RESEARCH ARTICLE

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

Samantha Thelijjagoda*, Yoshimasa Imai and Takashi Ikeda

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

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*), න්ව (*nwa*), එන් (*en*), කින් (*kin*), වලින්(*valin*), අක් (*ak*), එක් (*ek*), ඕ (*o*), එහි (*ehi*), ක (*ka*), වල (*wala*), Ø}. These are distinguished by using base-type rules of the system, which are dependent on the fact that the case frame pattern of a verb basically determines case markers.

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 "අවශ්‍යයයි(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[1] (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).
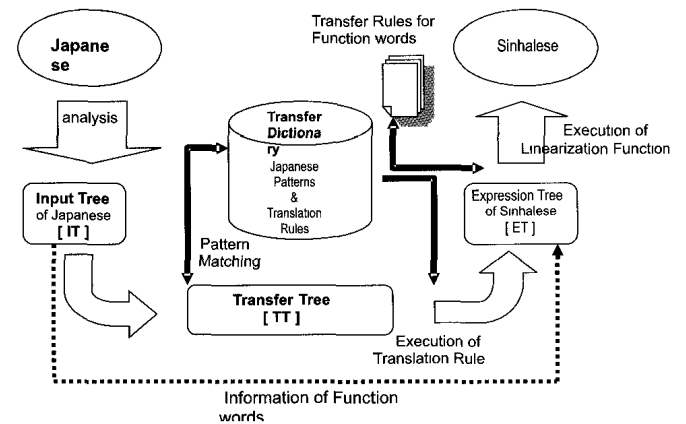

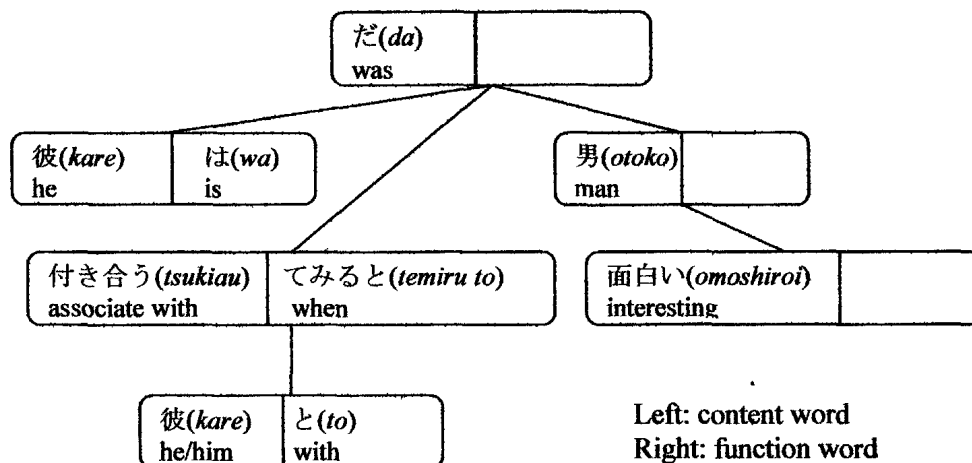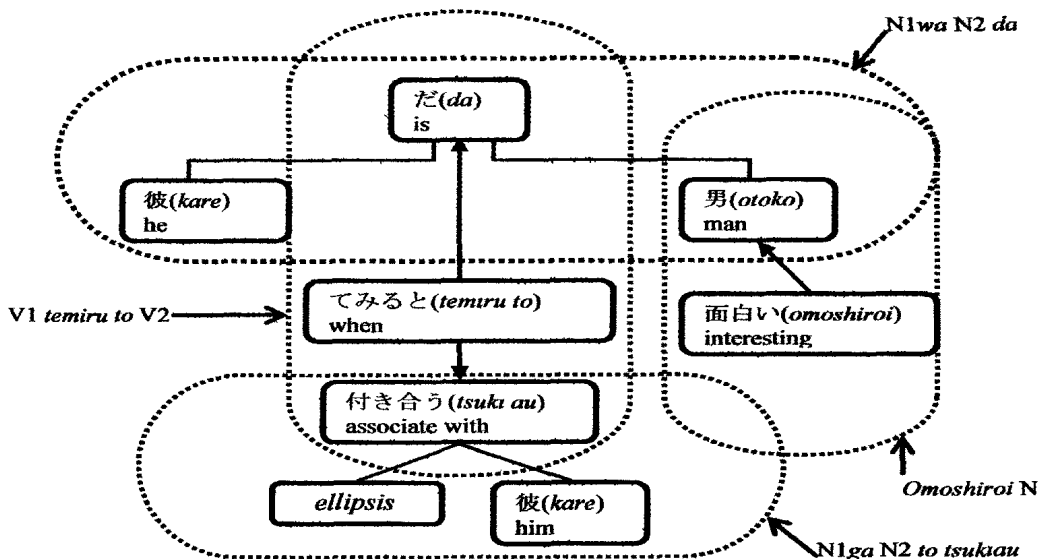**Figure 1:** Outline of *Jaw/Sinhalese* system

Figure 2: Input Tree (IT)

N1 *wa* N2 *da*

だ*(da)*
is

彼*(kare)*
he

男*(otoko)*
man

V1 *temiru to* V2

てみると*(temiru to)*
when

面白い*(omoshiroi)*
interesting

付き合う*(tsuki au)*
associate with

ellipsis

彼*(kare)*
him

Omoshiroi N

N1 *ga* N2 *to tsukiau*

Figure 3: Translation Tree (TT)

CProposition
m_subject
m_object1
m_centerW = "විය" [was]
m_CpConnection

CNoun
m_centerW="ඔහු"[he]

CpConnection
m_pSubordinate
m_connect="විට" [when]

CNoun
m_centerW=" පුරුෂයෙක් "
[a man]

CProposition
m_subject
m_object1
m_centerW = "ආශ්‍රය කරබලනවා"
[associate with]

Ellipsis

CNoun
m_centerW="ඔහු" [he]
m_caseMarker="ව"

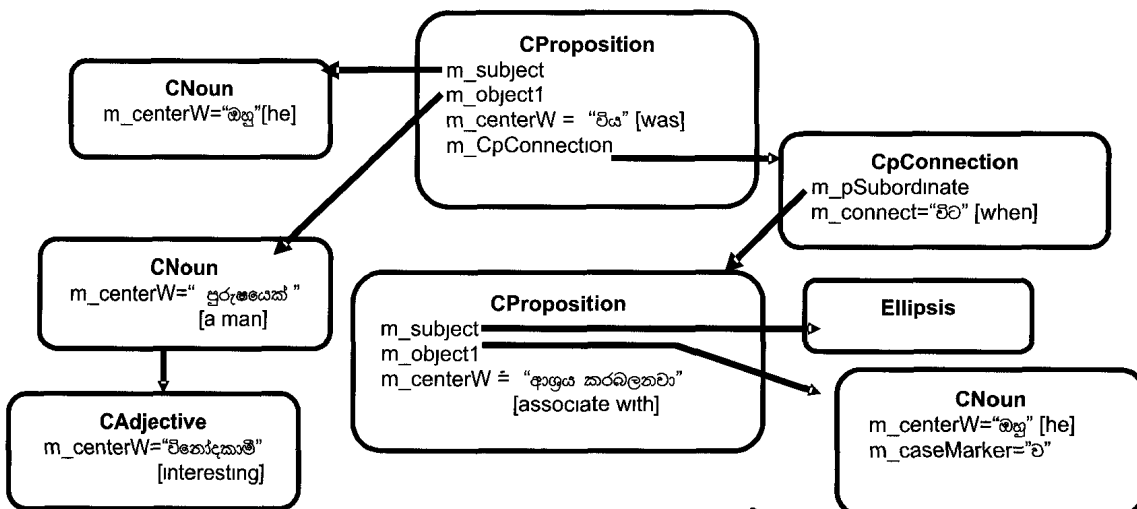**CAdjective**
m_centerW="විනෝදජනක"
[Interesting]

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 :ඔහුව ආශ්‍රයකර බැලුවිට (ඔහු) විනෝදජනක 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 *time, numerical, time-era, time-begin, time-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)[2] 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 conjunctional 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[1] 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[3]

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[4]

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 *ga*	b-rule	CProposition	m_subject	CNoun	1	Ø
2	N2 *to*			m_object	CNoun	2	ව
3	*tsukiau*			m_centerW	CString	ආශ්‍රයකරනවා	
	(N1associate with N2)						
1	N1 *wa*	b-rule	CProposition	m_subject	CNoun	1	Ø
2	N2			m_object	CNoun	2	
3	*da*			m_centerW	CString	වේ	
	(N1 is N2)						
1	V1 *temiru to*	a ruleFw	CpConnection	m_pSubordinate	CProposition	1	
2	V2			m_connect	CString	· බැලුවිට	
	(Vb2 when Vb1)						
1	*omoshiroi*	a ruleCw	CNoun	m_adjective	CAdjective	1	
2	N		CAdjective	m_centerW	CString	විනෝදකාම	
	(interesting N)						

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[5].

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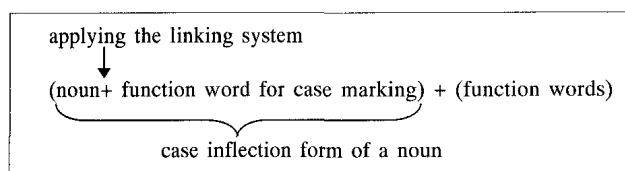
negative forms of Japanese verbs become adjectives, which are used to express a condition. Irregular Japanese verbs require different padding vowels.

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*), උනත් (*unath*), පවා (*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).
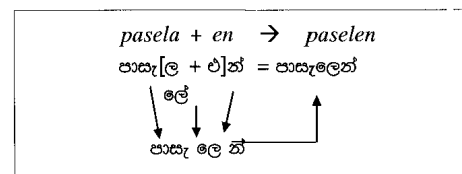

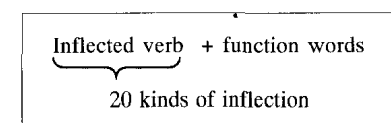**Figure 7:** Linking way of the noun "*pasela*"(school) and function word for case marking "*en*"

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

Figure 8: The form of Sinhalese verb *bunsetsu*

Table 2: Sinhalese diacritic symbols for vowels

Notation	Vowel (short)	Diacritic	Notation	Vowel (long)	Diacritic
ad_0	a (අ)	Ø	ad_8	$a:$ (ආ)	ා
ad_1	$æ$ (ඇ)	ැ	ad_9	$æ:$ (ඈ)	ෑ
ad_2	i (ඉ)	ි	ad_{10}	$i:$ (ඊ)	ී
ad_3	u (උ)	ු	ad_{11}	$u:$ (ඌ)	ූ
ad_4		ෘ	ad_{12}		ෲ
ad_5		ෟ	ad_{13}		ෳ
ad_6	e (එ)	ෙ	ad_{14}	$e:$ (ඒ)	ේ
			ad_{15}		ෛ
ad_7	o (ඔ)	ො	ad_{16}	$o:$ (ඕ)	ෝ

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

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

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 (උ), $u:$ (ඌ) and $e:$ (ඒ) consists of more than one symbol, depending on the consonant, and only the vowel "a (අ)" has no diacritic (Ø)}.

Figure 7 shows the example of a linking method according to the 7th 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 ලෙ (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 පාසෙල ($pasele$) and the inflected noun පාසෙලන් ($paselen$) 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 *yahaluwa* (Friend)]

Case forms	Singular Definite	Singular Indefinite	Plural
direct	යහළුවා + Ø *yahaluwa + Ø*	යහළුවා + එක් *yahaluwa + ek* යහළුවෙක් *yahaluwek*	යහළුවා +ඕ *yahaluwa + o* යහළුවෝ *yahaluwo*
accusative	යහළුවා + අව *yahaluwa + awa* යහළුවාව *yahaluwawa*	යහළුවා + එකුව *yahaluwa + ekuwa* යහළුවෙකුව *yahaluwekuwa*	යහළුවා + න්ව *yahaluwa + nwa* යහළුවන්ව *yahaluwanwa*
dative	යහළුවා + ආට *yahaluwa + a ta* යහළුවාට *yahaluwata*	යහළුවා + එකුට *yahaluwa + ekuta* යහළුවෙකුට *yahaluwekuta*	යහළුවා + අන්ට *yahaluwa + anta* යහළුවන්ට *yahaluwanta*
genitive	යහළුවා +අගේ *yahaluwa + age* යහළුවාගේ *yahaluwage*	යහළුවා + එකුගේ *yahaluwa + ekuge* යහළුවෙකුගේ *yahaluwekuge*	යහළුවා + අන්ගේ *yahaluwa + ange* යහළුවන්ගේ *yahaluwange*
instrumental	යහළුවා +අගෙන් *yahaluwa + agen* යහළුවාගෙන් *yahaluwagen*	යහළුවා + එකුගෙන් *yahaluwa + ekugen* යහළුවෙකුගෙන් *yahaluwekugen*	යහළුවා + අන්ගෙන් *yahaluwa + angen* යහළුවන්ගෙන් *yahaluwangen*
vocative	යහළුවා + ඒ *yahaluwa + e* යහළුවේ *yahaluwe*	-na-	යහළුවා + අනේ *yahaluwa + ane* යහළුවනේ *yahaluwane*

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

Inanimate noun [example *pasela* (School)]

Case forms	Singular Definite	Singular Indefinite	Plural
direct	පාසැල + Ø *pasela + Ø*	පාසැල + අක් *pasela + ak* පාසැලක් *paselak*	පාසැල + (-අ) *pasela +(a)* පාසැල් *pasel*
dative	පාසැල + අට *pasela + ata* පාසැලට *paselata*	පාසැල + අකට *pasela + akata* පාසැලකට *paselakata*	පාසැල් + වලට *pasel + walata* පාසැල්වලට *paselwalata*
genitive	පාසැල + එහි *pasela + ehi* පාසැලෙහි *paselehi*	පාසැල + අක *pasela + aka* පාසැලක *paselaka*	පාසැල් + වල *pasel + wala* පාසැල්වල *paselwala*
instrumental	පාසැල + එන් *pasela + en* පාසැලෙන් *paselen*	පාසැල + අකින් *pasela + akin* පාසැලකින් *paselakin*	පාසැල් + වලින් *pasel + walin* පාසැල්වලින් *paselwalin*

a The irregular plural nouns were not be inserted

$$n \times 27 \qquad \dots\dots\dots 1$$

$$n + 27 + 16 \qquad \dots\dots\dots 2$$

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.[6] 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[7].

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[8]. 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*), න්ව(*nwa*), එන්(*en*), කින්(*kin*), වලින්(*valin*), අක්(*ak*), එක්(*ek*), ඔ(*o*), එහි(*ehi*), ක(*ka*), වල(*wala*), Ø }. 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* <u>*wa*</u> *kare* <u>*ni*</u> *tegami* <u>*wo*</u> *yūbin* <u>*de*</u> *okuru.*
   මම<u>Ø</u> ඔහුට ලියුමක් තැපැල් මගින් යවනවා.
   I send a letter to him by post.

2. *kare* <u>*wa*</u> *komugiko* <u>*wo*</u> *mizu* <u>*de*</u> *mazeru.*
   ඔහු<u>Ø</u> පිටිගුලිටිව<u>Ø</u> වතුර වලින් අනනවා.
   He mixes corn flour using water.

3. *watashi* <u>*wa*</u> *Tokyo shiten* <u>*no*</u> *Sakai-san* <u>*wo*</u> *yoku shitteimasu.*
   මම<u>Ø</u> ටෝකියෝ ශාඛාවෙහි සකයි මහතාව 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		Past	plural	Non-past		Past
	present	Future			present	future	
First	බලමි *balami*	බලන්නෙමි *balannemi*	බැලිමි *belimi*	First	බලමු *balamu*	බලන්නෙමු *balannemu*	බැලීමු *beleemu*
Second	බලහි *balahi*	බලන්නෙහි *balannehi*	බැලිහි *belihi*	Second	බලහු *balahu*	බලන්නහු *balannahu*	බැලුහු *baluuhu*
Third	බලයි *balai*	බලන්නේය *balanneya*	බැලීය *beleya*	Third	බලති *balathi*	බලන්නාහු *balannaahu*	බැලුහ *beluuha*

Table 7: Examples for Inflected verbs and optional function words

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

Table 8: Inflected verb forms of Sinhalese verbs

verb	Perfect	Progress	Adno -present	Adno -past	Formal -rqst
go	ගිහින් *gihin*	යමින් *yamin*	යත *yana*	ගිය *giya*	යන්න *yanna*
take	ගෙන *gena*	ගමින් *gamin*	ගන්න *ganna*	ගත් *gat*	ගන්න *ganna*
give	දීලා *deela*	දෙමින් *demin*	දෙන *dena*	දුන් *dun*	දෙන්න *denna*
walk	ඇවිදලා *evidala*	ඇවිදිමින් *evidimin*	ඇවිදින *evidina*	ඇවිද්ද *evidda*	ඇවිදින්න *evidinna*
shoot	විදලා *vidala*	විදිමින් *vidimin*	විදින *vidina*	විද්ද *vidda*	විදින්න *vidinna*
rise	නැගලා *nagala*	නගිමින් *nagimin*	නගින *nagina*	නැග්ග *nagga*	නගින්න *naginna*
measure	මැනලා *menala*	මනිමින් *manimin*	මනින *manina*	මැන්න *menna*	මනින්න *maninnna*
jump	පැනලා *penala*	පනිමින් *panimin*	පනින *panina*	පැන්න *penna*	පනින්න *paninna*
hit	ගහලා *gahala*	ගහමින් *gahamin*	ගහන *gahana*	ගැහුව *gehuwa*	ගහන්න *gahanna*
eat	කාලා *kala*	කමින් *kamin*	කත *kana*	කැව *kewa*	කන්න *kanna*

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

Japanese sentence	Sinhalese translation	Sinhalese Function words
tsuma wo aisuru	බිරිඳට ආදරය කරනවා	ට
([I] love [my] wife)	*birinda̲ta̲ adarakaranawa*	*ta*
kodomo wo tsurete iku	ළමයාව එක්ක යනවා	ව
(accompany the child)	*lamaya̲va̲ ekka yanawa*	*va*
hashi wo wataru	පාලමෙන් එගොඩ වෙනවා	එන්
(cross the bridge)	*palame̲n̲ godawenawa*	*en*
gohan wo taberu	බත් කනවා	
(eat rice)	*bath∅ kanawa*	∅
inu wo kau	බල්ලෙක් හදනවා	එක්
(a dog is bred)	*balle̲k̲ hadanawa*	*ek*

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

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

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 (*Taro wa michi wo aruku*) Taro walks in the road.	太郎は道を歩いていた(*Taro wa michi wo aru̲ite̲ ita̲*) Taro was walking in the road
彼は来る(*kare wa kuru*) He will come.	彼は来るでしょう(*kare wa kuru̲ deshō̲*) He would probably come.
窓を開ける(*mado wo akeru*) Window is opened	窓を開けてください(*mado wo ake̲te̲ kudasai̲*) Please open the window.
熊はジャンプする(*kuma wa jumpu suru*) Bear jumps.	熊はジャンプするらしい(*kuma wa juumpu suru rashii̲*) It seems that a bear jumps)
私はワインを飲む(*watashi wa wain wo nomu*) I drink wine.	私はワインを飲みたい(*watashi wa wain wo nomi̲tai̲*) I want to drink wine.
本を読む(*hon wo yomu*) The book is read.	本を読みやすかったでしょう(*hon wo yomi yasukatta̲ deshō̲*) The book was easy to 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 (*rareru*)	た (*ta*) (past)	ろう (*rō*)	て (*te*)
させる (*saseru*)	ている (*teiru*)	だろう (*darō*)	が (*ga*)
たい (*tai*)	ていた (*teita*)	らしい (*rashii*)	と (*to*)
なる (*naru*)	だ (*da*)	でしょう (*deshō*)	ように (*youni*)
てくれる (*tekureru*)	だった (*datta*)	そうだ (*souda*)	で (*de*)
てもらう (*temorau*)	である (*dearu*)	なければならない (*nakerebanaranai*)	でも (*demo*)
やすい (*yasui*)	です (*desu*)	かもしれない (*kamoshirenai*)	から (*kara*)
てほしい (*tehoshii*)	てきた (*tekita*)	ね (*ne*)	ので (*node*)
なくなる (*nakunaru*)	ます (*masu*)	べき (*beki*)	ば (*ba*)
にくい (*nikui*)	なかった (*nakatta*)	な (*na*)	ような (*youna*)
...	...	...	...

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。

本は 読み | やすかっ | た | でしょう

hon wa yomi   yasukat ta   deshō.

The book <u>was</u> <u>easy</u> to read.   <u>Wasn't it?</u>

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 tabe<u>rareru</u>.* (The fish is eaten by the whale.)

   තල්මහ <u>විසින්</u> මාළුවාව කනු ලබනවා (*talmaha <u>visin</u> maluwa<u>wa</u> <u>kanu</u> <u>labanawa.</u>*)

2. *sakana ga kujira ni tabe<u>rareteiruyôda.</u>* (The fish is probably being eaten by the whale.)

   තල්මහ <u>විසින්</u> මාළුවාව කනු ලැබුවා වගෙයි (*talmaha <u>visin</u> maluwa<u>wa</u> <u>kanu</u> <u>labuwawagei</u>).*

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 <u>nomasetakatta yôda</u>.

ඔහුට වයින් <u>පොවන්න</u> <u>අවශ්‍යවුනා</u> <u>වගෙයි</u>

(I) <u>probably have</u> him <u>drink</u> 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 <u>nomitakatta yôda.</u>

ඔහුට වයින් <u>බොන්න</u> <u>අවශ්‍යවුනා</u> <u>වගෙයි.</u>

He <u>probably wanted</u> to drink wine.

The Japanese phrase "*-takatta yôda*" is divided into *tai/ta/yôda* (want/ed/probably). The Sinhalese translations are generated from the translation rules for "*-tai*" (Table 13), "*-ta*" (Table 14) and "*yôda*" (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: හිරුවා).

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 terum 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
rareru	Formal-comm	ලබනවා	විසින්	ව
saseru	C-Present	–		ව/චි/ලවා
tai	Formal-rqst	අවශ්‍යයයි	ව	අක්
naru	Formal-rqst	අවශ්‍ය වෙනවා	.	
tekureru	Conj	දෙනවා		ව
temorau	Conj	ගන්නවා		
yasui	Formal-rqst	පහසුයි		
tehoshii	Conj	ගන්න අවශ්‍යයි	ලවා	
nikui	Formal-rqst	අපහසුයි	ව	
sugiru	Past	වැඩියි		
tekudasai	Conj	දෙමින් සිටිනවා		
saserareru	C-Passive	–	ව	විසින්
temoraitai	Conj	ගන්න අවශ්‍යයි		
temoraeru	Conj	ගතහැකිය		
teyaru	Conj	දෙනවා		
takunaru	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
ta	Past	C-Past		
teiru	Progress	C-Progress	සිටිනවා	
da	-	-	වේ	
nai	-	-		නො
teita	Progress	C-Progress	සිටියා	
data	-	-	විය	
dearu	-	-	සේ වෙයි	
desu	-	-	වෙයි	
tekita	Progress	C-Progress	ආවා	
masu	Present	C-Present		
nakatta	Past	C-Past		නො
dewanai	-	-	නොවෙයි	
deoru	Progress	C-Progress	සිටිනවා	
teinai	Progress	C-Progress	සිටිනවා	නො
dekiru	-	-	පුලුවන්	
mashita	Past	C-Past		
teiku	Progress	C-Progress	යනවා	
teimasu	Progress	C-Progress	සිටිනවා	
tekuru	Progress	C-Progress	එනවා	
teshimau	Conj	C-Conj	ඉවරවෙනවා	
dekinai	-	-	නොහැකියි	
nanoda	-	-		
teshimatta	Conj	C-Conj	ඉවරවුනා	
masen	Present	C-Present		නො
deshita			වුනා	
temiru	Progress	C-Progress	බලනවා	

Table 15 Translation rules for the Japanese function word group "Judgments"

Japanese function words	Vb-inflection	Judge-Verb
rō	Will-Vb	නොවේද
darō	Present	නොවේද
souda	Present	වගේය
deshō	Present	නොවේද
rashii	Present	වගෙයි
yōda	Past	වගෙයි
nakerebanaranai	Ending-NE	කළ යුතුය
kamoshirenai	Present	ද කිව නොහැක
ne	Present	නේද
beki	Adno-Present	යුතුය
bekida	Adno-Present	යුතුවේ
na	Present	වගේ

Example:

gohan wo <u>*tabete kara*</u> *gakko e iku.* {After eating, (I) go to school.}
කෑම කෑවා ට පසු පාසල් යනවා *keme* <u>*kewa*</u> <u>*ta pasu*</u> *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)
te	1	*	Conj	
ga	3	වුවත්		
to	4	ෝ	Past	
youni	2	ෙලස	DropLL	DropLA
de	1	*	Conj	
demo	2	නමුත්	DropLL	DropLA
kara	4	ට පසු	past	
node	2	නිසා	DropLL	DropLA
ba	1	ත්	IfCondit	
youna	2	ආකාරයේ	DropLL	DropLA
ra	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.)	යාළුවෙකුත් කුඩයක් 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.)	තරඟය එය පරාජයෙන් සමඟට නිමවූ වගෙයි
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 ඇයේ දුවට ඇඳුමක් දුවට මැසුවනු
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 යාළුවෙකුත් ප්‍රවෘත්තිය අසා ඉතාමත් පුදුමවෙනවා බීඨා වගෙයි
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.)	ආතා පළතුරු වර්ගකිරීමේ වැඩවල මහන්සිවෙත්ත බීඨා වගෙයි
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?)	මේ බෙහෙත් බොනවිට වේදනාව නැතිවෙයි
彼の名前は故郷で有名だった。(His name was famous in (his) hometown.)	ඔහුගේ නම ඌරත් සමඟ ප්‍රසිද්ධිය
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 කුරු පැටවකුත් සිටත් ලස්සන හංසයෙකුට වැඩුනා
君の夢はいつか実現するでしょう。(Will your dream come true someday?)	ඔබගේ සිහිනය කවදාහොත් ඇත්තවෙයි
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 ඇලුත් වාහනයක් මිලට ගත්තනු
先生は私たちに地理を教えてくださっている。(The teacher is teaching us geography.)	ගුරුවරයා අපිට භූගෝලය ඉගන්වනදෙමින් 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 ළඟට ගිහින් එම දත ගලවා ගත්තෙ
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.	මගේ සගයාට නියෝජනය කරලා තැන්වෙම කවුරුත් ඔබෙන් කතාවක් කරඥනට කැමතියි
魚が鯨に食べられた。(The fish was eaten by the whale.)	මාළුවා තිමිත් මගාවා 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[a]
(a.1)私は彼にたくさんお金を借り<u>ている</u>。 *watashi wa kare ni takusan okane wo* *kari<u>teiru</u>* (I borrow money from him in large quantities.)	මම ඔහුගෙන් මුදල් ගොඩක් ණයට ගනිමින් සිටිනවා (X) (The translation of *"teiru"* is in present continuous tense. But the simple present tense is required.) මම ඔහුගෙන් මුදල් ගොඩක් ණයට ගන්නවා (0)
(a.2)私は彼を私の恩人と考え<u>ている</u>。 *watashi wa kare wo watashi no nanijin* *to kangae<u>te iru</u>* (I think that he is my companion.)	මම ඔහුව මගේ සහායකයා කියා සිතමින් සිටිනවා (X) (same as error description of (a.1)) මම ඔහුව මගේ සහායකයා කියා <u>සිතනවා</u> (0)
(b.1)肉は三日後に腐っ<u>てしまった</u>。 *niku wa mikkakango ni kusatte <u>shimatta</u>* (Meat were rotten after 3 days)	මේ දින 3කින් පසුව තරකවෙලා <u>ඉවරයි</u> (X) (Translation of *"kusatta"* should be used for Sinhalese translations.) මේ දින 3කින් පසුව තරකවිය (0)
(b.2)川に落ちて服が濡れ<u>てしまった</u>。 *kawa ni ochite fuku ga mure<u>te shimatta</u>* ((I) fell into the river, and clothes were soaked.)	ගඟට වැටිලා ඇඳුම් තෙමිලා <u>ඉවරයි</u> (X) (Translation of *"mureteshimmatta"* is not appropriate. Translation of *"mureta"* is needed.) ගඟට වැටිලා ඇඳුම් <u>තෙමුනා</u> (0)
(c) 人々は勇ましい彼<u>を</u>愛している。 *hitibito wa isamashii kare <u>wo</u> aishiteiru* (People love brave him) N<u>を</u>愛する。 N *wo aisuru* (love N)	මිනිසුන් වීර ඔහුව ආදරය කරනවා (X) (*"wo"* makes different translations) N<u>ව</u> ආදරය කරනවා N<u>ට</u> ආදරය කරනවා (indefinite) N<u>ැට</u> 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 65ᵗʰ National Convention of Information Processing Society of Japan (IPSJ)*. 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 9ᵗʰ Natural Language Processing Conference*. Yokohama, Japan. pp. 477-480.