

Generation of Sinhalese Units From Japanese Bunsetsu Structure

6 P-2

A.Herath*, Y.Hyodo*, T.Ikeda*, S.Herath**
 *Gifu University, **Aizu University

1. Introduction

Sinhalese, the major language of Shri Lanka, spoken by 16 million people, is more than two thousand years old and akin to Northern Indian languages. Sanskrit, one of the world existing oldest languages originated ten centuries before Christ, is the mother languages of Sinhalese. Japanese, an agglutinate language, is a member of Mongolian family of languages.

Though Sinhalese and Japanese are in different language families, there are some similarities. Both languages belong to S + O + V group, allow ellipses even in S and O, and have a more flexible word order than English like languages. Since Sinhalese word order is more flexible than Japanese, Japanese sentences can be translated in to Sinhalese in the same order, without affecting the meaning of the sentence.

This paper describes how Japanese bunsetsu structure can be mapped to Sinhalese unit structures.

2. Japanese Bunsetsu and Sinhalese Units

Japanese bunsetsu is a linguistic unit and consists of a content word with one or more functions words.

Sinhalese sentence consists of units. The unit is a component of the sentence written separately by space [1]. Japanese is different in this aspect since there is no segmentation between words.

A Japanese bunsetsu can be mapped to one or more units in a Sinhalese sentence.

犬が/食べています。
 {The dog is eating (something)}
 balla kamin sityi.

In the above example, a Japanese bunsetsu "犬が" is mapped to one sinhalese unit "balla", and a Japanese bunsetsu "食べています" is mapped to two sinhalese units "kamin sityi".

Sinhalese unit consists of a root with or without suffix/suffixes [1]. When forming a unit by adding suffixes to a root, changes can occur in the spelling. For example, "balla" is originated "balu" (root form) + "A*" (suffix). This change can be formalized as linking rule [1].

3. Mapping Rule of Verb Bunsetsu

Sinhalese verbs have six attributes; tense, person, gender, number, voice and causative [1]. Tense has present, past, and future; person has first, second and third; gender has masculine, feminine and neuter; number has singular and plural; voice has active and passive; and causative has causative and non causative. Japanese verbs do not have number, gender and person attributes as in Sinhalese.

In general Sinhalese verb part can be analyzed as follows.

- (1) V%ES
- (2) V%CS + OV%ES
- (3) V%CS%CS%... + OV%ES
- (4) V%EW
- (5) V%CS + EW

where V is a form of verb after some linking changes. ES is an ending suffix, CS is a combination suffix, OV is an optional verb and EW is an ending word. The notation "%" represents a non-space and "+" represents a space.

In the example of "kamin sityi", ka=V, min=CS, siti=OV

Suffix	Person	Gender	Number	Tense
mi/mu	first	-	sin/plu	pres
hi/hu	second	mascu	sin/plu	pres
ya/ya	third	femin	sin/plu	pres
yi/thi	third	mascu	sin/plu	pres
yemi/mu	first	-	sin/plu	past
yehi/yehu	second	mascu	sin/plu	past
yaya/yohu	third	mascu	sin/plu	past
yoya	third	mascu	plu	past
nnemi/nnemu	first	-	sin/plu	future
nnehi/nnehu	second	mascu	sin/plu	future

[Table.1] Some forms of ending suffix

Generation of Sinhalese Units From Japanese Bunsetsu Structure
 A.Herath*, Y.Hyodo*, T.Ikeda*, S.Herath**
 *Gifu University, 1-1 Yanagido, Gifu, 501-11, Japan
 **Aizu University, Aizu-Wakamatu, Fukushima, 965, Japan

and yi=ES. Ending suffixes varies depending on person, number, gender of the subject and tense, voice and causative of the verb. [Table 1] gives some of them.

We formalized the rules to map Japanese bunsetsu into Sinhalese units relating to Japanese function words using ES, CS, OV and EW. Some of them are shown in the [Table 2].

A Japanese verb bunsetsu which contains more than one function words can be mapped by chaining of these rules as the following example.

" 食べさせている (食べる%させる%ている) "
" kawamin sitiya (ka%wa%min+siti%[ES]) "

Japanese	Sinhalese	
tabete iru	kamin sitimi	V%min+siti%ES
oite aru	taba eta	V+EW
gakuseidearu	shisshsayekimi	N%ES
taberu	kami	V%ES
tabe rareru	kanna piliwan	V%nna+EW
tabe saseru	kawami	V%wa%ES
tabe nai	no kami	no+V%ES
tabeta	kewa	V%wa
tabetai	kanna ona	V%nna+EW
iku yoda	yanawalu	V+EW
iku daro	yawi	V%ES

[Table.2] Mapping rule of Verb Bunsetsu

Japanese	Sinhalese	Example
N%ga	N	inu-ga balla
N%wa	N	inu-wa balla
N%no	N%ge	inu-no ballage
N%ni	N%ta	inu-ni ballata
	N%wata	Tokyo-ni Tokyowata
N%to	N%th	inu-to ballath
	N+samaga	balla samaga
N%yori	Nta+ wada	inu-yori ballata wada
N%kara	N%gen	inu-kara ballagen
	N+sita	5ji-kara pahe sita
N%dake	N+pamanak	inu-dake balla pamanak
N%demo	N+wunath	inu-demo balla wunath
N%he	N%ata	gakko-he pasalata
N%de	N%ayen	basu-de basayen
	N%walin	ginza-de ginzawalin
	N%nisa	ame-de wessanisa
N%mo	N%th	inu-mo ballath
N%shika	N+hera	inu-sika balla hera
N%dake	N+pamanak	inu-dake balla pamanak

[Table.3] Mapping rule of noun bunsetsu

4. Mapping Rule of Noun Bunsetsu

Sinhalese noun unit has five attributes; gender, number, person, definiteness and case. Japanese noun bunsetsu can be mapped into Sinhalese in the following form.

- (1) N
- (2) NR%EW
- (3) NR%NS
- (4) N+CW

Where N is a noun, NR is a root form of noun, EW is an ending word, NS is a noun suffix and CW is a combining word.

[Table 3] shows some of the mapping rules of noun bunsetsu.

5. Mapping Rule of Complex Sentences

Japanese complex sentences consists of words such as "de", "kara", "nagara" etc. Sinhalese has corresponding words such as "wunath", "nisa" and "min" etc. Therefore, the Japanese complex sentences can be mapped to Sinhalese complex sentences by direct mapping as following example.

りんごは果物です。 apple palaturaki.

トマトは野菜です。 takkali elawaluwaki.

りんごは果物で、トマトは野菜です。

apple/ palaturak wunath/ takkali/ elawaluwaki.

6. Discussion

We formalized the rules for mapping Japanese bunsetsu into Sinhalese units. Sinhalese verbs and nouns have more attributes than Japanese and the attributes of verb should agree with the subject.

If the necessary information for the Sinhalese verbs (deep case) can be obtained from the Japanese analysis, then the Japanese sentences can be mapped to Sinhalese directly. Various basic Japanese sentences in [2] have been analysed and successfully translated into Sinhalese by applying the above rules.

References

- [1] S.Herath, T.Ikeda, S.Ishizaki, Y.Anzai, A.Aiso, Analysis System for Sinhalese Unit structure, Journal of Experimental and Theoretical Artificial Intelligence, 1(1992)29-48
- [2] Jido-gengo-kenkyuukai, Tanoshii Nihongo no bun sho, Ikkosha