E-46 Sinhalese Generation Rule and Machine Translation System j-aw/Sinhalese

Samantha Thelijjagoda*, Yoshimasa Imai*, Nayana Elikewala* and Takashi Ikeda*

1.Introduction

This paper presents about Sinhalese generation rules used for simple translations from Japanese language and the Sinhalese generation function including Sinhalese linking rules which is used to combine Sinhalese translated components. For the carrying out of above procedure, we designed a pilot model called j-aw/Sinhalese.

2. Sinhalese Structure

A Japanese sentence can be analyzed morphologically into one or more Bunsetsu and it is same for the Sinhalese language as they are agglutinative. Sinhalese and Japanese are thus similar, in many cases there are one to one correspondences between Japanese Bunsetsu and Sinhalese Bunsetsu-like unit independent in context. But at the same time there are many ambiguities on the correspondences. For examples, Japanese function words like が, を, は, に, の have many translation in Sinhalese. Table 1 and 2 shows some examples for them.

Japanese	case	Sinhalese	case
東京の叔父	の	Tokyo **wala** mama	wala
叔父の家	の	Mamage gedara	ge
服の色	の	Andumehi paata	ehi
家の庭	の	Gedara midula	∅
研究室の本	の	Labeke pota	eke
母のために	の	Amma wenuwen	∅

Table 1

Japanese	case	Sinhalese	case
日本に雪が降る	に	Japanayata hima wete	ta
岐阜に雪が降る	に	Gifu **walata** hima wete	walata
太陽は西に沈む	に	Ira batahirin basinawa	in
ご飯を食べる	を	Bath kanawa	∅
本を読む	を	Potak kiyawanawa	ak
犬を飼う	を	Ballek hadanawa	ek

Table 2

Not like Japanese, written Sinhalese verbs have to agree with the subject in number, gender and person, but in spoken Sinhalese, verbs in dictionary form are applied.

3. Transfer rule for Japanese to Sinhalese by j-aw

To disambiguate above one to many correspondence, j-aw/Sinhalese develops transfer rules from Japanese pattern to Sinhalese. The conditions for Japanese patterns give a way to these disambiguations, as shown in table 3, 4, 5 and 6.

ID	conW	funcW	keyW
1	N_1	は	
2	N_2	から	
3			出る

Table 3

ID	Modifier conW	keyW	Modifiee
1	N_1	の	
2			N_2

Table 4

ID	Modifier	keyW	Modifiee
1	V_1	て	
2			V_2

Table 5

As examples, the rules for above patterns with some conditions correspond to Sinhalese as in table 6. We currently use the base type rule and two addition type rules denoted by b-rule , a-ruleFw, a-ruleCw respectively.

Pattern	patternID	Rule type	Class	MemberName	MemberClass	Value	RoleName	cMarker
N_1 は N_2 から出る	1002	b-rule	CProposition	m_subject	CNoun	1		
	1002		CProposition	m_nounModifier	CNoun	2	place	ඉන්
	1002		CProposition	m_centerW	CString	බැහැරවෙ		
N_1 の N_2	1004	a-ruleFw	CNoun	m_nLinker	CNoun			
	1004		CNoun	m_caseMarker	CString	ගේ		
V_1 て V_2	1003	a-ruleCw	CpConnection	m_pSubordinate	CProposition	1		
	1003		CpConnection	m_connect	CString	ලා		

Table 6

The base type is for basic expression pattern for content word and the addition types are for additional expressions to the basic expression pattern.

* Department of Information Science, Faculty of Engineering, Gifu University.
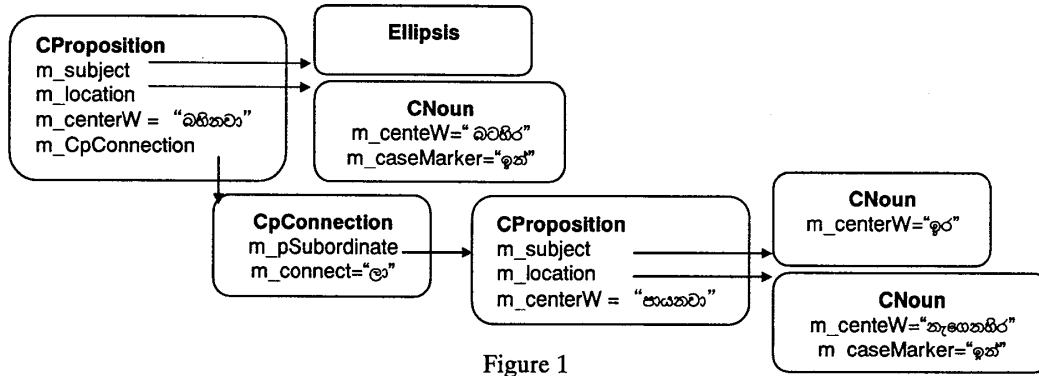
4. Generation Rule for Sinhalese

Figure 1

According to the above transfer rules, the expression tree is formed as the illustration of figure 1, and their linearizing function of each class generates the linear text When a Sinhalese sentence is generated, the components such as subject, object, adverb and noun modifiers with their roles and case markers are employed in the linearizing function as shown below.

Formation of linearizing function

[subject]+[time] + [numerical] + [time_era] + [TimeBegin] + [TimeLimit] + [material] + [deadline] + [object1] + [object2] + [direction] + [location1] +[location2] + [quantity] + [comparative] + [degree] + [purpose] + [joint-action] + [dependency] + [adverb] + [verb].

When we generate Sinhalese, it is necessary to change the form of verbs and nouns according to the gender number, person and tense. To change the verb, we refer verb forms in a database to form a meaningful accurate output. For example, take a verb of above Japanese sentence 太陽は東から出て西に沈む. For the Japanese 'て'form, here, for "出て", we had to pick-up the Sinhalese "て-like" form from a database where the all regular and irregular Sinhalese verbs are kept. A part of table is looks like table 7.

English	Present	Past	て-like	Progress	Ad-prsnt	And-past	Fml-rqst	Fml-cmd	Nounfrm	Agree
go	යනවා	ගියා	ගිහින්	යමින්	යන	ගිය	යන්න	යනු	යාම	යමු
take	ගන්නවා	ගත්තා	අරන්	ගමින්	ගන්න	ගත්	ගන්න	ගනු	ගැනීම	ගමු
give	දෙනවා	දුන්නා	දීලා	දෙමින්	දෙන	දුන්	දෙන්න	දෙනු	දීම	දෙමු

Table 7

The linking system is operated with regarding a component in a sentence to joint with it's case marker, the ending letter of the nouns and pronouns are quite different on the cases such as case marker is different. The suitable linking rules have been prepared and inserted to the system to get accurate linked words. For example, lets see linking method for "西に" in figure 2.
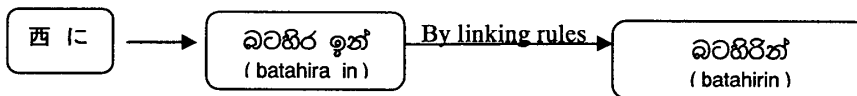
Example:

Figure 2

5. Examples of translation

Some example outputs generated on Japanese input tree.

太陽は東から出て西に沈む

仙台は日本の北部にある

北国は雪がたくさん降る

6. Future problems

Sinhalese verbs attribute distribution into Japanese verbs, the Sinhalese negation, the modality, method to identify the gender, number and person in a Japanese bunsetsu, and complex translations with paragraph are future problems to be considered.