인지과학, 제 15 권 제 2 호, 2004년

인접 단어들의 접속정보를 이용한 일한 활용어 번역

Japanese-to-Korean Inflected Word Translation Using Connection Relations of Two Neighboring Words

김 정 인* 이 강 혁**
(Jung-In Kim) (Kang-Hyuk Lee)

요 약 일본어와 한국어는 문법적으로 많은 유사점을 가지고 있다. 이러한 유사점을 잘 이용한다면 일한 기계번역 시스템에서 구문해석이나 의미해석의 상당한 부분을 생략할 수 있다. 몇 년 전부터 우리는 유사성을 이용하여 번역율을 높이는 방법으로 번역테이블을 이용한 일한기계번역 시스템을 연구해 왔다. 그러나 이 시스템은 활용어미의 번역, 다의성 단어의 처리 등 몇 가지 문제점을 가지고 있었다. 본 논문에서는 번역테이블을 이용하는 시스템을 개선하여 이웃 하는 단어들과의 관계 정보를 이용한 일한 기계번역 시스템을 제안한다. 현재 시스템의 문제점들을 해결하기 위하여 우선 조사, 조동사의 접속정보를 최대한 이용한다. 또한, 번역 테이블을 엔트리테이블과 접속정보 테이블로 나누어 설계하여 번역의 효율을 높인다. 즉, 하나의 역어만 가지는 단어인 경우, 우리는 일한 직접 대응 방법을 이용하여 바로 번역하고 2개 이상의 역어로 번역되어야 할 경우만 접속 정보 값을 평가하여 가장 가능성이 높은 번역어를 선택하도록 한다.

주제어 한국어, 일본어, 인접단어, 접속정보, 기계번역

Abstract There are many syntactic similarities between Japanese and Korean language. These similarities enable us to build Japanese-Korean translation systems without depending on sophisticated syntactic analysis and semantic analysis. To further improve translation accuracy, we have been developing a Japanese-Korean translation system using these similarities for several years. However, there still remain some problems with regard to translation of inflected words, processing of multi-translatable words and so on. In this paper, we propose a new method for Japanese-Korean machine translation by using the relationships of two neighboring words. To solve the problems, we investigate the connection rules of auxiliary verb priority. And we design the translation table, which consists of entry tables and connection form tables. For unambiguous words, we can translate a Japanese word to the corresponding Korean word in terms of direct-matching method by consulting the only entry table. Otherwise we have to evaluate the connection value computed from connection form tables and then we can select the most appropriate target word.

Keywords Korean, Japanese, Neighboring words, Connection Relations, Machine Translation

* 동명정보대학교 컴퓨터공학과
Department of Computer Engineering,
TongMyong University of Information Technology
** 동명정보대학교 게임공학과
Department of Game Engineering,
TongMyong University of Information Technology
투고분야: 자연어처리 (기계번역)
Natural Language Processing (Machine Translation)
주소 : 608-711 부산광역시 남구 용당동 535
동명정보대학교 공과대학 컴퓨터공학과
전화: 051-610-8393
팩스: 051-610-8393
휴대폰: 011-9067-2627
전자메일 : jikim@tmic.tit.ac.kr

1. Introduction

Due to the massive syntactic similarities between Korean and Japanese, many Japanese-to-Korean translation systems have been built by directly matching the result of morphological analysis of source words with the corresponding target words. The both languages also use particles to represent case relations between words, and thus many researches in Japanese-to-Korean machine translation depend on values of case-form pattern matching. However, since many J-to-K machine translation systems based on the linguistic similarity do away with syntactic and semantic analysis, they suffer from the problems with translation of conjugated words and disambiguation of multi-translatable words, which arises from insufficient information that the direct match method causes. It has been shown in the literature that it is possible to build a simplistic high-quality translation system if the information useful for translating conjugated and ambiguous words can be extracted and used.

A variety of methods have been proposed to deal with ambiguity in the MT literature: semantic disambiguation of simple sentences based on word relations [8] and verb disambiguation using case-frame patterns [13, 15], etc. Hwang *et al.* [19] also proposed a method to build a classified Korean lexical table from a Japanese lexical table, which is again depending on the similarity of the two languages.

Various methods have also been proposed to deal with inflected words: processing of predicates in terms of semantic correspondence of the two languages [9], and processing of inflected predicates in terms of phonological representations [14]. These methods, however, were not expressive enough to deal with irregular conjugation of Korean inflected words. In order to process irregular conjugation in a distributed way, a method using a translation table was proposed in which the table of translated words is prepared according to the semantic connective relations of inflected words [11, 12, 16, 17]. Kim *et al.* [21, 24, 25] proposed an extended translation table method to remedy the problems with the previous methods: lack of non-extendibility of translation tables, under-expressiveness of connection relations and inconsistency of translation systems.

In this paper, we propose an extended translation table method in which the structure of the translation table is modified to express the connection relations of two neighboring words, and inflected words can be translated in terms of connection relations of two neighboring words.

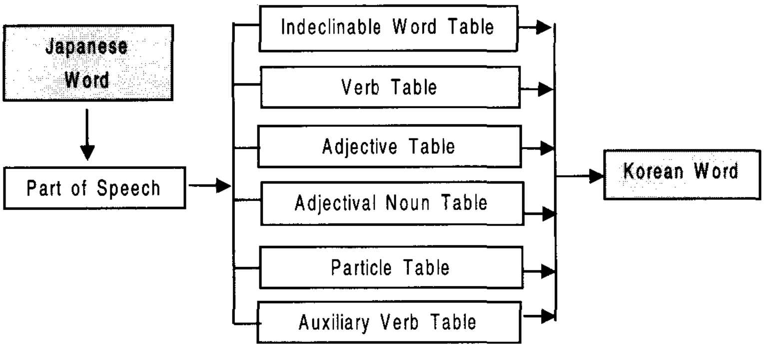
2. Japanese-Korean Machine Translation Using the Extended Translation Table Method

Machine translation systems must apply numerous linguistic rules that human speakers use. A translation processing system consists of the translation system that specifies the flow of process, translation rules, etc., and the translation dictionary that describes the properties of words, translatable words, etc. The two modules of the translation processing system are closely related, so the size of the dictionary can be minimized in terms of giving as much as detailed descriptions to linguistic rules and the translation system can be built in terms of putting as much as information into the translation dictionary. Ambiguity is so prevalent in processing natural language, and in order to deal with ambiguity, it is necessary to use massive translation rules. It is almost impossible to describe all the necessary translation rules, and thus it is much more effective to describe them in the form of translation dictionary.

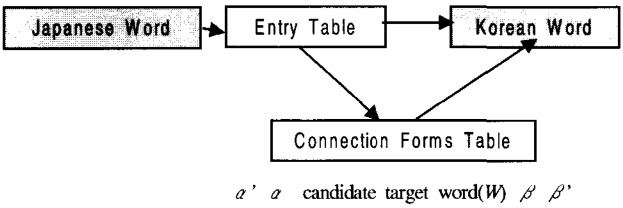
2.1 An Extended Translation Table Method

The method utilizing the extended translation table uses both base and inflected forms of words as entries in the table. That is, it is based on the one-to-one correspondence of Japanese-Korean morphemes, thus avoiding the process of converting base forms to inflected forms.

Whereas in the translation table method connection rules are described in the translation system and candidate target words are put into the translation table without introducing primary target words, in the extended translation table method primary target words and candidate target words are put into the entry table and the extended translation table, respectively. (Figure 2) depicts the conceptual structure of the extended translation table method. In the translation table method, translation is carried out by consulting different translation tables according to the part-of-speech of source words. The extended translation table translates Japanese source words into primary target words described in the entry table. For ambiguous source words, it chooses appropriate target words by using neighboring words and connection rules described in the connection information table.



(Figure 1) The Conceptual Structure of the Translation Table Method



(Figure 2) The Conceptual Structure of the Extended Translation Method

The connection information of a word with its two neighboring words described in the connection information dictionary is represented by α', α, β, and β', where α' represents the connection information of the second-preceding neighbor word, α, the connection information of the first-preceding neighbor word, β, the connection information of the first-following neighbor word, and β, the connection information of the second-following neighbor word. All of this information is specified in the dictionary when candidate target words are registered.

2.2 Connection Forms in the Extended Translation Table Method

The extended translation table method uses part-of-speech, semantic classifier of nouns, inflected form of declinable words, particles/auxiliary verbs and symbols as the connection forms that represent connection relations of inflected words.

- Part-of-Speech : noun, semi-nominal(の), verb (upper-column ru-ending, lower-column ru-ending, カ-row irregular inflection, サ-row irregular inflection, 5th-column), adjective, adjectival verb, auxiliary verb, particle (case, conjunctive, adverbial, ending), adverb, conjunction, exclamation, symbol, etc.
- Semantic Classifiers of Nouns: concrete nouns (animal, human, structure/organs, plants, parts of living organisms, natural matter, products, tools), phenomenal nouns, abstract nouns (motion and process, mind, linguistic process, character, relation, space/directedness, time, quantity), etc.
- Inflected Forms of Verbs: Incomplete-1, Incomplete-2, Incomplete-3, Verb-Followed-1, Verb-Followed-2, Verb-Followed-3, Ending, Subjunctive, Command-1, Command-2, Stem
- Particles, Auxiliary Verbs and Serial Numbers of Symbols: 58 particles, 19 auxiliary verbs, 30 symbols

By using the above-listed connection forms, the extended translation table method describes in the connection table the connection rules as to how an inflected word is connected with otherwords before and after it. In the area of natural language processing, a lot of research on part-of-speech classifications has been carried out to analyze Japanese more

<Table 1> Classification of Inflected Words

Form	Code	Verb					Adjective	Adjectival Verb
		upper-column ru-ending	lower-column ru-ending	カ-row irregular inflection	サ-row irregular inflection	5th-column		
Incomplete-1 Incomplete-2 Incomplete-3	1 2 3	起き	受け	こ	さ し せ	書か 書こ	美しかろ	元氣だろ
Verb-Followed-1 Verb-Followed-2 Verb-Followed-3	4 5 6	起き	受け	き	し	書き 書い	美しかっ 美しく	元氣だっ 元氣で 元氣に
Ending	7	起きる	受ける	くる	する	書く	美しい	元氣だ
Noun-Followed	8	起きる	受ける	くる	する	書く	美しい	元氣な
Subjunctive	9	起きれ	受けれ	くれ	すれ	書け	美しけれ	元氣なら
Command-1 Command-2	A B	起きろ 起きよ	受けろ 受けよ	こい	しろ せよ	書け		
Stem	C	起	受			書	美し	元氣

<Table 2> Ordered Number of Particles

Number	Particle	Number	Particle	Number	Particle	Number	Particle
1	が	16	から	31	さえ	46	か
2	の	17	し	32	でも	47	な
3	を	18	ても	33	しか	48	な
4	に	19	でも	34	まで	49	なあ
5	へ	20	けれど	35	ばかり	50	や
6	と	21	けれども	36	だけ	51	ぞ
7	から	22	て	37	ほど	52	とも
8	より	23	で	38	くらい	53	よ
9	で	24	ながら	39	ぐらい	54	の
10	や	25	たり	40	など	55	わ
11	ば	26	だり	41	きり	56	ね
12	と	27	ものの	42	ぎり	57	ねえ
13	が	28	は	43	なり	58	さ
14	のに	29	も	44	やら		
15	ので	30	こそ	45	か		

<Table 3> Ordered Number of Auxiliary Verbs

Number	Particle	Number	Particle	Number	Particle	Number	Particle
1	せ	6	ぬ	11	たい	16	ようだ
2	させ	7	ん	12	ます	17	らしい
3	れ	8	う	13	た	18	だ
4	られ	9	よう	14	だ	19	です
5	ない	10	まい	15	そうだ		

effectively [22]. To increase the effectiveness of Japanese analysis, we have also used the part-of-speech information of words and their inflected forms that are more subdivided than in school grammar. <Table 1> shows the classification of inflected forms according to the part-of-speech information of words. For the semantic classification of nouns, we have consulted the classification method of IPAL Verb Dictionary. The ordered number of particles and auxiliary verbs are shown in <Table 2 and 3>, respectively. Here, particles of the same form are multiply registered, but their meaning is different. Number 1 through 10 indicate case particles, number 11 through 27, conjunctive particles, number 28 through 45, adverbial particles, number 46 through 58, ending particles. For example, 1-numbered 「が」 and 13-numbered 「が」 plays the role of the subjective case particle and counterfactual conjunctive particle, respectively. 7-numbered 「から」 and 16-numbered 「から」 are used as the case particle meaning reference point/cause/reason and factual conjunctive particle, respectively. Also in the case of auxiliary verbs, 14-numbered 「た」 indicates past-tense whereas 18-numbered 「だ」 indicates declarative-ending.

<Table 4> Classification of Connection Forms

Type	Part-of-Speech	Semantic Classifier or Inflected Form	Ordered Number
Type 1	noun	Semantic classifiers of nouns	NULL
Type 2	verb	Inflected forms of declinable words	NULL
Type 3	auxiliary verb	Inflected forms of auxiliary verbs	Ordered Number
Type 4	particle	NULL	Ordered Number
Type 5	symbol	NULL	Ordered Number
Type 6	others	NULL	NULL

Our translation system presupposes that four kinds of information to be used as the connection forms for each word —that is, part-of-speech, noun classifier, inflected form and particle/auxiliary verb/symbol—are generated by morphological analysis. These four kinds of connection forms describe how a word is connected with other kinds of connection forms according to its part-of-speech information, as shown in <Table 4>. Type 1 through 6 indicate lexical information of noun, inflected word, auxiliary verb, particle, symbol and adverb/conjunction/ exclamation, respectively.

In <Table 4>, the first connection forms of each type indicates part-of-speech, and they are linked to the second connection forms, which describes semantic classifiers, inflected forms or NULL according to the first connection forms. The third connection forms are indicated by ordered numbers, if the first connection forms are auxiliary verbs, particles or symbols. Otherwise, they are marked NULL.

3. Translation Using Connection Relations of Two Neighboring Words

A Japanese-to-Korean machine translation system without syntactic and semantic analysis depends on scanty information resulting from morphological analysis. In such a system, the connection information of neighboring words serves greatly to translate source expressions into the target language in a more natural way or ambiguous words in a more optimal way.

3.1 Evaluation Method for Connection Relations of Two Neighboring Words

In order to choose target words by using the connection information of two neighboring words, it is necessary to evaluate which connection information is more appropriate. The evaluation of connection information is carried out by the comparison operator $\varDelta$. The operator $\varDelta$ compares the left-hand side of a word with the right-hand side. It both sides are identical, it returns 1. Otherwise, it returns -1. If both sides are NULL, the operator returns 0 without comparison. The connection information of two neighboring words are represented by the set C={ α', α, β, β'}, and a particular piece of connection information x is x∈C. The operation of $\varDelta$ can be formulated as:

$$x_i \varDelta x_{di} = \begin{cases} 1 \ (x_i = x_{di}) \\ 0 \ (x_i = \mathrm{NULL} \text{ or } x_{di} = \mathrm{NULL}) \\ -1 \ (x_i \neq x_{di}) \\ (\text{where } i=1,\ 2,\ 3) \end{cases}$$

Here, xi is the information of a source word, and xdi is the information of a source word prepared in the translation dictionary. Each connection form is given a weight value for considering its priority. x1 indicates part-of-speech, which is a relatively weak property, thus having the weight of 1. x_2 indicates semantic classifiers or inflected forms, which is a more concrete property than part-of-speech, thus having the weight of 2. x_3 is the ordered number of a word. It represents the word itself, so it has the weight of 3. The evaluation of connection relations is calculated by the following formula.

$$E_x = \begin{cases} -1 (x_i \varDelta x_{di} = -1, \text{ at least one } i) \\ \sum_{i=1}^{3} ((x_i \varDelta x_{di}) * w_i, \text{ otherwise}) \\ (\text{where } w_1=1,\ w_2=2,\ w_3=3) \end{cases}$$

The evaluation value of two neighboring words is the sum of four evaluation values.

$$E_t = \begin{cases} -1 (E_x < 0, \text{ at least one } x) \\ \sum_{x=c} E_x \quad (\text{otherwise}) \\ (\text{where } E_t \text{ is the evaluation value of } t\text{-th connection relation}) \end{cases}$$

If the result of evaluation turns out $E_t = -1$, it mean that the connection is impossible. If it turns out $E_t = 0$, it means that the connection form is null, that is, incomparable. $E_t > 0$ shows the strength of connection relations. That is, the connection relation in which Et has the maximum value is:

$$t^* = \arg\max_{t \in N} E_t$$

(n: number of candidate target words, $N = \{1, 2, \cdots, n\}$)

In terms of this, t^*-th connection information and its target word are chosen. If the same maximum values multiply exist, the system relies on the order of lexical entries registered in the dictionary and thus picks up the target word. The extended translation table method that introduces connection information of two neighboring words has the following characteristics:

- The detailed connection relations of preceding and following words can be described by separately processing attached words (particles, auxiliary verbs, etc.), which is not possible in the traditional translation table method.
- Whereas the number of connection relations and target words is fixed in the translation table of verbs, adjectives, adjectival verbs, etc., connection relations can be described in a more flexible way and the number of connection relations can also be adjusted in the extended translation table. Thus, the independent connection relations can be described for each inflected word.
- The primary target words for inflected words has the effect of eliminating multiply-registered target words.
- The extended translation table is designed to include all of the part-of-speech information in the two dictionaries, that is, the entry table and connection information table. This enables us to build a translation system in a more consistent way. That is, it makes it much easier to build and maintain a translation system.

The two neighboring words surrounding a source word have six possible forms, as shown in <Table 5>. The cases in which considering the first-preceding and first-following words must be considered simultaneously have not appeared in real translation process. If necessary, we can use a different form of connection.

<Table 5> Connection Forms with Two Neighboring Words

Type	Second-Preceding	First-Preceding	Source Word	First-Following	Second-Following
CF0			*W*		
CF1		*α*	*W*		
CF2			*W*	*β*	
CF3		*α*	*W*	*β*	
CF4	*α'*	*α*	*W*		
CF5			*W*	*β*	*β'*

3.2 Example of Translation Process Using Connection Relations of Two Neighboring Words

After the Japanese sentence 「学校に 行って来ました。」 is morphologically analyzed and the appropriate target words are chosen in order from the result of the morphological analysis, as shown in <Table 6>.

First of all, 「学校 (school)」 is translated into the Korean word "Hak-gyo" in terms of one-to-one word correspondence. However, the particle 「に(ni)」 should be translated into "eke(meaning 'to')" when used with human or animal, "(u)reo (semantically corresponding to 'to-infinitive of purpose' in English) when following an inflected word of the Verb-Followed-1 type, or "de(delivering the semantics of causal relations) when following a Noun-Followed form. In other words, the translation system must choose an appropriate target morpheme by using the information of connected words such as 「学校」, 「行っ」, and 「て」 because the possible

<Table 6> List of Translation Information of "学校に 行って来ました。"

Word	Type	Connection Form 1 (part-of-speech)	Connection Form 2 (semantic classifier, inflected form)	Connection Form 3 (ordered number)
学校	Type1	Noun	semantic classifier (location: LOC)	NULL
に	Type4	Particle	NULL	004
行っ	Type2	Fifth-Column Verb	inflected form (verb-followed 2: 5)	NULL
て	Type4	Particle	NULL	022
来	Type2	カ-row irregular verb	inflected form (verb-followed 1: 4)	NULL
まし	Type3	Auxiliary verb	inflected form (verb-followed 1: 4)	012
た	Type3	Auxiliary verb	inflected form (ending: 7)	017
。	Type5	Symbol	NULL	001

target morphemes for 「に」 multiply exist. The first-preceding, first-following and second-following word of 「に」 are shown below. *W* indicates the target word.

<Table 7> Entry Table and Connection Information Table of the Particle "に(ni)"

Base Form	POS	Entry	Word Information		Representative Target Word	Ambiguous?
			Inflected Form	No		
に	Particle	に		004	e	Yes

Entry	POS	Inflected Form	No	CF	Second-preceding word (α')			First-preceding word (α)			(β)			Second-following word (β')			Target Word
					POS	Inflected Form	No	POS	Inflected Form	No	POS	Inflected Form	No	POS	Inflected Form	No	
に	Particle		004	CF1				Noun	HUM								에게
				CF1				Noun	ANI								에게
				CF1				Verb	VF1								(으)러
				CF1				Particle	NF								데

学校(α) に(W) 行っ(β) て(β') 来ました。

In the extended translation table of 「に」, its two neighboring words and multiple target words according to the connection relations are registered. The connection form of the first connection information is CF1, and this leads to the comparison of α_d in the translation dictionary and the information α of its preceding word. In the example sentence, 「学校」 is the first preceding word α, 「行っ」 is the first-following word β, and 「て」 is the second-following word β'. Thus, the connection information α_d (α_{d1}:noun, α_{d2}:HUM α_{d3}:NULL) and the word information of 「学校」 α (α_1:noun, α_2:LOC α_3:NULL) are compared. Since connection form 1 is "α_1:noun" and "α_{d1}:noun" and their part-of-speech is identical, the evaluation value $x_i \varDelta x_{di}$ becomes 1. In the comparison of connection form 2, "α_2:LOC" and "α_{d2}:HUM" are not identical, and this connection is impossible. Thus, -1 is assigned to the evaluation value $x_i \varDelta x_{di}$ and the comparison halts. In order to choose a target word that satisfies the condition which requires 「学校」 to be the first-preceding word, the system tries to compare a series of connection information in the connection information table. Since the connection information that satisfies the condition does not exist in the table, the source word is translated into the primary target word "e" in the entry table. The next source word 「行っ」 has the two neighboring words as shown below, and the evaluation value is computed based on the connection forms in the extended translation table of 「行っ」.

学校(α') に(α) 行っ(W) て(β) 来(β') ました。

The extended translation table of 「行っ」 is given in <Table 8>. Since 「行っ」 has the target words according to eight connection relations, the evaluation value is computed from the first connection information. First of all, the evaluation value is computed by matching β with β_d because the connection form of the first connection information is CF2. From this result, the evaluation value E1 becomes 4 because the identical part-of-speech gets 1, NULL does not count, and the match of the ordered numbers gets 3. The second connection information being CF5, the evaluation value is computed by matching β with β_d and β' with β'_d, respectively.

<Table 8> Entry Table and Connection Information Table of "ITT"

Base Form	POS	Entry	Word Information: Inflected Form	Word Information: No	Representative Target Word	Ambiguous?
行く	Verb	行っ	VF2		ga	Yes

Entry	POS	Inflected Form	No	CF	Second-preceding word (α')			First-preceding word (α)			First-following word (β)			Second-following word (β')			Evaluation Value
					POS	Inflected Form	No	POS	Inflected Form	No	POS	Inflected Form	No	POS	Inflected Form	No	
行っ	Verb	VF2		CF2							Particle	022					1+0+3
				CF5							Particle	022		Particle	NULL	007	-1
				CF5							Particle	022		V-Cont	NULL	NULL	-1
				CF5							Particle	022		Verb	NULL	NULL	1+0+3+1
				CF5							Aux	013		Particle	NULL	011	-1
				CF5							Aux	013		Particle	NULL	013	-1
				CF5							Aux	013		Particle	NULL	015	-1
				CF5							Aux	013		Particle	NULL	016	-1

Although the evaluation value computed by matching β and β_d is 4 as in the case of the first connection information, E_2 also gets -1 and the comparison operator halts because the part-of-speech is not identical when matching β' and β'_d (カ-row irregular verb and particle). The matching process using the third connection information is the same as the second case, and thus E_3 becomes -1. The fourth connection information yields 5 for E_4, which is the result of adding 4 (from matching of β and β_d) and 1 (identical part-of-speech). From the fifth connection information, the first comparison of β and β_d, which is $x_1 \Delta x_{d1}$, shows that their part-of-speech is not identical. Thus, E_5 gets -1 and the matching process halts. Since the sixth through eighth connection information are the same as the fifth, E_6, E_7, and E_8 get -1 and the comparison process halts. Finally, by comparing the evaluation values from the matching processes, the system chooses the fourth target word (t*=4) since E_4 gets the maximum evaluation value. By using the above-described method to choose appropriate target words, the system can generate a more natural-sounding Korean translation (such as "학교에 갔다 왔습니다.").

4. Experimental Results

We have built a prototype translation system to identify the effectiveness of the Japanese-Korean machine translation method using connection relations of two neighboring words. To perform the relative evaluation of the system, we have used the same 474 examples sentences (11,695 words) as used in testing the traditional translation table method. The sample sentences were collected from S&T papers, manuals of machineries, newspaper articles, etc., which contain few grammatical errors [12, 16]. To make our prototype a more realistic application system, it is needed to put a variety of translation information into the dictionary including connection relations of words. However, it would be also possible to extend this kind of translation knowledge by automatically extracting it from parallel Japanese-Korean corpus.

<Table 9> summarizes the result of the experiment

performed with our prototype system. It shows the part-of-speech occurrence frequency and success rates of choosing target words. As can be seen from the table, the success rate of choosing appropriate particles is relatively low, and this result requires the system to classify the semantics of nouns and patterns of verbs in a more sophisticated way. It is also expected that more natural-sounding translations would be possible if the system uses grammatical knowledge and generation rules of Korean that are described taking connection relations into account [25]. When the word order of Japanese-Korean sentences does not match, it is hard to translate them in terms of one-to-one correspondence without losing the original meaning of source sentences. Thus, the system will need to improve translation quality by introducing the translation techniques such as post-processing, word compounding, etc. Nevertheless, our system using the extended translation table method has proven to be effective enough to increase the translation accuracy rate by 4% compared to the traditional translation table method. This also helps to improve the efficiency of the process of post-editing by 16%.

<Table 9> Success Rates Classified by Parts of Speech

Part-of-Speech		Freq.	# of Success	# of Failure	Success Rate
Noun		5063	4871	192	96.2
Verb	Upper-Column	307	261	46	85.0
	Lower-column	121	105	16	86.8
	カ-row	18	16	2	88.9
	サ-row	258	227	31	88.0
	5th Column	439	345	94	78.6
Adjective		120	97	23	80.8
Adjectival Verb		74	57	17	77.0
Noun-Following		124	109	15	87.9
Adverb		192	186	6	96.9
Conjunction		119	112	7	94.1
Auxiliary Verb		439	335	104	76.3
Particle	Case	1885	1328	557	70.5
	Conjunctive	294	260	34	88.4
	Adverbial	431	267	164	61.9
	Ending	36	22	14	61.1
Symbol		1775	1775	0	100.0
Total		11695	10373	1322	88.7

5. Conclusion

In this paper, we have proposed a Japanese-to-Korean translation method that uses the extended translation table, which defines the connection relations of neighboring words. The extended translation table enables us to maintain the consistency of translation processes by rendering the translation table in a generalized structure in which a word is independently defined according to its part-of-speech. Furthermore, by describing the connection relations of particles and auxiliary verbs, the system can use more elaborated connection relations, which had not been possible in the previous methods using a translation table. Since the connection relations can be freely described according to the form of inflected words, the tune-up of the existing translation system to improve translation accuracy can also be easily achieved. Our method also serves to save the size of the translation dictionary because the connection information table contains the entries only for multi-translatable words (Source words that can be translated into representative target words do not need special descriptions.).

Although our system succeeded in simplifying translation processes and improving word and sentence translation accuracy rates by doing away with syntactic and semantic analysis, there still remain the problems with processing of ambiguous auxiliary verbs, appropriate expressions for present/future tense, word order dismatch, etc., which we believe have to be seriously dealt with in the future research.

References

[1] Nomura, H., *Language processing and Machine Translation*, Koudansya, 1991.

[2] Tanaka, H., *Introduction to Analysis of Natural Language*, SanGyodosyo, 1989.

[3] Nomura, H. and H. Tanaka, *Machine Translation* (A Separate Volume of BIT), Koukyou-syuppan, 1988.

[4] Nagao, M., *Machine Translation Summit*, Ohum Sya, 1989.

[5] Nomura, H., *IPAL(Basic Verbs) Dictionary for Computers* (Commented Edition), IPA Technical Center, 1987.

[6] Tanabe, S., *Kuwasii Kokubunnpou*, Bunneitou, 1987.

[7] Choi, K. and N. Nomura, "A Consideration for Development of Korean Translation System under

Development Environment of Japanese Translation System", *NLP SIG Note of IPSJ*, Vol. 86, No. 4, 1988.

[8] Lee E, M Nakajima and A Takeshi, "A Simple Sentence Japanese Korean Machine Translation System Using Relationship with Word to Word by semantic analysis", *Journal of IEICE (D-II)*, Vol. 72, No. 10, 1989.

[9] Lee, H, M Nakajima and A.Takeshi, "A Japanese-Korean Predicate Translation System Using the Semantic Correspondence between Their Auxiliary Predicative Expressions", *Journal of IPSJ*, Vol. 31, 801-809, 1990.

[10] Kim, T. and S. Ura, "A Japanese-Korean Machine Translation based on Conjugated Words analysis", *ICEIC*, 199-203, 1991.

[11] Kim, T. and S. Ura, "Machine Translation Using Similarity between Japanese and Korean", Science & Engineering Research Dept. Technical Report, No 91001, Keio University, 1991.

[12] Kim, T. and S. Ura, "Translation Table Considering Semantic Connection in Japanese-to-Korean Machine Translation", Science & Engineering Research Dept. Technical Report, No 92002, Keio University, 1992.

[13] Kim, T., J. Kim, S. Okoma, and S. Ura, "Morphology Analysis of Japanese on Japanese-Korean Machine Translation by Tables for Translation of Words, *Proceedings of the 43th National Conference of IPSJ*, Vol.3, 201-202, 1991.

[14] Lee, S. and S. Ozawa, "Inflecting Process of Verbs by Phonological Form for Machine Translation System between Korean and Japanese", *Journal of IPSJ*, Vol.33, 1-12, 1992.

[15] Kim J and S. Okoma, "An Appropriate Selection Method for Multi Translatable verbs on Japanese Korean Machine Translation", *Proceedings of the 45th National Conference of IPSJ*, Vol. 3, 97-98, 1992.

[16] Kim, T. and S. Ura, "Generation of Korean Based on Connection Forms of the Semantics in Japanese-Korean Machine Translation", *Journal of IPSJ*, Vol. 33, No. 12, 1992.

[17] Kim, T. and S. Ura, "A Processing of Negative Sentences in Japanese Korean Machine Translation", *Journal of IPSJ*, Vol. 34, No. 5, 1993.

[18] Kim, J., T. Kim and S. Okoma, "Processing of Spoken Language in Japanese-Korean Machine Tranlation, *Proceedings of the 47th National Conference of IPSJ*, Vol. 3, 179-180, 1993.

[19] Hwang, D. and M. Nagao, "Construction of a thesaurus for Korean from a thesaurus for Japanese", *NLP SIG Note of IPSJ*, No. 094, 1993.

[20] Kim, J., T. Kim and S. Okoma, "A Processing of Polysemy and Multi-Translatable Verbs on Japanese Korean Machine translation", *Proceedings of ICCTA94*, 144-148, 1994.

[21] Kim, J., T. Kim and S. Okoma, "Japanese Korean Machine Translation by Using Extended Translation Table", *Proceedings of the 51th National Conference of IPSJ*, Vol. 3, 89-90, 1995.

[22] Miyazaki, M., S. Shirai and S. Ikehara, "A Japanese Syntactic Category System Based on the Constructive Process Theory and its Use", *Journal of Natural Language Processing*, Vol. 2, No. 3, 2-36, 1995.

[23] Kim, J. and S. Okoma, "A Method of Generating Korean Language by Using Extended Translation Tables on Japaese to Korean Machine Translation", *Journal of IPSJ*, Vol. 37, No. 9, 1697-1707, 1997.

[24] Kim, J., J. Lee and K. Lee, "Generation of Korean Predicates Based on Modality-Feature Ordering and Lexicalizing Table in Japanese-Korean Machine Translation", *Journal of Natural Language Processing*, Vol. 5, No. 2, 3-24, 1998.

[25] Kim, J., J. Lee, K. Moon and K. Yang, "Lexical Disambiguation Using Conjunctive Patterns of Words in Japanese-Korean Machine Translation", Journal of Korean Information Science Society (B), Vol. 25, No. 8, 1270-1280, 1998.

접 수	2004년 4월 12일
게재승인	2004년 5월 20일