Machine Translation Method Based on Non-Compositional Semantics (Word-Level Pattern-Based MT)

Jun Sakata*, Jin'ichi Murakami*, Masato Tokuhisa*, Masaki Murata*
**Information and Knowledge Engineering*
Tottori University
Tottori, Japan
{*d112004,murakami,tokuhisa,murata*}@*ike.tottori-u.ac.jp*

***Abstract*—The conventional machine translation method based on compositional semantics has a problem in that it cannot generate the semantics of the sentence when it generates the target sentence. To resolve this problem, Ikehara et al. proposed a machine translation scheme based on non-compositional semantics. This machine translation scheme requires many sentence patterns. The sentence patterns can preserve the semantics of the expression structure. To use this machine translation scheme for Japanese-English machine translation, a compound and complex sentence pattern dictionary, called "ToribankSPD", have been developed. This dictionary has three levels of sentence patterns: "word-level", "phrase-level", and "clause-level". In this paper, according to the machine translation scheme based on non-compositional semantics, we implemented the Japanese-English pattern-based machine translation method using the word-level sentence patterns of ToribankSPD. We carried out translation experiments with compound and complex sentences as inputs. In the experiments, the pattern matching rate was low (about 10%). However, 72 out of 100 evaluated sentences used the sentence patterns that had an appropriate expression structure, and the translation accuracy of 55 sentences was high.**

***Keywords*-Non-Compositional Semantics; Pattern-Based Machine Translation; Word-Level Sentence Pattern; Linear Component; Non-Linear Component;**

I. Introduction

The conventional machine translation method based on compositional semantics has a problem in that it cannot generate the semantics of the sentence when it generates the target sentence. To resolve this problem, Ikehara et al. proposed a machine translation scheme based on non-compositional semantics [1]. This machine translation scheme requires many sentence patterns that can preserve the semantics of the expression structure. The sentence patterns have linear components and non-linear components. Linear components are defined as components that don't change the semantics of the sentence when being replaced with other components. And non-linear components are defined as components that change the semantics of the sentence when being replaced with other components. In translation, we conduct local translation of matched morphemes for the linear components and insert these results into the target sentence pattern.

To use this scheme for Japanese-English machine translation, a compound and complex sentence pattern dictionary, called "ToribankSPD", have been developed [2]. The dictionary has 226,817 sentence pattern pairs from Japanese/English compound/complex sentences pairs. It also has three levels of sentence patterns: word-level (121,904 pattern pairs), phrase-level (79,438 pattern pairs), and clause-level (25,475 pattern pairs) [3].

In this paper, according to the machine translation scheme based on non-compositional semantics, we implement a Japanese-English pattern-based machine translation method that involves ToribankSPD, a structural pattern matching (SPM) system [4], and a generation system. We evaluate the translation accuracy of our word-level pattern-based machine translation method and describe the efficiency of and problems with the method. It is said that machine translation methods with sentence patterns have high translation accuracy when the sentence patterns match to the input sentence. Most of the machine translation methods with sentence patterns have only several hundreds or thousands sentence patterns [5]. This paper is the first attempt as the machine translation methods with about hundred thousands sentence patterns to compound and complex sentences. As compound and complex sentences have complicated structure, Our pattern-based machine translation method is considered effective for these sentences.

II. ToribankSPD:Compound and Complex Sentence Pattern Dictionary

To resolve the problem of the conventional machine translation method, Ikehara et al. proposed a machine translation scheme based on non-compositional semantics [1]. This machine translation scheme uses sentence patterns that have linear components and non-linear components. This scheme is available to various languages, and is especially effective in the language pairs that have different language structure. To use this machine translation scheme for Japanese-English machine translation, a compound and complex sentence pattern dictionary, called "ToribankSPD", have been developed [2]. The original sentences of ToribankSPD were collected from various Japanese-English and English-Japanese dictionary. In total, 226,817 sentence patterns have been created.

Table I shows an example of our sentence patterns. Sentence patterns have letters, variables, functions, and markers. Japanese/English word/phrase/clause alignments are replaced with variables that are equivalent to the linear components. Word-level patterns have word variables. Phrase-level patterns have word and phrase variables, and clause-level patterns have all of three variables. The variables in Japanese sentence patterns have semantic codes. These codes are used for sentence pattern selection.

Non-linear components are described as letters and functions. The terms ".hitei" and ".kako" are modality and tense functions, respectively. ".hitei" matches negative expressions, and ".kako" matches the past tense. English functions have a role to assign a word form. "*N*2^poss" means *N*2 is the possessive case in English. "*V*5^past" means *V*5 is the past tense. In Japanese, a subject is often omitted, so "<*N*1 は >" means whether the subject is omitted in the pattern. In English patterns, <I|*N*1> is "*N*1" if Japanese matches *N*1, or "I" if not.

The semantics of the expression structure is composed from non-linear components, word forms, parts of speech of linear components, and an order of components. The sentence patterns preserve the semantics of the expression structure of sentence.

Table I
EXAMPLE OF SENTENCE PATTERN PAIRS IN TORIBANKSPD

Japanese Sentence	彼 の お母さん が ああ 若い と は 思わ なかっ た 。
English Sentence	I never expected his mother to be so young.
Word-Level Japanese Pattern	<*N*1 は >*N*2(NI:23,NI:24) の *N*3(NI:80,NI:49) が ああ *AJ*4(NY:5) と は *V*5(NY:32,NY:31).hitei.kako。
Word-Level English Pattern	<I|*N*1> never *V*5^past *N*2^poss *N*3 to be so *AJ*4.
Phrase-Level Japanese Pattern	< *N*1 は > *NP*2(NI:49,NI:80) が あ あ *AJ*3(NY:05) と は *V*4(NY:31,NY:32).hitei.kako。
Phrase-Level English Pattern	<I|*N*1> never *V*4^past *NP*2 to be so *AJ*3.

III. JAPANESE-ENGLISH WORD-LEVEL PATTERN-BASED MACHINE TRANSLATION METHOD (PROPOSED METHOD)

In this section, we describe our word-level pattern-based machine translation method. Figure 1 shows an overview of the method. The translation steps are as follows.

Step 1) Sentence pattern matching is conducted.

Step 2) Pattern selection with semantic codes is conducted.

Step 3) Candidates of English sentences are generated using the word-level generation system.

Step 4) Only one English sentence is selected with use of the translation probability and word trigram.

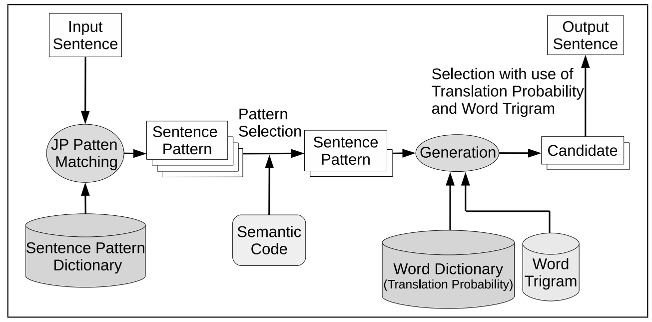

Figure 1. Word-level pattern-based machine translation method

A. Japanese Sentence Pattern Matching

We developed the structural pattern matching (SPM) system [4]. It implements the augmented transition network (ATN) algorithm [6] with breadth-first search and uses sentence patterns. The input sentence for the SPM is already morphological and has semantic codes added. It conducts pattern matching between the input sentence and sentence patterns. Moreover, it outputs the pattern matching results. Figure 2 shows an example of an input sentence.

1. /彼 (1710,{NI:23,NI:48})
2. +の (7410)
3. /お母さん (1100,{NI:80,NI:49})
4. +が (7410)
5. /ああ (1110)
6. /若い (3106,{NY:5})
7. +と (7420)
8. +は (7530)
9. /思わ (2392, 思う, 思わ,{NY:32,NY:31})
10. +なかっ(7184, ない, なかっ)
11. +た (7216)
12. +。(0110)
13. /nil

Figure 2. Example of input sentence

In the first line in Figure 2, "彼"[he] is a Japanese morpheme, "1710" is the tagging code, and "NI:23,NI:48" are indeclinable semantic codes [7]. In the sixth line, "NY:5" is a declinable semantic code. Each line shows a Japanese morpheme and semantic information.

Table II shows the SPM results. "Japanese Pattern" is a matched Japanese pattern for input sentence. "English Pattern" is a corresponding English pattern to that Japanese pattern. In "Matched Morpheme", for example, "*N*2=彼" shows that the morpheme "彼 [he]" matches the noun variable *N*2.

Table II
RESULTS OF SPM

Japanese Pattern	*N*2, の,*N*3, が, ああ,*AJ*4, とは *V*5,.hitei,.kako,。
English Pattern	<I|*N*1> never *V*5^past *N*2^poss *N*3 to be so *AJ*4.
Matched Morpheme	*N*2=彼 {NI:23,NI:24}, *N*3=お母さん {NI:80,NI:49}, *AJ*4=若い {NY:5}, *V*5=思わ (思う){NY:32,NY:31}

B. Sentence Pattern Selection with Semantic Code

If several sentence patterns match, we select sentence patterns with semantic codes. We use the semantic codes in "Nihongo-Goi-Taikei" [7]. These codes consist of indeclinable and declinable semantic codes, which have a hierarchical structure. The steps of sentence pattern selection are as follows.

Step 1) Semantic codes of the input sentence and matched sentence patterns are extracted.

Step 2) A range for sentence pattern selection is extracted from a matched morpheme for a variable. A parent code of a matched morpheme and all its children codes are determined as this range.

Step 3) If all semantic codes of variables in the Japanese sentence pattern are included within those ranges, these sentence pattern pairs are used for translation.

C. Word-Level Pattern-Based Machine Translation

1) Word Dictionary: In the word-level generation system, word translation of a morpheme is done with a word bilingual dictionary. ToribankSPD has alignments of words, phrases, and clauses. We use this word alignments for creating a word dictionary.

In this dictionary, translation probabilities are given to each Japanese-English word pair. Table III lists examples of the word dictionary.

Table III
EXAMPLE OF WORD DICTIONARY

Japanese word (POS)	English word (POS)	Translation probability
蹴る [kick] (V)	hit (V)	0.01
	kick (V)	0.29
	stamp out (V)	0.05
	turn down (V)	0.06
翻訳 [translation] (N)	translation (N)	0.53
	translate (V)	0.20

Translation probabilities in Table III are calculated using Equation (1).

$$P = \frac{C(e,j)}{C(j)} \times \frac{C(e,j)}{C(e)} \quad (1)$$

where $C(e,j)$ is the number of co-occurrences of each word pair in the alignments list, $C(j)$ is the number of occurrences of each Japanese word, and $C(e)$ is the number of occurrences of each English word. The first to fourth rows in Table III show English verbs "hit", "kick", "turn down", and "stamp out" coupled with the Japanese verb "蹴る"[kick], and the translation probabilities are "0.01", "0.29", "0.05", and "0.06", respectively.

2) Word-Level Generation System: An English sentence is translated with matched sentence pattern pairs with the word-level generation system. The word-level generation system performs word translation for the Japanese morphemes of the SPM results. Word translation is done using the word dictionary. Several results of word translation are inserted into the English pattern, and a single maximum likelihood sentence is selected with use of the translation probability and English word trigram. The translation steps of the word-level generation system are as follows.

Step 1) Word translations of Japanese morphemes are done.

Step 2) Word translation results are changed to the assigned form.

Step 3) Candidates of word translation are inserted into the English sentence pattern.

Step 4) The maximum likelihood combination of words is selected with use of the translation probability and English word trigram.

3) Example of English Sentence Generation: In the last generation step, the maximum likelihood combination of words is selected with use of the translation probability and English word trigram. Figure 3 shows an example of English sentence generation.

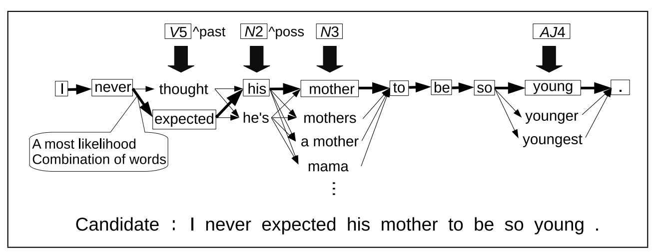

Figure 3. Example of generation

4) Selection of an Output Sentence from Candidates: If several sentence patterns are obtained in pattern matching and pattern selection, all matched sentence patterns are used for translation. As one translation candidate is obtained from one sentence pattern, in this case, several translation candidates are obtained. One translation candidate with the maximum translation probability and English word trigram (Section III-C2) is selected as an output sentence.

IV. EXPERIMENTS

A. Experimental Conditions

We carried out an open-test to investigate the effectiveness of the proposed method. We used 100,000 sentence patterns for translation experiments and created a word dictionary from word alignments extracted from these 100,000 sentence patterns. Word trigrams were trained from about 280,000 English sentences. In these sentences, 100,000 were original English sentences of these 100,000 English patterns and about 180,000 were English sentences from an other Japanese-English parallel corpus [8]. We used 5,000 original Japanese sentences of the remaining sentence patterns as input sentences.

B. Baseline System

We used the phrase-based SMT (MOSES) [9] as the baseline system for comparison. We used 100,000 sentence pairs (same as sentence pairs of sentence patterns for pattern matching) for training of the translation model and about 280,000 English sentences (the same as in Section IV-A) for training of the language model. We conducted parameter tuning with MERT.

C. Evaluation Method

Translation accuracy was measured using automatic metrics and manual evaluation. We used the evaluation tools BLEU [10], TER [11], METEOR [12], and RIBES [13]. Manual evaluation was done from the point of view of adequacy [14], and Table IV shows the scoring criteria.

Table IV
CRITERIA OF MANUAL EVALUATION

Eval. 5	The meaning of the input sentence is correctly recognized.
Input	きみ は お父さん の 意見 に 従っ た ほう が いい 。
Output	You had better follow your father's opinion.
Reference	You should take your father's advice.
Eval. 4	Although a part is not appropriate grammatically, the meaning of the input sentence is correctly recognized.
Input	彼 は 彼女 を きびしい 目 で 見 た 。
Output	He looked at her with harsh regard.
Reference	He regarded her sharply.
Eval. 3	The meaning of the input sentence is recognized somewhat.
Input	彼女 は とても 涼しい 目 を し て いる 。
Output	She is very bright eyes.
Reference	She has very bright eyes.
Eval. 2	Only a part of the input sentence is recognized or output sentence is grammatically correct but the meaning is significantly different from the input sentence.
Input	彼 は 時計 を 出し て 時間 を 見 た 。
Output	He set the clock to look at the time own .
Reference	He took out his watch and glanced at the time.
Eval. 1	Nothing is recognized.
Input	彼 は なに が 起こっ て も 無神経だ 。
Output	He is might happen.
Reference	He is insensible of anything that happens around him.

D. Results

Out of the 5,000 input sentences, 502 sentences had at least one matched sentence pattern. Table V lists the automatic evaluation results of these 502 sentences. In all metrics, the proposed method outperformed the baseline.

Table V
AUTOMATIC EVALUATION RESULTS

	BLEU	TER	METEOR	RIBES
Proposed	0.358	0.489	0.642	0.806
Baseline	0.307	0.549	0.583	0.780

We carried out manual evaluation for 100 randomly extracted sentences from the 502 sentences. Table VI lists the manual evaluation results of these 100 sentences. The value of each evaluation criteria is the number of sentences. In Table VI, the number of Eval. 1 and Eval. 2 of proposed method are smaller than baseline, and the number of Eval. 4 and Eval. 5 are larger than baseline. And the average of proposed method is higher than baseline. These results show that the proposed method outperformed the baseline.

Table VI
MANUAL EVALUATION RESULTS

	Eval. 1	Eval. 2	Eval. 3	Eval. 4	Eval. 5	Average
Proposed	4	26	15	11	44	3.65
Baseline	21	37	11	9	22	2.74

An example of the translation results is shown in Table VII. "Japanese Sentence Pattern" and "English Sentence Pattern" is the sentence pattern pair used for translation of output sentence. "Original Japanese Sentence" is the original sentence of this Japanese sentence pattern, and "Original English Sentence" is the original sentence of this English sentence pattern.

Table VII
EXAMPLE OF TRANSLATION RESULTS

<table>
<tr><td>Input</td><td>彼 が 帰国する 可能性 は ない 。</td></tr>
<tr><td>Reference</td><td>There is no question of his going home.</td></tr>
<tr><td>Output of proposed method</td><td>There is no possibility of his coming home.</td></tr>
<tr><td>Value of manual evaluation</td><td>5</td></tr>
<tr><td>Japanese Sentence Pattern</td><td>$N1$ が ($V2$^rentai|$ND2$ をする)$N3$ は' 無い'#1(.genzai|.kako)。</td></tr>
<tr><td>English Sentence Pattern</td><td>There @be#1(^present|^past) no $N3$ of $N1$^poss $V2$^ing .</td></tr>
<tr><td>Matched morpheme</td><td>$N1$=彼，$V2$=帰国する，$V3$=可能性</td></tr>
<tr><td>Original Japanese Sentence</td><td>彼 が 当選する 可能性 は ない 。</td></tr>
<tr><td>Original English Sentence</td><td>There is no possibility of his winning the election.</td></tr>
<tr><td>Output of baseline</td><td>There is no possibility of his home country.</td></tr>
<tr><td>Value of manual evaluation</td><td>2</td></tr>
</table>

V. Discussion

The translation accuracy of the proposed method was high (3.65 : the average of human evaluation), but the pattern matching rate was low (about 10% : 502/5,000). In this section, we discuss the reasons for these results and describe the efficiency of and problems with the proposed method.

A. *Efficiency of Proposed Method*

We assumed that a expression structure can be preserved in a sentence pattern. We discuss whether a used sentence pattern for output is appropriate. In the 100 sentences used for manual evaluation, 72 sentences were appropriate and 28 sentences were not. Table VIII lists the causes of not using an appropriate sentence pattern in the 28 sentences.

Table VIII
Causes of not using appropriate sentence pattern

Cause	Number of sentences
(a) Problem in way of making sentence patterns	5
(b) Multi-word expression is decomposed and matched for variable	3
(c) Failure in pattern selection	14
(d) Error in sentence pattern	4
(e) Error in morphological analysis and SPM	2

In Table VIII, (c), (d) and (e) were caused by an error in each process. The error of annotations of semantic codes to morphemes causes the problem of (c). We can improve this problem by adding the words to the annotation list used by the annotation program. To improve the problem of (d), We need to correct these sentence pattern pairs manually. We omit the discussion about the problem of (e).

If we assume these processes were correctly performed, only 8 sentences did not use an appropriate sentence pattern. These results suggest that 72 out of the 80 sentences matched an appropriate pattern when all processes were correctly performed. Fifty-five out of these 72 sentences had high translation accuracy (Eval. 4 or 5). The other 17 sentences had problems in word translation, and the translation accuracy of those sentences was low. Therefore we confirmed that the proposed method obtained high translation accuracy by using appropriate sentence patterns.

1) Problem in Way of Making Sentence Patterns: Table IX shows an example of cause (a) in Table VIII. This example has two problems. One is concerning English functions. English functions have a role to determine word form of word translation results. This input sentence is an imperative sentence, but the verb variables in the used English pattern are not added a function which assign an original form of verb. This results in the wrong form in the output, such as "gets". This problem can be resolved by adding appropriate English functions. The other problem concerns markers for a zero pronoun. In general Japanese imperative sentence, the object ordered by speaker at input is not included in that sentence. This input is a unique sentence that includes the object ("子供" [children]) of order as a subject ("子供は" [children are]). The Japanese pattern in Table IX matches the input by a marker for a zero pronoun ("$< N1$ は $>$" matches "子供は"). In this case, adding this marker to this sentence pattern is not necessary. This problem can be resolved by deleting unnecessary markers from each sentence pattern.

Table IX
Example of problem in way of making sentence patterns

<table>
<tr><td>Input</td><td>子供 は 早く 寝 て 早く 起き なさい 。</td></tr>
<tr><td>Output</td><td>Lie down its speed and gets up early .</td></tr>
<tr><td>Reference</td><td>Children should go to bed early and get up early.</td></tr>
<tr><td>Value of manual Eval.</td><td>1</td></tr>
<tr><td>Japanese Sentence Pattern</td><td>< N1 は > ADV2 V3(て | で)ADV4(V5ˆmeirei|V5.meireigo)。</td></tr>
<tr><td>English Sentence Pattern</td><td><i>V</i>3 <your|<i>N</i>1ˆposs> <i>N</i>2 and <i>V</i>5 <i>ADV</i>4 .</td></tr>
<tr><td>Matched morpheme</td><td>N1=子供, ADV2=早く, V3=寝 (寝る) ADV4=早く, V5=起き (起きる)</td></tr>
<tr><td>Original Japanese Sentence</td><td>十分 食べ て 十分 眠り なさい 。</td></tr>
<tr><td>Original English Sentence</td><td>Eat your fill and sleep well.</td></tr>
</table>

2) Problem of Multi-Word Expression: Table X shows an example of cause (b) in Table VIII. In the input, "情の細やかな" is a multi-word expression. The meaning of this expression is closest to "sensitive". However, a corresponding expression in the output is "tinier sensibilities". This result is caused by splitting this expression and matching each morpheme ("情" and "細やかな") to each variable ($N2$ and $AJ3$). If we use word-level sentence patterns, we need a sentence pattern that describes "情の細やかな" as letters to obtain an appropriate result. However, it is possible to obtain an appropriate result by using phrase-level sentence patterns.

Table X
Example of multi-word expression decomposed and matched for variables

<table>
<tr><td>Input</td><td>彼女 は 情 の 細やかな 人 だ 。</td></tr>
<tr><td>Output</td><td>She has tinier sensibilities.</td></tr>
<tr><td>Reference</td><td>She is a warm and considerate woman.</td></tr>
<tr><td>Value of manual Eval.</td><td>2</td></tr>
<tr><td>Japanese Sentence Pattern</td><td>N1 は N2 の AJV3ˆrentai(人 | ひと | ヒト) だ。</td></tr>
<tr><td>English Sentence Pattern</td><td>N1 @have AJ3 N2 .</td></tr>
<tr><td>Matched morpheme</td><td>N1=彼女, N2=情, AJV3=細やかな (細やかだ)</td></tr>
<tr><td>Original Japanese Sentence</td><td>彼 は 心 の まっすぐな 人 だ 。</td></tr>
<tr><td>Original English Sentence</td><td>He has an upright mind.</td></tr>
</table>

3) Low Translation Accuracy of Word Translation: With the proposed method, word translation is done using the word dictionary, and a single maximum likelihood sentence is selected with use of the translation probability and English word trigram. In Section V-A, 55 out of 72 sentences that had an appropriate sentence pattern had high translation accuracy. However, the remaining 17 sentences had low translation accuracy due to problems in word translation.

If a matched morpheme is not in the word dictionary, the morpheme is outputted as an unknown word. Nine out of the 17 sentences had this problem. This problem can be resolved by adding new words to the word dictionary.

The remaining 8 sentences exhibited a problem of not selecting an appropriate word from the translation probability and word trigram. Table XI shows an example of the problem in word selection. "Fruitless" in the output is different from the meaning of "無意味だ"[nonsense]. There are appropriate candidates in the word dictionary, for example "meaningless", but "fruitless" is selected with use of the translation probability and word trigram.

Table XI
EXAMPLE OF PROBLEM IN WORD SELECTION

Input	彼 の 言う こと は 無意味だ 。
Output	He is fruitless.
Reference	There is no meaning in what he says.
Value of manual Eval.	2
Japanese Sentence Pattern	*N*1 の (言う ｜ いう ｜ 謂う ｜ 云う) ことは *AJV*2。
English Sentence Pattern	*N*1 @be *AJ*2 .
Matched morpheme	*N*1=彼, *AJV*2=無意味だ
Original Japanese Sentence	君 の 言う こと は 本当 だ 。
Original English Sentence	You are truthful.

B. Pattern Matching Rate

The pattern matching rate was about 10%, which means that 100,000 word-level sentence patterns are not enough to cover compound and complex sentences for translation. Our sentence pattern dictionary has word-level, phrase-level, and clause-level sentence patterns. We have already carried out a pattern matching test with phrase-level patterns in the same situation as Section IV-A, and the pattern matching rate was about 40%. In the future, we will evaluate the translation accuracy of our pattern-based machine translation method with those phrase-level sentence patterns.

VI. CONCLUSION

The conventional machine translation method based on compositional semantics has a problem in that it cannot generate the semantics of the sentence when it generates the target sentence. To resolve this problem, Ikehara et al. proposed a machine translation scheme with sentence patterns based on non-compositional semantics. In this paper, according to this machine translation scheme, we implemented a Japanese-English pattern-based machine translation method using our word-level sentence patterns. We carried out translation experiments with compound and complex sentences as inputs. In the experiments, the pattern matching rate was low (about 10%). However, 72 out of the evaluated 100 sentences used the sentence patterns that had an appropriate expression structure, and the translation accuracy of 55 sentences was high. For future work, we will evaluate the translation accuracy of the machine translation method, using our phrase-level sentence patterns.

VII. ACKNOWLEDGMENTS

The development of a compound and complex sentence pattern dictionary, called "ToribankSPD", was supported under the CREST program of the Japan Science and Technology Agency.

REFERENCES

[1] S. Ikehara, M. Tokuhisa, J. Murakami, M. Saraki, M. Miyazaki, and N. Ikeda, "Pattern Dictionary Development based on Non-Compositional Language Model for Japanese Compound and Complex Sentences," *The 21st International Conference on Computer Processing of Oriental Languages(ICCPOL-06)*, pp. 509–519, 2006.

[2] S. Ikehara, M. Tokuhisa, and J. Murakami, "Non-compositional language model and pattern dictionary development for Japanese compound and complex sentences," in *Proc. 22nd International Conference on Computational Linguistics-Volume 1*, Manchester, UK, Aug. 2008, pp. 343–360.

[3] S. Ikehara, "Tori-Bank," http://unicorn/toribank, 2007.

[4] M. Tokuhisa, J. Murakami, and S. Ikehara, "Pattern Search by Structural Matching from Japanese Compound and Complex Sentence Pattern Dictionary," *IPSJ SIG Technical Report*, 2004-NL-176, pp. 9–16, 2006 (in Japanese).

[5] D. YIN and D. ZHANG, "Construct Chunk-level Templates for Improving Rule-based Machine Translation," *Journal of Computational Information System*, pp. 5505–5512, 2013.

[6] S. C. Shapiro, "Generalized Augmented Transition Network Grammars For Generation From Semantic Networks," *The American Journal of Computational Linguistics*, Vol. 8, pp. 12–25, 1982.

[7] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi, "Goi-Taikei:A Japanese Lexicon," Iwanami Shoten, 1997 (in Japanese).

[8] J. Murakami and S. Hujinami, "Japanese-English Parallel Sentences Collection from Digital Media," *JCL Workshop2012*, 2012 (in Japanese).

[9] P. Koehn, M. Federico, B. Cowan, H. Hoang, N. Bertoldi, W. Shen, A. Birch, C. Moran, C. Callison-Burch, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation," in *Proc. ACL 2007 Demo and Poster Sessions*, Prague, Czech Republic, Jun. 2007, pp. 177–180.

[10] P. Kishore, S. Roukos, T. Ward and W. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," *40th Annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.

[11] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, R. Weischedel, R. Schwartz, L. Micciulla, and J. Makhoul, "A Study of Translation Edit Rate with Targeted Human Annotation," in *Proc. 7th Conference of the Association for Machine Translation in the Americas*, Boston, USA, Aug. 2006.

[12] B. Satanjeev and L. Alon, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," in *Proc. 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)*, Ann Arbor, USA, Jun. 2005, pp. 65–72.

[13] I. Hideki, H. Tsutomu, D. Kevin, S. Katsuhito, and T. Hajime, "Automatic Evaluation of Translation Quality for Distant Language Pairs," in *Proc. 2010 Conference on Empirical Methods on Natural Language Processing (EMNLP)*, Massachusetts, USA, Oct. 2010, pp. 944–952.

[14] P. Koehn and C. Monz, "Manual and Automatic Evaluation of Machine Translation Between European Languages," in *Proc. Workshop on Statistical Machine Translation*, New York, USA, Jun. 2006, pp. 102–121.