Web-based English to Sinhala Selected Texts Translation system

B. Hettige[1], A. S. Karunananda[2]
[1]Department of Statistics and Computer Science, Faculty of Applied Science,
University of Sri Jayewardenepura, Sri Lanka.
[2]Faculty of Information Technology, University of Moratuwa, Sri Lanka.
budditha@yahoo.com[1] , asoka@itfac.mrt.ac.lk[2]

***Abstract* – This paper presents English to Sinhala Machine Translation system that can translate selected English text into Sinhala through the web. This Translation system contains two modules, namely; web-based machine translation system and java based user interface. Core of the translation system runs on a web server and can be accessed by user interface. The core of the translation system contains seven modules, namely, English Morphological analyzer, English Parser, Translator, Sinhala Morphological generator, Sinhala parser, Transliteration module and three Lexicon Databases. Java based user interface provides a mechanism for on-demand translation of selected texts from an English document. This enables users to get translated a selected set of English sentences while reading a document.**

1. Introduction

Sinhala is a member of Indo Aryan family of languages and is the spoken language of the majority of Sri Lankans. Most Sri Lankan people use Sinhala as the spoken and written language. Their ability in Sinhala language usage is at a reasonable level. However their understanding and writing ability of the English language is not comprehensive. This is known as the language barrier that affects both acquisition and dissemination of knowledge.

Machine Translation (MT) is a process that translates one natural language into another. MT is a complex and signal task because it can be used as the solution to the language barrier. Therefore, we have been working on the development of English to Sinhala machine translation system.

In general, a machine translation system contains a source language morphological analyzer, a source language parser, translator, a target language morphological analyzer, a target language parser and several lexical dictionaries. The source language morphological analyzer analyzes a source language word and provides morphological information. The source language parser is a syntax analyzer that analyzes the source language sentences. A translator is used to translate a source language word into the target language. The target language morphological analyzer works as a generator and generates appropriate target language words for given grammatical information. Also the target language parser works as a composer and composes suitable target language sentences. Furthermore, any MT system needs a minimum of three dictionaries such as the source language dictionary, the bilingual dictionary and the target language dictionary. The source language morphological analyzer needs a source language dictionary for morphological analysis. A bilingual dictionary is used by the translator to translate the source language into the target language; and the target language morphological generator uses the target language dictionary to generate target language words.

Many Asian and European countries have already taken steps to develop machine translation systems. In the Asian region, Indians have developed a variety of machine translation systems, including Mantra(Machine assisted translation tool) [6], Matra [8], Anusaaraka [2], AngalaBarathi [4] and Angalahindi [3], Shakit[7] and UNL Based MT system [9]. METIS-II [10] , PLOENG [11], and MANOS [12] being some of the European Machine Translation systems. Among others, EDR [28] by the Japanese is one of the most successful machine translation systems in the world.

These translation systems use various approaches to machine translation, including, Human-Assisted Translation, Rule based Translation, Statistical Translation, Example-based and Knowledge-based Translation etc. However, due to various reasons associated with complexity of languages, for more than fifty five years, Machine Translation (MT) has been identified as one of the least achieved areas in computing. Most of these issues are associated with semantic handling in MT systems.

As for the English to Sinhala machine translation system, we have already developed the Sinhala parser [13], Sinhala morphological analyzer [14], Transliteration module [15] and three lexical databases [16]. The Sinhala parser and morphological analyzer have been tested through various applications such as Sinhala Chatbot [15] and Sinhala Sentence generator [13]. To test and update the English to Sinhala Machine Translation system we have introduced three prototype systems, namely; Human-Assisted machine translation system[18], human-interaction machine

translation system for online dictionary update[20] and web-based English to Sinhala MT system[21]. The human-assisted machine translation system uses intermediate-editing approach [19] to semantic handling. This system also introduces an intermediate editor to handle semantics of the sentence. The human-interaction system also uses intermediate-editing approach to semantic handling and it uses human interaction methods to update dictionaries within the translation. Web-based English to Sinhala MT system is developed to translate English text into Sinhala through the web.

This paper reports the development of a mechanism for on-demand translation of selected texts from an English document. This enables users to get translated a selected set of English sentences while reading a document. This system is an extension of our core machine translation system and it enables students and the general public to use our translation system.

The rest of this paper is organized as follows. Section 2 describes the overview of some existing machine translation systems. Section 3 gives design of the developed core machine translation system. Then section 4 briefly describes existing version of the English to Sinhala Machine translation system. Section 5 introduces the selected text translation system. Then section 6 elaborates on how the system works in practice. Finally, Section 7 concludes the paper with a note on further work.

2. Some Existing MT Systems

Machine Translation systems use various approaches for translation; including Human-Assisted translation, Rule based translation, Statistical translation and Example-based translation [24]. Human-Assisted machine translation shares the translation task between man and the machine. The rule based approach translates rule maps from source to the target language representations. The example-based machine translator uses the extended idea of translation memories and reuses existing translation fragments. The statistical machine translation approach is a popular approach that gives alternative possible translations and finds the most probable one in the target language. This method needs a large corpus of the target language. Compared with the existing approach, Human-Assisted machine translation is the most fundamental approach for the machine translation. This approach uses human knowledge to solve translation problems like semantic and multiword-expressions.

However, due to various reasons associated with the complexity of languages, over the past fifty five years, MT has been identified as one of the least developed areas in computing. Most of these issues are associated with semantic handling in machine translation systems. A large number of MT systems have been developed for many languages all over the world. Until recently, there were no initiatives taken to attempt machine translations for Sinhala language. Sinhala is an Indo Aryan language and certain Indian languages like Pali, Sanskrit and Tamil are close to the Sinhala language. Therefore, we need to study some existing MT systems; especially the ones developed for Indian languages.

At present Indians have developed a variety of machine translation systems such as Anusaaraka, Mantra, Angalahindi, Shakti, etc. The Anusaaraka [2] is a popular machine-aided translation system for Indian languages that makes text in one Indian language accessible in another Indian language. Also, this system uses Paninian Grammar (PG) model [1] for its language analysis. The Anusaaraka project has been developed to translate Punjabi, Bengali, Telugu, Kannada and Marathi languages into Hindi. The approach and the lexicon is general, but the system has mainly been applied to children's stories.

MANTRA [5] is one of the web-enabled machine translation systems that translate the English text into Hindi in a specified domain of personal administration, specifically gazette notifications, office orders, office memorandums and circulars. It uses Tree Adjoining Grammar (TAG) [24] for Parsing and Generation and bottom-up parsing algorithm to speed up the parser and online word addition and grammar creation, updating facilities. Angalahindi [3] is a web-based, English to Hindi Machine-Aided translation system. Its translation methodology has been developed by the author who translates all Indian languages to English.

Shakti[7] and UNL Based MT system [9] are two other web based machine translation systems used for translating English to Indian languages. Among others, Electronic Dictionary Research (EDR)[30], by Japanese, is the most successful machine translation system. This system has taken a knowledge-based approach in which the translation process is supported by several dictionaries and a huge corpus. While using the knowledge based approach, EDR is governed by a process of statistical MT. As compared with other MT systems, EDR is more than a mere translation system but provides lots of related information.

3. Design of the English to Sinhala MT System

Brief description of the core English to Sinhala machine translation system is given below. The core English to Sinhala MT system contains seven modules, namely; English morphological analyzer, English parser, translator, Sinhala morphological analyzer, Sinhala Parser, Transliteration module and three Lexical dictionaries namely the English dictionary, the Sinhala dictionary and the English Sinhala bilingual dictionary. Fig 1 shows the design of the English to Sinhala core machine

translation system with these modules. The main input of the system is English sentences and output of the system is translated Sinhala sentences. After reading the input sentence it analyzes the input by using English morphological analyzer and the English parser. Then the system finds the corresponding Sinhala word for the given input word. After that, the system generates suitable Sinhala words by using Sinhala morphological generator. Finally the system generates correspondent Sinhala sentences by using Sinhala parser. This is the basic design of the English to Sinhala machine translation system and it does not handle semantics of the sentence. Each component of the core system describes below.

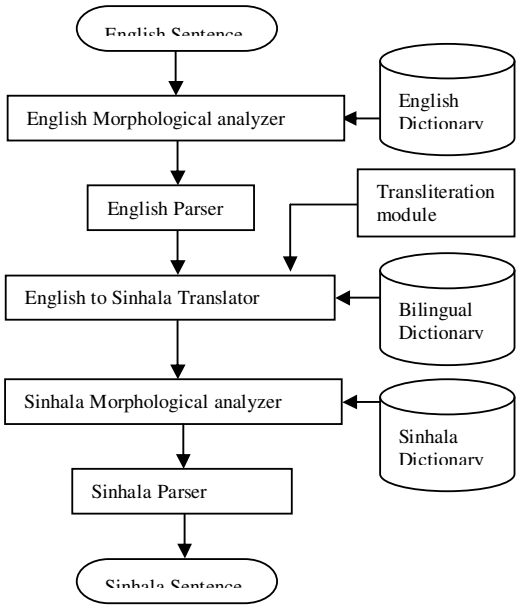

Fig. 1: Design of core English to Sinhala MT system

3.1 English Morphological analyzer

The English Morphological analyzer reads the given English sentence word-by-word and identifies morphological and lexical information on each word such as type of the word (Noun, Verb, Adjective, Adverb etc.), grammatical information for each word types such as (person, number and case for Noun, tense for verb etc.) There are many morphological analyzers available for the English language. Therefore, in this development, we have customized an existing English morphological analyzer. The morphological analyzer in our MT system has linked up with an English dictionary to get grammatical information on the words in the input sentence. SWI-Prolog [25]has been used to implement this morphological analyzer. The present English Morphological analyzer can identify all the Basic English word patterns (inflections). The prolog file named 'eng_morp_analyzer.pl' contains prolog based English Morphological analyzer and analyzeWords/1 prolog predicate is used to analyze a given word list. Prolog based English dictionary named 'eng_word_dic.pl' is used to analyzed the English words. Output result of the English Morphological analyzer is stored in a file named 'eng_morp_analyzer_out.pl". Result of the Morphological analysis is stored in a file and that can be read by others.

3.2 English Parser

The English parser receives source English sentences and the output result of the English morphological analyzer. This parser works as a syntax analyzer. Since there are many English parsers, we have customized an existing parser for our purpose. The current version of the parser used in our MT system is mainly concerned only with simple sentences. At present we have assumed that the input English sentence is grammatically correct therefore we do not need complex parser. The parser has also been implemented using SWI-PROLOG. The prolog file named 'eng_parser.pl' contains all in the English parser and 'analys_eng_sen/2' prolog predicate is used to analyze a given English sentence. Results of the parsing for a given English sentence is stored in a file named 'eng_parser_out.pl'. Note that, in the machine translation point of view, English parser identifies the English sentence and categorizes the sentence into sub parts such as subject, object verb and complement. This information is stored in a file named 'eng_parser_out.pl'.

3.3 Translator

The Translator is used to translate an English base word into a Sinhala base word with the help of the bilingual dictionary. The translator translates subject, object and verb in the English sentence separately. This is a method we have used to reduce the complexity of the translation process. This translator is a simple one and it does not automatically handle the semantics of sentences. Note that, this stage can be supported by human intervention to generate the most appropriate translation for some words in a sentence. As such, handling semantic, pragmatic and Multiword expressions must be addressed with the support of humans, for which we introduce an intermediate-editor. The prolog file named 'eng_sin_translator.pl' contains prolog based translator. Translator reads all the output results from the English Morphological analyzer and the English parser. Finally, result of the translation is stored in a file named as 'eng_sin_translator_out.pl'.

3.4 Sinhala Morphological Analyzer

The Sinhala Morphological analyzer works as a morphological generator. This morphological analyzer reads the words from the Translator word by word, all other information from English Morphological analyzer and English parser. For each word, the Sinhala morphological analyzer generates the appropriate Sinhala word with full grammatical information such as nama (nouns), kriya (verb) and nipatha (preposition) in the Sinhala language[22]. This morphological analyzer works with the help of three dictionaries, namely, Sinhala rule dictionary, Sinhala word dictionary and Sinhala concepts dictionary. All these dictionaries (prolog databases) and the morphological analyzer are implemented using SWI-Prolog.

3.5 Sinhala Parser

The Sinhala parser works as a Sentence composer. It receives tokenized words from the morphological analyzer and composes grammatically correct Sinhala sentences. Generally, a Sinhala sentence contains 5 components, namely, Ukktha visheshana (adjunct of subject), Ukkthya (Subject), karma visheshanaya (attributive adjunct of object), karmaya (object) and akkyanaya (verb) [28][29]. These five components of a Sinhala sentence are the building blocks for the design and implementation of a Sinhala parser. The parser is also one of the key modules of this English to Sinhala machine translation System and is also implemented using SWI-PROLOG.

3.6 Lexical Dictionaries

The translation system uses six dictionaries such as an English word dictionary, English concepts dictionary, English-Sinhala bilingual dictionary, Sinhala word dictionary, Sinhala rule dictionary and Sinhala concept dictionary. The English word dictionary contains English words and lexical information. The English concept dictionary contains synonyms, antonyms and general knowledge about English words. The English to Sinhala bilingual dictionary is used to identify the appropriate Sinhala base word for a given English word and contains the relation between English and Sinhala words. The Sinhala word dictionary stores Sinhala regular base words and lexical information. Similar to the English dictionary, the Sinhala concept dictionary stores semantic information. The Sinhala rule dictionary stores rules required to generate various word forms. These are the inflection rules for the formation of various forms of verbs and nouns from their base words. The rule dictionary also stores vowels, consonants, upasarga (prefixes) and vibakthi (case).

3.7 Transliteration module

The MT system needs to solve Out-of-vocabulary problems and handle technical terms. Machine transliteration can be used as a resalable solution for this. Transliteration is the practice of transcribing a word or text written in one writing system into another writing system [24]. In other words, machine transliteration is a method of automatic conversion of words in one language into phonetically equivalent ones in another language. At present we have developed two types of transliteration models. One of these models transliterates Original English text into Sinhala Transliteration and the other transliterate Sinhala words that are written in English which are transliterated into Sinhala. Finite State Transducers (FST) have been used to develop these two modules.

4. Development Stages of the English to Sinhala Machine Translation

The core of our MT system has seven modules, namely, English morphological analyzer, English parser, Translator, Sinhala morphological analyzer, Sinhala parser, Transliteration module and Lexical dictionaries. Our project has introduced the first ever parser and morphological analyzer for Sinhala language. Fig 2 shows the basic interface of our stand-alone MT system.

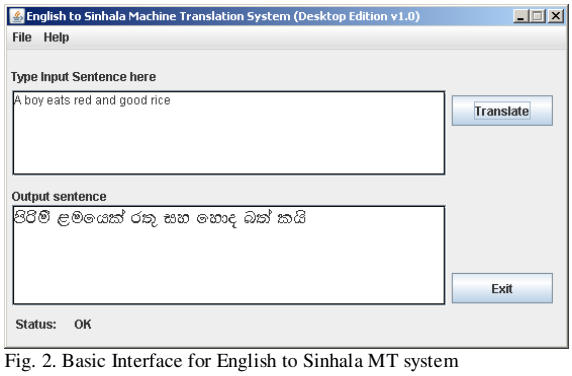

Fig. 2. Basic Interface for English to Sinhala MT system

This project has gone through various development stages. For instance, our English to Sinhala MT system has been improved to enable human-assisted translation [14]. This is a useful approach to improve the lexicon databases [13] and to identify limitations of the system, especially at the early developments. We have also improved our MT system with facilities to apply transliteration to handle pronouns and out-of-vocabulary problems during a translation process. This is essential for handling names of persons and cities, etc, for which there are corresponding Sinhala terms. In addition, we have gone beyond the traditional pre-editing and post-editing concepts in MT systems and introduced a new concept of intermediate-editing for MT

system. This concept effectively handles ambiguities in semantic, pragmatic and multiword expressions before proceeding to Sinhala linguistic modules in the MT system.

Fig 3 shows a screen shot of the Intermediate editor. This editor provides facilities such as display of synonym and antonym and related words. The intermediate-editor is linked with both English and Sinhala dictionaries in the MT system. The process of intermediate-editing, before composing a Sinhala sentence, drastically reduces computational costs of running a Sinhala morphological analyzer and parser. In addition, the requirement for post editing [23] can be reduced by the process of intermediate editing. On the other hand, intermediate-editing can be used as a means of continuous capturing of human expertise for machine translation. This knowledge can be reused for subsequent translations. It should be noted that the knowledge used for pre-editing and post-editing cannot be readily captured by a MT system, as these processes are generally done outside an MT system. In contrast, intermediate-editing will be an integral part of the MT system, in which humans directly interact with the system.

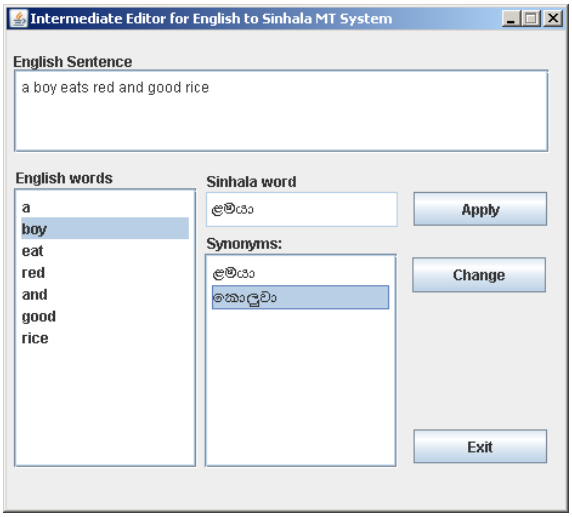

Fig 3: Intermediate editor for the English to Sinhala translation system

We have also improved our MT system with the report on the latest development to introduce web access for our MT system. The extension is primarily based on the use of Prolog sever pages. The architecture of the web-based English to Sinhala machine translation system is shown in Fig. 4.

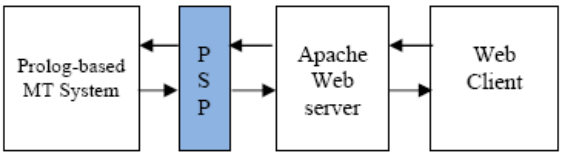

Fig 4. Architecture of the Web-based MT system

The web-based system contains four modules, namely; web client; Apache web server[27], PSP[26] module and Prolog based core translation system. Web browser is the user interface of the system. Apache web server handles all the web-based transactions of the system. PSP provides facilities to run Prolog-based system through the web. Prolog-based system is the core of the machine translation system. Through the PSP scripts, the core system reads input English sentence that comes from the web client. After the translation, the core MT system returns the output Sinhala sentence to the web client.

At This stage we need to test our machine translation system through the prototype working module. The present Parsers and Translator in the MT system have the following features; Handle Simple present tense, Handle determinants, Handle adjectives, Handle adverbs, Handle Compound subject ,Handle Compound adjectives, Handle Transliteration of proper nouns etc.

5. English to Sinhala Selected Text Translation System

English to Sinhala selected text translation system is designed by using web-based English to Sinhala MT System and java-based text selection tool. Web-based English to Sinhala translation system is the core system of the Machine translation system. The Java-based text selection tool is designed to translate English texts in to Sinhala by the help of web based English to Sinhala MT system. This selected text translation tool is an interface of the web based translation system. If the system is running, a translation tool appears in the system tray. To translate the selected text, you need copy it and click the translation icon on the system tray. Then text selection tool reads the clipboards and send the text into web based English to Sinhala MT system. Actually it reduces the access time of the system. Selected text translation tool is developed by using java and it is a simple java application. This java application accesses the English to Sinhala MT system through the internet.

6. How System works

In this section we describe, how the translation system works for a given input sentence. For example, suppose that the system accepts the input sentence: “A boy eats red rice for his lunch”. Then the English Morphological analyzer identifies each

word and returns the lexicon information of each word through the following Prolog predicates.

```
eng_detm([e1000001], id, 'a').
eng_noun([e1000006], td, sg, ma, sb, 'boy').
eng_verb([e1000009], sp, 'eats').
eng_verb([e1000014], pt, 'red').
eng_verb([e1000014], pp, 'red').
eng_adjv([e1000008], p, 'red').
eng_noun([e1000013], td, sg, no, sb, 'rice').
eng_prep([e1000027], v5, 'for').
eng_noun([e1000029], td, uc, ma, ob, 'his').
eng_noun([e1000028], td, uc, no, sb, 'lunch').
```

Note that eng_detm/3, eng_noun/6, eng_verb/3, eng_adjv/3 and eng_prep/3 prolog predicates are used to store English lexical information such as determination, noun, verb, adjective and preposition respectively. Furthermore, English morphological analyzer provides all the grammatical information for each word. For example, the English morphological analyzer identifies the word 'red' as an adjectival form, and past tense and past participle of the verb 'read'.

After that, the English parser reads the original English sentence together with the output of the Morphological analyzer. After this analysis, the parser returns the following information

```
eng_sentence_type(simple,sp).
eng_sen_verb([e1000009]).
eng_sen_complement([e1000027, e1000029, e1000028,
e1000008, e1000013]).
eng_sen_subject([e1000001, e1000006]).
eng_sen_ekeys([e1000001,..]).
```

Rest of the information is subject, verb and complement of the English sentence. Eng_sen_ekeys/1 is the key list that assigns for the each word for the sentence. These keys (Tokenized ID for English words) are used for further operation for the translation system.

Then tokenized IDs of English words are forwarded to the translator. The translator identifies the Sinhala base word for each English word in the sentence, with the help of the bilingual dictionary. The output of the translator is as follows.

```
estrwords(1001, e1000001, s1000000, dt).
estrwords(1002, e1000006, s1000014, na).
estrwords(1003, e1000009, s1000011, vb).
estrwords(1004, e1000027, s1000023, pp).
estrwords(1005, e1000029, s1000025, na).
estrwords(1006, e1000028, s1000024, na).
estrwords(1007, e1000008, s1000006, aj).
estrwords(1008, e1000013, s1000015, na).
```

estrwords/4 prolog predicate stores bilingual information for the each word. It stores the tokenized id, English base word key, Sinhala based word key and word type. After this, the Sinhala morphological analyzer reads these words and generates appropriate Sinhala words with all the grammatical information. Sinhala morphological analyzer plays a key role in handling Sinhala language specific features in the MT system.

When executing the Sinhala morphological analyzer, the system uses subject, verb and object based translation to generate the appropriate Sinhala words. This is mainly because of the effect of the determination and proposition in the sentence is different from these two languages and Sinhala sentence has strong subject verb relationship than English (Example: I eat rice 'මම බත් කමි', we eat rice 'අපි බත් කමු') Furthermore, English determinations come with a separate word which is not so in Sinhala.(Sinhala Noun shows person, number, case, sex, live, direct/indirect and case forms for word inflection) Some prepositions work seperately and some affect only the case of a noun (generally, 'to boy' is translate as 'ළමයාට' that affects only for the case). The output of the Sinhala Morphological analyzer is as follows.

```
snoun([s1000014], td, sg, ma, li, id, v1,'msrsus
<ufhla').
sin_sub_info([s1000014]).
sin_sub_word(['msrsus <ufhla', []]).
sin_fverb([s1000011], td, sg, pr,'lhs').
sin_veb_info([s1000011]).
sin_veb_word(['lhs', []]).
snoun([s1000025], td, uc, ma, li, dr, v2,'Tyqf.a').
snoun([s1000024], td, uc, no, nl, dr, v5,'osjd
wdydrh ioyd').
sin_adjv([s1000006],'r;=').
snoun([s1000015], td, sg, no, nl, dr, v1,'n;a').
sin_cmp_info([s1000025, s1000024, s1000006,
s1000015]).
sin_cmp_word([Tyqf.a', osjd wdydrh ioyd, r;=, n;a,
[]]).
```

Finally, the Sinhala parser composes the corresponding Sinhala sentence 'පිරිමි ළමයෙක් ඔහුගේ දිවා ආහාරය සඳහා රතු බත් කයි'. Figure 5 shows output of the web-based English to Sinhala machine translation system.

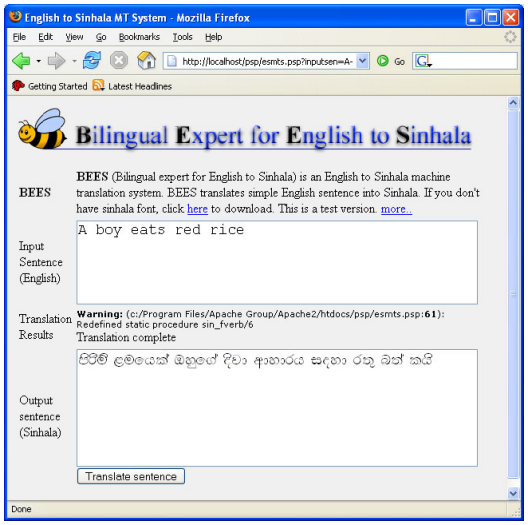

Fig. 5: web-based English to Sinhala machine translation system.

5. Conclusion and Further Works

This paper has reported on the ongoing project on English to Sinhala machine translation system and its current extension to provide selected text translation capabilities through web-based English to Sinhala Machine translation system. We have briefly described the developments of the previous work of this project by giving an emphasis on the modules of the core of our MT system. In this sense, we have reported on the first ever Sinhala Parser, Morphological analyzer, Intermediate editor and Human Assisted translator for English to Sinhala machine translation. More importantly we reported on the expansion to the core system thereby enabling the access to our MT system to a wider audience. This system will be useful for the general public, and students and teachers in particular. Further, work of this project has many directions. A system implemented with a powerful English parser and the WordNet lexicon database will be yet another further work of this project. In addition, the system will be improved to handle more complex English sentences.

References

[1] Askhar B, Chaitanya V, Sangal R, "Natural Language Processing:
A Paninian Perspective", Prentice Hall of India, New Delhi, India, 1995.

[2] Bharathi A, Chaitanya V, Kulkarni A. P, Sangal R., "Anusaaraka: Overcoming language barrier in India", to appear in "Anuvad: Approaches to Translation", Rukmini Bhaya Nair, (editor), Sage, New Delhi, 2001.

[3] Sinha R.M.K, Jain A., "AnglaHindi: an English to Hindi machine-aided translation system", MT Summit IX, New Orleans, USA, 23-27 September 2003; pp.494-497.

[4] Sinha R.M.K, "Integrating CAT and MT in AnglaBhart-II architecture", 10th EAMT conference, May. 2005, pp. 235-244.

[5] Durgesh R., "Machine Translation in India: A Brief Survey", National Centre for Software Technology, Mumbai, India. http://www.elda.org/en/proj/scalla/scalla2001/scalla2001Rao.pdf

[6] Mantra: web based Machine assisted translation tool for English to Hindi, http://www.cdac.in/html/aai/mantra.asp

[7] Shakti: A web based Translation system, http://shakti.iiit.net

[8] MaTra : English to Hindi MT system, http://202.141.152.9/ matra/index.jsp

[9] UNL Based MT System http://www.cfilt.iitb.ac.in /machinetranslation/eng-hindi-mt

[10] P. Dirix, I. Schuurman, V. Vandeghinste, "METIS-II: example based machine translation using monolingual corpora - system description", MT Summit X, Phuket, Thailand, September, Proceedings of Second Workshop on Example-Based Machine Translation; 2005, pp.43-50.

[11] J. Krzysztof, F. Gralinski, G. Krynicki: "POLENG - adjusting a rulebased Polish-English machine translation system by means of text corpus analysis", 5th EAMT Workshop "Harvesting existing resources", Ljubljana, Slovenia, pp. 47-50. 2000.

[12] B. Xu, Z.B. Chen, W. Wei, W.Pan, Z.D.Yang, "Phrase-based statistical machine translation for MANOS system". MT Summit X, Phuket, Thailand, September 13-15, Conference Proceedings: the tenth Machine Translation Summit: invited paper; 2005, pp.23-26.

[13] Hettige B., Karunananda A. S., "A Parser for Sinhala Language – First Step Towards English to Sihala Machine Translation", International Conference on Industrial and Information Systems (ICIIS2006), IEEE, Peradeniya, Sri Lanka, 8-11 Aug 2006, pp 583-587.

[14] Hettige B., Karunananda A. S., "A Morphological analyzer to enable English to Sinhala Machine Translation", Proceedings of the 2nd International Conference on Information and Automation (ICIA2006), Colombo, Sri Lanka, 2006 pp. 21-26.

[15] Hettige B., Karunananda A. S., "First Sinhala chatbot in action", Proceedings of the 3rd Annual Sessions of Sri Lanka Association for Artificial Intelligence(SLAAI), University of Moratuwa, 2006.

[16] Hettige B., Karunananda A. S., "Transliteration System for English to Sinhala Machine Translation", Industrial and Information Systems 2007, (ICIIS 2007)International Conference on Volume, Issue, 9-11 Aug. 2007, pp 209 – 214.

[17]Hettige B., Karunananda A. S., "Developing Lexicon Databases for English to Sinhala Machine Translation", Industrial and Information Systems 2007, (ICIIS 2007)International Conference on Volume, Issue, 9-11 Aug. 2007, pp 215 – 220.

[18] B. Hettige, A. S. Karunananda, "Using Human Assisted Machine Translation to overcome language barrier in Sri Lanka", proceedings of 4th Annual session of Sri Lanka Association for Artificial Intelligence, University of Moratuwa, 2007.

[19] Hettige B., Karunananda A. S., "Intermediate-Editing Approach for English to Sinhala Machine Translation", Proceedings of the 3rd International Conference on Information and Automation (ICIAFS07), Melborn, Australia, 2007

[20]Hettige B., Karunananda A. S., "Human Interactive NLP System for English to Sinhala Machine Translation", paper submitted to Industrial and Information Systems 2008, (ICIIS 2008) International Conference, India, 2008..

[21] Hettige B., Karunananda A. S., "Web-based English-Sinhala translator in action", paper submitted to 4th International Conference on Information and Automation (ICIAFS07), Colombo , Sri Lanka 2008

[22] A. M. Gunasekara, A Comprehensive Grammar of the Sinhalese Language", Asian Educational Services, New Delhi, Madras, India., 1999.

[23] Jeff L., Christopher H, "Toward the development of a post-editing module for Machine Translation raw output", Presented at the Third International Controlled Language Applications Workshop (CLAW2000), Washington, 2000.

[24]wikipedia, the free encyclopedia, http://en.wikipedia.org.

[25] SWI-PROLOG home page; http://www.swi-prolog.org

[26] Prolog server pages: http://193.226.6.174/roedunet2003/ site/conference/papers/SUCIU_A-Prolog_Server_Pages.pdf

[27] Apache web server: Apache software foundation,http://www.apache.org/

[28] S. Karunarathna, Sinahala Viharanaya, Washana prakasakayo, Dankotuwa, Sri Lanka., 2004.

[29] W. S. Karunathilaka, Sinahala Basha Viyakaranaya, M.D. Gunaseena & Company, Clolombo 11, Sri Lanka.,2003.

[30] Toshio Y, "The EDR electronic dictionary", Communications of the ACM, Volume 38, Issue 11 (Nov. 1995), pp. 42 – 44.