Existing Systems and Approaches for Machine Translation: A Review

B. Hettige¹, A. S. Karunananda²

¹Department of Statistics and Computer Science, Faculty of Applied Science, University of Sri Jayewardenepura, Sri Lanka.
budditha@yahoo.com

²Faculty of Information Technology, University of Moratuwa, Moratuwa
asoka@itfac.mrt.ac.lk

Abstract: The Machine Translation has been a branch of Natural Language Processing, which comes under the broad area of Artificial Intelligence. Machine Translation system refers to computer software that translates text or voice from one natural language into another with or without human assistance. Worldwide, large number of machine translation systems have been develop by using several approaches including human-assisted, rule-based, statistical, example-based, hybrid, agent based and etc. Among others, Statistical machine translation approach is by far the most widely-studied machine translation method in the field of machine translation. The multi-agent approach is a modern approach to handle complexity of the systems in past five years.

This paper reports on existing approaches for machine translation including human-assisted, rule-based, statistical, example-based, hybrid and agent based. Human-assisted machine translation approach is an approach for the machine translation particularly Indian families of machine translation. In the Indian region a number of machine translation systems have used this approach, including Anusaaraka, ManTra, MaTra, Angalabarathi etc. The Rule-based approach is yet another approach for machine translation. A number of machine translation systems have been designed through the rule-based approach. Among others Apertium is a rule-based Machine Translation system, which translates related languages. Statistical machine translation approach is by far the most widely-studied machine translation method in the field of natural language processing. Moses, Babel Fish, Bing Translator, Google Translator are the existing Statistical machine translation systems.

During the past few years many Sri Lankan researchers contributed to develop Machine Translation systems for local languages. As a first attempt Weersinghe and others have been researching to develop Sinhala to Tamil machine translation system through the corpus based approach. In addition to the above they have designed translation tool named OpenTM, which is based on the translation memories. Further, many other local researchers have developed several prototype English to Sinhala machine translation systems through several approaches. In 2003, Vithanage and others have developed English to Sinhala machine translation systems for weather forecasting domain. In 2008, Fernando and others have developed English to Sinhala machine translation system using Artificial Neural Networks and etc.

BEES (Bilingual Expert for English to Sinhala machine translation) is a rule-based English to Sinhala machine translation system which is runs through the concept of Varanegeema (conjugation) in Sinhala language. BEES uses 85 grammar rules for Sinhala nouns and 18 rules for Sinhala verbs for its word generation. The translation system contains seven modules, namely, English Morphological Analyzer, English Parser, English to Sinhala Base Word Translator, Sinhala Morphological Generator, Sinhala Parser, Transliteration module and Intermediate Editor. In addition to the above, system uses four lexical dictionaries namely, English dictionary, Sinhala dictionary, English-Sinhala Bilingual dictionary and Concept dictionary. The BEES successfully translates English sentences with simple or complex subjects and objects. The translation system successfully handles most commonly used patterns of the tenses including active and passive voice forms.