AN INTELLIGENT APPROACH IN LANGUAGE TRANSLATION

M. I. Weerawarna
Postgraduate Institute of Science
University of Peradeniya
Peradeniya
Sri Lanka

Sinhala is the language of around 13 million people living in Sri Lanka. It is not spoken in any other country. This study was conducted to find a method that can be used to translate the Sinhala sentence to its equivalent meaning in English sentence. In Sri Lanka many University students, employers find great difficulty in expressing their ideas in English and they are not confident about the using foreign languages. This will generally lead to a huge gap in the knowledge, usage and familiarity between their mother tongue and English language. In a small country, where the population and resources may be comparatively less, this language gap could have devastating effects on development. An accurate, effective and easy-to-use translating method could be one of the easiest solutions to this problem. Translation can play a vital role in bringing together the society by improving the understanding among the peoples.

Sri Lanka also faces this problem to a considerable extent. Sinhala is the native language of 74% of the Sri Lankan population and only used by Sri Lankans. It is more pronounced in rural areas than in main cities. An effective translating package may allow user to be under less stress, without worrying too much about linguistic errors, and focus more on the core requirement of the involved subject. To overcome this problem I devoted my time to find method that translates Sinhala to English.

Sinhala language belongs to large family of Indo-Aryan languages. The language has its unique writing script, which in turn has descended from Indian "*Brahmi*" script. Maldives, Dhivehi is considered as the closest relative of Sinhala. The oldest of Sinhala inscriptions belong to 6^{th} and 7^{th} century BCE [1]. Languages like Sinhala loosely structured and most of the time follows subject-object-verb (S-O-V) structure and it is very rich in synonyms and loan words (words from other languages) because of that Sinhala language is ambiguous to. Language is ambiguous when it can be understood in two or more possible

senses or ways. If the ambiguity is in a single word it is called lexical ambiguity. Ambiguity in a sentence or clause called structural ambiguity. Because of the ambiguity Sinhala translating is difficult.

In traditional language translation direct translation method is used. What it mean is translate the given input sentence to output sentence by translating word by word to Target Language (TL) words in input language (IL) find the matching meaning words in Target language and combined the words to form a sentence in Target language. It used word by word translation problem in this approach is, it rarely gives grammatically correct sentence and it does not convey the intended message to the recipient. Most of Translational systems are not accurate as professional translators.

To develop language translation system, **Machine Translation (MT)** techniques are used. These techniques will translate words or sentences in one natural language to another. The ideal aim of MT is to produce best possible translation without human aid. Machine Translation is one of the oldest applications that utilize **Natural Language Processing (NLP).** "Natural Language processing is the computerized approach analyzing text that is based on both a set of theories and set of technologies" [2]. The goal of NLP is to "accomplish human like language processing" [3]. Various levels of NLP have been utilized in Machine Translation systems, ranging from word based approach to higher levels [4]. For several decades dating back to the late 1940s, NLP has been one of the most active areas of research. Machine Translation was the first computer-based application developed under the field of NLP. The task of MT system is to translate the text or speech from one language into text or speech in another language. There are many approaches to MT that are roughly classified in two paradigms: Rule-based and Data-driven. In a classical Rule-base systems deep analysis of linguistic phenomenon of the given language pair is performed. Rule-base systems usually consist of a set of transformation rules written by human expert and MT engine, where linguistic knowledge is represented through that set of rules. Rule-based system involves three phases: analysis, transfer, and generation. Source sentence is analyzed using parsers and morphological tools, gets transformed into intermediate representation using the transfer rules, and then target language sentence is generated from the intermediate representation [5].

There are several techniques in Machine Translation. **Direct Translation, Interlingua Translation** and **Transfer Based Translation** are historical approaches that have been

used in Machine Translation [6]. Several new approaches were introduced during late 1980s and early 1990s.

In most modern methods in language translation to increase the accuracy of translational result, two or more combined approaches are used. The main method used in modern language translation is Rule based. When there are many words in a sentence pattern matching (Example–based) also used. I use Rule-based and Example-based methods to machine translation. If both methods are not successful system capable of identify in the most important words (nouns, proper nouns, verbs) in input sentence to construct output sentence with higher accuracy.

First input sentence was broken into words then corresponding English words were found with their classified groups (noun, verb and tense in bilingual database) if it was possible. It was checked by existing rules to construct output sentence. When input sentence was complicated it tries to identify most important words given in input sentence by omitting some words that does not harm to the meaning of input sentence or not much effect to its intended meaning in the input sentence.

The developed system was tested with simple sentences. According to the results obtained the system functions accurately when input sentences are simple. However the accuracy is not satisfactory for complicated sentences.