

# Japanese-Chinese Machine Translation for Japanese *Suru* Expressions

Yiou Wang

School of Engineering, Gifu University  
[y\\_wang@ikd.info.gifu-u.ac.jp](mailto:y_wang@ikd.info.gifu-u.ac.jp)

Takashi Ikeda

Faculty of Engineering, Gifu University.  
[ikeda@info.gifu-u.ac.jp](mailto:ikeda@info.gifu-u.ac.jp)

## 1 Introduction

Although Japanese and Chinese both use Chinese characters, they belong to different language families. Chinese is an isolating language without inflexion, while Japanese is an agglutinative language. In the study of Japanese-Chinese machine translation, there are many unresolved problems and underdeveloped areas; one of these is the translation of *suru* expressions.

*Suru*, which is one of the most frequently used Japanese words, is an irregular verb with extraordinarily rich usage. Given the diversity and the complexity of its usage, many syntactic studies concerning the verb *suru* have been conducted. However, few researchers have discussed this from the viewpoint of machine translation, and as yet there have been no studies on machine translation of *suru* from Japanese to Chinese. Because of the high occurrence frequency of *suru*, the quality of its translation has a great impact on the overall Japanese-Chinese MT system. Such factors as numerous semantic variants and the syntactic complexity of corresponding Chinese expressions of *suru* constructions put great difficulties in the Japanese-Chinese translation. It is very necessary and important to make full analysis of it. In Japanese-Chinese translation, *suru* can be rendered as [是 *shi* (to be); 有 *you* (to have); 将要 *jiangyao* (to be about to); 做 *zuo* (to do), etc.]. However, because current commercially available translation software does not have a sufficiently deep analysis of these translation ambiguities, many mistranslations occur in areas such as word selection and word order determination.

In this paper, we study the usage of the verb *suru* systematically, analyze complicated correspondences with Chinese, propose translation rules and discuss the application of this analysis to the real MT system.

## 2 A solution for translating *suru* expressions

There are many sentence structures in Japanese that utilize the verb *suru*. We conducted a statistical investigation of more than 20,000 Japanese sentences in a Japanese-English parallel corpus (Murakami, 2002), and found 2,899 sentences containing *suru* (nearly 15%). Among these 2,899 *suru* constructions, 1,007 are

instances of *suru* used in combination with a noun in order to convert it to a verb (this is called a verbal noun); it is also used to verbify other parts of speech, such as adjectives, verbal adjectives and adverbs. The corresponding Chinese expressions are also diversified: constructions using *suru* can be translated into parts of speech such as verbs, adverbs, and conjunctions, as well as causative sentences, passive sentences, and so on.

We extracted and manually translated 500 sentences containing the structure [*~suru*] from the corpus as investigating materials. We classified the *suru* sentences into 21 types from the viewpoint of semantic usage. Based on the semantic analysis, we compared the [*~suru*] structures with the corresponding Chinese expressions from the viewpoint of syntactical features (sentence patterns, tense, aspect and so on) as well as semantic features, taking into account limitations in the possible Chinese combinations, and used these comparisons to develop translation rules for the conversion of Japanese *suru* expressions into Chinese. which can be processed by *jaw/Chinese* (Section 3) ( see Table 1 ) .

Table 1. Translation rules for *suru* expressions (Partial)

	Conditions and attributions					Translation rules
	Japanese Pattern	Attribution of N1	Attribution of N2	Attribution of N3	Structure and constituent of X	
(1)	N1 <i>ga</i> N2 <i>ga suru</i>	subject	feeling			N1 感觉 N2
(2)	N1 <i>ga</i> N2 <i>wo</i> N3 <i>ni suru</i>	Noun, pronoun	Descendant, offspring	occupation		N1 要把 N2 培养成为 N3
(3)	N1 <i>ga</i> X <i>koto ni suru</i>	Human being			verb-equivalent words	N1 决定 X

### 3 The comparison with Japanese -English patterns

*Goi-Taikei--A Japanese Lexicon* (NTT 1997) is a translation pattern dictionary for Japanese-English translation. We made comparisons of the Japanese-Chinese translation patterns for “*suru*” expressions mentioned (92 patterns)in section 2 with the Japanese-English translation patterns(319 patterns) listed in *Goi-Taikei--A Japanese Lexicon* .We analyze the differences and summarize the reasons for them.

### 4 Outline of the *jaw/Chinese* MT system

*Jaw/Chinese* is a machine translation system which translates from Japanese to Chinese based on a pattern transfer paradigm. *Jaw* (from Japanese to Asian and World languages) is the translation engine, which is applicable to any target language. Figure 1 shows a rough outline of *jaw/Chinese*, which contains three sub-processes: parsing a Japanese sentence, transferring it to a Chinese expression structure, and then generating a Chinese sentence.

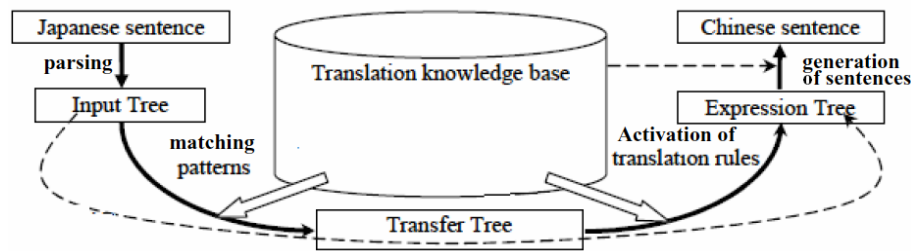


Figure 1. Outline of jaw/Chinese

## 5 Experiments and Evaluation

We first conducted manual experiments using 200 sentences to evaluate these rules; furthermore implemented our translation rules for *suru* expressions into jaw/Chinese and verified their applicability in real MT system by experiments using 100 sentences. We evaluated the translation results manually, focusing on the *suru* expressions and judging for each individual sentence whether a suitable word for *suru* was used in the translation and whether the word was placed in the correct order. Each translated sentence was graded as follows:

○ = In terms of grammar, as accurate as natural Chinese translation.

△ = A little unnatural grammatically, but the overall meaning of the sentence is correct.

× = Incorrect both grammatically and in terms of overall meaning.

Evaluation of the results was carried out using two alternative criteria: (1) ○ is counted as a correct translation and (2) ○ and △ are counted as correct translations. The results are shown in Table 2. From the results, we can conclude that our method is valid and could achieve a rather high accuracy, compared with commercially available MT software.

Table 2. Evaluation of experiments

	Evaluated sentences	Correct (○)		Correct rate (○)		Correct (○ and △)		Correct rate (○ and △)	
		A	B	A	B	A	B	A	B
		manual experiments	200	162	82	81%	41%	164	102
Experiments by jaw/Chinese	100	80	40	80%	40%	82	48	82%	48%

(A: translation using our rules; B: translation by commercially available MT software)

## References

- Murakami. Situation of Japanese-English parallel corpus database.(2002) The seventh annual seminar of "Language, recognition, and expression"  
 NTT Communication Science Basic Research. (1997). *Goi-Taikei--A Japanese Lexico* Iwanami Publisher.