

## RESEARCH ARTICLE

# Japanese-Sinhalese\*\* machine translation system *Jaw/Sinhalese*

Samantha Thelijagoda\*, Yoshimasa Imai and Takashi Ikeda

Department of Information Science, Graduate School of Engineering, Gifu University, Gifu, Japan.

Revised: 05 October 2006 ; Accepted: 16 March 2007

**Abstract:** This paper describes a machine translation system, *Jaw/Sinhalese*, that translates Japanese into Sinhalese (Sinhala). This is the first Japanese-Sinhalese machine translation system. Both Japanese and Sinhalese are agglutinative languages. The Japanese language contains *bunsetsu* (Japanese basic linguistic units), which consist of a content word (a lexical root) with one or more function words. Sinhalese also has *bunsetsu*-like units, but the grammatical structure is not necessarily fully revealed. The paper proposes a method of analysis for the *bunsetsu*-structure of Sinhalese which shares features with Japanese, and translation solutions for Japanese function words after predicates and nouns. Function words after predicates express tense, modality, conjunctions, etc., while function words after nouns express the case, topic, thematic roles, etc. This paper clarifies the *bunsetsu*-structure of Sinhalese. It also distinguishes case marker (a leading group of function words after nouns) correspondences using three types of pattern-based translation rules and solves the multi-layered problem of the translation of function words after a predicate by means of translation rules in a table format. Translations were implemented on the pilot machine translation system, *Jaw/Sinhalese*. As an experiment, 200 sample sentences were evaluated. The results (72% rate of success) indicated that this approach is within an acceptable accuracy range.

**Key words:** *Bunsetsu*, case marker, expression elements, function word after predicate, *Japanese-Sinhalese* machine translations, Sinhalese *bunsetsu*, Sinhalese grammatical structure, transfer rules

## INTRODUCTION

Machine translation (MT) developments in Japan over the past decade have increased extensively. Most projects consider translations between Japanese and English. Other Asian languages such as Chinese and Korean have also been considered, but languages like Sinhalese (Sinhala) have received less attention. *Jaw/Sinhalese* has

been developed as a pilot machine translation system to translate Japanese into Sinhalese. This is the first attempt at a MT system for this language pair.

Japanese and Sinhalese are typologically classified as Subject Object Verb (SOV) languages and both are agglutinative languages. (In agglutinative languages some function words agglutinate to a content word. Function words after the noun, work as case markers, topic markers, etc, and function words after the verb, express tense, modality, etc.) Japanese has *bunsetsu* (Japanese basic linguistic units) that consists of a content word with one or more function words. The function words of *bunsetsu* may be divided into two types. The first type includes the function words that come after a predicate, which express grammatical categories like tense-aspect, modality, conjunctions, etc. The second type includes the function words that come after nouns, which express grammatical categories such as case, topic, focus, etc. Sinhalese also has *bunsetsu*-like units, and for the ease of comparison, they will be referred to as Sinhalese *bunsetsu*.

In this paper, a method is proposed to analyze the Sinhalese *bunsetsu* structure for machine translations. More than 20 kinds of verb inflection words (base parts of the verb *bunsetsu*) are defined and a set of function words that follows them is reorganized. Likewise, for the case inflection word of a noun, a new organization of function words is proposed. The case inflection is formed by a combination of a noun stem and a function word for case marking. 27 function words are isolated for case marking.

Since Sinhalese and Japanese are very similar in this respect, in many cases there are one-to-one correspondences between Japanese *bunsetsu* and Sinhalese *bunsetsu*. But at the same time, these correspondences are

\* Corresponding author

\*\* Currently, the term 'Sinhala' is in use

not perfect. The translation of Japanese case markers (a leading group of function words after a noun) into Sinhalese is ambiguous; that is, one Japanese case marker often corresponds to several Sinhalese case markers. For example, the Japanese case marker “を” “wo”, which is usually used to indicate the object case, corresponds to 16 case markers in Sinhalese {ට (*ta*), කට (*kata*), වලට (*valata*), ව (*va*), ක්ව (*kva*), ත්ව (*nwa*), එන් (*en*), කින් (*kin*), වලින් (*valin*), අක් (*ak*), එක් (*ek*), ඕ (*o*), එහි (*ehi*), ක (*ka*), වල (*wala*), ∅}

One Japanese function word after a predicate often affects more than one expression element in a Sinhalese sentence. For example, the Japanese function word “[食べ]たい” “[*tabe*]tai”, which expresses “hope/eagerness” on the part of the subject, is translated into “ට(*ta*)”, “අක් (*ak*)”, “කන්න (*kanna*)”, and “අවශ්‍යයි(*awashyayai*)” (after a subject, after an object, after an inflected verb, and after a verb) in a Sinhalese sentence. This problem is called a multi-layered problem of translation and is solved by describing the translation rules in a table format.

The first stage of the *Jaw/Sinhalese* system processes the translations of propositional content. The clarification of case marker correspondence is handled in this stage. The second stage covers the translation of tense and modality, which are expressed by function words after the predicate in Japanese. The multi-layered problem of the translation of Japanese function words after a predicate is resolved at this stage. To estimate the accuracy of these rules, 200 Japanese sentences were evaluated. The results (72% rate of success) show this approach is effective and acceptable within machine translations.

## METHODS AND MATERIALS

### The translation system *Jaw/Sinhalese*

*The outline of Jaw/Sinhalese:* The *Jaw/Sinhalese* system has been developed as a pilot model for Japanese-to-Sinhalese machine translations. *Jaw* is the translation engine from Japanese to other languages<sup>1</sup> (Japanese to Asian and World languages). The other translation systems that adopt *Jaw* at present are *Jaw/Chinese*, *Jaw/Vietnamese*, *Jaw/Myanmarese*, and *Jaw/SL* (sign language). The method of generating the target language in each system is different. The *Jaw/Sinhalese* system is currently in its basic developmental stage. Figure 1 illustrates the flow of *Jaw/Sinhalese*. After analyzing a Japanese sentence by means of *IBUKI* (a system developed in the Ikeda laboratory at Gifu University, Japan for segmentation of Japanese sentence into *bunsetsu*), the translation engine *Jaw* puts them into a tree (input tree: IT) (Figure 2.) The IT will be explained with the help of the Japanese sentence J1 given below. Then *Jaw* searches Japanese patterns in the transfer dictionary for the IT and makes a tree of transfer rules (transfer tree: TT) (Figure 3).

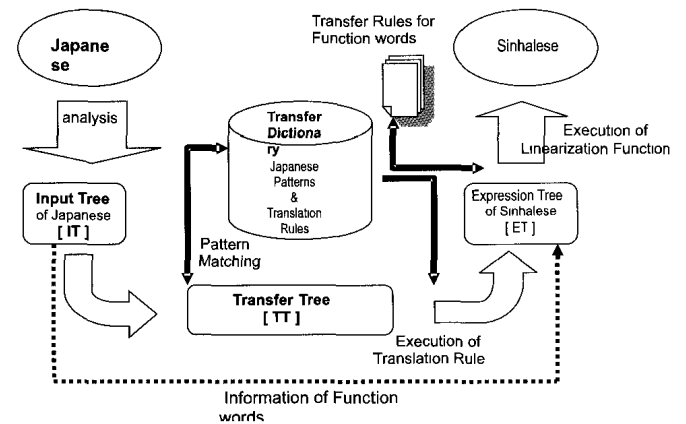
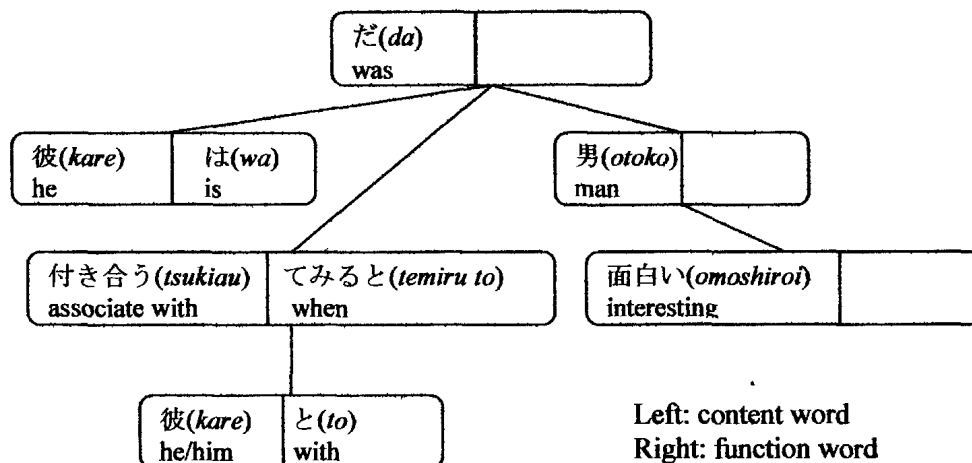


Figure 1: Outline of *Jaw/Sinhalese* system



Left: content word  
Right: function word

Figure 2: Input Tree (IT)

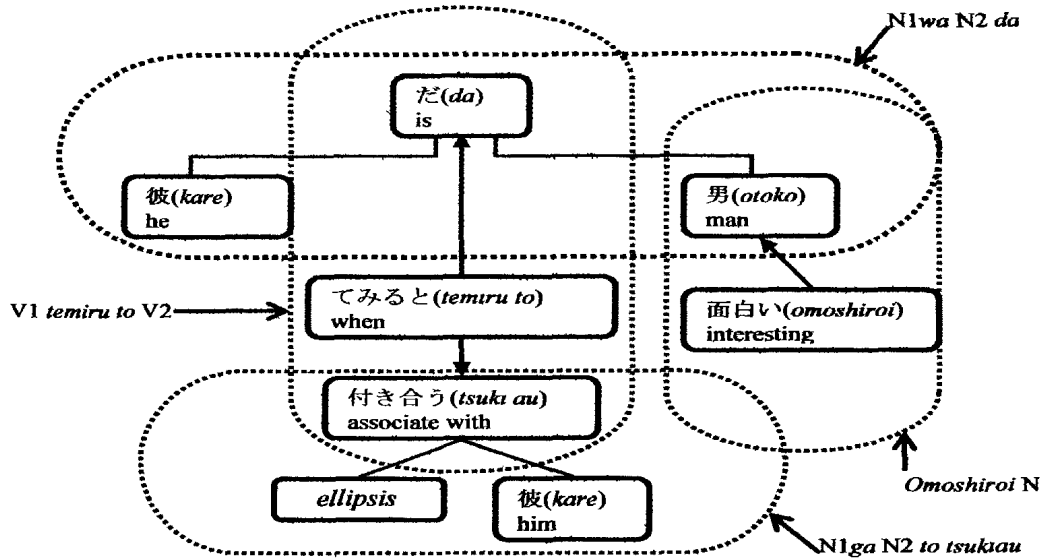


Figure 3: Translation Tree (TT)

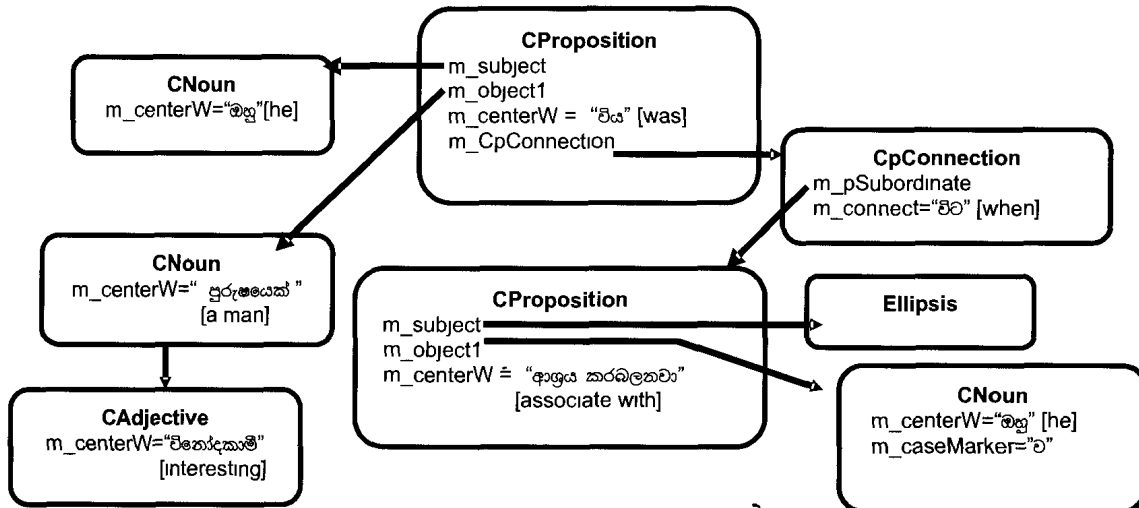


Figure 4: Expression tree  
(Each box shows C++ objects with members)

Subject + [After-Sub] + time + numerical + time\_era + time-begin + time-limit + material + deadline + Object1 + [After-Obj] + Object2 + direction + location1 + location2 + quantity + comparative + degree + purpose + joint-action + dependency + Adverb + [Before-Vb] + Verb + [After-Verb] + [Judge-Verb]

Figure 5: Formation of the linearizing function for the generation of Sinhalese

J1 :彼と付き合ってみると(彼は)面白い男だった。  
 J1 :kare to tsukiatte miruto (kare wa) omoshiroi otoko datta.  
 E1 :After (I) associated with him, ([I found that] he) was an interesting man.  
 S1 :ඔහුට ආශ්‍රයකර බැලුවේ (ඔහු) විනෝදකාමී පුරුෂයෙක් විය.

The expression tree and generation of Sinhalese:  
 The system is implemented in C++ and the transfer rule is, in fact, a program stored as a dynamic link library (DLL). The execution of the transfer rules in the transfer tree (TT) produces a network for Sinhalese expressions (expression tree: ET) (Figure 4). A linearizing function is defined for each object as a class method of C++. The execution of the linearizing function on the ET orders the members of the ET to make a Sinhalese output sentence.

In the case of translating Japanese function words after predicates, the information of the function word is directly provided to the ET from the IT, and this activates the transfer rules for function word translations while executing the linearizing function

Every object of the ET has its own generation procedure. The linearizing function of each object linearizes the components, which are represented as members of the C++ objects such as subject, object 1, object 2, adverb and verb. Besides them, noun modifiers such as *tune*, *numerical*, *tune-era*, *tune-begin*, *tune-end*, and the expression components for function words after predicates such as After-Subj, After-Obj, Before-Vb, and After-Vb perform a major role in the linearizing function. The order of the linearization is programmed according to Sinhalese word order. The arrangement of all components for the generation of Sinhalese is shown in Figure 5

*Jaw/Sinhalese* is a pattern-based machine translation system. It performs the translations of propositional contents and Japanese tense-aspect and modality (Japanese function words that come after a predicate). The propositional contents are translated through the IT, TT, and ET. The ET holds the propositional contents in Sinhalese. The translations of Japanese tense-aspect and modality were introduced to the ET through an appropriate method, which will be described later.

Three types of rules are employed for Japanese expression patterns and transfer, a base-type rule (b-rule) and two addition-type rules (a-ruleFW, a-ruleCW)<sup>2</sup>. A base-type rule is a case-frame-like rule and deals with the translation of basic propositional contents. Addition types deal with adverbial expressions and conjunctive expressions, which are optionally added to the base-type

expressions. Some examples of these three types are shown below (Table 1)

Table 1 also shows the corresponding translation rules for Sinhalese. The classes, member classes, member names, values, and case markers are used to build a C++ programme for the translation rule. The C++ programme is automatically constructed with this. At present, a database of 370,000 Japanese expression patterns is used<sup>1</sup>. The *Jaw*-editor has been developed to write the patterns and transfer rules.

*Sinhalese and Japanese grammatical structures*

*The outlines of Sinhalese and Japanese* Sinhalese is a member of the Indo-Aryan family of languages and its script bears a close structural resemblance to Thai and Malayalam scripts. The Sinhalese writing system is a syllabary system (a set of written symbols that represent syllables, which make up words) derived from ancient North and South Indian scripts. The traditional literary Sinhalese alphabet consists of 58 symbols. Of these, only 42 symbols (12 vowels and 30 consonants) are necessary to represent the writing system<sup>3</sup>.

Japanese belongs to the Altaic language family and employs a combination of three different types of writing systems: *hiragana*, *katakana*, and *kanji*. *Hiragana* is a system of Japanese words for which *kanji* cannot be easily provided. *Katakana* is used similarly for transcribing foreign loan words (other than Chinese) and some onomatopoeic words<sup>4</sup>.

In the agglutinative language structures, Japanese and Sinhalese words are formed by joining meaningful units. The words should be separated by spaces in

**Table 1:** Japanese expression pattern and translation rules for Sinhalese

Japanese Pattern	Rule Type	Class	Member Name	Member Class	Value (Sinhalese)	Case Marker
1 N1 <i>ga</i>	b-rule	CProposition	m_subject	CNoun	1	∅
2 N2 <i>to</i>			m_object	CNoun	2	ඉ
3 <i>tsukiauu</i> (N1 associate with N2)			m_centerW	CString	ආශ්‍රයකරනවා	
1 N1 <i>wa</i>	b-rule	CProposition	m_subject	CNoun	1	∅
2 N2			m_object	CNoun	2	
3 <i>da</i> (N1 is N2)			m_centerW	CString	එව්	
1 V1 <i>temuru to</i>	a ruleFw	CpConnection	m_pSubordinate	CProposition	1	
2 V2 (Vb2 when Vb1)			m_connect	CString	බැඳුවිට	
1 <i>omoshiroi</i>	a ruleCw	CNoun	m_adjective	CAdjective	1	
2 N (interesting N)		CAdjective	m_centerW	CString	විනෝදකාමී	

Sinhalese. If the space is not placed correctly, the sentence becomes meaningless or assumes a different meaning. But the Japanese writing system does not separate each word. The Japanese sentence, as described earlier, can be partitioned into several meaningful linguistic units called “*bunsetsu*”.

A *bunsetsu* is a chunk of words consisting of a content word [noun, verb, etc.] accompanied by some function words [particle, auxiliary, etc.]. A function word is based on a prime root that has a lexical meaning. This root word, which is a lexical item defined in the dictionary, includes a verb, a noun, an adjective, and an adverb. A function word consists of post-positional particles and auxiliary verbs. The concept of *bunsetsu* is common for Sinhalese, too<sup>5</sup>.

In literary Sinhalese, the subject agrees with the verb in number and person, whereas in spoken Sinhalese, there is no agreement between the subject and the verb. Japanese uses neither number nor person agreement. Japanese is a topic-comment prominent language with a basic word order of SOV. Beyond being verb-final, Japanese word order places a modifier (such as adjectives and clausal modifiers) before the noun being modified. The notation of the topic plays a vital part in organizing information to form an utterance or a sentence. The topic is marked by the topic-marking particles “*wa*” and “*mo*”. The Japanese language has particles or post-positions that express not only a grammatical relationship, but also interpersonal feelings. Non-specification of the topic, the subject, the object, and the particles is common. Unlike Sinhalese, Japanese has a rich system of respectful and humble forms, as well as a variety of polite expressions.

The main parts of speech of Sinhalese are the noun, the verb, and the particle. Adjectives and adverbs are not considered parts of speech in Sinhalese, as they are derived from nouns and verbs. But in Japanese, all these components are regarded as parts of speech.

The prefix “*no*” is added to a verb base to yield the corresponding negative meaning in Sinhalese. In the case of compound verbs, “*no*” is added to the last element, which is a simple verb. In Japanese, adding the suffix “*-nai*” to the stem of a verb creates its negative form. The

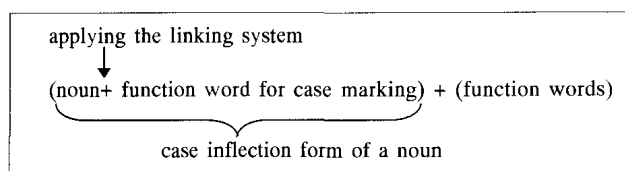


Figure 6: Sinhalese noun *bunsetsu* with function word for case inflection

negative forms of Japanese verbs become adjectives, which are used to express a condition. Irregular Japanese verbs require different padding vowels.

*The Sinhalese bunsetsu structure:* The Sinhalese *bunsetsu* structure depends on a basic Sinhalese grammatical unit. More research is needed to analyze its internal structure. In this paper, the linguistic structure of Sinhalese *bunsetsu* was observed and a sound description is proposed.

In Sinhalese, a *bunsetsu* consists of one content word and some function words. Content words are nouns, verbs, adverbs, etc. Function words are added to the content word in noun *bunsetsu* and verb *bunsetsu*. Other *bunsetsu* consist of a content word only. The structure of the noun and verb *bunsetsu* will be discussed here.

The Sinhalese noun makes a distinction between definite/indefinite, singular/plural, and animate/inanimate. An animate noun has six cases (direct, accusative, dative, genitive, instrumental, and vocative), and an inanimate noun has four cases (direct, dative, genitive, and instrumental). These case forms are distinct for singular-definite nouns, singular-indefinite nouns, and plural nouns. Sinhalese function words such as සමග (*samaga*), පසුව (*pasuwa*), උනන් (*unath*), පටා (*pawa*), හැර (*hera*), සඳහා (*sandaha*), and නිසා (*nisa*) are agglutinated after these case forms.

A noun inflects its ending part according to case. Instead of the idea of case inflections in nouns, a set of function words for case marking is proposed. The proposal contains a method of constructing a case inflection word by linking a noun stem and the function word for case marking. For this purpose, a linking system has been devised. The linking system handles the changes of the ending letter of the noun and the beginning letter of the function word for case marking. Other function words precede this case inflection (Figure 6).

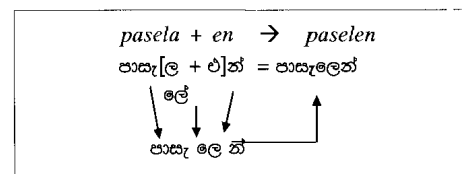


Figure 7: Linking way of the noun “*pasela*”(school) and function word for case marking “*en*”

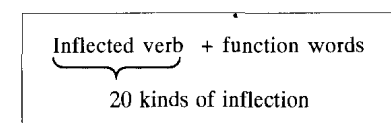


Figure 8: The form of Sinhalese verb *bunsetsu*

**Table 2:** Sinhalese diacritic symbols for vowels

Notation	Vowel (short)	Diacritic	Notation	Vowel (long)	Diacritic
ad <sub>0</sub>	a (අ)	∅	ad <sub>8</sub>	a: (ආ)	ා
ad <sub>1</sub>	æ (ඇ)	ෑ	ad <sub>9</sub>	æ: (ඈ)	ෑ
ad <sub>2</sub>	i (ඈ)	ඞ	ad <sub>10</sub>	i: (ඉ)	ඞ
ad <sub>3</sub>	u (ඊ)	ඬ	ad <sub>11</sub>	u: (උ)	ඬ
ad <sub>4</sub>		ඬ	ad <sub>12</sub>		ඬ
ad <sub>5</sub>		ඬ	ad <sub>13</sub>		ඬ
ad <sub>6</sub>	e (එ)	ඞ	ad <sub>14</sub>	e: (ඒ)	ඞ
			ad <sub>15</sub>		ඞ
ad <sub>7</sub>	o (ඔ)	ඞ	ad <sub>16</sub>	o: (ඌ)	ඞ

**Table 3:** The linking methods of a noun stem and function word for case marking

	function word for case marking		Linking of noun stem <i>NZ</i> and function word for case marking ( <i>Z</i> is the last letter of noun )	adding	dropping
	In roman	In Sinhalese			
1	<i>a:ta</i>	ආට	$N + [ Z + (ad_8) ] + [ ta ]$	ad <sub>8</sub>	<i>a:</i>
2	<i>ekwa</i>	එක්ව	$N + [ Z + (ad_6) ] + [ kwa ]$	ad <sub>6</sub>	<i>e</i>
3	<i>o</i>	ඔ	$N + [ Z + (ad_{16}) ] + [ ∅ ]$	ad <sub>16</sub>	<i>o</i>
4	<i>akin</i>	අකින්	$N + [ Z ] + [ kin ]$	-	<i>a</i>
5	<i>ekuge</i>	එකුගේ	$N + [ Z + (ad_6) ] + [ kuge ]$	ad <sub>6</sub>	<i>e</i>
6	<i>aka</i>	අක	$N + [ Z ] + [ ka ]$	-	<i>a</i>
7	<i>en</i>	එන්	$N + [ Z + (ad_6) ] + [ n ]$	ad <sub>6</sub>	<i>e</i>

Twenty seven linking methods are introduced and some of them are listed (Table 3). To explain the linking method simply, the function words for case marking are denoted in Roman letters instead of Sinhalese script. These linking methods are implemented with the help of the linking rules that are employed for Sinhalese consonant-vowel combinations.

Sinhalese characters are generally composed of consonants and vowels. When a Sinhalese consonant is joined to a vowel, a diacritic (a small mark that added to a letter to alter its pronunciation) is added to the consonants instead of the vowel. This is known as the linking rule of a consonant and a vowel. These vowel diacritics will be called “Sinhalese additions” hereafter. They are placed to the top, bottom, left, and right of the consonants. Sinhalese has 12 vowels and each of them has a diacritic or diacritics (Table 2). In total, 16 linking rules were developed for constant-vowel combinations {each vowel *u* (ඊ), *u:* (උ) and *e:* (ඌ) consists of more than one symbol, depending on the consonant, and only the vowel “*a* (අ)” has no diacritic (∅)}.

Figure 7 shows the example of a linking method according to the 7<sup>th</sup> method “(*NZ=pasela*) + *en*”, as illustrated (Table 2). When the function word for case marking is එන් (*en*), the final letter of the noun ල (*la*), changes to ලෑ (*le*), and the එ (*e*) of එන් (*en*) is dropped. The last letter න් (*n*) of එන් (*en*) is joined to the changed noun stem පාසැලෑ (*pasele*) and the inflected noun පාසැලෑන් (*paselen*) emerges.

Sixteen (16) function words for case marking were defined for animate nouns and eleven (11) were defined for inanimate nouns. These are shown in Tables 4 and 5 respectively with the sample nouns “friend” and “school”. In this method, the noun stem is always in the singular-definite-direct case. This noun form will be called “the noun” hereafter. In this way, all inflected forms can be formed for each noun from 27 function words and 16 linking rules.

In Sinhalese, these combinations are formed individually and necessarily listed in a dictionary (1). But with this method, the number of combinations is reduced considerably (2).

**Table 4:** Linking ways of a noun and function words for case marking on animate noun

Animate noun [example <i>yahaluwa</i> (Friend) ]			
Case forms	Singular		Plural
	Definite	Indefinite	
direct	යහළුවා + ∅ <i>yahaluwa + ∅</i>	යහළුවා + එක් <i>yahaluwa + ek</i> යහළුවෙක් <i>yahaluwek</i>	යහළුවා + ඕ <i>yahaluwa + o</i> යහළුවෝ <i>yahaluwo</i>
accusative	යහළුවා + අට <i>yahaluwa + awa</i> යහළුවාට <i>yahaluwawa</i>	යහළුවා + එකුට <i>yahaluwa + ekuwa</i> යහළුවෙකුට <i>yahaluwekuwa</i>	යහළුවා + අන්ට <i>yahaluwa + nwa</i> යහළුවන්ට <i>yahaluwanwa</i>
dative	යහළුවා + අට <i>yahaluwa + a ta</i> යහළුවාට <i>yahaluwata</i>	යහළුවා + එකුට <i>yahaluwa + ekuta</i> යහළුවෙකුට <i>yahaluwekuta</i>	යහළුවා + අන්ට <i>yahaluwa + anta</i> යහළුවන්ට <i>yahaluwanta</i>
genitive	යහළුවා + අගේ <i>yahaluwa + age</i> යහළුවාගේ <i>yahaluwage</i>	යහළුවා + එකුගේ <i>yahaluwa + ekuwe</i> යහළුවෙකුගේ <i>yahaluwekuwe</i>	යහළුවා + අන්ගේ <i>yahaluwa + ange</i> යහළුවන්ගේ <i>yahaluwange</i>
instrumental	යහළුවා + අගෙන් <i>yahaluwa + agen</i> යහළුවාගෙන් <i>yahaluwagen</i>	යහළුවා + එකුගෙන් <i>yahaluwa + ekugen</i> යහළුවෙකුගෙන් <i>yahaluwekugen</i>	යහළුවා + අන්ගෙන් <i>yahaluwa + angen</i> යහළුවන්ගෙන් <i>yahaluwangen</i>
vocative	යහළුවා + ඒ <i>yahaluwa + e</i> යහළුවේ <i>yahaluwe</i>	-na-	යහළුවා + අනේ <i>yahaluwa + ane</i> යහළුවනේ <i>yahaluwane</i>

**Table 5:** Linking way of a noun and function words for case marking on inanimate noun

Inanimate noun [ example <i>pasela</i> (School) ]			
Case forms	Singular		Plural
	Definite	Indefinite	
direct	පාසැල + ∅ <i>pasela + ∅</i>	පාසැල + අක් <i>pasela + ak</i> පාසැලක් <i>paselak</i>	පාසැල + (-අ) <i>pasela + ( a )</i> පාසැල් <i>pasel</i>
dative	පාසැල + අට <i>pasela + ata</i> පාසැලට <i>paselata</i>	පාසැල + අකට <i>pasela + akata</i> පාසැලකට <i>paselakata</i>	පාසැල් + වලට <i>pasel + walata</i> පාසැල්වලට <i>paselwalata</i>
genitive	පාසැල + එහි <i>pasela + ehi</i> පාසැලෙහි <i>paselehi</i>	පාසැල + අක <i>pasela + aka</i> පාසැලක <i>paselaka</i>	පාසැල් + වල <i>pasel + wala</i> පාසැල්වල <i>paselwala</i>
instrumental	පාසැල + එක් <i>pasela + en</i> පාසැලෙන් <i>paselen</i>	පාසැල + අකින් <i>pasela + akm</i> පාසැලකින් <i>paselakin</i>	පාසැල් + වලින් <i>pasel + walm</i> පාසැල්වලින් <i>paselwalm</i>

a The irregular plural nouns were not be inserted

$$\begin{array}{ll} n \times 27 & \dots\dots 1 \\ n + 27 + 16 & \dots\dots 2 \end{array}$$

where  $n$  = number of nouns in dictionary

In the case of a Sinhalese verb, they are generally conjugated according to person, number, and tense. Sinhalese has conjugated single verbs (Table 6) and conjugated inflected verbs. The inflected verbs are preceded by some function words. If there is a function word after the inflected verb, the function word at the end is conjugated for inflected verb. The most general way to form an inflected verb form is by the addition of suffixes to the verb-base.<sup>6</sup> The function words are optional after them. The verb-base is derived by dropping “*nawa*” from the general verb form (Table 7).

In the Sinhalese verb *bunsetsu*, the inflected verb and function words have no clear definition as of yet. The inflected verb and function words are represented as a single verb block. But this has some subdivisions, which must be separated. Here, a solid separation is defined between the inflected verb and the function words after the verb as Sinhalese verb *bunsetsu* (Figure 8). More than 20 kinds of inflected verb forms emerge initially. The Sinhalese function words can be easily placed after the inflected verb with the help of these separations in machine translations.

All inflected verb forms are listed in a dictionary because of their irregular formations. Table 8 shows only a part of the dictionary. All inflected verbs are derived from either the present form or past form of Sinhalese verbs. Adno-present, Adno-past, Formal-rqst, etc. are the abbreviations for the inflected verb forms adnominal present, adnominal past, and formal request, etc. respectively.

The verb bases of Sinhalese causative verbs are different from general verb bases. It is impossible for the machine to distinguish between causative verbs and the general verb. Another parallel dictionary was developed and used for Sinhalese causative verbs.

#### Translation methods for Japanese function words

One characteristic of the Japanese language that is shared with Sinhalese is that the meaning of a sentence cannot be understood entirely until one reaches the very end. In an extreme case, it is possible to alter the meaning of a sentence

by changing the last word. This is a characteristic of an agglutinative language, which does not possess the prefixes and suffixes of an inflectional language\* or the strict word order of an isolated language\*\*. In Japanese, grammatical functions are performed by function words<sup>7</sup>.

A function word is an inflecting agent that agglutinates to indeclinable words (nouns) and to declinable words (verbs). Here, a solid trial of translations has been made for Japanese function words after the noun and Japanese function words after the predicate while minimizing the ambiguity problems and multi-layered translation problems.

#### The translation of Japanese function words after a noun:

Japanese function words for case marking are considered the leading group of Japanese function words after the noun. This section discusses the translation methods of Japanese function words for case marking into Sinhalese. Japanese function words for case marking such as を (*wo*), は (*wa*), が (*ga*), に (*ni*), and の (*no*) often yield several function words for case marking in Sinhalese<sup>8</sup>. For example, the Japanese function word for case marking “を (*wo*)”, which usually indicates an object, corresponds to no fewer than 16 function words for case marking in Sinhalese: { ට (*ta*), කට (*kata*), වලට (*valata*), ව (*va*), ක්ව (*kva*), ක්ව (*nwa*), එන් (*en*), කින් (*kin*), වලින් (*valin*), අක් (*ak*), එක් (*ek*), ඔ (*o*), එහි (*ehi*), ක (*ka*), වල (*wala*), ∅ }. Some examples of the multi-correspondence of function words for case marking を (*wo*) and の (*no*) are shown in Tables 9 and 10 respectively.

Basically, the function words for case marking are determined by verbs. Therefore, they can be distinguished by using the base-type rule (the b-rule).

Examples:

1. *watashi wa kare ni tegami wo yūbin de okuru.*  
මමගේ මහුවල ලියුමක් තැපෑල මගින් යවනවා.  
I send a letter to him by post.
2. *kare wa komugiko wo mizu de mazeru.*  
මහුවල කිරිගුපිරිලි වතුර වලින් අනනවා.  
He mixes corn flour using water.
3. *watashi wa Tokyo shiten no Sakai-san wo yoku shitteimasu.*  
මමගේ ටෝකියෝ ශාඛාවෙහි සකයි මහතාට හොඳින් දන්නවා.  
I know Mr. Sakai at the Tokyo Office well.

\* In inflectional languages inflectional morphemes are added to a word, which may indicate grammatical information (case, number, tense, person, etc.) as in the case of English.

\*\* In isolated languages like Chinese, the word expresses just substantial meaning and there is no inflection of word forms. The grammatical relations are mainly expressed by the word order.

In these Japanese sentences, the function words for case marking *wa*, *wo*, *ni*, and *de* are used with different verbs. Corresponding Sinhalese function words for case marking are different in each sentence and can be successfully

translated using the above-mentioned base-type rule through the benefits of pattern-based translations.

*The translation of Japanese function words that follow a predicate:* Japanese has several function words used to

**Table 6:** Conjugation of the Sinhalese verb

singular	Non-past			plural	Non-past		
	present	Future	Past		present	future	Past
First	බලමි	බලන්නෙමි	බැලිමි	First	බලමු	බලන්නෙමු	බැලිමු
	<i>balami</i>	<i>balannemi</i>	<i>belimi</i>		<i>balamu</i>	<i>balannemu</i>	<i>beleemu</i>
Second	බලහි	බලන්නෙහි	බැලිහි	Second	බලහු	බලන්නහු	බැලුහු
	<i>balahi</i>	<i>balannehi</i>	<i>belihi</i>		<i>balahu</i>	<i>balannahu</i>	<i>baluuhu</i>
Third	බලයි	බලන්නේය	බැලිය	Third	බලති	බලන්නාහු	බැලූහ
	<i>balai</i>	<i>balanneya</i>	<i>beleya</i>		<i>balathi</i>	<i>balannaahu</i>	<i>beluaha</i>

**Table 7:** Examples for Inflected verbs and optional function words

Verb	General form	Verb base	Inflected verb suffix=මින් <i>min</i>	Optional function words		
Give	දෙනවා	දෙ	දෙමින්	සිටිනවා	විය හැකියි	වගෙයි
	<i>denawa</i>	<i>de~</i>	<i>demin</i>	<i>sitinawa</i>	<i>wiyahekii</i>	<i>wagei</i>
Go	යනවා	ය	යමින්	ලබනවා	පුලුවන්	නේද
	<i>yanawa</i>	<i>ya~</i>	<i>yamin</i>	<i>labanna</i>	<i>puluwan</i>	<i>neda</i>
Eat	කනවා	ක	කමින්	අවශ්‍යයි	වෙයි	යුතුය
	<i>kanawa</i>	<i>ka~</i>	<i>kamin</i>	<i>awashyai</i>	<i>wei</i>	<i>yutuya</i>
Run	දුවනවා	දුව	දුවමින්	වැඩියි	නොවෙයි	වගේවිය
	<i>duwanawa</i>	<i>duwa~</i>	<i>duwamin</i>	<i>wediyi</i>	<i>nowei</i>	<i>wagewiya</i>

**Table 8:** Inflected verb forms of Sinhalese verbs

verb	Perfect	Progress	Adno -present	Adno -past	Formal -rqst
go	ගිහින් <i>gihin</i>	යමින් <i>yamin</i>	යන <i>yana</i>	ගිය <i>giya</i>	යන්න <i>yanna</i>
take	ගෙන <i>gena</i>	ගමින් <i>gamin</i>	ගන්න <i>ganna</i>	ගත් <i>gat</i>	ගන්න <i>ganna</i>
give	දීලා <i>deela</i>	දෙමින් <i>demin</i>	දෙන <i>dena</i>	දුන් <i>dun</i>	දෙන්න <i>denna</i>
walk	ඇවිදලා <i>avidala</i>	ඇවිදමින් <i>avidimin</i>	ඇවිදන <i>avidina</i>	ඇවිද්ද <i>avidda</i>	ඇවිද්දක් <i>avidinna</i>
shoot	විදලා <i>vidala</i>	විදමින් <i>vidimin</i>	විදන <i>vidina</i>	විද්ද <i>vidda</i>	විද්දක් <i>vidinna</i>
rise	නැගලා <i>nagala</i>	නගිමින් <i>nagimin</i>	නගින <i>nagina</i>	නැග්ග <i>nagga</i>	නගින්න <i>naginna</i>
measure	මැනලා <i>menala</i>	මනිමින් <i>manimin</i>	මනින <i>manina</i>	මැනන <i>menna</i>	මනින්න <i>maninnna</i>
jump	පැනලා <i>penala</i>	පනිමින් <i>panimin</i>	පනින <i>panina</i>	පැනන <i>penna</i>	පනින්න <i>paninna</i>
hit	ගහලා <i>gahala</i>	ගහමින් <i>gahamin</i>	ගහන <i>gahana</i>	ගැහුව <i>gehuwa</i>	ගහන්න <i>gahanna</i>
eat	කාලා <i>kala</i>	කමින් <i>kamin</i>	කන <i>kana</i>	කෑව <i>kewa</i>	කන්න <i>kanna</i>

**Table 9:** Multi-correspondence of the Japanese function word “*wo*” with Sinhalese function words

Japanese sentence	Sinhalese translation	Sinhalese Function words
<i>tsuma wo aisuru</i> ([I] love [my] wife)	බිරිඳට ආදරය කරනවා <i>birindata adarakaranawa</i>	ට <i>ta</i>
<i>kodomo wo tsurete iku</i> (accompany the child)	ලමයාට එක්ක යනවා <i>lamayava ekka yanawa</i>	ට <i>va</i>
<i>hashi wo wataru</i> (cross the bridge)	පාලමෙන් එහොඩි වෙනවා <i>palamen godawenawa</i>	එක් <i>en</i>
<i>gohan wo taberu</i> (eat rice)	ඛන් කනවා <i>bath kanawa</i>	∅
<i>inu wo kau</i> (a dog is bred)	බල්ලෙක් හදනවා <i>ballek hadanawa</i>	එක් <i>ek</i>

**Table 10:** Multi-correspondence of the Japanese function word “*no*” with Sinhalese function words

Japanese sentence	Sinhalese translation	Sinhalese Function words
<i>Tokyo no oji</i> (Uncle in Tokyo)	ටෝකියෝ ටල මාමා <i>Tokyo wala mama</i>	ටල <i>wala</i>
<i>oji no ie</i> (Uncle's home)	මාමාගෙ ගෙදර <i>mamage gedara</i>	ගෙ <i>ge</i>
<i>fuku no iro</i> (Color of the clothes)	ඈඳුමෙහි පාට <i>endumehi paata</i>	එහි <i>ehi</i>
<i>uchi no niwa</i> (Home garden)	ගෙදර මිදුල <i>gedara midula</i>	∅
<i>gakkō no hon</i> (Book belongs to the school)	පාසැල සතු පොත <i>pasela satu pota</i>	සතු <i>satu</i>
<i>haha no tameni</i> (On behalf of mother)	අම්මා වෙනුවෙන් <i>amma wenuwen</i>	∅

**Table 11:** Japanese sentences without and with function words

Sentences without function words	Same sentences with function words
太郎は道を歩く ( <i>Taro wa michi wo aruku</i> ) Taro walks in the road.	太郎は道を歩いて <u>いた</u> ( <i>Taro wa michi wo aruite ita</i> ) Taro <u>was walking</u> in the road
彼は来る ( <i>kare wa kuru</i> ) He will come.	彼は来る <u>でしょう</u> ( <i>kare wa kuru deshō</i> ) He <u>would probably</u> come.
窓を開ける ( <i>mado wo akeru</i> ) Window is opened	窓を開けて <u>ください</u> ( <i>mado wo akete kudasai</i> ) <u>Please open</u> the window.
熊はジャンプする ( <i>kuma wa jumpu suru</i> ) Bear jumps.	熊はジャンプする <u>らしい</u> ( <i>kuma wa jumpu suru rashii</i> ) It <u>seems</u> that a bear jumps)
私はワインを飲む ( <i>watashi wa wain wo nomu</i> ) I drink wine.	私はワインを飲む <u>みたい</u> ( <i>watashi wa wain wo nomitai</i> ) I <u>want to</u> drink wine.
本を読む ( <i>hon wo yomu</i> ) The book is read.	本を <u>読みやすかった</u> <u>でしょう</u> ( <i>hon wo yomi yasukatta deshō</i> ) The book <u>was easy to</u> read. Wasn't it?

append special meanings to ordinary verbs. They are restricted to dependent usage and always follow independent words or independent phrases. In other words, Japanese function words are postpositional. At the second stage of Japanese-Sinhalese translations, the translations of Japanese function words after a predicate are carried out covering tense-aspect and modality. The multi-layered problem of the translation of a Japanese function word after a predicate into Sinhalese is

solved by translation rules presented here in a table format. As an introduction, Table 11 illustrates some examples of various Japanese function words in various environments.

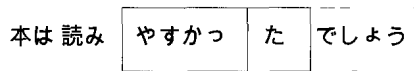
For ease of translation, Japanese function words after a predicate are divided into four groups by *IBUKI* system: (1) voice, etc., (2) tense-aspect, etc., (3) judgments, and (4) conjunctions (Table 12).

Table 12: Groups of Japanese function words

[1]voice, etc	[2]tense-aspect, etc	[3]judgments	[4]conjunctions
られる( <i>rareru</i> )	た ( <i>ta</i> ) (past)	ろう ( <i>rō</i> )	て ( <i>te</i> )
させる( <i>saseru</i> )	ている ( <i>teiru</i> )	だろう ( <i>darō</i> )	が ( <i>ga</i> )
たい ( <i>tai</i> )	ていた ( <i>teita</i> )	らしい ( <i>rashii</i> )	と ( <i>to</i> )
なる ( <i>naru</i> )	だ ( <i>da</i> )	でしょう ( <i>deshō</i> )	ように ( <i>youni</i> )
てくれる( <i>tekureru</i> )	だった ( <i>datta</i> )	そうだ ( <i>souda</i> )	で ( <i>de</i> )
てもらう ( <i>temorau</i> )	である ( <i>dearu</i> )	なければならない ( <i>nakerebanaranai</i> )	でも ( <i>demo</i> )
やすい ( <i>yasui</i> )	です ( <i>desu</i> )	かもしれない ( <i>kamoshirenai</i> )	から ( <i>kara</i> )
てほしい ( <i>tehoshii</i> )	てきた ( <i>tekita</i> )	ね ( <i>ne</i> )	ので ( <i>node</i> )
なくなる ( <i>nakunaru</i> )	ます ( <i>masu</i> )	べき ( <i>beki</i> )	ば ( <i>ba</i> )
にくい ( <i>nikui</i> )	なかった ( <i>nakatta</i> )	な ( <i>na</i> )	ような ( <i>youna</i> )
...	...	...	...

For example, the formation of the above groups of function words can be described as follows.

Sample sentence: 本は読みやすかったでしょう。



*hon wa yomi yasukat ta deshō.*

The book was easy to read. Wasn't it?

According to the above sentence, “*hon wa*” is the object with a case marker and “*yomi*” is an inflection of the verb “*yomu*”. The other parts are function words which are picked up from each column of Table 12 (the relevant parts are underlined). The word “*yasukatta* (was easy)” is a Japanese expression for the past of “*yasui* (easy)”. This information is collected from column[1] and “*ta*” in column[2]. And the ending word “*deshō*” comes from column[3].

A Japanese function word is translated into several expression elements at different places in a Sinhalese sentence. They cannot be translated through a simple one-to-one correspondence. This is identified as the multi-layered problem in Japanese-Sinhalese translations.

Examples:

1. *sakana ga kujira ni taberareru.* (The fish is eaten by the whale.)  
 කල්මහ විසින් මාලුවාට කනු ලබනවා (*talmaha visin maluwawa kanu labanawa.*)
2. *sakana ga kujira ni taberareteiruyōda.* (The fish is probably being eaten by the whale.)  
 කල්මහ විසින් මාලුවාට කනු ලැබුවා වගෙයි (*talmaha visin maluwawa kanu labuwawagei.*)

The translation rules for these function words are described in a table format (Tables 13-16) with expression elements in its each column (Japanese function words were taken from Japanese newspaper articles that were published in the past ten years and listed in a descending sequence. Only the top rows of each table, with rules, are presented here). The expression elements are Vb-Inflection, CVb-Inflection, After-Verb, After-Sub, After-Obj, and Judge-Verb.

“Vb-Inflection” indicates the inflected form of the verb for Sinhalese. A Sinhalese verb has more than 20 inflected forms (Table 8).

“CVb-Inflection” indicates the inflected form of the causative verb. The causative verb means the causative form of the verb (see the example below).

Example for a causative verb:

*kare ni wain wo nomasetakatta yōda.*

ඔහුට වයින පොවන්න අවශ්‍යවුනා වගෙයි

(1) probably have him drink wine.

The “After-Verb” and “Before-Verb” indicate the elements to be put in just after the verb and before the verb respectively.

The “After-Sub” indicates the elements that always follow the subject, while the “After-Obj” indicates the elements that follow the object in the output sentences. The “Judge-Verb” indicates the elements at the very end of the translated sentences.

A simple example for the translation of function words:

*kare wa wain wo nomitakatta yōda.*

ඔහුට වයින බොන්න අවශ්‍යවුනා වගෙයි.

He probably wanted to drink wine.

The Japanese phrase “-*takatta yōda*” is divided into *tai/ta/yōda* (want/ed/probably). The Sinhalese translations are generated from the translation rules for “-*tai*” (Table 13), “-*ta*” (Table 14) and “*yōda*” (Table 15), applying the linearizing function for Sinhalese.

The rules for the fourth group, “conjunctions”, are different from the rules of the other groups. Table 16 shows the rules for conjunctions. There are four types for building Sinhalese sentences with conjunctions.

1. Subordinate Sentence [with inflected verb] + Main Sentence

The inflected verb form of the subordinate sentence is indicated in column “V-change (Present)” for type 1 in Table 16.

Example:

*taiyou wa higashi kara dete nishi ni shizumu* (The sun rises in the east and sets in the west).

ඉර නැගෙනහිරින් පායලා බටහිරින් බහිනවා *Ira negenahirin payala batahirin bahinawa*. (When the conjunction is “-*te*”, the Sinhalese verb “*payanawa*” is changed to “*payala*” according to the inflected verb form “Conj”).

2. Subordinate Sentence [with tail-changed verb] + conjunction + Main Sentence

This type is prepared for cause-effect and temporal conjunctions like *node* (since), *youni* (as if), *demo* (although), *nagara* (while), etc. The special effect is that the tail of the verb in a subordinate sentence is changed according to the tense. The last letter of the verb is dropped in present tense [DropLL] (example: වහිනවා) and

the last addition of the verb is dropped in the past tense [DropLA] (example: ලියුවා).

Example:

*ame ga futte iru node kuruma de iku*. (As it is raining, I travel by car.)

වැස්ස වහින(වා) නිසා මෝටර් රථයෙන් යනවා.

*wessa wahina(wa) nisa moter ratayen yanawa*.

The “*wa*” is the last letter that is dropped from the verb in the present.

3. Subordinate Sentence+conjunction+ Main Sentence

This type is for conjunctions such as *ga* (though), *nara* (if), *to* (and), *ka* (or), etc. In this case there is no change in the verb of the subordinate sentence and it is simply connected to the conjunction.

Example:

*kono hon wa nando mo yonda ga, yoku wakarimasen*. {Though I read this book several times, (I) cannot understand it.}

*me pota keepavitak ma kiyawuwath terem ganna apahasui*.

මේ පොත කීපවිටක්ම කියවුවත් තේරුම්ගන්න අපහසුයි.

4. Subordinate Sentence (with inflected verb) + conjunction + Main Sentence

Both the inflected verb and the conjunction are employed for this type. *Toshite* (and next), *made* (till), *kara* (after), etc. are included in this type.

Table 13: Translation rules for the Japanese function word group “Voice, etc.”

Japanese function words	Vb-inflection	After-Verb	After-Sub	After-Obj
<i>rareru</i>	Formal-comm	ලබනවා	විසින්	ට
<i>saseru</i>	C-Present	-		ට/ව/ලවා
<i>tai</i>	Formal-rqst	අවශ්‍යයි	ට	අක්
<i>naru</i>	Formal-rqst	අවශ්‍ය වෙනවා		
<i>tekureru</i>	Conj	දෙනවා		ට
<i>temorau</i>	Conj	ගන්නවා		
<i>yasui</i>	Formal-rqst	පහසුයි		
<i>tehoshii</i>	Conj	ගන්න අවශ්‍යයි	ලවා	
<i>nikui</i>	Formal-rqst	අපහසුයි	ට	
<i>sugiru</i>	Past	වැඩියි		
<i>tekudasai</i>	Conj	දෙමින් සිටිනවා		
<i>saserareru</i>	C-Passive	-	ට	විසින්
<i>temoraitai</i>	Conj	ගන්න අවශ්‍යයි		
<i>temoraeru</i>	Conj	ගතහැකිය		
<i>teyaru</i>	Conj	දෙනවා		
<i>takunaru</i>	Formal-rqst	අවශ්‍ය වෙනවා		

**Table 14:** Translation rules for the Japanese function word group “Tense and Aspect, etc.”

Japanese function words	Vb-inflection	CVb-inflection	After-Verb	Before-Verb
<i>ta</i>	Past	C-Past		
<i>teiru</i>	Progress	C-Progress	සිටිනවා	
<i>da</i>	-	-	වේ	
<i>nai</i>	-	-		නො
<i>teuta</i>	Progress	C-Progress	සිටියා	
<i>data</i>	-	-	විය	
<i>dearu</i>	-	-	ස් වෙයි	
<i>desu</i>	-	-	වෙයි	
<i>tekuta</i>	Progress	C-Progress	ආවා	
<i>masu</i>	Present	C-Present		
<i>nakatta</i>	Past	C-Past		නො
<i>dewanai</i>	-	-	නොවෙයි	
<i>deoru</i>	Progress	C-Progress	සිටිනවා	
<i>teunai</i>	Progress	C-Progress	සිටිනවා	නො
<i>dekiru</i>	-	-	පුලුවන්	
<i>mashta</i>	Past	C-Past		
<i>teiku</i>	Progress	C-Progress	යනවා	
<i>teimasu</i>	Progress	C-Progress	සිටිනවා	
<i>tekuru</i>	Progress	C-Progress	එනවා	
<i>teshuma</i>	Conj	C-Conj	ඉවරවෙනවා	
<i>dekinai</i>	-	-	නොහැකියි	
<i>nanoda</i>	-	-		
<i>teshumatta</i>	Conj	C-Conj	ඉවරවුනා	
<i>masen</i>	Present	C-Present		නො
<i>deshita</i>			වුනා	
<i>temuru</i>	Progress	C-Progress	බලනවා	

**Table 15** Translation rules for the Japanese function word group “Judgments”

Japanese function words	Vb-inflection	Judge-Verb
<i>rō</i>	Will-Vb	නොවේද
<i>darō</i>	Present	නොවේද
<i>souda</i>	Present	වගේවිය
<i>deshō</i>	Present	නොවේද
<i>rashu</i>	Present	වගෙයි
<i>yōda</i>	Past	වගෙයි
<i>nakerebanaranai</i>	Ending-NE	කල යුතුය
<i>kamoshirenai</i>	Present	ද කිව නොහැක
<i>ne</i>	Present	නේද
<i>beki</i>	Adno-Present	යුතුය
<i>bekida</i>	Adno-Present	යුතුවේ
<i>na</i>	Present	වගේ

Example:

*gohan wo tabete kara gakkō e iku.* {After eating, (I) go to school.}

කෑම කෑවා ව පසු පාසල් යනවා *keme kewa ta pasu pasal yanawa*

“Kewa” is the inflected verb and “ta pasu” is the conjunction.

## RESULTS AND DISCUSSION

The MT system from Japanese to Sinhalese, *Jaw/Sinhalese*, is proposed, including a clear definition of the Sinhalese *bunsetsu* structure and solutions for Japanese function words that follow a noun and a predicate.

**Table 16:** Translation rules for the Japanese function word group “Conjunctions”

Japanese Conjunctions	Type	Sinhalese Conjunctions	V-change (Present)	V-change (Past)
<i>te</i>	1	*	Conj	
<i>ga</i>	3	වුවන්		
<i>to</i>	4	මි	Past	
<i>youni</i>	2	ලෙස	DropLL	DropLA
<i>de</i>	1	*	Conj	
<i>demo</i>	2	නමුත්	DropLL	DropLA
<i>kara</i>	4	ට පසු	past	
<i>node</i>	2	නිසා	DropLL	DropLA
<i>ba</i>	1	ත්	IfCondit	
<i>youna</i>	2	ආකාරයේ	DropLL	DropLA
<i>ra</i>	4	ට පස්සේ	Past	

\* the system itself generates these conjunctions

**Table 17:** Japanese input sentences and translated Sinhalese sentences by Jaw/Sinhalese

Japanese sentences (input sentences)	Sinhalese sentences (translated sentences)
友達に傘を貸してもらった。 (I borrowed an umbrella from a friend.)	යාලුවාගෙන් කුවියක් ණයට ඉල්ලා ගත්තා
試合は引き分けに終わったらしい。 (It seems the game was finished in a draw.)	තරගය එය පරාජයෙන් කෙරවූ නිමවූවා වූහෙයි
彼女は娘に新しい洋服を縫ってやった。 (She made new clothes for the daughter.)	ඇය ඇඳුමක් දුටුව මසාදන්නා
彼は友達から知らせを聞いて非常に驚いていたようだった。 (It seems that he was awfully surprised when he heard the news from the friend.)	ඔහු යාලුවාගෙන් ප්‍රවෘත්තිය අඩු මහත්කම දුටුමගටපත්වී සිටියා වූහෙයි
祖父は果物の出荷の分別作業で疲れているようだった。 (It seemed that grandfather is working on distributions during the fruit season.)	ආහා පළතුරු වර්ගකිරීමේදී වැඩවලට මහත්කිවීමක් සිටියා වූහෙයි
この薬を飲むと痛みがなくなるでしょう。 (This medicine will relieve your pain, won't it?)	මෙම බෙහෙම බොහෝවිට වේදනාව නැතිවේවි
彼の名前は故郷で有名だった。 (His name was famous in (his) hometown.)	ඔහුගේ නම උපත් ගමේදී ප්‍රසිද්ධවිය
あひるはひよこから美しい白鳥に成長した。 (It grew up to be a beautiful swan from an ugly duckling.)	තාරාවා කුඩා පැටවුන්ට පත්වූ ලක්ෂ්‍ය භංගයෙන්වූ වැඩුණා
君の夢はいつか実現するでしょう。 (Will your dream come true someday?)	ඔබගේ සිහිනය කවදාහෝ සැබෑවේවි
彼は一人息子に新車を買ってやった。 (He bought a new car for his only son.)	ඔහු එකම පුතාට ඇත් වාහනයක් මිලට ගත්තේවි
先生は私たちに地理を教えてください。 (The teacher is teaching us geography.)	ගුරුතුමා අපට භූගෝලය උගන්වනුදෙමින් සිටිනවා
歯医者に行ってその歯を抜いてもらった。 (I went to the dentist and got that tooth pulled out.)	දත් දොස්තර වෙත ගියින් එම දත් ගලවා ගනු ලබාගත්තා
私の同僚を代表して君が皆にスピーチしていただきたい。 (I would like you to represent my colleague and make a speech on behalf of all.)	මමගේ සහකාරාට නියෝජනය කරනු ලැබීමේ ඉඩකඩවක් සිටියවා කතාවක් කරවාගන්නා කැමතියි
魚が鯨に食べられた。 (The fish was eaten by the whale.)	කල්මස විසින් මාලාවට කනු ලැබුවා

Although the usage of Japanese function words in translation is difficult to master, it is an important part of translations from Japanese into Sinhalese. In many cases, there are one-to-one correspondences between Japanese *bunsetsu* and Sinhalese *bunsetsu*. But these correspondences are far from perfect: for example, a single Japanese case marker often corresponds to several Sinhalese case markers. The translation of a Japanese function word after a predicate affects more than one

expression element in a Sinhalese sentence. This paper has proposed a solution to these problems, culminating in the *Jaw/Sinhalese* Machine Translation System.

Two hundred example sentences, which cover most Japanese grammatical characteristics, were chosen manually from 1000 sentences that were taken from a Japanese-English dictionary. The translation rules for function words after a predicate and after a noun were

Table 18: Errors at translations

Japanese sentence	Sinhalese sentence and error description <sup>a</sup>
(a.1)私は彼にたくさんお金を借りている。 <i>watashi wa kare ni takusan okane wo kariteiru</i> (I borrow money from him in large quantities.)	මම ඔහුගෙන් මුදල් ගොඩක් ණයට ගනිමින් සිටිනවා (X) (The translation of “ <i>teiru</i> ” is in present continuous tense. But the simple present tense is required.) මම ඔහුගෙන් මුදල් ගොඩක් ණයට ගන්නවා (O)
(a.2)私は彼を私の恩人と考えている。 <i>watashi wa kare wo watashi no nanijin to kangaete iru</i> (I think that he is my companion.)	මම ඔහුව මගේ සහයකයා කියා සිතමින් සිටිනවා (X) (same as error description of (a.1) ) මම ඔහුව මගේ සහයකයා කියා සිතනවා (O)
(b.1)肉は三日後に腐ってしまった。 <i>niku wa mikkakango ni kusatte shimatta</i> (Meat were rotten after 3 days)	මස් දින 3කින් පසුව තරක්ඉවලා ඉවරවුනා (X) (Translation of “ <i>kusatta</i> ” should be used for Sinhalese translations.) මස් දින 3කින් පසුව තරක්විය (O)
(b.2)川に落ちて服が濡れてしまった。 <i>kawa ni ochite fuku ga nurete shimatta</i> (I fell into the river, and clothes were soaked.)	ගඟට වැටිලා ඇඳුම් ඝෙම්ඟා ඉවරවුනා (X) (Translation of “ <i>nurete shimatta</i> ” is not appropriate. Translation of “ <i>nureta</i> ” is needed.) ගඟට වැටිලා ඇඳුම් ඝෙම්ඟා (O)
(c)人々は勇ましい彼を愛している。 <i>hitibito wa isamashii kare wo aishiteiru</i> (People love brave him) Nを愛する。 N wo aisuru (love N)	මිනිසුන් වීර ඔහුව ආදරය කරනවා (X) (“wo” makes different translations) Nට ආදරය කරනවා Nඊක්ව ආදරය කරනවා (indefinite) Nඉකව ආදරය කරනවා (inanimate)

a: O =correct translations and X =incorrect translations.

examined. A translation result of 72% success was achieved. Table 17 shows examples for correctly translated sentences. The underlined parts are the function words (single underlines show function words after a predicate and double underlines show function words after a noun. Japanese function words with dot underlines have no translation for Sinhalese and dark underlines show conjunctions). Table 18 shows some examples for incorrectly translated sentences.

Function words after a predicate, “-*teiru*”, “-*te shimatta*” in Table 18 (a, b), have several Sinhalese correspondences and they are not yet disambiguated through the present rules. The error of (c) in Table 18 shows that some of case markers of Sinhalese that correspond to Japanese “wo” are not completely disambiguated. They varied depending on the attributes of “N” (i.e., definite/indefinite, singular/plural, and animate/inanimate). For a complete disambiguation, the related attribute of a Japanese noun should be identified. The problem recognized here is that the Japanese noun has no such attributes, therefore the verb of the translated Sinhalese sentences is in dictionary form (*yanawa*, *giya*, etc.). If the above-mentioned attributes of a Japanese noun, especially for the subject, can be identified, the Sinhalese conjugated verbs (as in Table 6) can also be formed for the translations.

In a Japanese sentence the subject is often omitted, and it is common in Sinhalese too. For the translations of

this type of sentence, the Jaw/Sinhalese system produces a special object for the subject called “ellipsis” (Figure 3 and 4). When there is no subject, the verb is kept in dictionary form according to Sinhalese grammar.

### CONCLUSION

The *Jaw/Sinhalese* system is described in this paper, including a definition of the Sinhalese *bunsetsu* structure and solutions for Japanese function words. This study on Japanese function words in a Japanese-Sinhalese machine translation system is the first attempt at constructing syntactical noun and verb formats for Sinhalese. The formation of nouns and verbs in this language pair is virtually unexplored from the machine translation perspective. This work provides the first test data on the evaluation of a practical pattern-based transfer approach. A test was conducted for 200 sample sentences using the *Jaw/Sinhalese* system, and it achieved a good translation result. The sample is small as it is the first trial of MT for this language pair. The sample size is to be increased in the next development stage, and then the forthcoming problems have to be solved.

For the growth and improvement of the system, a thorough contrast of this language pair is required. To improve the system and its accuracy, the size of the dictionary and translation rules have to be increased.

Compared to Sinhalese, the Japanese language has a rich set of function words after the predicate and noun. The translations in this study so far have been limited to a sample, which needs to be expanded with more investigations. Furthermore, the grammatical aspects such as definite/indefinite, singular/plural and animate/inanimate in Japanese nouns have to be solved with a meaning (semantic) analysis to determine the correct formation of Sinhalese verbs and case markers.

---

## References

- 1 Imai Y., Shieh J., Thelijjagoda S., Elikewala N., Chau M., Bu Z. & Ikeda T. (2003). Machine translation system for Japanese to Asian languages. *Proceedings of the 65<sup>th</sup> National Convention of Information Processing Society of Japan (IPSI)*. Tokyo University of Technology, Tokyo, pp. 363-366.
- 2 Thelijjagoda S., Imai Y., Elikewala N. & Ikeda T. (2002). Sinhalese generation rule and machine translation system Jaw/Sinhalese. *Proceedings of Forum of Information Technology 2002 Conference*. Tokyo, pp. 173-174.
- 3 Karunatilake W.S. (1990). *An Introduction to Spoken Sinhalese*. pp.6-16. M.D. Gunasena and Company Ltd. Colombo.
- 4 Sugawara M.(1989). *A Japanese approach to Japanese*. pp.537-593. The East Publications, Inc., Minato-ku, Tokyo.
- 5 Herath A. (1997). A study on natural language processings for Sinhalese. *PhD dissertation*. Gifu University, Japan.
- 6 Karunatilake W.S. (2001). *Sinhala basha viyaranaya*. pp.173-210. M.D. Gunasena and Company Ltd, Colombo.
- 7 Maynard S.K. (1990). *An introduction to Japanese grammar and communication strategies*. pp.134-148. The Japan Times Ltd. Tokyo, Japan.
- 8 Thelijjagoda S., Imai Y., Elikewala N. & Ikeda T. (2003). Machine translation system Jaw/Sinhalese. *Proceedings of the 9<sup>th</sup> Natural Language Processing Conference*. Yokohama, Japan. pp. 477-480.