

Machine Translation Using Recursive Chain-Link-Type Learning Based on Translation Examples

Hiroshi Echizen-Ya,¹ Kenji Araki,² Yoshio Momouchi,¹ and Koji Tochinal³

¹Department of Electronics and Information Engineering, Hokkai-Gakuen University, Sapporo, 064-0926 Japan

²Division of Electronics and Information Engineering, Hokkaido University, Sapporo, 060-8628 Japan

³Division of Business Administration, Hokkai-Gakuen University, Sapporo, 062-8605 Japan

SUMMARY

Rule-based machine translation analyzes source-language sentences using large-scale linguistic knowledge that is given by the developer beforehand. However, it is difficult to give complete linguistic knowledge to the system *ex ante* because natural language has various linguistic phenomena. Therefore, we worked to develop learning-based machine translation. In learning-based machine translation, a system acquires translation rules automatically from translation examples that are pairs of source and target language sentences. However, existing learning-based machine translation presents the problem that it requires a large number of similar translation examples. Consequently, it cannot acquire enough useful translation rules from sparse translation examples. This paper proposes a method of machine translation using Recursive Chain-Link-type Learning, which can acquire many useful translation rules from sparse translation examples. Our system, based on this method, efficiently acquires translation rules from each translation example without requiring two similar translation examples. Translation rules are acquired by extracting corresponding parts between source and target language

sentences in translation examples. Our system determines those corresponding parts using previously acquired translation rules. Therefore, the system engenders a chain reaction in acquisition of new translation rules. Evaluation experiments using our system demonstrated an effective translation rate of 61.1%. Moreover, the effective translation rate was 85.0% when sufficient learning data were given to our system. © 2004 Wiley Periodicals, Inc. Syst Comp Jpn, 35(2): 1–15, 2004; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.10511

Key words: machine translation; learning; recursive; chain-link; translation examples.

1. Introduction

Many machine translation application systems have been developed in recent years, such as speech machine translation systems and multilingual machine translation systems [1]. However, the quality of their product translations remains insufficient. Therefore, users cannot obtain adequate translation results.

The mainstream method in existing machine translation systems is rule-based machine translation [2, 3]. In rule-based machine translation, all translation knowledge

Contract grant sponsors: Partially supported by grants from the High-Tech Research Center of Hokkai-Gakuen University and a Government subsidy for aiding scientific research (No. 14658097) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

(e.g., grammar rules, rules to solve word ambiguity) is given by a developer *ex ante*. As a result, rule-based machine translation presents some problems. For example, it is difficult for the developer to provide complete translation knowledge that can address all linguistic phenomena *ex ante*. Moreover, adding new translation knowledge sometimes causes new problems in the translation process because rule-based machine translation is quite *ad hoc* [3].

A corpus-based machine translation has been proposed to solve these problems of rule-based machine translation. The advantage of corpus-based machine translation is that it can generate high-quality translation results by using a corpus directly as translation knowledge. Moreover, the system can easily improve the correct translation rate through mere addition of new sentences. However, corpus-based machine translation requires a large corpus to achieve a high rate and quality of translation. Therefore, many corpus-based studies have utilized static analytical knowledge [4–11]. This means that a corpus-based machine translation shares the inherent problem of rule-based machine translation, namely, the system becomes very *ad hoc*. On the other hand, learning-based machine translation has been proposed in recent years [12–16]. It automatically acquires translation rules from translation examples by using a learning algorithm without static analytical knowledge. However, in existing learning-based machine translation, a large number of similar translation example pairs are required to produce many useful translation rules.

We have studied machine translation systems to realize one that possesses high learning ability. This means that the system acquires many high-quality translation rules from only the provided translation examples without using static linguistic knowledge. The system acquires various translation rules and uses them. Abstract translation rules, which correspond to grammar rules in rule-based machine translation, can be applied to various source language sentences. Moreover, by using concrete translation rules, the system can correctly translate the source language sentences because those concrete translation rules are close to source language sentences, although they are applied only to specific source language sentences like translation knowledge in corpus-based machine translation. Therefore, the system offers the advantages of rule-based and corpus-based machine translation.

We have already proposed the method of **Machine Translation Using Inductive Learning with Genetic Algorithms (GA-ILMT)** [18, 19] as the first step to realize such a learning-based machine translation. In GA-ILMT, translation rules are acquired from two similar translation examples using inductive learning. Similar translation examples are pairs of translation examples in which the system can easily decide corresponding parts between source language (SL) sentences and target language (TL) sentences in character strings. For example,

the system can acquire high-quality translation rules by extracting different and common parts between two translation examples by inductive learning when two translation examples with one different part between two SL sentences and with also just one different part between two TL sentences are given. We do not use static analytical knowledge for determination of different and common parts. The reason is that the system has the same problem as rule-based machine translation: the system becomes very *ad hoc*. Moreover, in GA-ILMT, by applying genetic algorithms [20], translation examples with one different part between two SL sentences and with also just one different part between two TL sentences are automatically generated. However, correct translation examples are generated only when the sentences of the two translation examples are short and have similar grammatical structure [18, 19]. It is difficult for the system to generate similar translation examples when various translation examples are given. As a result, the translation rate decreases because the system cannot acquire enough high-quality translation rules. This means that GA-ILMT has the same problem as other learning-based machine translation systems [12–16]: the systems require a large number of similar translation examples.

This paper presents a machine translation method that uses Recursive Chain-Link-type Learning, which is a learning algorithm to acquire translation rules efficiently from each translation example, not pairs of similar translation examples. In this method, many new translation rules are acquired by using previously acquired translation rules. As a result, this system using recursive chain-link-type learning causes a chain reaction in acquisition of translation rules. For example, when a translation rule **A** exists in the dictionary, the system can acquire a translation rule **B** using translation rule **A** from a translation example. Using the acquired translation rule **B**, the system can acquire a translation rule **C** from another translation example. In this method, the acquisition process of the translation rules forms a pattern like a chain in which each rule is linked. Moreover, the extraction of the corresponding parts between SL sentences and TL sentences of translation examples is performed recursively with the same translation examples. As a result, the system can acquire several translation rules from one translation example. We call such an algorithm, which acquires translation rules, **Recursive Chain-Link-type Learning (RCL)**. This paper presents a method of machine translation using RCL and describes the effectiveness of this system using RCL from performance evaluation results.

2. Basic Ideas

2.1. Acquisition of translation rules based on recursive chain-link-type learning

RCL is an algorithm to acquire knowledge in a computer that has no knowledge. This system using RCL possesses the ability to extract the corresponding parts from pairs of objects. We applied RCL to translation examples: pairs of SL sentences and TL sentences. Figure 1[†] shows the schema for translation rule acquisition.

Figure 1 shows the process whereby translation rule B, translation rule C, and translation rule D are acquired one after another using RCL when translation rule A exists. In this paper, a translation rule is a pair comprising a source part and a target part. Source parts are parts acquired from SL sentences of translation examples; target parts are parts acquired from TL sentences of translation examples. This system using RCL acquires two types of translation rules. In Fig. 1, translation rules A and C express parts of SL and TL sentences. Hereafter, we call this type of translation rule a *partial translation rule*. In contrast, translation rules B and D express the entirety of SL sentences and TL sentences. We call this type of translation rule a *sentence translation rule*. On the other hand, in this paper, translation rules that are used as starting points in the acquisition process of new translation rules, like translation rule A in Fig. 1, are acquired using GA-ILMT. GA-ILMT is also a method that automatically acquires translation knowledge from the perspective of learning. This system using RCL can acquire many translation rules based only on learning ability from various translation examples as shown in Fig. 1. In this paper, this system using RCL indicates a system that uses both RCL and GA-ILMT.

This system using RCL causes a chain reaction in acquisition of partial translation rules and sentence translation rules reciprocally as shown in Fig. 1. In process No. 1 of Fig. 1, using the partial translation rule A, this system using RCL acquires the sentence translation rule B. First, the system extracts “Z” from the SL sentence of translation example No. 1, and “ζ” from the TL sentence of translation example No. 1. The system replaces the extracted parts “Z” and “ζ” with variables “@*.” Using the acquired sentence translation rule B, this system using RCL acquires new partial translation rules. In process No. 2 of Fig. 1, this system using RCL first extracts “NΩ” from the SL sentence of translation example No. 2. “NΩ” corresponds to variable “@*” in the SL part of sentence translation rule B because “E” and “H” adjoin variable “@*,” and they are the com-

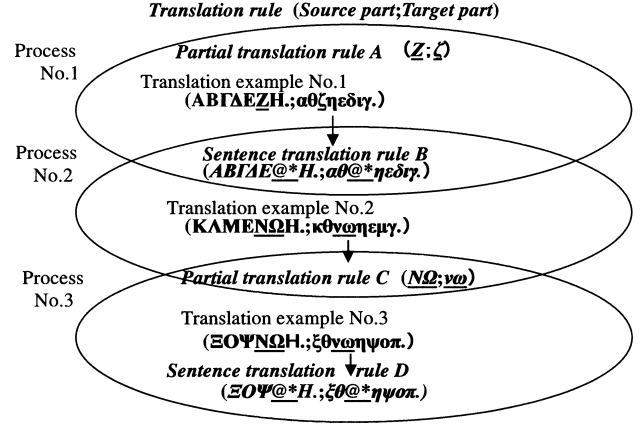


Fig. 1. Schema in acquisition process of translation rules using Recursive Chain-Link-type Learning.

mon parts between the source part of sentence translation rule B and the SL sentence of translation example No. 2. Moreover, this system using RCL extracts “νω” from the TL sentence of translation example No. 2. “νω” corresponds to variable “@*” in the TL part of sentence translation rule B because “θ” and “η” adjoin variable “@*,” and they are the common parts between the target part of sentence translation rule B and the TL sentence of translation example No. 2. As a result, this system using RCL can acquire the partial translation rule C (NΩ; νω) as corresponding parts to variables “@*” in sentence translation rule B. Moreover, in process No. 3 of Fig. 1, this system using RCL acquires the partial translation rule D using the sentence translation rule C. In this paper, translation examples are also registered in the dictionary as sentence translation rules without variables. Moreover, this system using RCL acquires translation rules by determining common parts and different parts based on character strings of translation examples. Therefore, this system using RCL possesses the ability to decide similarity or difference of parts in two objects [17]. In the decision process of extraction parts, we do not use parts-of-speech or a bilingual dictionary because using them would introduce the same problems that are inherent in rule-based machine translation to this system using RCL. Therefore, this system using RCL performs extraction process using a learning algorithm that is based on two heuristics. Applying these heuristics enables the system to efficiently acquire translation rules by focusing on parts of each translation example or each translation rule [21]. Two heuristics are as follows:

(1) In between sentence translation rules with source parts and target parts corresponding to each other and

[†] In Fig. 1, the use of a Greek character means that all language characters correspond to unknown character strings for a computer when no linguistic knowledge is given at all.

translation examples, or between sentence translation rules with source parts and target parts corresponding to each other and other sentence translation rules, when translation examples or other sentence translation rules have the same parts as parts that adjoin variables in sentence translation rules, this system using RCL can acquire the partial translation rules by extracting parts that adjoin the same parts from translation examples or other sentence translation rules. Those partial translation rules have source parts and target parts that correspond to each other.

(2) There are partial translation rules, for which source parts and target parts correspond to each other, and translation examples, or partial translation rules, for which source parts and target parts correspond to each other, and sentence translation rules, when translation examples or sentence translation rules have the same character strings as the partial translation rules. This system using RCL can acquire new sentence translation rules from cases where source parts and target parts correspond to each other. It does so by replacing parts that have the same character strings as the partial translation rules with variables, with translation examples, or with sentence translation rules.

2.2. Preliminary experiments for heuristics

We performed preliminary experiments to confirm the effectiveness of the above-mentioned heuristics. First, this system using RCL acquired translation rules from 678 translation examples that were pairs of English sentences and Japanese sentences taken from textbooks for first grade junior high school students [25]. Next, we investigated the number of partial translation rules for which source parts and target parts corresponded to each other among all partial translation rules acquired using RCL to confirm the effectiveness of heuristic (1). As a result, 179 acquired partial translation rules had source parts and target parts that corresponded to each other. In contrast, 70 acquired partial translation rules had source parts and target parts that did not correspond to each other. Therefore, the precision was 71.9%.

Moreover, we investigated the number of sentence translation rules for which source parts and target parts corresponded to each other among all sentence translation rules acquired using RCL to confirm the effectiveness of heuristic (2). As a result, 1044 acquired sentence translation rules had source parts and target parts that corresponded to each other. In contrast, 14 acquired sentence translation rules had source parts and target parts that did not correspond to each other. Therefore, the precision was 98.7%. These preliminary experiments confirmed that this system using RCL can acquire partial translation rules and sentence translation rules for which source parts and target parts correspond to each other.

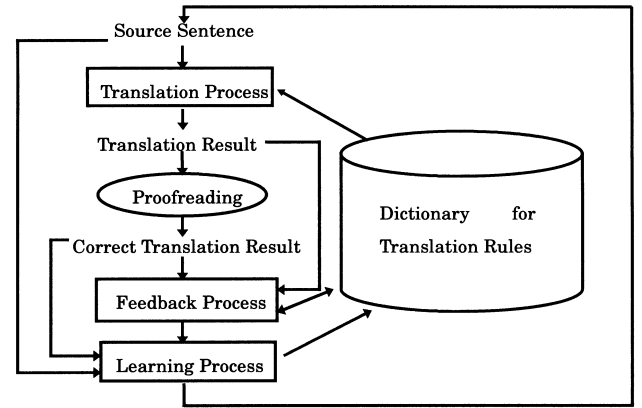


Fig. 2. Process flow.

3. Outline

Figure 2 shows the outline of an English-to-Japanese machine translation system using RCL. A user inputs an English sentence. In the translation process, this system using RCL generates a Japanese sentence using acquired translation rules in the dictionary. The user then proofreads translation results checking for errors. In the feedback process, this system using RCL evaluates whether used translation rules are correct or not. In the learning process, our system acquires translation rules from given translation examples using RCL and GA-ILMT.

On the other hand, “/” are inserted into Japanese sentences after each morpheme because Japanese is an agglutinative language. This process is performed automatically, using a system based only on its learning method [22] without requiring any static linguistic knowledge.

4. Process

4.1. Translation process

In the translation process, this system using RCL generates translation results using acquired translation rules. First, this system using RCL compares input English sentences with source parts of translation rules. As a result, when source parts of translation rules, except variables, have the same character strings as parts in the input English sentences, those translation rules are selected to generate translation results. This system using RCL generates sentence translation rules that do not have variables. It does so by combining selected sentence translation rules and selected partial translation rules. Target parts of generated

sentence translation rules that do not have variables become translation results.

Translation results are sorted so that translation results that have the highest concrete degree are ranked at the top when several different translation results are obtained for the same English sentences. The degree of correctness is defined as follows:

Concrete degree (%) =

$$\frac{\text{The number of words in source parts except variables}}{\text{The number of words in source sentences}} \times 100.0$$

Moreover, translation results are sorted so that translation results that have the highest degree of correctness are ranked at the top when several translation results that have the same concrete degree are obtained. The degree of correctness is defined as follows:

Correct degree (%) =

$$\frac{\text{Correctness frequency}}{\text{Correctness frequency} + \text{Erroneous frequency}} \times 100.0$$

The detail of correctness frequency and erroneous frequency is described in Section 4.2. Figure 3 shows an example translation process; three sentence translation rules and one partial translation rule are selected because they are translation rules in which source parts of translation rules, except variables, have the same character strings as parts in input English sentences. In Fig. 3, the system selects (He is my @0.; *Kare wa watashi no @0 desu.*)[†] that has the highest degree of concreteness among three sentence translation rules, and generates the sentence translation rule (He is my friend.; *Kare wa watashi no tomodachi desu.*) which do not have variables by combining the selected sentence translation rule with the partial translation rule (friend; *tomodachi*). As a result, the source part of “*Kare wa watashi no tomodachi desu.*” is obtained as the translation result.

4.2. Feedback process

In the feedback process, this system using RCL evaluates translation rules used in the translation process.

In that case, this system using RCL uses two kinds of evaluation methods. In concrete translation rules without variables, this system using RCL utilizes the result of combination of translation rules [18]. In abstract translation rules with variables, this system using RCL utilizes the process of combination of translation rules [23, 24].

Evaluation for concrete translation rules without variables is performed by comparing generated translation results with proofread Japanese sentences. As a result, this system using RCL adds one point to the correctness frequency of used concrete translation rules when generated

Source sentence: He is my friend.

(1) Selection of translation rules

Sentence translation rules
(He is @0.; *Kare wa @0 desu.*)

(@0 is my @1.; @0 wa watashi no @1 desu.)

(He is my @1.; *Kare wa watashi no @1 desu.*)

Partial translation rule

(friend; *tomodachi*)

(2) Generation of translation result

(friend; *tomodachi*)

(He is my @0.; *Kare wa watashi no @0 desu.*)

(He is my friend.; *Kare wa watashi no tomodachi desu.*)

Translation result: *Kare wa watashi no tomodachi desu.*

Fig. 3. Example of translation process.

translation results have the same character strings as the proofread Japanese sentences. In contrast, this system using RCL adds one point to the erroneous frequency of used concrete translation rules when generated translation results have different character strings from the proofread Japanese sentences.

In evaluation of abstract translation rules with variables, this system using RCL memorized the generation process of the correct translation rules, that is, the process of combination of translation rules. This system using RCL increases correctness frequency or erroneous frequency of translation rules by applying heuristics addressing the combination of translation rules. We reserve details of this evaluation method for explanation elsewhere [23, 24].

4.3. Learning process

4.3.1. Acquisition of translation rules based on pairs of translation examples

This study utilizes GA-ILMT for acquiring sentence translation rules or partial translation rules that are used as starting points in the process of RCL. In GA-ILMT, translation rules are acquired by performing phrased extraction of common parts and different parts from two translation examples using inductive learning. Figure 4 shows examples of acquisition of translation rules based on pairs of translation examples. In Fig. 4, our system first acquires translation rules by extracting common parts and different parts from two translation examples. Next, more abstract translation rules are acquired by performing phrased extrac-

[†]Italics express the pronunciation in Japanese.

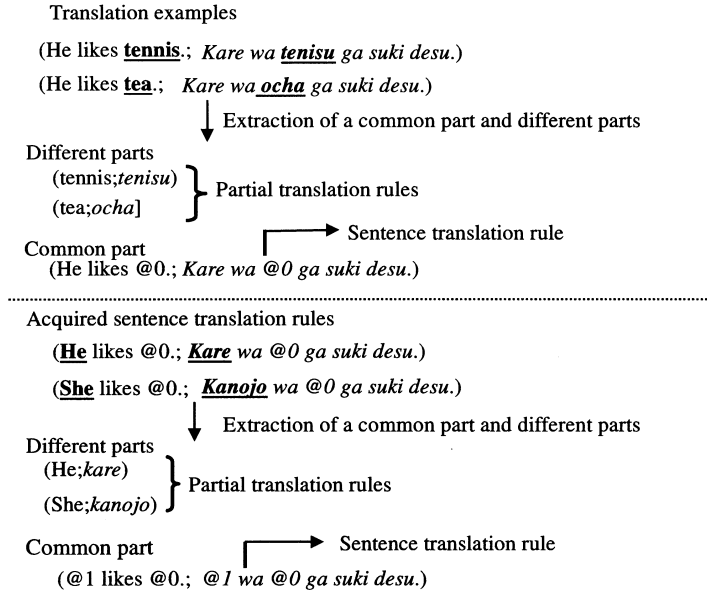


Fig. 4. Examples of acquisition of translation rules based on pairs of translation examples.

tion from two acquired translation rules. In this paper, a common part and a different part mean a word or a part that is composed of more than one word.

4.3.2. Acquisition of partial translation rules using sentence translation rules

Partial translation rules and sentence translation rules are acquired reciprocally in RCL. As a result, this system using RCL causes a chain reaction in acquisition of translation rules. This section describes details of the acquisition process of partial translation rules using sentence translation rules. This process is based on heuristic (1) described in Section 2.1. Details of the acquisition process of partial translation rules using sentence translation rules are:

(1) This system using RCL selects translation examples that have the same parts as parts that adjoin variables in sentence translation rules.

(2) This system using RCL acquires partial translation rules by extracting parts that adjoin common parts from translation examples. These parts are the same parts as parts in sentence translation rules. This means that parts extracted from translation examples correspond to variables in sentence translation rules. In this extraction process, there are three patterns from the view of the position of variables and their adjoin words in sentence translation rules:

Pattern 1: When common parts exist on both the right side and the left side of variables in source parts or target parts of sentence translation rules, this system using RCL

extracts parts between two common parts from SL sentences or TL sentences of translation examples.

Pattern 2: When common parts exist only on the right side of variables in source parts or target parts of sentence translation rules, this system using RCL extracts parts from words at the beginning to words on the left of common parts in SL sentences or TL sentences of translation examples.

Pattern 3: When common parts exist only on the left side of variables in source parts or target parts of sentence translation rules, this system using RCL extracts parts from words on the right of common parts to words at the end in SL sentence or sentence parts of translation examples.

(3) This system using RCL obtains partial translation rules by which pairs of parts are extracted from SL sentences and parts extracted from TL sentences in translation examples.

(4) This system using RCL gives a degree of correctness that is the same as used sentence translation rules to the acquired partial translation rules. This means that partial translation rules acquired using sentence translation rules that have a high degree of correctness also yield a high degree of correctness.

Figure 5 shows an example of acquisition of partial translation rule using sentence translation rule. In Fig. 5, “in” and “minutes” adjoin the variable in the source part of the sentence translation rule. They are the same parts as parts in the English sentence of a translation example. Therefore, “thirty” between common parts “in” and “minutes” is extracted from the English sentence of the translation example. Moreover, “wa” and “pun” adjoin the

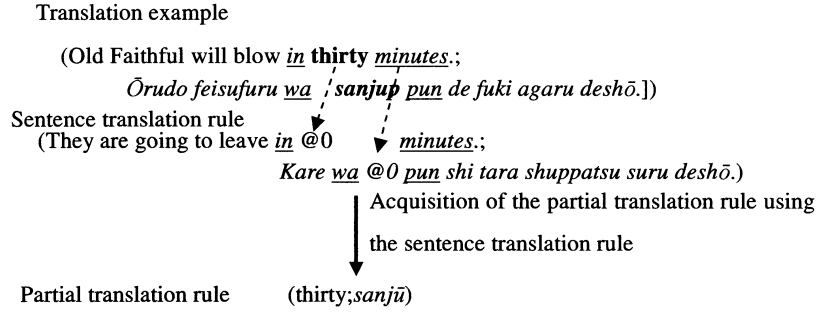


Fig. 5. Example of acquisition of a partial translation rule using a sentence translation rule.

variable in the target part of the sentence translation rule; they are the same parts as parts in the Japanese sentence of the translation example. Therefore, “sanjū” between common parts “wa” and “pun” is extracted from the Japanese sentence of the translation example. Eventually, (thirty;sanjū) is acquired as the partial translation rule.

4.3.3. Acquisition of sentence translation rules using partial translation rules

After partial translation rules are acquired using sentence translation rules described in Section 4.3.2, new sentence translation rules are acquired using partial translation rules. This process is based on heuristic (2) described in Section 2.1. Details of the acquisition process of sentence translation rules using partial translation rules are as follows:

(1) This system using RCL selects partial translation rules in which source parts have the same character strings

as parts in SL sentences of translation examples or SL parts of sentence translation rules, and in which target parts have the same character strings as parts in TL sentences of translation examples or TL parts of sentence translation rules.

(2) This system using RCL acquires sentence translation rules by replacing common parts that are the same parts as partial translation rules with variables, with SL sentences and TL sentences of translation examples or SL parts and TL parts of sentence translation rules.

(3) This system using RCL yields a degree of correctness that is the same as that of used partial translation rules to the acquired sentence translation rules. This means that sentence translation rules acquired using partial translation rules that have a high degree of correctness also have a high degree of correctness.

Figure 6 shows examples of acquisition of the sentence translation rule using the partial translation rule

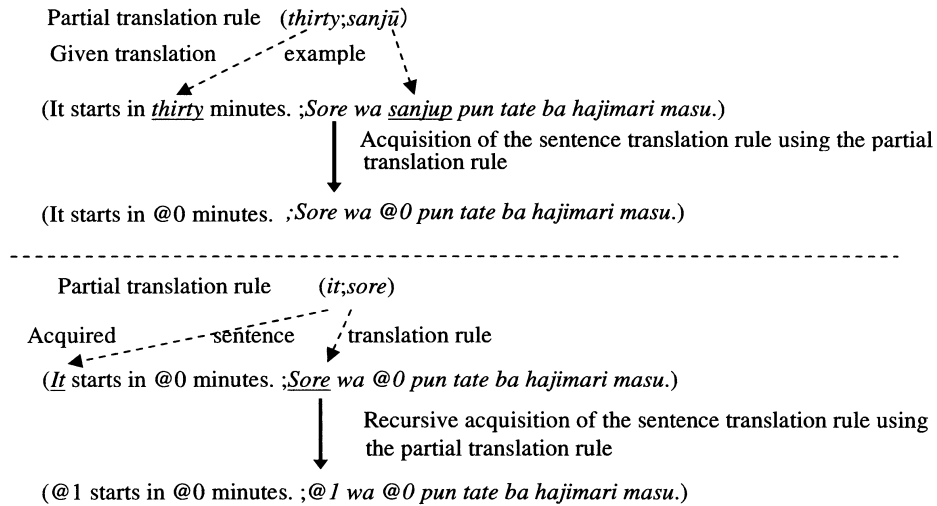


Fig. 6. Examples of acquisition of sentence translation rules using partial translation rules.

(thirty;*sanjū*) acquired in Fig. 5. In Fig. 6, the source part and the target part of the partial translation rule (thirty;*sanjū*) have the same character strings as parts in SL sentences and TL sentences of translation examples. Therefore, (It starts in @0 minutes.;*Sore wa @0 pun tate ba hajimari masu.*) is acquired as the sentence translation rule. Moreover, (@1 starts in @0 minutes.;*@1 wa @0 pun tate ba hajimari masu.*) is acquired recursively as a more abstract sentence translation rule.

5. Experiments for Performance Evaluation

5.1. Experimental procedure

Experimental data are of two kinds. One is learning data and the other is evaluation data. Translation examples (1759; number of characters: 82,327) that are pairs of English sentences and Japanese sentences were used as learning data. These translation examples are taken from textbooks for first [25] and second [26, 27] grade junior high school students. The average number of words of English sentences is 5.5 (first grade: 4.5; second grade: 6.2) among 1759 translation examples. Moreover, 1097 translation examples (number of characters: 60,265) that are pairs of English sentences and Japanese sentences are used as evaluation data. These translation examples are taken from other textbooks [28, 29] for second grade junior high school students. The average number of words of English sentences is 6.4 among 1097 translation examples. All translation examples are processed by English-to-Japanese machine translation system based on the process described in Section 4. The initial condition of the dictionary is empty. This is done to indicate that this system using RCL translates using translation rules acquired automatically from only translation examples without any static linguistic knowledge. Next, this system using RCL performed process of translation and learning to sentences one by one for all evaluation data. The reason is that this system using RCL can use all previously acquired translation rules.

5.2. Evaluation standards

Generated translation results are grouped into two categories: effective translation results and ineffective translation results. Moreover, effective translation results are grouped into the following two categories:

(1) Correct translation results without unknown words

This means that generated Japanese sentences correspond to Japanese sentences taken from textbooks [28, 29].

(2) Correct translation results with unknown words

This means that Japanese sentences that correspond to Japanese sentences taken from textbooks are obtained by substituting nouns and adjectives for generated Japanese sentences. In this case, it is easy for the user to substitute nouns and adjectives for generated Japanese sentences.

The effective translation rate is the rate of effective translation results for all evaluation data. Translation results are ranked when several different translation results are obtained for the same source sentences. Translation results are sorted so that translation results that have the highest degree of concreteness and highest degree of correctness described in Section 4.1 are ranked at the top. The user evaluates three translation results ranked from No. 1 to No. 3 among ranked translation results.

5.3. Experimental results

The number of effective translation results is 670 in these experiments with 1097 evaluation data. Therefore, the effective translation rate is 61.1%. Table 1 shows details of the effective translation rate; “(1)” refers to correct translation results without unknown words and “(2)” to correct translation results with unknown words. Values in parentheses show the number of effective translation results.

On the other hand, the number of effective translation results is 355 for the system using only GA-ILMT that acquires translation rules based on pairs of translation examples. Therefore, that effective translation rate is 32.2%. Using RCL, the effective translation rate improved from 32.4% to 61.1%. Table 2 shows examples of effective translation results by this system using RCL.

5.4. Discussion

5.4.1. Effective translation rate

This system using RCL automatically acquires translation rules in which source parts and target parts mutually correspond. We investigated the rate of effective translation results obtained using correct translation rules by which source parts and target parts correspond to each other in all

Table 1. Effective translation rate in this system using RCL

Effective translation rate	Details	
	(1)	(2)
61.1% (670)	41.6% (279)	58.4% (391)

Table 2. Examples of effective translation results

Examples of effective translation results without unknown words	
English sentences	Japanese sentences
Will you close the door?	<i>Doa o shimete ku re masen ka?</i>
This bag was made in France.	<i>Kono baggu wa furansu sē desu.</i>
We went there to play baseball.	<i>Watashi tachi wa yakyū wo suru tame ni soko e iki mashi ta.</i>
Examples of effective translation results with unknown words	
English sentences	Japanese sentences
Shall I take you to the amusement park?	<i>@0 e ture te itte age mashō ka?</i>
How far is it from Kyoto to Hiroshima?	<i>@0 kara Hiroshima made dono kurai no kyorī ga ari masu ka?</i>
He is the tallest in his class.	<i>Kare wa @0 de mottomo se ga takai desu.</i>

effective translation results. As a result, 572 effective translation results were based on correct translation rules in 670 effective translation results. The rate of effective translation results based on correct translation rules is, therefore, 85.4%. In 98 effective translation results that excluded 572 effective translation results from 670 effective translation results, 90 effective translation results were generated using the translation rules in which English sentences have identical character strings as English sentences of translation examples. Eight effective translation results were generated by combining erroneous translation rules by which source parts and target parts did not correspond to each other. The number of effective translation results that were generated by the system using only GA-ILMT was 355. Therefore, using RCL, the number of effective translation results increased from 355 to 670.

Next, we describe the effectiveness of this system using RCL from the learning ability perspective. In general, a system possesses high learning ability when it does not require any translation examples to acquire correct translation rules. Therefore, in 572 effective translation results that are based on correct translation rules, we investigated the number of translation examples that this system using RCL needed in order to acquire correct translation rules. In this case, we investigated the frequency of appearance of verbs in English sentences of translation examples. By evaluation standards described in Section 5.2, this system using RCL cannot translate SL sentences when verbs are unknown words. Therefore, investigating the appearance frequency of verbs is important. From investigation results, in 322 (56.3%) effective translation results among 572 effective translation results, this system using RCL was able to acquire correct translation rules that included new verbs after the translation examples that have the new verbs were given only one time. In 108 (18.9%) effective translation

results, this system using RCL was able to acquire correct translation rules that included new verbs after the translation examples that have the new verbs were given twice. In 142 (24.8%) effective translation results, this system using RCL was able to acquire correct translation rules that included new verbs after the translation examples that have the new verbs were given over three times. Therefore, in 75.2% of 572 effective translation results, this system using RCL was able to acquire correct translation rules that included new verbs from the translation examples in which new verbs were given only once or twice. This means that this system using RCL demonstrates a high ability which acquires useful translation rules.

Moreover, we describe the relation between effective translation rate and rate of SL sentences including unknown words in evaluation data. Figure 7 shows the changes of the effective translation rate and the rate of SL sentences including new verbs for every 100 SL sentences. In Fig. 7, the

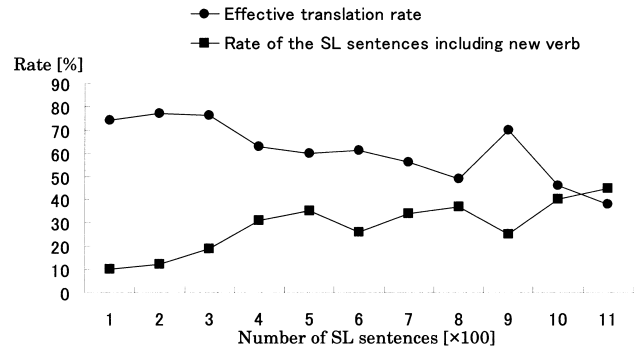


Fig. 7. Effective translation rate and rate of SL sentences including new verb.

effective translation rate is high when the rate of SL sentences including new verbs is low. In contrast, the effective translation rate is low when the rate of SL sentences including new verbs is high. This result indicates that the change of the effective translation rate depends on the number of SL sentences including new verbs. The number of ineffective translation results including new verbs was 309. Therefore, the effective translation rate is 85.0% when SL sentences which have unknown words do not exist in evaluation data.

5.4.2. Acquired translation rules

The number of sentence translation rules acquired from 1759 translation examples used as learning data was 16,222; the number of partial translation rules was 2101. Figure 8 shows changes of the number of sentence translation rules and partial translation rules for every 100 translation examples. It is seen that partial translation rules increased slowly, not rapidly. This means that the number of partial translation rules increases in proportion to the increase in the number of new words. On the other hand, sentence translation rules increased rapidly. In this system using RCL, the number of acquired sentence translation rules increases as sentences of translation examples lengthen because various sentence translation rules with variables are acquired recursively, as described in Section 4.3.3. Figure 8 shows that the average number of words of English sentences of translation examples between translation example Nos. 1 and 100 is 3.8; the average number of words between translation example Nos. 1700 and 1759 is 6.3. The number of sentence translation rules increases rapidly between translation example Nos. 1500 and 1600. The reason is that the number of average words of English sentences in the translation examples is highest in those. When the number of words of English sentences in the translation examples is high, many sentence translation

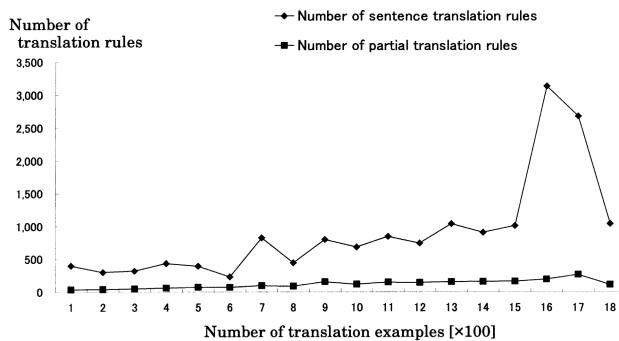


Fig. 8. Numbers of acquired translation rules using Recursive Chain-Link-type Learning.

rules are acquired by performing the processing described in Section 4.3.3 because the acquired partial translation rules have the same character strings as parts of many translation examples. The average number of words of English sentences in the translation examples between translation example Nos. 1500 and 1600 is 7.2. However, after that, the number of sentence translation rules decreased rapidly. Therefore, an explosive increase in translation rules did not occur.

Next, we describe the precision of acquired translation rules using RCL. In 1759 translation examples used as learning data, the precision of correct partial translation rules by which source parts and target parts correspond to each other was 52.1%. In erroneous partial translation rules by which source parts and target parts do not correspond to each other, the rate of cases for which this system using RCL acquired them using erroneous sentence translation rules was 30.9%. Moreover, the rate of cases for which this system using RCL acquired them using correct sentence translation rules was 69.1%. In most cases, acquisition of erroneous sentence translation rules was caused by different structure between sentence translation rules and translation examples. For example, from the correct sentence translation rule (Kate₁ @0 is Jim₁; Keito₂ @0 wa jimu desu.) which is an affirmative sentence and the translation example (Aki₁ who is that boy₁?; Aki₁ ano shōnen wa dare desu ka?) which is an interrogative sentence, (who ; ano shōnen) is acquired as the erroneous partial translation rule by extracting parts between common parts because “who” exists between “,” and “is” which are common parts in English sentences, and “ano shōnen” exists between “,” and “wa,” which are common parts in Japanese sentences.

The precision of correct sentence translation rules by which source parts and target parts correspond to each other was 61.8%. In the erroneous sentence translation rules, by which source parts and target parts do not correspond to each other, the rate of cases in which this system using RCL acquired them using erroneous partial translation rules was 96.2%. Moreover, the rate of cases in which this system using RCL acquired them using the correct partial translation rules was 3.8%. In most cases, acquisition of erroneous partial translation rules occurs because this system using RCL could not recognize various expressions in Japanese. For example, there are several equivalent expressions for “sister.” Both “imōto” and “imōto san” are correct equivalents for “sister” although the expression is different. This system using RCL cannot recognize such cases. Therefore, from the translation example (Is she your sister?; Kanojo wa anata no imōto san desu ka?) and the partial translation rule (sister; im{ō}to), the system acquires (Is she your @0?; Kanojo wa anata no @0 san desu ka?) as the erroneous sentence translation rule. In this case, the correct sentence translation rule is (Is she your @0?; Kanojo wa anata no @0 desu ka?) without “san” in the Japanese part. However,

our system does not use these erroneous translation rules in the translation process because their degree of correctness decreases using evaluation methods in the feedback process [23, 24]. We investigated the precision of evaluation for erroneous translation rules. As a result, the precision that this system using RCL could decide as the erroneous translation rules for acquired erroneous translation rules was 70.0%. Therefore, we confirmed that this system using RCL can evaluate erroneous translation rules generally. Erroneous translation rules determined by this system using RCL indicate acquired translation rules whose correct degree is under 50.0% because a value less than 50.0% implies that the erroneous frequency is higher than the correct frequency.

5.4.3. Comparative experiments with rule-based machine translation

Next, we describe the effectiveness of our system through comparative experiments with three other commercial English-to-Japanese machine translation systems. We used three commercial English-to-Japanese machine translation systems that are available on the market from companies A, B, and C in those experiments. We designate these three machine translation systems as systems A, B, and C, respectively. In these three machine translation systems, 788 English sentences were used as evaluation data. These are English sentences that exclude 309 English sentences that were not translated into correct Japanese sentences in experimental results of RCL from 1097 English sentences used as evaluation data. This system using RCL was unable to translate into correct Japanese sentences because the 309 English sentences include new verbs. In contrast, all equivalent for verbs in the 309 English sentences were given beforehand in the other three rule-based machine translation systems. Therefore, by excluding the 309 English sentences that included new verbs in this system using RCL from 1097 English sentences, we were able to perform fair comparative experiments with our system and rule-based machine translation systems. Table 3 shows comparative experiment results; “(1)” refers to correct translation results without unknown words and “(2)” to correct translation

results with unknown words described in Section 5.2. Values in parentheses indicate the number of effective translation results.

In Table 3, we confirm that the effective translation rate of this system using RCL is almost identical to those of the other three rule-based machine translation systems. However, the rate of correct translation results with unknown words is very high in this system using RCL. This means that this system using RCL is not comfortable because the user must give the equivalent for unknown words which are nouns or adjectives. However, the number of unknown words decreases by adding learning data.

Moreover, from the view of high-quality translation, we evaluated translation results of this system using RCL and the three rule-based machine translation systems more strictly. Only translation results that have nearly identical character strings to Japanese sentences taken from textbooks are evaluated as effective translation results. Translation results are not effective translation results when translation results have different phrasing from Japanese sentences in the textbooks. For example, “*Sore wa oyoso jup pun o yō shi masu.*” is not an effective translation result when the correct translation result for “It takes about ten minutes.” is “*Yaku jup pun kakari masu.*” because “*Sore wa oyoso jup pun o yō shi masu.*” uses the different phrasing “*sore wa*” from the correct translation result. Therefore, we excluded translation results that have different phrases from correct translation results. Table 4 shows a comparison of effective translation rates in terms of translation quality. This system using RCL can generate more high-quality translation results than the other three rule-based machine translation systems in Table 4.

Next, we describe the effectiveness of this system using RCL by generated translation results. For example, our system translated “How long does it take by Shinkansen?” into “*@0 de dono kurai jikan ga kakari masu ka?*”. Systems A, B, and C translated “How long does it take by Shinkansen?” into “*Sore wa Shinkansen no soba ni dore gurai fōi desu ka?*”, “*Dono kurai (jikan) , sore wa Shinkansen ni yotte nori masu ka?*”, and “*Sore wa dore hodo nagaku Shinkansen no tame tori masu ka?*”, respectively. These

Table 3. Results of comparative experiments

Systems	Effective translation rate	Details	
		(1)	(2)
System using RCL	85.0% (670)	41.6% (279)	58.4% (391)
System A	85.8% (676)	84.0% (568)	16.0% (108)
System B	81.7% (644)	83.7% (539)	16.3% (105)
System C	76.9% (606)	82.7% (501)	17.3% (105)

Table 4. Comparison of effective translation rates in terms of translation quality

Systems	Effective translation rate	Details	
		(1)	(2)
System using RCL	73.7% (581)	47.5% (276)	52.5% (305)
System A	69.8% (550)	84.2% (463)	15.8% (87)
System B	63.3% (499)	85.0% (424)	15.0% (75)
System C	58.2% (459)	82.8% (380)	17.2% (79)

translation results are erroneous translation results because they do not make sense. In our system, using RCL, from the translation example (How long does it take by bus?;*Basu de dono kurai jikan ga kakari masu ka?*), (bus;*basu*) was acquired as the correct partial translation rule that the source part and the target part correspond to each other. As a result, (How long does it take by @0?;*@0 de dono kurai jikan ga kakari masu ka?*) was acquired as the correct sentence translation rule. In general, the equivalent of “it” is “*sore*.” However, in the case of “How long does it take by Shinkansen?”, the system must omit “*sore*” to get the correct translation result. Translation results of systems A, B, and C do not omit “*sore*.” Thereby, they generate erroneous translation results which do not make sense. This result shows that this system using RCL can generate high-quality translation results by acquiring useful translation rules based on translation examples.

5.4.4. Advantage of RCL

This system using RCL can acquire many translation rules efficiently from each translation example. The reason is that this system using RCL induces a chain reaction in acquisition of new translation rules. Figure 9 shows an example of a chain reaction in acquisition of translation rules using RCL. In Fig. 9, this system using RCL eventually acquires (I want to be @0.;*Watashi wa @0 ni nari tai.*) as the sentence translation rule that is used to translate the source sentence “I want to be a baseball player.” First, the sentence translation rule (This is @0.;*Kore wa @0 desu.*) that is used as starting point for the acquisition process of translation rules in RCL is acquired by GA-ILMT in (1) of Fig. 9. Using this sentence translation rule, the system

extracts the part from “a” on the right of “is” to the word “singer” at the end in SL sentences of translation examples; it also extracts the part between the right of “*wa*” and the left of “*desu*” in TL sentences of translation examples. Therefore, in (2) of Fig. 9, the system automatically acquires (a singer;*kashu*) as the partial translation rule from (She is a singer.;*Kanojo wa kashu desu.*). This means that “a singer” and “*kashu*” in the translation example (She is a singer;*Kanojo wa kashu desu.*) correspond to variables “@0” in the sentence translation rule (This is @0.;*Kore wa @0 desu.*). Using the partial translation rule (a singer;*kashu*), the system acquires (Yumi will be @0.;*Yumi wa @0 ni naru deshō.*) as the sentence translation rule from the translation example (Yumi will be a singer.;*Yumi wa kashu ni naru deshō.*) in (3) of Fig. 9. Moreover, (an English teacher;*eigo no sense*) as the partial translation rules in (4) of Fig. 9, (I want to be @0.;*Watashi wa @0 ni nari tai.*) as the sentence translation rule in (5) of Fig. 9, are acquired one after another. The advantage of RCL is that such acquisition of translation rules enables acquisition of other new translation rules. Therefore, this system using RCL causes a chain reaction in acquisition of translation rules automatically.

This system using RCL can acquire various translation rules that correspond to word, phrases, and clauses using previously-acquired translation rules. In (4) of Fig. 9, the acquired partial translation rule (an English teacher;*eigo no sense*) corresponds to a noun phrase. Moreover, the system can acquire a clause level translation rule by using the sentence translation rule (This is @0.;*Kore wa @0 desu.*), which has the basic structure of English and Japanese sentences. For example, the system can acquire (the book which he bought;*kare ga katta hon*) as the partial

(1) Acquisition of the sentence translation rule using GA-ILMT

(This is a nice house.; *Kore wa subarashiie desu.*) → (This is @0.;*Kore wa @0 desu.*)

(This is my book.; *Kore wa watashi no hon desu.*)

(2) Acquisition of the partial translation rule using the sentence translation rule

(This is @0.;*Kore wa @0 desu.*) → (a singer;*kashu*)

(She is a singer.; *Kanojo wa kashu desu.*)

(3) Acquisition of the sentence translation rule using the partial translation rule

(a singer;*kashu*)

(Yumi will be a singer.; *Yumi wa kashu ni naru deshō.*) → (Yumi will be @0.
; *Yumi wa @0 ni naru deshō.*)

(4) Acquisition of the partial translation rule using the sentence translation rule

(Yumi will be @0.; *Yumi wa @0 ni naru deshō.*)

(I want to be an English teacher.; *Watashi wa eigo no sense ni nari tai.*) → (an English teacher
; *eigo no sense*)

(5) Acquisition of the sentence translation rule using the partial translation rule

(an English teacher;*eigo no sense*)

(I want to be an English teacher.; *Watashi wa eigo no sense ni nari tai.*) → (I want to be @0.
Watashi wa @0 ni nari tai.)

Fig. 9. Example of acquisition process of translation rules using Recursive Chain-Link-type Learning.

translation rule that corresponds to a clause from translation example (This is the book which he bought.;*Kore wa kare ga katta hon desu.*). The acquired partial translation rule (the book which he bought;*kare ga katta hon*) is the part that corresponds to variables in the sentence translation rule (This is @0.;*Kore wa @0 desu.*), with the translation example (This is the book which he bought.;*Kore wa kare ga katta hon desu.*). This system using RCL can acquire translation rules of various units without any analytical knowledge.

5.4.5. Comparison with related work

We describe the effectiveness of RCL from comparison with related work that acquires translation rules based on learning algorithms without depending on analytical knowledge. Malavazos and colleagues [12, 13] proposed a learning method that acquires partial translation rules by extracting different parts in character strings of translation examples that are pairs of English and Greek sentences. It then acquires abstract translation rules by replacing different parts with variables. However, in two translation examples, this method acquires translation rules only when the number of different parts between two SL sentences is one and the number of different parts between two TL sentences is also just one. This implies a very strict condition of acquisition of translation rules. Moreover, experiments for performance evaluation have not been performed sufficiently; only the precision of the acquired translation rules has been evaluated. Güvenir and Cicekli [14] propose a learning method that acquires translation rules using previously-acquired partial translation rules. Using partial translation rules in this method, the system decides corresponding parts when several different parts exist in two translation examples that are pairs of English and Turkish sentences. The system can acquire translation rules even when the number of different parts between two SL sentences is more than two and the number of different parts between two TL sentences is also more than two. However, the number of different parts between two SL sentences must be the same as the number of different parts between two TL sentences. Therefore, this method also means that it is difficult to acquire the translation rules from sparse translation examples because the condition of acquisition of translation rules is strict. Experiments for performance evaluation have not been performed. McTait [15] proposes a learning method that acquires translation rules by replacing common parts with variables to two translation examples that are pairs of English and French sentences when the number of different parts between two SL sentences differs from the number of different parts between two TL sentences. However, it is difficult to acquire the translation rules that correspond to partial translation rules because almost all parts differ between two translation examples in

general. Performance evaluation experiments indicate an approximately 30% correct translation rate [16]. The rate presented in this paper is lower than 61.1%, but we cannot compare the previous method with our system because data that were used to evaluate this method differ from data in this paper.

In existing related works, the system cannot efficiently acquire translation rules because it requires pairs of similar translation examples at all times. As a result, existing learning methods present the obstacle that the system requires a large number of similar translation examples to acquire many high-quality translation rules. In contrast, this system using RCL can acquire translation rules efficiently because it decides the corresponding parts for each translation example, unlike existing works that depend on different parts between two translation examples. This means that this system using RCL can acquire translation rules from sparse translation examples without requiring pairs of similar translation examples. Moreover, we implemented RCL to an English-to-Japanese machine translation system and confirmed the effectiveness of RCL with performance evaluation experiments.

6. Conclusions

This paper has proposed a method of machine translation system using Recursive Chain-Link-type Learning that efficiently acquires translation rules for each translation example. This system using RCL determines corresponding parts between SL sentences and TL sentences of translation examples using previously acquired translation rules. This means that this system using RCL causes a chain reaction in acquisition of translation rules. As a result, this system using RCL can reciprocally acquire partial translation rules and sentence translation rules in a pattern like a linked chain. Moreover, this system using RCL can acquire various translation rules by recursively performing the acquisition process of translation rules. Experimental results demonstrate an effective translation rate by this system using RCL: 61.1%. The effective translation rate in this system using RCL was 85.0% when sufficient learning data were given. Consequently, we confirmed that this system using RCL can generate high-quality translation results than other commercial rule-based machine translation systems.

In the future, we need to implement a new algorithm to combine acquired translation rules effectively toward translation of longer sentences. Moreover, we plan to confirm RCL's effectiveness by applying it to other language sentences because RCL is the learning algorithm that does not depend on specific linguistic knowledge.

Acknowledgments. This work was partially supported by grants from the High-Tech Research Center of Hokkai-Gakuen University and a Government subsidy for aiding scientific research (No. 14658097) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1. Asia-Pacific Association for Machine Translation. Vision of the 21st century—Machine translation. Tokyo, 2000. (in Japanese)
2. Nomura H. Language processing and machine translation. Kōadansha; 1991. (in Japanese)
3. Tanaka H. Natural language processing and its applications. IEICE; 1999. (in Japanese)
4. Brown P, Cocke J, Della Pietra S, Della Pietra VJ, Jelinek F, Lafferty JD, Mercer RL, Roossin PS. A statistical approach to machine translation. *Computational Linguistics* 1990;16:79–85.
5. Sato S. MBT2: A method for combining fragments of examples in example-based translation. *J Jpn Soc Artif Intell* 1991;6:861–871. (in Japanese)
6. Kaji H, Kida Y, Morimoto Y. Learning translation templates from bilingual text. *Proc Coling'92*, p 672–678, Nantes, France.
7. Kitamura M, Matsumoto Y. Automatic acquisition of translation rules from parallel corpus. *Inf Process Soc Japan* 1996;37:1030–1040. (in Japanese)
8. Furuse O, Iida H. Cooperation between transfer and analysis in example-based framework. *Proc Coling'92*, p 645–651, Nantes, France.
9. Watanabe H, Takeda K. A pattern-based machine translation system extended by example-based processing. *Proc Coling-ACL'98*, p 1369–1373, Montreal.
10. Brown RD. Transfer-rule induction for example-based translation. *Proc Workshop on EBMT, Machine Translation Summit VIII*, Santiago de Compostela, Spain, 2001.
11. Carl M. Inducing translation grammars from bracketed alignments. *Proc Workshop on EBMT, Machine Translation Summit VIII*, Santiago de Compostela, Spain, 2001.
12. Malavazos C, Piperidis S. Application of analogical modelling to example based machine translation. *Proc Coling* 2000, p 516–522, Saarbrücken, Germany.
13. Malavazos C, Piperidis S, Carayannis T. Towards memory and template based translation synthesis. *Proc Machine Translation and Multilingual Applications in the New Millennium*, p 1-1–1-8, Exeter, England, 2000.
14. Güvenir HA, Cicekli I. Learning translation templates from examples. *Inf Syst* 1998;23:353–363.
15. McTait K. Memory-based translation using translation patterns. *Proc 4th Annual CLUK Colloquium*, p 43–52, Sheffield, England, 2001.
16. McTait K. Linguistic knowledge and complexity in an EBMT system based on translation patterns. *Proc Workshop on EBMT, Machine Translation Summit VIII*, p 23–34, Santiago de Compostela, Spain, 2001.
17. Araki K, Takahashi Y, Momouchi Y, Tochinai K. Non-segmented kana–kanji translation using inductive learning. *Trans IEICE* 1996;J79-D-II:391–402. (in Japanese)
18. Echizen-ya H, Araki K, Momouchi Y, Tochinai K. Application of genetic algorithms for example-based machine translation method using inductive learning and its effectiveness. *Inf Process Soc Japan* 1996;37:1565–1579. (in Japanese)
19. Echizen-ya H, Araki K, Momouchi Y, Tochinai K. Machine translation method using inductive learning with genetic algorithms. *Proc Coling'96*, p 1020–1023, Copenhagen.
20. Goldberg DE. Genetic algorithms in search, optimization, and machine learning. Addison–Wesley; 1989.
21. Echizen-ya H, Araki K, Momouchi Y, Tochinai K. Recursive acquisition method of translation rules by focusing on local parts: GA-ILMT2. *Tech Rep IEICE* 2001;NLC 2001-10:9–16. (in Japanese)
22. Araki K, Tochinai K. Acquisition words by inductive learning and recognition words using certainty. *Trans IEICE* 1992;J75-D-II:1213–1221. (in Japanese)
23. Echizen-ya H, Araki K, Momouchi Y, Tochinai K. Using state transition on GA-ILMT based on learning capability. *Tech Rep IEICE IPSJ* 2000;NL135-23:173–180. (in Japanese)
24. Echizen-ya H, Araki K, Momouchi Y, Tochinai K. Effectiveness of layering translation rules based on transition networks in machine translation using inductive learning with genetic algorithms. *Proc Machine Translation and Multilingual Applications in the New Millennium*, p 5-1–5-8, Exeter, England, 2000.
25. Nihon-Kyōzai(1). One World English course 1 new edition. Tokyo, 2001. (in Japanese)
26. Nihon-Kyōzai(2). One World English course 2 new edition. Tokyo, 2001. (in Japanese)
27. Shuppan H. System English course 2 new edition. Tokyo, 2001. (in Japanese)
28. Bunri. Work English course 2 new edition. Tokyo, 2001. (in Japanese)
29. Shuppan S. Training English course 2 new edition. Osaka; 2001. (in Japanese)

AUTHORS (from left to right)



Hiroshi Echizen-Ya received his B.E. and M.E. degrees from Hokkai-Gakuen University in 1991 and 1996. From 1996 to 1998, he was a doctoral student at Hokkaido University. Currently he is a research associate on the Faculty of Engineering, Hokkai-Gakuen University. His research interests include natural language processing and machine translation. He is a member of IPSJ, JSAI, NLP, and ACL.

Kenji Araki received his B.E. and Ph.D. degrees from Hokkaido University in 1982 and 1988. He was an associate professor from 1991 to 1998, and a professor in 1998 at Hokkai-Gakuen University. He then joined Hokkaido University as an associate professor, and is currently a professor in the Graduate School of Engineering. His research interests include natural language processing, morphological analysis, machine translation, and speech dialogue processing. He is a member of ACL, IEEE, IPSJ, and NLP.

Yoshio Momouchi received his B.E. and D.Eng. degrees from Hokkaido University in 1965 and 1973. After serving there as a research associate, lecturer, and associate professor, he joined the Department of Electronics and Information Engineering, Hokkai-Gakuen University, as a professor in 1988. His research interests include understanding and generation of natural language. He is a member of IPSJ, NLP, MLSJ, JSAI, JCSS, and ACL.

Koji Tochinal received his B.S., M.S., and D.Eng. degrees from Hokkaido University in 1962, 1964, and 1986. He was formerly a professor of electronics and information engineering at the Graduate School of Engineering, Hokkaido University, and is currently a professor on the Faculty of Business Administration, Hokkai-Gakuen University. His research interests are in the areas of natural language processing and spoken language processing. He is a member of IEICE, IPSJ, and the Acoustic Society of Japan.