りである: 1) 技術者(平成4年度の調査と文献検索)。2) の動向、技術者(個々の技術者に対するアンケート調査)の実態と意識に関する調査。
unreranked	調査項目は以下のようである。
reranked	調査項目は、1) 技術者の動向(平成11年度の研究、文献調査)、2) 技術者の実態と意識(個々の技術者ごとのアンケート調査)。
BLEU_f	0.0019
BLEU_r	0.5796
remarks	Under-translation of unreranked is corrected in reranked.

snt. #	44
src	By measuring Josephson vortex flow resistance of slightly over doped ( Tc = 86.0K ) and intermediately doped ( Tc = 82.5K ) Bi2Sr2CaCu2O8 + δ ( Bi - 2212 ) single crystals , the magnetic phase diagram was studied .
ref	僅かに過剰添加 ( Tc = 86.0 K ) 及び中程度に過剰添加 ( Tc = 82.5 K ) した Bi 2 Sr 2 CaCu 2 O 8 + δ ( Bi - 2212 ) 単結晶の Josephson 渦系フロー抵抗を測定して磁気状態図を調べた。
smt	わずかにドーピングした ( Tc = 86.0 K 以上 ) と中間ドーブ ( Tc = 82.5 K ) Bi 2 Sr 2 CaCu 2 O の Josephson 渦系流抵抗 + δ ( Bi - 2212 ) 単結晶を測定し、磁気相図を研究した。
unreranked	磁気状態図を調べた。
reranked	わずかにドーブした ( Tc = 86.0 K ) と中間ドーブ ( Tc = 82.5 K ) Bi 2 Sr 2 CaCu 2 O 8 + δ ( Bi - 2212 ) 単結晶の Josephson 渦流抵抗を測定し、磁気相図を研究した。
BLEU_f	0.0003
BLEU_r	0.5689
remarks	Under-translation of unreranked is corrected in reranked.

snt. #	1357
src	The electric power generation was the 380 micro watt .
ref	発電量は380マイクロワットであった。
smt	発電は380マイクロワットであった。
unreranked	発電は380 μ W であった。
reranked	発電は380マイクロワットであった。
BLEU_f	0.1703
BLEU_r	0.729
remarks	"Micro watt" is translated to "マイクロワット" in ref, smt and reranked and it is translated to "μ W" in unreranked.

snt. #	253
src	To begin with , this paper explains atomic energy level and transition between levels .
ref	先ず原子のエネルギー準位と準位間の遷移から説明した。
smt	本稿では、まず、原子エネルギー準位と準位間の遷移について解説した。
unreranked	まず、原子カレベルとレベル間の推移について解説した。
reranked	まず、原子のエネルギー準位と準位間の遷移について解説した。
BLEU_f	0.077
BLEU_r	0.6119
remarks	"Atomic energy level" is incorrectly translated in reranked (原子カレベル), but in smt and reranked it is correctly translated (原子のエネルギー準位).

snt. #	303
src	The author also considered that on condition that the jointing part of two rows bonded with van der Waals force between main chains is a segment , Tg of linear polymer is the temperature when thermal expansion of the C-C bond from 0 °C becomes 1 % .
ref	また、線型ポリマのTgは主鎖間の van der Waals 力で結合された2列の結合部分をセグメントと考えて、そのC-C結合の0°Cからの熱的な伸び率が1%の温度とした。
smt	また、主鎖間のファンデルワールス力による結合した2列の接合部はsegmentであることを条件に、線状重合体のTgは熱expansion C-C結合の0°Cから1%になると温度であると考えた。
unreranked	また、主鎖間のvanderWaals力に結合した2列の接合部が偏析する条件として、線形高分子のTgは、C-C結合の熱膨張が1%になると温度になると考えた。
reranked	また、主鎖間のvanderWaals力に結合した2列の接合部が偏析する条件として、線状高分子のTgは、C-C結合の熱膨張が1%になると温度は温度であると考えた。
BLEU_f	0.2003
BLEU_r	0.2052
remarks	Reranked includes over-translation "温度は温度である".

Table 4: Comparison of translation results with reranking and without reranking (case that the BLEU score of translation with reranking is higher than the BLEU score of translation without reranking)

item	value
snt. #	1527
src	This paper shows the optical system .
ref	光学系を示した。
smt	本稿では、光学系を示した。
unreranked	光学系を示した。
reranked	本稿では、光学系を示した。
BLEU_f	1
BLEU_r	0.4463
remarks	"This paper" is not translated in ref and unreranked.

snt. #	817
src	Therapy is mentioned afterwards .
ref	その後の治療法について言及した。
smt	その後療法が挙げられる。
unreranked	その後の治療法に言及した。
reranked	その後療法が挙げられる。
BLEU_f	0.6102
BLEU_r	0.0585
remarks	Worse translation of smt makes bad affects to reranked.

snt. #	509
src	Telomerase activity is high in reproductive cells and cancer cells .
ref	テロメラーゼ活性は生殖細胞と癌細胞で高い。
smt	テロメラーゼ活性は、生殖細胞と癌細胞に高い。
unreranked	テロメラーゼ活性は生殖細胞と癌細胞で高い。
reranked	テロメラーゼ活性は、生殖細胞と癌細胞に高い。
BLEU_f	1
BLEU_r	0.466
remarks	Mistranslation of case marker "に" in smt affects reranked.

snt. #	624
src	In the case of banking sites , the growth of large gully by water erosion was observed .
ref	一方、盛り土では水食による大きなガリの成長が観察された。
smt	盛土サイトの場合には、水による侵食により、大きなガリーの成長を観測した。
unreranked	盛土サイトでは、水侵食による大きなガリの成長が観察された。
reranked	盛土サイトの場合、水侵食による大きなガリーの成長を観測した。
BLEU_f	0.6342
BLEU_r	0.1135
remarks	Src, ref and unreranked are passive voice. Smt and reranked are active voice.

snt. #	1320
src	This paper described views on MEMS ( micro electro mechanical system ) studies in Singapore .
ref	シンガポールにおけるMEMS ( 微小電気機械システム ) 研究の展望を述べた。
smt	本稿は、MEMS ( マイクロエレクトロメカニカルシステム ) についての見解、シンガポールの研究について述べた。
unreranked	シンガポールにおけるMEMS ( マイクロエレクトロメカニカルシステム ) 研究の展望を述べた。
reranked	シンガポールのMEMS ( マイクロエレクトロメカニカルシステム ) 研究についての見解を述べた。
BLEU_f	0.6739
BLEU_r	0.1762
remarks	"Views" is translated to "展望" in ref and unreranked and it is translated to "見解" in smt and reranked.

snt. #	1146
src	This was a case of a male , 66 years old .
ref	症例は66歳男性。
smt	症例は66歳女性であった。
unreranked	症例は66歳男性。
reranked	症例は66歳女性。
BLEU_f	1
BLEU_r	0.5373
remarks	Mistranslation of "male" in smt "女性" affects reranked.

snt. #	544
src	The collection survey was performed on the reef coasts of the northern region of Ibaraki Prefecture .
ref	茨城県北部地域の岩礁海岸で採集調査を行った。
smt	採集調査礁北部地域の茨城県の海岸で行った。
unreranked	茨城県北部の礁海岸で採集調査を行った。
reranked	採集調査は茨城県北部の礁海岸で行った。
BLEU_f	0.5974
BLEU_r	0.1971
remarks	Smt has no meaning however reranked has similar meaning to ref and unreranked.

snt. #	1179
src	This paper surveys the safety of robot , and introduces the issues on safety of robot taken up through the EXPO .
ref	ロボットの安全性を概観し、万博を通じて取り上げられたロボットの安全性に関する課題を紹介した。
smt	本論文では、ロボットの安全性について調査し、ロボットの安全性に関する問題を取り上げ愛を紹介した。
unreranked	ロボットの安全性を概観し、ロボットの安全性に関する課題を紹介した。
reranked	ロボットの安全性について概観し、ロボットの安全性に関する問題点を紹介した。
BLEU_f	0.6615
BLEU_r	0.424
remarks	Both unreranked and reranked include under-translation of " taken up through the EXPO (万博を通じて取り上げられた)".

snt. #	866
src	By the mometasone furancarboxylate ointment , the pit was improved a little , and the pit derived from the normal skin was continued .
ref	フランカルボン酸 モメタゾン 軟膏により、やや陥凹は軽快したが、正常皮膚よりの陥凹は継続した。
smt	furancarboxylate モメタゾン 軟膏外用により、ピットはやや改善し、正常皮膚由来のピットを継続した。
unreranked	シス etasone フランカルボキシ 軟膏 軟膏により、陥凹はやや改善し、正常皮膚からの陥凹が続いた。
reranked	モメタゾンフランカルボン酸 エステル 軟膏により、ピットはやや改善し、正常皮膚からの陥凹が続いた。
BLEU_f	0.2211
BLEU_r	0.1897
remarks	Unreranked includes over-translation "軟膏 軟膏".

snt. #	1313
src	Based on such design guidelines , engine and television of the DLP TMsystem are produced experimentally .
ref	このような設計指針に基づき、DLPTM方式のエンジンとテレビを試作した。
smt	このような設計指針に基づき、エンジン、DLP TMsystemのテレビを試作した。
unreranked	この設計指針に基づき、DLPTMシステムのエンジンとテレビを試作した。
reranked	このような設計指針に基づき、DLPシステム のエンジンとテレビを試作した。
BLEU_f	0.7448
BLEU_r	0.7448
remarks	Reranked includes under-translation of "TM".

Table 5: Comparison of translation results with reranking and without reranking (case that the BLEU score of translation with reranking is lower than the BLEU score of translation without reranking)

#### 4.2 Comparison of evaluation results by BLEU and pairwise evaluation

In this subsection, we examine the contradiction between BLEU score and pairwise evaluation score in ASPEC en-ja task. Several examples that  $\text{diff BLEU} \geq 0$  and pairwise score  $< 0$  are listed in Table 6. Several examples that  $\text{diff BLEU} < 0$  and pairwise score  $\geq 0$  are listed in Table 7.

In these tables, “baseline” means baseline translation result, “BLEU\_b” means BLEU score of baseline and “pairwise” means pairwise evaluation score (sum of five evaluator’s score).

Other abbreviations are explained in the previous subsection.

item	value
snt. #	365
src	Toxic substances in heavy oil is a cause of failure.
ref	重油中に含まれる有害物質が障害の原因となる。
baseline	重油中の有害物質は故障の原因である。
reranked	重質油中の有害物質が障害の原因となっている。
BLEU_b	0.1027
BLEU_r	0.4261
pairwise	-1
remarks	Important information "heavy oil" is incorrectly translated to "重質油" in reranked. Long phrase "有害物質が障害の" is same in ref and reranked, and "有害物質は故障の" in baseline has similar meaning.

snt. #	157
src	During the follow-up period 176 people died by malignant tumor.
ref	追跡期間中に176名が悪性腫瘍により死亡した。
baseline	追跡期間中176人が悪性腫瘍で死亡した。
reranked	フォローアップ期間中、悪性腫瘍により死亡した。
BLEU_b	0.21
BLEU_r	0.3191
pairwise	-5
remarks	Important information "176 people" is not translated in reranked. Long phrase "悪性腫瘍により死亡した。" is same in ref and reranked, and "悪性腫瘍で死亡した。" in baseline has similar meaning.

snt. #	277
src	In addition, the neutronic characteristics of the demonstration reactor are introduced.
ref	このほか実証炉の核特性について紹介した
baseline	さらに、実証炉の中性子特性を紹介した。
reranked	さらに、実証炉の核特性を紹介した。
BLEU_b	0.1836
BLEU_r	0.4035
pairwise	-4
remarks	"Neutronic characteristics" is literally translated to "中性子特性" in baseline and non literally translated to "核特性" in ref and reranked.

snt. #	118
src	A fractal analysis result of potentials related to events was applied to the survey of fluctuating conditions of intelligent activities.
ref	事象関連電位のフラクタル分析を行い、その結果を知的な活動の変動的な状態の調査に使った。
baseline	事象に関連したポテンシャルのフラクタル解析結果を、知的活動の変動条件の調査に適用した。
reranked	事象関連電位のフラクタル解析結果を、知的活動の変動条件の調査に適用した。
BLEU_b	0.1413
BLEU_r	0.2825
pairwise	-3
remarks	"Potentials related to events" is literally translated to "事象に関連したポテンシャル" in baseline and non literally translated to "事象関連電位" in ref and reranked.

Table 6: Comparison of baseline translation and reranked translation (case that  $\text{diff BLEU} \geq 0$  and pairwise score  $< 0$ )

From Table 6, we can see two reasons why  $\text{diff BLEU} \geq 0$  and pairwise score  $< 0$ . First is that important information is differ in ref and reranked and not so important information that is expressed

in long phrase is differ in ref and baseline and is same in ref and reranked. Because of it, diff BLEU  $\geq 0$  and pairwise score  $< 0$ . Second reason is that literal translation in baseline and non-literal translation in ref and reranked.

item	value
snt. #	128
src	No side effect was noted during treatment.
ref	治療中、副作用は認めなかった。
baseline	副作用は認めなかった。
reranked	治療中副作用は認められなかった。
BLEU_b	0.6065
BLEU_r	0.4033
pairwise	5
remarks	Important information "during treatment" is not translated in baseline. Long phrase "副作用は認めなかった。" is same in ref and baseline, and "副作用は認められなかった。" in reranked has similar meaning.

snt. #	76
src	Cutting, patterning polishing, and metalizing work of diamond films with laser beams are reviewed.
ref	レーザーによるダイヤモンド膜の切断、パターニング、研磨、金属化加工について総説した。
baseline	レーザービームを用いたダイヤモンド膜の切断、パターニング、およびメタライゼーション作業についてレビューした。
reranked	レーザービームを用いたダイヤモンド膜の切削、パターニング、研磨、金属化についてレビューした。
BLEU_b	0.3412
BLEU_r	0.1592
pairwise	3
remarks	Important information "metalizing work" is literally translated to "メタライゼーション作業" in baseline and non literally translated to "金属化" in ref and reranked. Long phrase "ダイヤモンド膜の切断、パターニング" is same in ref and baseline and "ダイヤモンド膜の切削、パターニング" in reranked has similar meaning.

snt. #	36
src	Through the detailed analysis of the intensity dependence of excited light, this is judged to be due to photoionization.
ref	励起光強度依存性の詳細解析により、光イオン化によるものと判断した
baseline	励起光の強度依存性の詳細な解析により、光イオン化によると判断した。
reranked	励起光の強度依存性を詳細に解析して、これは光イオン化によるものと判断した。
BLEU_b	0.592
BLEU_r	0.4533
pairwise	3
remarks	Ref, baseline and reranked almost have similar meaning, Ref and baseline are literally near but less fluent than reranked.

Table 7: Comparison of baseline translation and reranked translation (case that diff BLEU  $< 0$  and pairwise score  $\geq 0$ )

The reason why diff BLEU  $< 0$  and pairwise score  $\geq 0$  of example 1 and 2 (snt. # 128 and 76) is similar to the opposite case. The case of snt. # 36 is a little delicate. Three translations: ref, baseline and reranked have almost similar meaning but reranked has rather clear meaning and fluent expression.

## 5 Conclusion

We participate four tasks of WAT2018 with the technique of SMT reranked NMT. This technique can reduce under-translation that is frequently found in NMT. For over-translation, the technique, however, cannot reduce it.

## References

- Terumasa Ehara. 2017. SMT reranked NMT. *Proceedings of the 4th Workshop on Asian Translation (WAT2017)*, pages 119–126.
- Isao Goto, Ka Po Chow, Bin Lu, Eiichiro Sumita and Benjamin K. Tsou. 2013. Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop. *Proceedings of the 10th NTCIR Conference*, pages 260–286.
- Toshiaki Nakazawa, Shohei Higashiyama, Chenchen, Ding, Raj Dabre, Anoop Kunchukuttan, Win Pa Pa, Isao Goto, Hideya Mino, Katsuhito Sudoh and Sadao Kurohashi. 2018. Overview of the 5th Workshop on Asian Translation, *Proceedings of the 5th Workshop on Asian Translation (WAT2018)*.
- Philipp Koehn, Franz J. Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation. *Proceedings of HLTNAACL 2003*, pages 48–54.
- 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.