Two stage Machine Translation System using Pattern-based MT and Phrase-based SMT

Jin'ichi Murakami
Department of Information and Knowledge Engineering,
Faculty of Engineering,
Tottori University, Japan
murakami@ike.tottori-u.ac.jp

Takuya Nishimura
Department of Information and Knowledge Engineering,
Faculty of Engineering,
Tottori University, Japan
s062044@ike.tottori-u.ac.jp

Masato Tokuhisa
Department of Information and Knowledge Engineering,
Faculty of Engineering,
Tottori University, Japan
tokuhisa@ike.tottori-u.ac.jp

Abstract

We have developed a two-stage machine translation (MT) system. The first stage consists of an automatically created pattern-based machine translation system (PBMT), and the second stage consists of a standard phrase-based statistical machine translation (SMT) system. We studied for the Japanese-English simple sentence task.

First, we obtained English sentences from Japanese sentences using an automatically created Japanese-English pattern-based machine translation. We call the English sentences obtained in this way as *"English"*. Second, we applied a standard SMT (Moses) to the results. This means that we translated the *"English"* sentences into English by SMT. We also conducted ABX tests(Clark, 1982) to compare the outputs by the standard SMT (Moses) with those by the proposed system for 100 sentences.

The experimental results indicated that 30 sentences output by the proposed system were evaluated as being better than those outputs by the standard SMT system, whereas 9 sentences output by the standard SMT system were thought to be better than those outputs by the proposed system. This means that our proposed system functioned effectively in the Japanese-English simple sentence task.

1 Introduction

Machine translation (MT) systems have been extensively studied, and there are now three generations of this technology. The first generation consists of rule-based MT (RBMT) systems. A pattern-based MT (PBMT) system is a kind of RBMT system. The second generation consists of example-based machine translation systems, and the third generation consists of statistical machine translation (SMT) systems, which have become very popular. Many versions of SMT systems have been introduced. An early SMT system was based on word-based models (IBM 1 $\sim$ 5(Brown et al., 1993)). Recent statistical MT systems have usually used phrase-based models.

However, some problems arise with phrase-based SMT. One problem is the language model. Generally, an N-gram model is used as the language model. However, this kind of model includes only local language information and does not include grammatical information. To solve this problem, we developed a two-stage MT system. The first stage consists of an automatically created PBMT system, and the second stage consists of a standard SMT system.

For Japanese-English translation, the first stage consists of Japanese-English PBMT. In this stage, we obtain *"English"* sentences from Japanese sentences. Our aim is to produce grammatically correct *"English"* sentences. However, these *"English"* sentences sometimes have low levels of fluency because they were obtained using an automatically created PBMT. In the second stage, we use a standard SMT system. This stage involves *"English"* to English machine translation. With this stage, our aim is to revise the outputs of the first stage in order to improve fluency.

We developed a PBMT system for the first stage using "train-model.perl"(Koehn et al., 2007). We also developed a standard SMT system for the second stage using general SMT tools such as "Moses" (Koehn et al., 2007). We used these data and tools to translate Japanese-English simple sentences.

We obtained a Bilingual Evaluation Understudy (BLEU) score(Papineni et al., 2002) of 0.1821 with our proposed system. In contrast, we obtained a BLEU score of 0.2218 in the Japanese-English simple sentences using a standard SMT system (Moses). This means that the proposed system was not effective for automatic evaluation in the Japanese-English simple sentence task.

However, we conducted ABX tests(Clark, 1982) to compare the output of the standard SMT system (Moses) and the output of the proposed system for 100 sentences. The results indicated that 30 sentences of the proposed system were thought to be better than those of the standard SMT system, and 9 sentences of the standard SMT system were thought to be better than those of the proposed system. This means that our proposed system was effective in the Japanese-English simple sentence task for human evaluation.

2 Concept of Two-Stage Machine Translation

One problem with phrase-based statistical machine translation is the language model. Generally, an N-gram model is used as the language model. However, this model includes only local language information and does not include grammatical information. We studied hierarchical phrase-based statistical machine translation (HSMT) (Li et al., 2009) as a way to include grammatical information. However, HSMT analysis is similar to that of context-free grammars (CFG). We believe that such analysis complicates statistical machine translation by adding too many parameters. Therefore, it is unreliable and does not perform well, especially for the small amount of training data. On the contrary, PBMT is well known and has been extensively studied. Normally, PBMT is simple and has few parameters compared to CFG-based MT, and the output of PBMT contains grammatical information. However, there is a trade-off between the coverage of input sentences and the translation quality in the PBMT results. If we obtain good translation quality, then the coverage of RBMT for input sentences is low in the translation. If we obtain high coverage for input sentences, the translation quality is low.

We propose a two-stage MT system to overcome these problems. We developed a PBMT system for the first stage. This PBMT system had low coverage and high quality. When Japanese sentences were translated using this system, the quality of the output was good, and the outputs contained grammatical information. When not using the PBMT system to translate Japanese sentences, we used a standard SMT system. Therefore, we can obtain good quality from the entire system. Also, PBMT systems are usually created manually, which results in a huge labor cost. Therefore, we developed an automatically created PBMT system. However, this automatic PBMT output sometimes had less fluency, so we added SMT after PBMT to improve the fluency. In this system, we used PBMT in the pre-processing stage of SMT.

3 Related Work

Two-stage MT systems have been proposed before (Xu and Seneff, 2008), (Ehara, 2007), (Dugast et al., 2007), (Simar et al., 2007). L. Dugast, et al. (Dugast et al., 2007) and M. Simard, et al. (Simar et al., 2007) applied SYSTRAN and SMT for Japanese-English translation. Their concept was to use SMT as a post-process for SYSTRAN. The results of these studies indicated that these systems are more effective than using SYSTRAN or SMT alone. In M. Simard's research (Simar et al., 2007), the BLEU score was 0.2598 for SMT and 0.2880 for SYSTRAN + SMT in English-Japanese translation, and 0.2517 for SMT and 0.2679 for SYSTRAN + SMT in Japanese-English translation. Ehara(Ehara, 2007) reported on the same system for Japanese-English translation of a patent task. The BLEU score was 0.2821 for SMT and 0.2921 for RBMT + SMT. Ehara's RBMT system was a commercial Japanese-English system. For these systems, SMT was used in the post-process for RBMT, which means that SMT was used as a means of language adaptation. Also, these RBMT systems were created by hand, so they were expensive to build.

4 Pattern-Based Machine Translation

We developed an automatically created Japanese-English pattern-based machine translation system using “train-model.perl” (Koehn et al., 2007). Our system is divided into two processes. One is a process to form Japanese-English patterns, and the other is a decoding process. The details of these two processes are described below.

4.1 Japanese-English Patterns

We developed the following process for forming Japanese-English patterns.

1. Parallel Japanese-English Corpus

   We prepare Japanese-English parallel sentences for training. Example sentences are listed in Table 1.

Table 1: Parallel Japanese-English Corpus

Japanese sentence	信号 は 赤 だ 。
English sentence	The light was red .

2. Japanese-English Phrase Table

   We construct a Japanese-English phrase table using train-model.perl (Koehn et al., 2007). An example Japanese-English phrase table is given in Table 2.

Table 2: Example of Japanese-English Phrase Table

Ex.1	信号 |||The light |||0.5 0.07 0.5 0.2
Ex.2	信号 は |||lights |||0.01 0.06 0.03 0.04
Ex.3	赤 |||red |||0.2 0.1 0.2 0.3
Ex.4	赤 だ |||red |||0.3 0.2 0.2 0.2

3. Japanese-English High Probability Phrase Table

   We deleted the low-probability Japanese-English phrase table (Table 2), in which the threshold was 0.1. We call the resulting table a Japanese-English high-probability phrase table (HPPT). An example of an HPPT is presented in Table 3.

Table 3: Example of Japanese-English High Probability Phrase Table

Ex.1	信号 |||The light |||0.5 0.07 0.5 0.2
Ex.3	赤 |||red |||0.2 0.1 0.2 0.3
Ex.3	赤 だ |||red |||0.3 0.2 0.2 0.2

4. Japanese-English Patterns

   We used Japanese-English parallel sentences (Table 1) and the Japanese-English HPPT (Table 3) to form Japanese-English patterns. Note that all possible Japanese-English patterns were generated. Therefore, one or more Japanese-English patterns were generated from one Japanese-English parallel sentence. Example Japanese-English patterns are listed in Table 4.

Table 4: Japanese-English Patterns

Ex.1	Japanese pattern	X1 は X2 だ 。
	English pattern	X1 was X2 .
Ex.2	Japanese pattern	X1 は X2 。
	English pattern	X1 was X2 .

Figure 1 shows a flowchart for forming Japanese-English patterns.

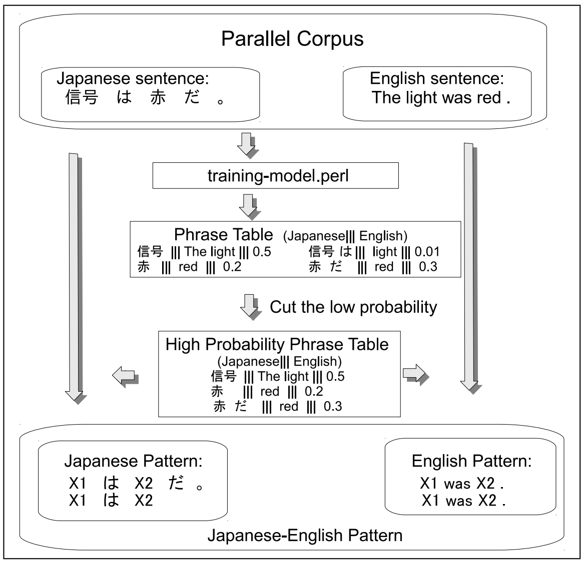

Figure 1: Formation of Japanese-English Patterns

4.2 Decoding Pattern

The decoding process is as follows.

1. Input Japanese Sentences

   We prepare input Japanese sentences. An example sentence is given in Table 5.

Table 5: Japanese Sentence

郵便 局 は どこ に 有り ます か ？

2. Search Japanese Pattern and Output English Pattern

   We search for a Japanese pattern that is matched with the input Japanese sentence using Japanese patterns and the HPPT (section 4.1). Then we obtain English patterns. Example Japanese-English patterns are listed in Table 6. Also, an example Japanese-English HPPT is shown in Table 7.

Table 6: Japanese-English Patterns

Ex.1	Japanese Pattern	X1 X2 は どこ に 有り ます か ？
	English Pattern	Where's the X2 X1 ?
Ex.2	Japanese Pattern	X2 は どこ に 有り ます か ？
	English Pattern	Where is a X2 ?

Table 7: Japanese-English High-Probability Phrase Table

Ex.1	局 |||post |||0.5 0.07 0.5 0.21
Ex.2	郵便 |||postal |||0.4 0.031 0.2 0.11
Ex.3	郵便 局 |||postal service |||0.1 0.07 0.1 0.01

3. Generate English Sentences

   We generate English sentences using the English pattern and Japanese-English High-Probability phrase tables. Note that all possible English sentences are generated. Therefore, multiple English sentences are generated from an input Japanese sentence. Example English sentences are listed in Table 8.

Table 8: Generated English Sentences

Ex.1	Where's the post office ?
Ex.2	Where is a post station ?

4. Select English Sentence.

   We select one English sentence from the multiple generated English sentences using 3-gram. We used the n-gram-count in the Stanford Research Institute Language Model (SRILM) toolkit (Stolcke, 2002) and used "-ukndiscount -interpolate" as the smoothing parameter.

   An example of an English sentence that might be selected is shown in Table 9.

Table 9: Select English Sentence

Where is a post station ?

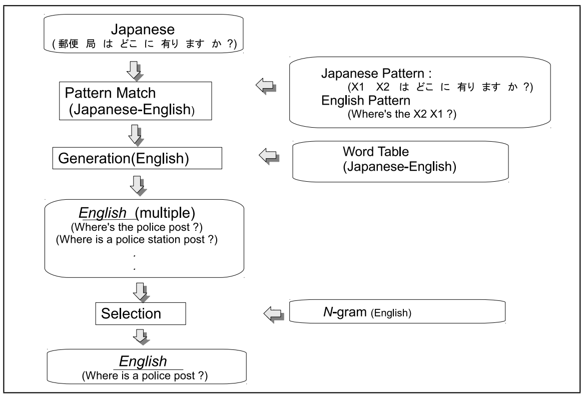

Figure 2: Decoding of Pattern-Based Machine Translation

Figure 2 shows the process of decoding English sentences for PBMT. We refer to the output of the proposed pattern-based machine translation as *"English"* sentences.

5 Overview of Proposed Machine Translation System for Training

The training model of our proposed machine translation system has three parts. The first process involves constructing an *"English"*-English phrase table, the second process involves constructing a Japanese-English phrase table, and the third part involves constructing a language model (N-gram).

5.1 *"English"*-English phrase table

"English"-English phrase tables are constructed as follows.

1. Parallel Corpus

   We prepare a Japanese-English parallel corpus.

2. Pattern-Based Machine Translation

   We use Japanese-English PBMT. Thus, we obtain *"English"* sentences from Japanese sentences. These *"English"* sentences are pairs of English sentences.

3. *"English"*-English phrase tables

   We construct *"English"*-English phrase tables using Giza++ (Och and Ney, 2003) and train-model.perl (Koehn et al., 2007) from the *"English"* sentences (outputs of Japanese-English PBMT) and English sentences (from the parallel corpus).

Figure 3 is a flow chart that shows how *"English"*-English phrase tables are constructed.

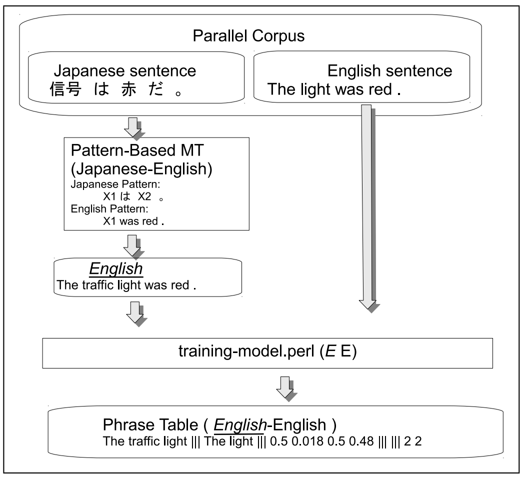

Figure 3: Flowchart for constructing *"English"*-English Phrase Tables

5.2 Japanese-English Phrase Table

We construct a Japanese-English phrase table using Giza++ (Och and Ney, 2003) and train-model.perl (Koehn et al., 2007) using the Japanese-English parallel corpus. Figure 4 shows a flow chart for constructing Japanese-English phrase tables.

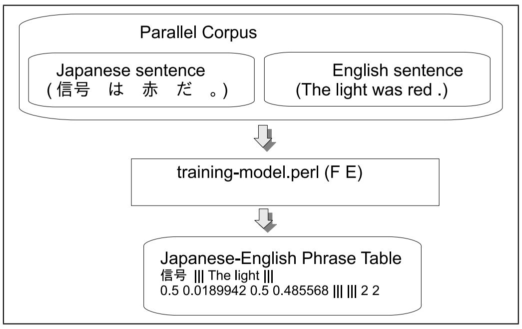

Figure 4: Flowchart for Constructing Japanese-English Phrase Tables

5.3 Language Model (N-gram).

We calculated the 5-gram model using the n-gram-count in the SRILM toolkit (Stolcke, 2002) and used "-ukndiscount -interpolate" as the smoothing parameter.

6 Overview of Proposed Machine Translation System for Decoding

The decoding process is as follows.

1. Test Corpus

   We prepare Japanese test sentences.

2. Pattern-Based Machine Translation

   We use a Japanese-English Pattern-Based Machine Translation. If an input Japanese sentence matches the Japanese patterns, we can obtain a translated *"English"* test sentence.

3. *"English"*-English Statistical Machine Translation

   Using an *"English"*-English phrase table, N-gram model, and Moses (Koehn et al., 2007), we decode the *"English"* test sentence. This involves *"English"*-English translation, resulting in an English sentence.

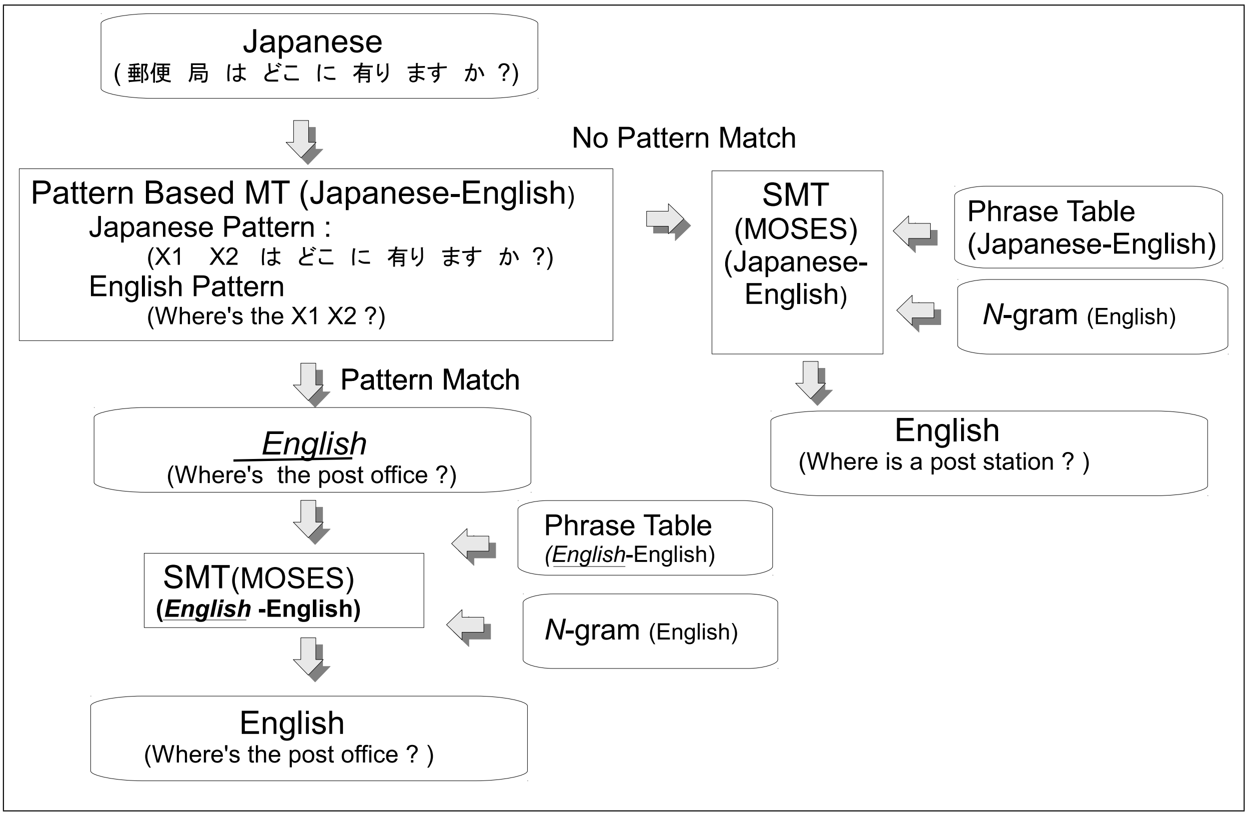

Figure 5: Flowchart of Decoding Process

4. Japanese-English Statistical Machine Translation System

   If an input Japanese sentence does not match the Japanese patterns, we conduct a standard Japanese-English SMT using a Japanese-English phrase table and N-gram model to obtain an English sentence.

Figure 5 shows a flowchart of the decoding process.

7 Experiments with our Machine Translation System

7.1 Japanese-English Simple Sentence Task

We collected a large number of Japanese-English parallel sentences from many electronic media sources. Then we selected simple sentences from these Japanese-English parallel sentences(Murakami et al., 2007). We used these Japanese-English simple sentences for training and development and test data.

Also we formed these sentences as follows. We used the English punctuation system, which means we changed “,” and “.” to “ ” and “ ’’. And, we did not take into account English case forms. Also, we used Chasen (Asahara and Matsumoto, 2000) as the Japanese morphological analyzer.

7.2 Data Sets

1. Training Data

   A total of 100,000 Japanese-English simple sentences were used for training data.

2. Development Data

   We used 3,000 sentences for development data for Japanese-English SMT. Of these 3,000 sentences, 375 matched the Japanese-English patterns. We therefore obtained 375 *“English”* sentences for the results. These 375 *“English”* sentences were used as development data for the *“English”*-English translation.

3. Test Data

We used 10,000 Japanese-English simple sentences as test sentences.

7.3 *"English"*-English Phrase Tables

For the second stage, we constructed an *"English"*-English phrase table using Giza++ (Och and Ney, 2003) and "train-model.perl"(Koehn et al., 2007). We set default values for the parameters. Also, 60,000 of the 100,000 training sentences matched the Japanese-English patterns. Thus, we used these 60,000*"English"* sentences to make an *"English"*-English phrase table.

7.4 N-gram Language Model

We built an N-gram language model using 100,000 sentences.

7.5 Decoder

We used "Moses"(Koehn et al., 2007) as a decoder. We also used parameter tuning (MERT) and reordering models. Note that in Japanese-English translation, the position of the verb is sometimes significantly changed from its original position. Thus, we used the unlimited word reordering for a standard SMT. So, we set the "distortion-limit" set to "-1" for a standard SMT. However, our system consists of two-stage machine translation, and the output of the first stage is *"English"*. Consequently, word positions did not dramatically change. Therefore, we set the "distortion-limit" to "6" for the second-stage SMT for our system.

8 Results of our Machine Translation

8.1 Examples of output sentences

Table 10 lists example sentences from our proposed system for the Japanese-English simple sentences. These example sentences are matched with the Japanese-English patterns. In this table, "Input" indicates the input Japanese sentence, "Proposed" indicates the output of our proposed system (PBMT+SMT), "Reference" indicates a correct sentence, and "Moses" indicates the output of a standard SMT.

Table 10: Example Outputs for Japanese-English simple sentences

Input	土手 が 切れ た 。
Proposed	We are out of dikes .
Reference	The bank gave way .
Moses	The bank broke .
Input	この 薬 は 歯痛 に 効く 。
Proposed	This medicine for A toothache .
Reference	This medicine helps a toothache .
Moses	This medicine acts on the toothache .
Input	火 は 台所 から 出 た 。
Proposed	The fire started in the kitchen .
Reference	The fire started in the kitchen .
Moses	The fire started in the kitchen .
Input	内閣 が つぶれる 。
Proposed	The Cabinet collapses .
Reference	The cabinet is dissolved .
Moses	The Cabinet goes bankrupt .
Input	彼女 は フランス へ 行っ た 。
Proposed	She went to France .
Reference	She went over to France .
Moses	She went to France .

8.2 Automatic Evaluations

We used 10,000 test sentences in this experiment. Among these 10,000 sentences, 1,143 sentences matched the Japanese-English patterns. The results of *"English"*-English translation revealed that 725 out of the 1,143 sentences were different compared to the standard SMT system (Moses). The other 8,857 sentences (10,000 - 1,143) did not match the Japanese-English pattern.

We used the BLEU (Papineni et al., 2002) and NIST (NIST, 2003) and METEOR (Banerjee and Lavie, 2005) for evaluation tools. Table 11 summarizes the automatic evaluation results of our machine translation evaluation for the Japanese-English simple sentences. This table shows the results of 1,143 sentences that were matched with the Japanese-English patterns. "Proposed" indicates our proposed system (PBMT+SMT), and "Moses" indicates a standard SMT system.

We obtained a BLEU score of 0.1821 in the Japanese-English simple sentences using our proposed system. In contrast, we obtained a BLEU score of 0.2218 in the Japanese-English simple sentences using the standard SMT system (Moses). This means that our proposed system was not effective for automatic evaluation in the Japanese-English simple sentences.

Table 11: Experimental Results (1,143 sentence)

	BLEU	NIST	METEOR
Proposed	0.1821	4.817	0.4426
Moses	0.2218	5.239	0.4363

Table 12 shows the all test sentences(10,000 sentences). The 1,143 sentences were translated with the proposed method. The rest of 8,857 sentences were translated with the standard SMT system (Moses).

Table 12: Experimental Results (10,000 sentence)

	BLEU	NIST	METEOR
Proposed	0.1101	4.4511	0.3175
Moses	0.1130	4.5131	0.3160

8.3 Human evaluation

We conducted an ABX test(Clark, 1982), which is a human evaluation method, in order to compare the outputs by the proposed method with those by Moses.

8.3.1 Evaluation Criteria

We organized the outputs into four categories according to the following evaluation criteria. Also, we converted unknown words into the "romaji" characters.

1. Proposed $>$ Moses

   This refers to the case when the output of the proposed method was better than that of Moses. Example sentences are listed in Table 13.

Table 13: Example of "Proposed $>$ Moses"

Input	私 は 彼女 に 結婚 を 申し込ん だ 。
Proposed	I made a proposal of marriage to her .
Reference	I proposed to her .
Moses	I He asked her for her hand .
Input	彼女 は 5 人 の 子供 を 育て た。
Proposed	She brought up five children .
Reference	She has brought up five children .
Moses	She is five children .

2. Proposed $<$ Moses

   This is the case when the output of Moses was better than that of the proposed method. Example sentences are listed in Table 14.

Table 14: Example of "Proposed $<$ Moses"

Input	仕事 は 山場 に 入っ た 。
Proposed	work went into the labor-management .
Reference	Work has reached the critical point .
Moses	The work is appear to have entered the final stage .
Input	農園 は 道路 に 接し て いる 。
Proposed	The farm is roads are .
Reference	The farm abuts on the road .
Moses	Farm adjoins the road .

3. Proposed $\approx$ Moses

   In this case, the output of the proposed method is the same quality as those by Moses. Example sentences are listed in Table 15.

Table 15: Example of "Proposed $\approx$ Moses"

Input	豊作 に なり そう だ 。
Proposed	It looks like rejoicing .
Reference	The harvest looks promising .
Moses	Hopes looks like .
Input	彼 に よろしく お 伝え ください 。
Proposed	Please send him my best wishes .
Reference	Give him my good wishes .
Moses	Please give my best regards to him .

4. Proposed $=$ Moses

   This refers to when the output of the proposed method and the output of Moses were exactly the same. Examples of such sentences are listed in Table 16.

Table 16: Example of "Proposed $=$ Moses"

Input	彼 は 故郷 を 恋し がっ て いる 。
Proposed	He is homesick .
Reference	He is sick for home .
Moses	He is homesick .

8.3.2 Results of Human Evaluation

We randomly selected 100 sentences from the 1,143 output sentences that were matched with the

Japanese-English patterns. Then we evaluated these 100 sentences. The results are listed in Table 17.

Table 17: Results of Human Evaluation

Proposed > Moses	30 / 100
Proposed < Moses	9 / 100
Proposed ≈ Moses	50 / 100
Proposed = Moses	11 / 100

As the table indicates, the proposed method achieved better evaluation than Moses. The p-value was exceeded for 0.95. This means that the proposed method is effective for human evaluation.

9 Discussion

9.1 Analysis of Our Proposed System

Our aim with this system is to reduce the number of ungrammatical sentences produced in machine translation systems. Thus, we analyzed the outputs based on this factor. We compared the output of Moses and the output of our proposed system. And we found that the output of our proposed system affected the output of PBMT, that our system produces more grammatically correct sentences compared to a standard SMT.

9.2 Comparison with Hierarchical phrase-based MT

The pattern acquisition process in the proposed method was similar to the rule extraction of hierarchical statistical phrase-based MT (HSMT). Only, the confident rules are extracted in the proposed method. The reason are discussed follows.

Hierarchical SMT (HSMT) is similar to statistic CFG decoder. So, the number of HSMT parameters is very large. However the number of training data was limited. As the results, they are unreliable and does not perform well, especially for the small amount of training data. Contrast, the proposed method is pattern based. Pattern based approach is similar to network grammar. And it has little parameters compared CFG. So we might obtained these parameters with high reliability.

Also, HSMT has the problem of limiting reordering. The number of spans that are filled during chart decoding is quadratic with respect to sentence length. Hence, it gets worse according as the sentence length increases.

The number of spans that are combined into a span grows linear with sentence length for binary rules, quadratic for trinary rules, and so on. In short, long sentences become a problem. To solved this problem, the size of internal spans has a maximum number. Reordering is limited in hierarchical phrase-based models and should limit reordering for the same reason. On the other hand, the proposed method does not face with such problems because it used patterns. In this reason, we studied the proposed method.

9.3 Improved Pattern Based Statistical Machine Translation

There are many things to improve in PBMT. For example, there is a trade-off between the coverage of input sentences and the translation quality in PBMT. When we made the "high probability phrase table", we set the threshold to 0.1. This was a completely heuristic value. If this value sets low, we obtained many word pairs and many patters. However the reliability of these value was decrease. So we must cut and try this value.

Moreover, there were many bugs in our system. There were 10,000 test sentences in this experiment. Of these 10,000 sentences, 1,143 sentences matched the Japanese-English patterns. We think this number is small for our experience. One possible cause is that we might not have obtained all the possible Japanese-English patterns. We will work on improving the performance of our pattern-based MT system.

10 Conclusion

We developed a two-stage MT system. The first stage consists of an automatically created pattern-based machine translation system. The second stage consists of an phrase-based SMT system. Our goal with this system is to obtain fewer ungrammatical sentences. We performed ABX tests between the output of a standard SMT system (Moses) and the output of the proposed system for 100 sentences. The results indicated that 30 sentences output by the proposed system were evaluated as better than those output by the standard SMT system. In contrast, 9 sentences output by the standard SMT system were thought to be better than those output by the proposed system. This means that our proposed system functioned effectively in the Japanese-English simple sentence task.

We need to overcome several difficulties in order to improve the proposed methods. Moreover, there were many bugs in our system. We will focus on how to solve such difficulties in the future.

References

Masayuki Asahara and Yuji Matsumoto. 2000. Extended models and tools for high-performance part-of-speech tagger. *Proceedings of COLING 2000*.

Satanjeev Banerjee and Alon Lavie. 2005. 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*.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer, M. Laignelet, and F. Rioult. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

David Clark. 1982. High-resolution subjective testing using a double-blind comparator. *J. Audio Eng. Soc*, 30(5):330–338.

Loic Dugast, Jean Senellart, and Philipp Koehn. 2007. Statistical postediting on systran's rule-based translation system. *in Second Workshop on SMT*, pages 179–182.

Terumasa Ehara. 2007. 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*, pages 13–18.

Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 177–180.

Zhifei Li, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren Thornton, Jonathan Weese, and Omar Zaidan. 2009. Joshua: An open source toolkit for parsing-based machine translation. *Proceedings of the Workshop on Statistical Machine Translation*.

Jin'ichi Murakami, Masato Tokuhisa, and Satoru Ikehara. 2007. Statistical machine translation using large j/e parallel corpus and long phrase tables. *International Workshop on Spoken Language Translation 2007*, pages 151–155.

NIST, editor. 2003. *Automatic Evalation of Machine Translation Quality Using n-gram Co-Occurence Statistics*, http://www.itl.nist.gov/iad/mig/test/mt/.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. *40th Annual meeting of the Association for Computational Linguistics*, pages 311–318.

Michel Simar, Nicola Ueffing, Pierre Isabelle, and Roland Kuhn. 2007. Rule-based translation with statistical phrase-based post-editing. *Second Workshop on SMT*, pages 203–206.

Andreas Stolcke. 2002. Srilm - an extensible language modeling toolkit. *Intl. Conf. Spoken Language Processing, Denver, Colorado*.

Yushi Xu and Stephanie Seneff. 2008. Two-stage translation: A combined linguistic and statistical machine translation framework. *Proceedings of the Eighth Conference of the Association for Machine Translation*.