# Statistical Machine Translation adding Rule based Machine Translation

Jin'ichi Murakami
Department of Information and
Knowledge Engineering
Faculty of Engineering
4-101 koyamachou Tottori City
Tottori 680-8552, Japan
murakami@ike.tottori-u.ac.jp

Masato Tokuhisa
Department of Information and
Knowledge Engineering
Faculty of Engineering
4-101 koyamachou Tottori City
Tottori 680-8552, Japan
tokuhisa@ike.tottori-u.ac.jp

Satoru Ikehara
Department of Information and
Knowledge Engineering
Faculty of Engineering
4-101 koyamachou Tottori City
Tottori 680-8552, Japan
ikehara@ike.tottori-u.ac.jp

## ABSTRACT

We have developed a two-stage machine translation (MT) system. The first stage is a rule-based machine translation system. The second stage is a normal statistical machine translation system. For Japanese-English machine translation, first, we used a Japanese-English rule-based MT, and we obtained "ENGLISH" sentences from Japanese sentences. Second, we used a standard statistical machine translation. This means that we translated "ENGLISH" to English machine translation. We believe this method has two advantages. One is that there are fewer unknown words. The other is that it produces structured or grammatically correct sentences.

From the results of experiments, we obtained a BLEU score of 0.2565 in the Intrinsic-JE task using our proposed method. In contrast, we obtained a BLEU score of 0.2165 in the Intrinsic-JE task using a standard method (moses). And we obtained a BLEU score of 0.2602 in the Intrinsic-EJ task using our proposed method. In contrast, we obtained a BLEU score of 0.2501 in the Intrinsic-EJ task using a standard method (moses).

This means that our proposed method was effective for the Intrinsic-JE and Intrinsic-EJ task. For the future study, we will try to improve the performance by optimizing parameters.

## Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: Machine translation

## General Terms

Languages

## Keywords

SMT Rule-Based MT Hybrid System

## 1. INTRODUCTION

Many machine translation systems have been studied for long time and there was three generations of this technology.

The first generation was a rule-based translation method, which was developed over the course of many years. This method had translation rules that were written by hand. Thus, if the input sentence completely matched the rule, the output sentence had the best quality. However, many expressions are used for natural language, this technology had very small coverage. In addition, the main problem are that the cost to write rules was too high and that maintaining the rules was hard.

The second generation was example-based machine translation method. This method finds a similar sentence from corpus and generates a similar output sentence. The problem with this method is calculating the similarity. Many methods like dynamic program (DP) are available. However, they are very heuristic and intuitive and not based on mathematics.

The third generation was a statistical machine translation method and this method is very popular now. This method is based on the statistics, and it seems very reasonable. There are many versions of statistical machine translation models available. An early model of statistical machine translation was based on IBM1 $\sim$ 5[1]. This model is based on individual words, and thus a "null word" model is needed. However, this "null word" model sometimes has very serious problems, especially in decoding. Thus, recent statistical machine translation systems usually use phrase based models. This phrase based statistical machine translation model has translation model and language model. The phrase table is a translation model for phrase-based SMT and consists of Japanese language phrases and corresponding English language phrases and these probabilities. And word $N$-gram model is used as a language model. By the way, there are two points to evaluate English sentences for Japanese to English machine translation. One is adequacy, and the other is fluency. We believe adequacy is related to translation model $P(English|Japanese)$ and fluency is related to language model $P(English)$.

However some problems arise with phrase-based statistical machine translation. One problem is as follows. Normally, a translation model requires a large parallel corpus. However, if we use a smaller parallel corpus, it results in many unknown words in the output translation. The second problem is that normally, an $N$-gram model is used as a language model. However, this model consists of local language information and does not have grammatical information.

In Japanese-English translation, the first stage consists of Japanese-English rule-based machine translation. In this stage, we obtained "ENGLISH" sentences from Japanese sentences. We aim to achieve "ENGLISH" sentences that contain few unknown words and that are generally grammatically correct. However, these "ENGLISH"

sentences have low levels of fluency and naturalness because they were obtained using rule-based machine translation. In the second stage, we used a normal statistical machine translation system. This stage involves "ENGLISH" to English machine translation. With this stage, we aim to revise the outputs of the first stage improve the naturalness and fluency.

We used IBM King504 (　　　　　　in Japanese) for the first stage. We used general statistical machine translation tools for the second stage, such as "Giza++"[5], "moses" [7], and "training-phrase-model.perl" [10]. And, we could not use all data for restrict of computer memory and computational costs. We used only NTCIR-7 data. It means we used only 1798571 sentences. Also, the score was not optimized, and our method was still very promising. We used these data and these tools and participated in the Intrinsic-JE, Intrinsic-EJ, and Extrinsic-JE. at NTCIR-8.

From the results of experiments, we obtained a BLEU score of 0.2565 in the JE task using our proposed method. In contrast, we obtained a BLEU score of 0.2165 in the Intrinsic-JE task using a standard method (moses). And we obtained a BLEU score of 0.2602 in the Intrinsic-EJ task using our proposed method. In contrast, we obtained a BLEU score of 0.2501 in the Intrinsic-EJ task using a standard method (moses). This means that our proposed method was effective for all task.

As the results, the proposed method was effective for all task. Even though we used only NTCIR-7 database(1798571 sentences), our system had average performance for NTCIR-7 Patent Translation task [14]. For example, our system was the 11th place in 20 system for Intrinsic-JE task and the 19th place in 22 system for Intrinsic-EJ task.

For the future study, we will try to improve the performance by using all NTCIR-8 and NTCIR-7 database and by optimizing parameters. So, we will continue to develop the method and try again in the future.

## 2. RELATED WORKS

Our system has a two-stage machine translation system. The first stage is a rule-based machine translation system, and the second stage is a normal statistical machine translation system. This idea was based on paper[3],[4],[5]. Similar studies were on paper[12],[13],[15][11]. [12] and [13] was Fresh-English translation and used SYSTRAN. [11] was Chineses-English translation for patent task and used SYSTRAN. [15] was Japanese-English translation for patent task.

## 3. CONCEPTS OF OUR STATISTICAL MACHINE TRANSLATLION SYSTEM

Our statistical machine translation consists of a two-stage translation system. The first stage is rule-based machine translation, and the second stage is statistical machine translation. We describe our system by dividing it into two processes, training and decoding. These processes are assumed to be Japanese-English translation.

### 3.1 Training

The training process is as follows.

1. Parallel Corpus

   We prepare a Japanese-English parallel corpus.

2. Rule-based Machine Translation

   We used a Japanese-English rule-based machine translation. Thus, we obtain "ENGLISH" sentences from Japanese sentences. These "ENGLISH" sentences are pairs of English sentences.

3. Make "ENGLISH"-English phrase table

   We make an "ENGLISH"-English phrase table using training-phrase-model.perl[10].

4. English $N$-gram model

   We make an $N$-gram model from English sentences using SRILM [6].

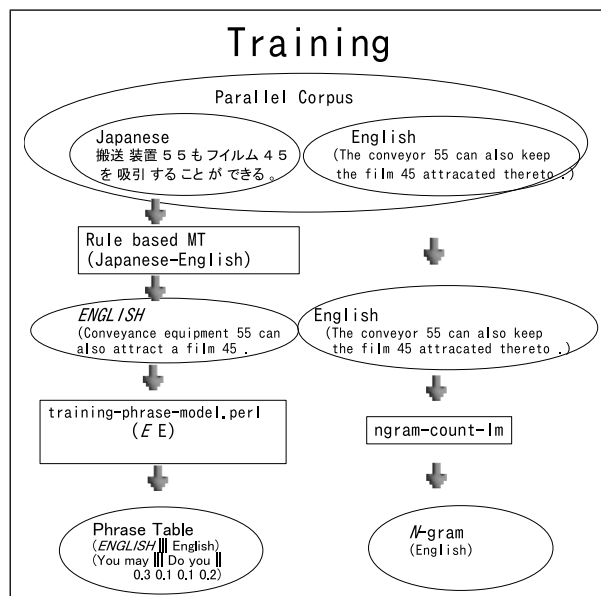Fig. 1 shows the flow chart of the training process.



**Figure 1: Flowchart of Training**

### 3.2 Decoding

The decoding process is as follows.

1. Test Corpus

   We prepare the Japanese test sentences.

2. Rule-based Machine Translation

   We used a Japanese-English rule-based machine translation. Thus, we obtain "ENGLISH" test sentences.

3. Statistical Machine Translation System

   Using phrase table in Section 3.1, $N$-gram model in Section 3.1, and moses[7], we decode the "ENGLISH" sentences. This involves "ENGLISH"-English translation. In this way, we obtain English sentences.
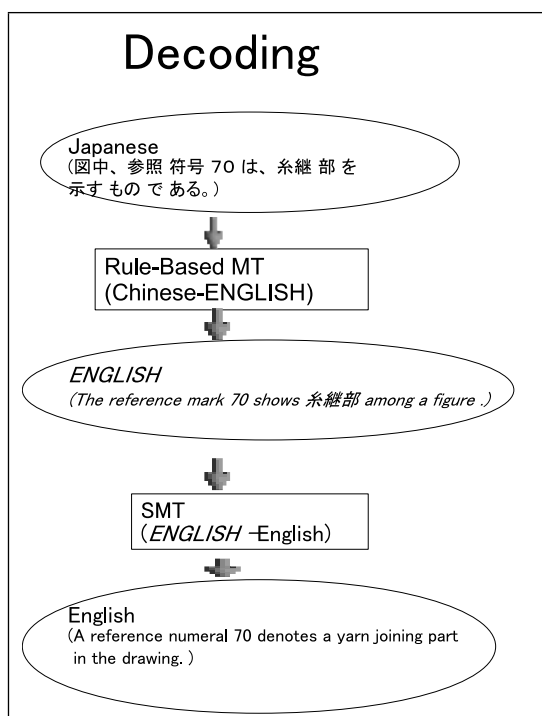
Fig.2 shows the flow chart of the decoding process.

**Figure 2: Flowchart of Decoding**

## 4. EXPERIMENTS WITH OUR MACHINE TRANSLATION

### 4.1 Training Data

We used the English punctuation procedure, which means that we changed "," and "." to " , " and " . ". Also, we did not handle English case forms. Also, we could not use all data for restrict of computer memory and computational costs. We used only NTCIR-7 data. It means we used only 1798571 sentences.

### 4.2 First Stage

We used IBM King504 (             in Japanese) for the first stage. For terms of IBM in Table 5 shows examples of the first stage (IBM) output.

### 4.3 "ENGLISH"-"English" Phrase Tables

For the second stage, we made an <u>ENGLISH</u>-English phrase table. To make this table, we used "train-phrase-model.perl[10]" in "training-release-1.3.tgz". We set parameters to default values.

Table 1 lists examples of phrase tables for the second stage of our MT. This phrase table represents an "ENGLISH" "English" phrase table. As seen in this table, some English phrases are natural, although some of them are unnatural.

### 4.4 $5$-gram Language Model

We calculated the 5-gram model using ngram-count in the Stanford Research Institute Language Model (SRILM) toolkit [6], and used " -ukndiscount -interpolate" as the smoothing parameter.

### 4.5 Decoder

We used "Moses[7]" as a decoder. In Japanese to English translation, the position of the verb is sometimes significantly changed

**Table 1: Examples of phrase-tables**

| |
|---|
| Extremely appropriate . \|\|\|It fits very well . \|\|\| |
|         1 0.0037774 1 0.000165701 |
| Extremely appropriate \|\|\|It fits very well \|\|\| |
|         1 0.00394828 1 0.000167943 |
| Extremely attractive . \|\|\|It is very beautiful . \|\|\|1 |
|         0.00468009 0.5 0.000167226 |
| Extremely attractive . \|\|\|Very beautiful . \|\|\| |
|         1 0.121764 0.5 0.0529012 |
| Extremely attractive \|\|\|It is very beautiful \|\|\| |
|         1 0.00489181 0.5 0.000169488 |
| Extremely attractive \|\|\|Very beautiful \|\|\| |
|         1 0.127273 0.5 0.053617 |
| want to go to eat meal . \|\|\|like to have dinner . \|\|\| |
|         1 4.70488e-06 0.5 0.00340606 |
| want to go to eat meal . \|\|\|want to go to the restaurant . \|\|\| |
|         1 1.02487e-05 0.5 4.7193e-06 |
| want to go to eat meal \|\|\|like to have dinner \|\|\| |
|         0.333333 4.91772e-06 0.5 0.00345215 |
| want to go to eat meal \|\|\|want to go to the restaurant \|\|\| |
|         1 1.07123e-05 0.5 4.78316e-06 |
| want to go to eat \|\|\|like to have \|\|\| |
|         0.0222222 3.18012e-05 1 0.0191019 |
| you eaten ? \|\|\|you tried ? \|\|\| |
|         1 0.0705182 1 0.0519143 |
| you eaten \|\|\|you tried \|\|\| |
|         1 0.0714764 1 0.0524031 |

from its original position. Thus, we set the "distortion weight (weight-d)" to "0.2" and "distortion-limit" to "-1" for standard statistical machine translation. However, our system has 2 stage machine translation and the output of first stage is "ENGLISH". In this case, the position of word did not move so widely. So, we set the "distortion-limit" to "-6" for second stage statistical machine translation for our system.

Table 2 indicates the other parameters. We did not optimize these parameters nor use a reordering model.

**Table 2: Parameters of moses.ini**

| ttable-limit | 40 | 0 | | | |
|---|---|---|---|---|---|
| weight-d | 0.1 | | | | |
| weight-l | 1.0 | | | | |
| weight-t | 0.5 | 0.0 | 0.5 | 0.1 | 0.0 |
| weight-w | -1 | | | | |
| distortion-limit | (-1 or 6) | | | | |

## 5. RESULTS OF OUR MACHINE TRANS-LATION

Table 3 summarizes the results of our machine translation evaluation for the Intrinsic-JE and Intrinsic-EJ tasks. In this table, "Proposed" indicates our proposed system, "Baseline" indicates the normal statistical machine translation (moses).

**Table 3: Results**

| | task | bleu |
|---|---|---|
| Proposed (TORI) | Intrinsic-JE | 0.2565 |
| Baseline (moses) | Intrinsic-JE | 0.2156 |
| Proposed (TORI) | Intrinsic-EJ | 0.2602 |
| Baseline (moses) | Intrinsic-EJ | 0.2501 |

As seen in these results, our method was so effective, as indicated by the BLEU score.

## 6. DISCUSSION

With our system, we aim to reduce the number of unknown words and ungrammatical sentences. Thus, we analyze the outputs according to these factors.

### 6.1 Unknown Words

If we compare the outputs of moses and of our system, we find very few unknown words. Therefore, the proposed method is effective in reducing the number of unknown words.

### 6.2 Grammatical Correctness

We analyzed the outputs of our MT system. However, there were no native speakers of Enlish to check the inputs and outpus. Therefore, it was impossible to analyze these results for detail and could not determine what was wrong. However, by comparing the output of moses and the output of our system, we found that our system produced more grammatically correct sentences. Therefore, the BLEU score was so good.

## 7. CONCLUSION

We have developed a two-stage machine translation system. The first stage is a rule-based machine translation system. The second stage is a statistical machine translation system. Our goal with this system was to obtain fewer unknown words and fewer ungrammatical sentences. The results that we obtained in this experiments were so good.

We did not use all training data nor optimize parameters nor did we use a reordering model. In future experiments, we will try these data and these techniques, which we expect will enable our system to perform better.

## 8. APPENDIX: EXPERIMENTS WITH ALL TRAINING DATA

We tried to used all training data ( 2920434 sentence ) to improve these results. Also we did not use reordering models or not optimize these parameters using MERT in this experiments.

Table4 shows the results of these experiments. As can be seen this table, proposed method was so effective for BLEU score. Theses BLEU score was the best group of NTCIR-8.

**Table 4: Appendix: Results with All data**

|                  | task         | bleu   | nist   | meteor |
|------------------|--------------|--------|--------|--------|
| Proposed         | Intrinsic-JE | 0.2924 | 7.2904 | 0.6216 |
| Baseline (moses) | Intrinsic-JE | 0.2229 | 6.1266 | 0.5842 |
| Proposed         | Intrinsic-EJ | 0.3276 | 7.5638 |        |
| Baseline (moses) | Intrinsic-EJ | 0.3232 | 7.2663 |        |

## 9. ACKNOWLEDGEMENTS

## 10. REFERENCES

[1] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. "The machinematics of machine translation: Parameter estimation", Computational Linguistics, 19(2): pp. 263-311. (1993).

[2] Philipp Koehn, Franz J. Och, and Daniel Marcu. " Statistical phrase-based translation". In Marti Hearst and Mari Ostendorf, editors, HLT-NAACL 2003: Main Proceedings, pages 127.133, Edmonton, Alberta, Canada, May 27 -June 1. Association for Computational Linguistics. (2003).

[3] Pierre Isabelle, Cyril Goutte, and Michel Simard., "Domain Adaptation of MT systems through automatic post-editing", MT Summit XI, 102, 2007.

[4] Yushi Xu and Stephanie Seneff, "Two-Stage Translation: A Combined Linguistic and Statistical Machine Translation Framework", Proceedings of the Eighth Conference of the Association for Machine Translation (AMTA) 2008.

[5] Jason Katz-Brown and Michael Collins, " Syntactic Reordering in Preprocessing for Japanese English Translation: MIT System Description for NTCIR-7 Patent Translation Task", Proceedings of the 7th NTCIR Workshop Meeting, 2008.

[5] Franz Josef Och, Hermann Ney. "A Systematic Comparison of Various Statistical Alignment Models", Computational Linguistics, volume 29, number 1, pp. 19-51 March 2003.

[6] Andreas Stolcke, "SRILM - An Extensible Language Modeling Toolkit", in Proc. Intl. Conf. Spoken Language Processing, Denver, Colorado, September 2002

[7] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondej Bojar, Alexandra Constantin, Evan Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation", Proceedings of the ACL 2007 Demo and Poster Sessions, pages 177180, June 2007.

[8] NIST Open Machine Translation, http://www.nist.gov/speech/tests/mt

[9] Banerjee, S. and A. Lavie, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments", Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), June 2005.

[10] NAACL 2006 WORKSHOP ON STATISTICAL MACHINE TRANSLATION Shared Task: Exploiting Parallel Texts for Statistical Machine Translation Shared Task Baseline System, training-release-1.3.tgz, http://www.statmt.org/wmt06/shared-task/baseline.html

[11] Jin'ichi Murakami, Masato Tokuhisa, Satoru Ikehara, "Statistical Machine Translation adding Pattern-based Machine Translation in Chinese-English Translation", International Workshop on Spoken Language Translation 2009, (IWSLT 2009), pp.107-112, October 2009.

[12] L.Dugast, J.Senellart, and P.Koehn, "Statistical postediting on SYSTRAN's rule-based translation system", in Second Workshop on SMT, 2007, pages.179-182

[13] M.Simard, N.Ueffing, P.Isabelle, and R.Kuhn, "Rule-based translation with statistical phrase-based post-editing", in Second Workshop on SMT, 2007, pages.203-206

[14] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro, Terumasa Ehara, Hiroshi Echizen-ya, Sayori Shimohata, "Overview of the Patent Translation Task at the NTCIR-8 Workshop. Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access", 2010.

[15] Terumasa EHARA, "Rule Based Machine Translation Combined with Statistical Post Editor for Japanese to English Patent Translation", Proceedings of Machine Translation Summit XI, Workshop on Patent Translation, pp.13-18, Sept., 2007.

**Table 5: Outputs for Intrinsic-JE**

| 107 | Input | |
|---|---|---|
| | Proposed | In FIG . 11 , when high-pressure pump discharge quantity is larger , the amount of change is small with respect to the amount of change in the amount of discharge of the high pressure pump spill valve drive timing . |
| | Baseline | In FIG . 11 , the high-pressure pump discharge quantity is large or small , or when the spill valve drive timing variations with respect to variation of the high-pressure pump discharge quantity is small . |
| | IBM | In Fig . 11 , when high-pressure pump　　　　is size or smallness , the amount of change of high-pressure pump　　　　is small to the amount of change of　　　　valve drive timing . |
| | REFERENCE | In FIG . 11 , when the high-pressure pump discharge quantity is large or small , the change quantity in the high-pressure pump discharge quantity is small with respect to the change quantity in the spill valve drive timing . |
| 108 | Input | |
| | Proposed | The amount of each cylinder is calculated based on the target fuel pressure PFo and the fuel pressure PF to the PID control unit 200 , and the feedback amount is set as the feed forward from the fuel injection valve in an amount equal the amount of fuel supplied to the fuel feed forward operation section 100 , QFinj QFspl2 . |
| | Baseline | The amount of fuel QF is set as a final fuel injection amount QF based on the fuel pressure PF and the target fuel pressure PFo and the PID control unit 200 as a feedback signal . 2 pl is calculated from the fuel injection valve feed forward amount calculating section 100 is supplied to each cylinder of the feed forward amount . |
| | IBM | In the amount operation part 100 of　　　　forwards , the fuel of the amount of fuel supplied to each cylinder from a fuel injection valve and the amount of said is set up as an amount QFinj of　　　　forwards , and the amount QFspl2 of feedbacks is computed based on target　　　　PFo and　　　　PF in the PID control part 200 . |
| | REFERENCE | In the feedforward quantity computing unit 100 , the same quantity of fuel as the quantity of fuel supplied from the fuel injection valve to each cylinder is set as a feedforward quantity QFinj , and in the PID control unit 200 , a feedback quantity Qfspl2 is calculated on the basis of the target fuel pressure PFo and the fuel pressure PF . |
| 109 | Input | |
| | Proposed | First , as shown in FIG . 7 , when the predetermined range is changed based on the engine rotational speed N , and the threshold value 1XPFH 2XPHL are set . |
| | Baseline | First , the engine speed is within a predetermined range based on a change in the threshold value 1 is set , for example , as shown in FIG . 2 , the threshold value H XPF XP HL . |
| | IBM | First , as shown in Fig . 7 , when the predetermined range is changed based on organization rotation speed , threshold 1XPFH and threshold 2XPHL are set up . |
| | REFERENCE | First , when the predetermined range is changed on the basis of the engine rotation speed , the threshold value 1XPFH and the threshold value 2XPFL are set as shown in FIG . 7 , for example . |
| 111 | Input | |
| | Proposed | Next , the operation of the second feedback quantity of the fuel pressure control apparatus for a direct injection type internal combustion engine combustion chamber , shown in FIG . 1 will now be described with reference to the flow chart of FIG . 14 . |
| | Baseline | A description will now be given , with reference to FIG . 14 of the direct cylinder fuel injection control device for an internal combustion engine having the above structure will now be described as a second feedback amount of the behavior of the fuel pressure . |
| | IBM | Next , the action of the second amount of feedbacks in the fuel pressure control device of the charge injection formula internal combustion engine of pipe internal combustion constituted as mentioned above is explained , referring to the diagram of Fig . 14 of operation . |
| | REFERENCE | Next , the behavior of the second feedback quantity in the fuel pressure control apparatus of the cylinder fuel injection type internal combustion engine configured as described above will be described with reference to FIG . 14 . |
| 112 | Input | |
| | Proposed | When the integral term I is larger than the target amount of one or more of threshold values is calculated using the following equation at step S409 . XPFHs QFspl1 except for the discharge amount ( i.e. , YES ) , PFFB equation ( A ) |
| | Baseline | The integral term I / F _ B is larger than the threshold ( that is , YES ) , the routine proceeds to step 409 , the amount of the target discharge quantity Q is calculated . 1 excluding at least 1 according to the following equation ( A ) PF type XPF pl Fs . |
| | IBM | When integration clause PFFB _ I is larger than a formula ( A ) ( namely , YES ) , target　　　　QFspl1 except the quantity of 1 or more XPFHs of thresholds is computed by the following formula at Step S409 here . |
| | REFERENCE | Here , when the integral term PFFB _ I is larger than the expression ( A ) ( i.e. , YES ) , the target discharge quantity QFspl1 excluding a quantity equal to or greater than a threshold value 1XPFH is calculated by the following expression in step S409 . |

Input means input Japanese sentence. Proposed means the output our proposed method. Baseline means the output of moses.
IBM meas the output of IBM King504 (　　　　　　). REFERENCE means the correct sentence and handmade.

## Table 6: Outputs for Intrinsic-EJ

| 5 | Input | Through the temperature control circuit ( CPU ) , the operation of the heater is controlled so that the temperature to be detected by each thermistor becomes the target temperature . |
|---|---|---|
| | Proposed | |
| | Baseline | |
| | IBM | ( C P U ) |
| | REFERENCE | |
| 6 | Input | As the abnormal temperature rise observed in this embodiment is up to approximately 190 & # xb0 ; C. , the high-temperature offset and downtime are not caused . |
| | Proposed | |
| | Baseline | ° |
| | IBM | 1 9 0 & # x b 0 ;   C |
| | REFERENCE | |
| 8 | Input | The fixing roller 51 rotates clockwise as indicated by the arrow . |
| | Proposed | |
| | Baseline | |
| | IBM | 5 1 |
| | REFERENCE | |
| 13 | Input | FIG . 8 shows the belt , seen from the fixing roller . |
| | Proposed | |
| | Baseline | |
| | IBM | 8 |
| | REFERENCE | |
| 15 | Input | The pressure belt 53 is brought into contact with the fixing roller 51 . |
| | Proposed | |
| | Baseline | |
| | IBM | 5 3       5 1 |
| | REFERENCE | |
| 17 | Input | Referring now to FIG . 3 , the arrangement of the fixing unit 9 in the longitudinal width direction is described . |
| | Propoased | |
| | Baseline | |
| | IBM | 3      9 |
| | REFERENCE | |
| 19 | Input | Next , the fixing unit 9 in accordance with the first embodiment is described in conjunction with FIG . 2 . |
| | Proposed | |
| | Baseline | |
| | IBM | 9       2 . |
| | REFERENCE | |
| 22 | Input | The width of the conductor patterns 11a & # x2013 ; 11d and lead conductors 13 , 14 after fired is set , for example , to about 40 & # x3bc ; m . |
| | Proposed | μ |
| | Baseline | μ |
| | IBM | 4 0 & # x 3 b c ; m     1 1 a & # x 2 0 1 3 ; 1 1 d    1 3   1 4 |
| | REFERENCE | μ |

Input means input Eapanese sentence. Proposed means the output our proposed method. Baseline means the output of moses.
IBM means the output of IBM King504 ( ). REFERENCE means the correct sentence and handmade.