1T-05 Generating of Sinhalese Expressions in Japanese- Sinhalese Machine Translation System

Samantha Thelijjagoda, Nayana Elikewala, Jun Shieh, Yasuaki Hyodo and Takashi Ikeda

Email: samanta,nayana,shieh,hyodo,ikeda(@ikd.info.gifu-u.ac.jp)

1. Introduction

This paper presents an introductory step on Japanese – Sinhalese machine translation system, especially on generating phase of Sinhaleses. With the pattern transfer method and object oriented method, we create Sinhalese expression objects and linearize them into Sinhalese sentences.

2. Sinhalese language and it's relation with Japanese language.

Both of these languages are basically agglutinative. A Japanese sentence can be analyzed morphologically into one or more Bunsetsu, a linguistic unit, and it is same for the Sinhalese language. Unlike English, their grammatical word order is relatively free and the ellipsis are widely accepted. Sinhalese and Japanese are thus similar, and in many cases there are one to one correspondences between Japanese Bunsetsu and Sinhalese Bunsetsu-like unit independent in context. But this is not common for all the cases. There are often ambiguous cases in which the correspondence depends on the context. To overcome this problem, we designed a pattern based MT system.

3. MT system based on pattern transferring.

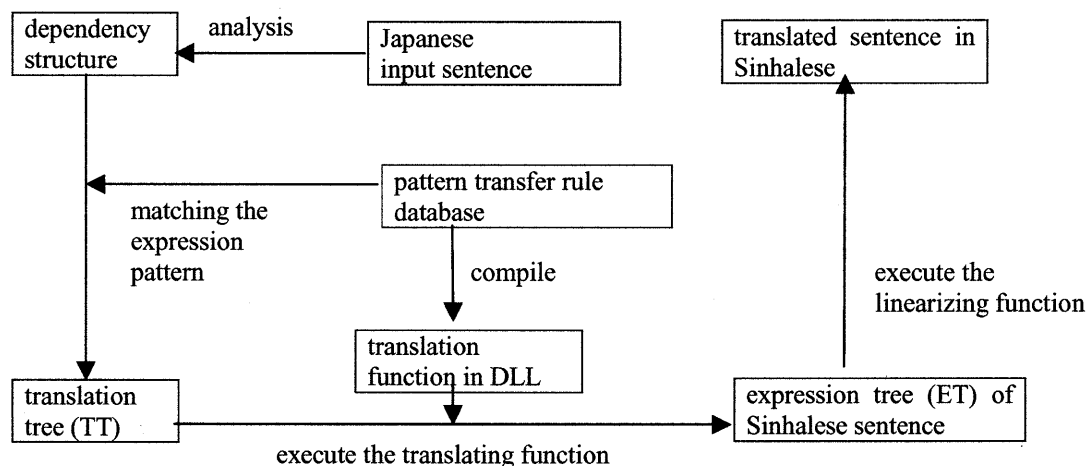The following processing chart shows the stream of our MT system.

Figure 1: The stream of the MT system

The analysis result of Japanese is matched to the condition of expression pattern in pattern transfer database and produce the TT(translation tree). The pattern types and identity numbers of pattern transfer rule are kept in the translation tree and the system calls the translation rule (which is beforehand compiled into the translation function in the DLL) to produce the ET(expression tree). Executing of the linearizing functions in the ET produces translated Sinhalese sentence.

4.Generating Sinhalese sentences

Table 1 is pattern transfer rule table for a simple example which consists the Japanese patterns and the Sinhalese translation rule with it's class name.
Figure 2 shows the translation tree (TT) for this example.

Table 1: Pattern transfer rule table (The input: 「白い熊が道をゆっくり歩いていた。」)

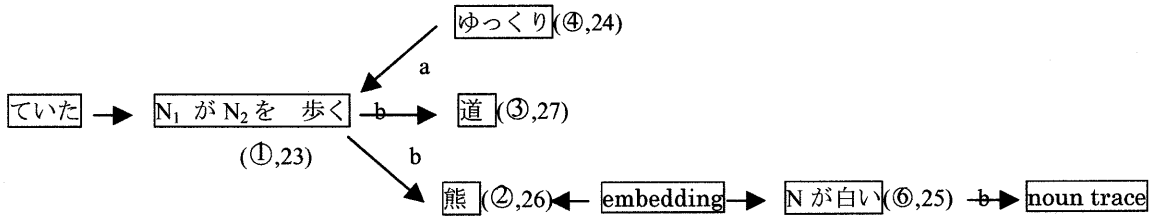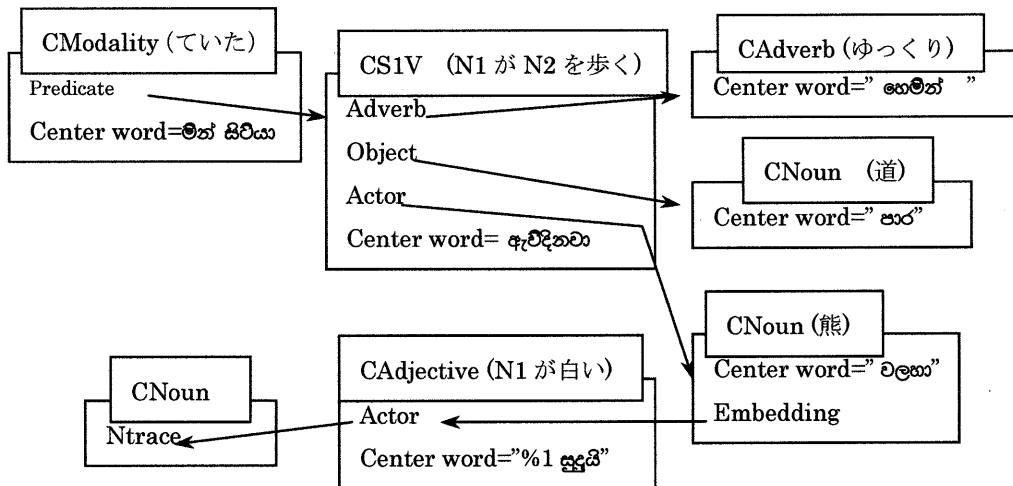| Pid | Bid | pType | Modifier | | Key word | Modifiee | | Class | Transfer |
			TypW	Cond.		TypW	Cond.		Rule
23	1	base	熊(N1)	が					
23	2		道(N2)	を					
23	3				歩く			CS1V	ඇවිදිනවා %1 0 %2 එහි
24	1	addition			ゆっくり	verb		CAdverb	හෙමින්
25	1		N	が					
25	2	base			白い	noun		CAdjective	සුදු
26	1	base			熊			CNoun	වලහා
27	1	base			道			CNoun	පාර
28	1	modality			ていた			CModality	මින් සිටියා

Figure 2: Translating tree

Figure 3: Expression tree

The Figure 3 shows the expression tree for the same example. Each box shows the object of each class and relevant class members. By executing the linearizing function of the most upper box, recursively calling the linearizing function of the following boxes, the system generate the Sinhalese sentence.

"සුදු වලහා පාරෙහි හෙමින් ඇවිදිමින් සිටියා"   (SuDu   WaLaHa   PaReHi   HeMiN   AWiDiMiN SiTiYa)

5.Conclusion

In this paper we described the implementation of a pattern based MT system which generates Sinhalese expression. We are planning to expand the system to realize the larger domains of this languages.