Sri Lanka Association for Artificial Intelligence (SLAAI)
Proceeding of the eighth Annual Sessions
20th December 2011 – Moratuwa

Existing Systems and Approaches for Machine Translation: A Review

B. Hettige[1], A. S. Karunananda[2], G. Rzevski[3]
[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 **–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 and agent based. 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 reviews existing machine translation approaches and systems including existing English to Sinhala machine translation systems.**

1. Introduction

The Natural Language Processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (Natural) languages [1]. It is also a sub field of Artificial Intelligence (AI) in the area of Computer Science. According to many electronic resources, the history of the Natural language processing began with the Turing article named "Computing Machinery and Intelligence" [2]. It is known as the Turing test as a criterion of intelligence. After that, In 1957 Noam Chomsky in the academic and scientific community as one of the fathers of modern linguistics, introduced the Syntactic Structures for grammar [3]. It is recognized as a most important text in the field of linguistics. After that, it becomes fundamental theory for Natural Language Processing and many of these Machine Translation systems use this syntactic structure.

The Natural language processing has come under broad area of the field of Artificial Intelligence. The NLP is used to do several tasks including machine translation, automatic summarization, Information retrieval, optical character recognition, speech recognition, text-to-speech and etc. Based on the task, the Natural Language Processing systems reserved several issues such as Natural language understanding, Natural language generation, Speech and text segmentation, Part-of-speech tagging and the Word sense disambiguation [4].

At present there are large number of machine translation systems have been develop to translate many related and none-related language fairs. These system are used several approaches including human-assisted, rule-based, statistical, hybrid and agent based etc. Ausaaraka, Google translator, SYSTRAN are the some existing successful machine translation systems in the world.

This paper reports overview of the existing machine translations including their approaches. The rest of the paper is organized as follows. Section 2 reports brief summary of the historical development of the machine translations. Section 3 reviews existing machine translation system under the machine Translation approaches. Section 4 reports existing English to Sinhala machine translation system including their features and limitations. Finally section 5 reports conclusions.

2. History: Machine Translation

Machine Translation system is a computer software to translate text or speech from one natural language to another [5]. The Machine translation is a sub area of the Natural language processing which is identified during early days of Artificial Intelligent (AI). Due to various reasons associated with complexity of languages, for more than last sixty years, Machine Translation has been identified as one of the least achieved areas in computing [6]. These issues range from Morphological to semantics of source and target languages.

The history of Machine Translation dates back to late 1940s. A look-up dictionary at Birkbeck College in London has been cited as an early work of machine translation in 1948. After that, 1950 to 1960 many researchers attended to develop Machine Translation systems by using trial-and-error approach [7] especially for Russian to English language. In 1950 first machine

translation system was developed to translate Russian sentences into English.

In 1958 first practical machine translation system was implemented by the IBM Corporation to US Air force under direction of Gilbet King [8]. This system translates Russian text into English and it successfully works until 1970. In the meantime RAND cooperation distributed current linguistic theory and emphasized the Statistical analysis. They were prepared bilingual glossaries with grammatical information and the grammar rules with the first parser based on the dependency of grammar.

In 1970, SYSTRAN [9] implemented a new Russian-English machine translation system which is the replacement of the previous system of the US Air force. This system translated more than 100000 pages per year. In the mean time, many researchers were attempting to develop machine translation systems. Among others, syntactic transfer system for English-French is one of the strong researches in the field. Further, principal experimental effect focused on the Interlingua approaches with more attention pays to the syntactic aspects [7].

In 1980, many computer companies attempted to develop computer-aided translations especially for Japanese-English. These systems are low level direct translation systems that are confined to morphological and syntactic analysis. After 1980 Machine translation researches were developed through many areas. Corpus- based machine translation approach is the most popular approach until now.

However, due to the complexity of the natural languages, development of the machine translation systems has become a research challenge. In addition, many researchers have also noted that, Operational syntax, idioms and Universal syntactic categories are some completely unsolved linguistic problems in the machine translation

2. Existing Approaches to Machine Translation

Considering the translation approaches, machine translation system can be classified into seven categories, namely, Human-assisted, Rule-based, Statistical, Example-based, Knowledge-based, Hybrid and Agent-based. Statistical, Example based, Knowledge based and Hybrid approaches are used copra for the machine translation. Therefore, these approaches are named as corpus-based approach. All of these machine translation approaches have their own strengths and weakness. Obviously, the success rate of a translation is depended on the approach. Each approach for the machine translation is discussed below.

A Human Assisted approach

Human-assisted machine translation approach is an approach for the machine translation particularly Indian families of machine translation. The human assisted approach uses human interaction for the pre editing, post editing and/or intermediate editing stages [10]. This approach uses human support for the semantic handling in the machine translation. Using this human assisted approach, numbers of machine translation systems have been developed In the Indian region a number of machine translation systems have used this approach, including Anusaaraka, ManTra, MaTra, Angalabarathi etc.

Anusaaraka [13] is a popular Human-assisted translation system for Indian languages that makes text in one Indian language accessible to another Indian language. This system uses Paninian Grammar model [14] to its language analysis. The Anusaaraka project has been developed to translate Punjabi, Bengali, Telugu, Kannada and Marathi languages into Hindi. English-Hindi Anusaaraka translates English text into Hindi. The approach and lexicon is general, but the system has mainly been applied for children's stories. [12][13]

MaTra [15] is a human-assisted transfer-based translation system for English to Hindi [11]. This System uses general-purpose lexicons and applied mainly in the domains of news. MaTra follows a structural and lexical transfer approach for its machine translation. The MaTra aims to produce understandable output for wide coverage, rather than perfect output for a limited range of sentences [16].

Mantra [17] is a machine assisted translation tool that, translates English text into Hindi in several domains. ManTra is based on the Tree Adjoining Grammar (TAG). The Mantra system was started with the translation of administrative documents such as appointment letters, notification and circular issued in central government from English to Hindi.

Angalabharti [18] is also a human-assisted machine translation system used in India. Since India has many languages, there are a variety of machine translation systems. For example, Angalahindi [19] translates English to Hindi using machine- aided translation methodology. Human-aided machine translation approach is a common feature of most Indian machine translation systems. In addition, these systems also use the concepts of both pre-editing and post-editing as the means of human intervention in the machine translation system.

Chandrashekhar Research Centre [20] has developed a machine aided translation system for Tamil to Hindi .

Tamil to Hindi translator is based on Anusaaraka Machine Translation System and the input text is in Tamil and the output can be seen in a Hindi text. Stand-

alone, API and Web-based on-line versions are developed. Tamil morphological analyzer and Tamil-Hindi bilingual dictionary are the byproducts of this system [19].

In addition to the above, KSHALT is a human assisted Machine Translation system that translates English to Korean language [21]. This translation system contains four phrases namely English Parser, English Analyzer, English to Korean transfer and the Korean generation

B Rule Based Approach

The Rule-based approach is yet another approach for machine translation. This approach gives grammatical correct translation by using set of rules. Basically, the rule-based machine translation system contains a source language morphological analyzer, a source language parser, translator, target language morphological analyzer, target language parser and several lexicon dictionaries. Source language morphological analyzer analyzes a source language word and provides the morphological information. Source language parser is a syntax analyzer that analyzes source language sentences. Translator is used to translate a source language word into target language. Target language morphological analyzer works as a generator and it generates appropriate target language words for the given grammatical information. Also target language parser works as a composer and it composes a suitable target language sentence. Furthermore, this type of machine translation system needs minimum of three dictionaries namely the source language dictionary, the bilingual dictionary and the target language dictionary. Source language morphological analyzer needs a source language dictionary for morphological analysis. Bilingual dictionary is used by the translator for translating source language into target language; and the target language morphological generator uses the target language dictionary to generate target language words.

A number of machine translation systems have been designed through the rule-based approach. Among others Apertium [22] is a rule-based Machine Translation system, which translates related languages. This is an open–source system that can be used to translate any related two languages. The Apertium engine follows a shallow transfer approach and consists of the eight pipelined modules, such as de-formatter, A morphological analyzer, A parts-of-speech (PoS) tagger, A lexical transfer module, A structural transfer module, A morphological generator, A post-generator, and A re-formatter

Toshiba [23] is another Rule-based Machine translation system for English to Japanese vice versa. To translate a given source text, system uses Morphological analysis, Syntax analysis, translation word selection and structural transformation, syntax transformation and morphological generation steps. This system can translate open-domain written texts by using rule-based. This system uses three dictionaries namely common word dictionary, a technical-term dictionary and a user-defined dictionary. The common word dictionary includes both English-Japanese and Japanese- English translation. The technical term dictionary includes domain-specific technical terms. They have used user defined dictionary to store user provided information such as unknown word information. Further, rule-based machine translation approaches can be categorized as three groups namely transfer-based, Interlingua and dictionary based. The transfer based and Interlingua approach has same idea for translation. Both two approaches used intermediate representation that captures the "meaning" of the original sentence [24]. The difference between both approaches is the interlingua-based system uses language independent intermediate representation and transfer-based system uses language dependent intermediate representation. Most of these machine translation systems include Morphological analysis, lexical categorization, lexical transfer, Structural transfer and Morphological generation. The dictionary based machine translation system uses dictionary for its machine translation with or without Morphological or syntax analysis. These type of Machine Translation systems ideally suitable to translate long lists of phrases. Numbers of machine translation systems have been developed under the above three border headings

Lavie and others [25] have applied transfer based approach to the Hindi-to-English translation system named Xferand. It trained under the extremely limited data scenario. This Xfer system uses IIITMorpher (Morphological analyzer) [26] to analyze Hindi words with the root and the other features such as gender, number, and tense. The Xfer system uses 70 transfer rules including a rather large verb paradigm, with 58 verb sequence rules, ten recursive noun phrase rules and two prepositional phrase rules. They have noted that, this approach is particularly suitable for languages with very limited data resources. Arabic to English machine translation system has been developed through the Transfer-based approach [27]. This system is named as Npae-Rbmt. The Npae-Rbmt is used an intermediate representation that captures the “meaning” of the original sentence in order to generate the correct translation. This system has evaluated through the 88 thesis titles and journals from the computer science domain. The accuracy of the result was 94.6%.

Apertium platform follows a transfer-based machine translation model. Using these shallow-transfer approach Swedish to Danish machine translation system has been developed [28]. Swedish to Danish machine translation system uses two morphological

dictionaries to analysis and generation. This is the first free software translator of Swedish to Danish. Using Affix-Transfer-based approach, Tagalog-to-Cebuano [29] Unidirectional Machine Translator system has been developed. The morphological analysis is based on TagSA (Tagalog Stemming Algorithm) and is focused on an affix correspondence-based POS (parts-of-speech) tagger.

Opentrad is an open source transfer based Machine translation system intended for related language pairs and not so similar pairs [30]. The Opentrad uses different translation methods according to each language pair. For related languages it uses shallow transfer, even though for nonrelated pairs the system uses deep transfer. Opentrad also uses open-source machine translation engine[31] (Matxin) as the translation engine.

OpenLogos is the Open Source version of the Logos Machine Translation System [32]. It is one of the earliest and longest running commercial machine translation products in the world. This system accepts documents in various formats and produces high quality translations [33]. OpenLogos translates from English and German to the major European languages, including Spanish, Italian, French and Portugese.

C Interlingua Machine Translation

The Interlingua approach gives language independent meaning representation for the source language to target language translation. The Interlingua gives one single meaning representation for all the languages and it has been reserved as an extremely difficult task in practice [34]. However, there are several advantages in the Interlingua approach. Among others Interlingua gives more easy way to adding new language than all other methods. Also it seems several disadvantages. Meaning representation is the critical approach in Interlingua. If the meaning is too simple then meaning will be lost in the translation. On the other hand it is too complex and analysis and generation will be too difficult.

Numbers of Machine translation system have been developed through the Interlingua approach. Abdelhadi and others have been developed English to Arabic machine translation system based on Interlingua approach [35]. They have used mapping system to Arabic to intermediate representation. This mapping system contains three steps namely, selecting lexical items for each Interlingua concepts, mapping the semantic roles and mapping the semantic features for each Interlingua concept to appropriate syntactic feature in the feature structure.

Among others ICENT is the interlingua-based Chinese-English natural language translation system [36]. This system introduces the realization mechanism of Chinese language analysis, which contains syntactic parsing and semantic analyzing and gives the design of Interlingua in details.

Tai to English machine translation system is another successful machine translation system for Tai to English [37]. This system translates the Thai sentences into Interlingua of a Thai LFG tree using LFG grammar and a bottom up parser.

D Dictionary based Machine Translation

The dictionary based machine translation systems are commonly used for cross-language retrieval systems [38]. This dictionary based approach uses dictionary-based method to generate the equivalent target query for the given source language query.

Mandal and others [39] have been developed a cross-language retrieval system for the retrieval of English documents in response to queries in Bengali and Hindi. This dictionary-based machine translation system uses to generate the equivalent English query out of Indian language topics.

Thenmozhi and Aravindan have been developed Tamil-English Cross Lingual Information Retrieval System for Agriculture Society [40]. This system developed for the Farmers of Tamil Nadu which helps them to specify their information need in Tamil and to retrieve the documents in English. It uses a Morphological Analyzer to obtain the root terms of source query. This Machine Translation approach retrieves the pages with mean average precision of 95%

Statistical machine translation approach is by far the most widely-studied machine translation method in the field of natural language processing. This approach tries to generate translations using statistical methods based on bilingual text corpora [41]. Using this statistical approach, large numbers of machine translation systems have been developed.

Moses is a Statistical machine translation system that allows automatically train translation models for any language pair [42]. The Moses system has several features. It offers two types of translation models namely, phrase-based and tree- based. Moses system uses factored translation models, which enable the integration linguistic and other information at the word level.

Babel Fish [43] is a web-based application developed by AltaVista which translates text or web pages from one language into another. The translation technology for Babel Fish is provided by SYSTRAN [44], whose technology also powers the translator at Google and a number of other sites. It can translate among English, Simplified Chinese, Traditional Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Portuguese, Russian, and Spanish. A number of sites have sprung up that used the Babel Fish service to

translate back and forth between one or more languages.

Bing Translator [45] is a service provided by Microsoft as part of its Bing services which allow users to translate texts or entire web pages into different languages. All translation pairs are powered by Microsoft Translation, developed by Microsoft Research; it uses Microsoft's own syntax-based statistical machine translation technology.

Google Translator [46] translates a section of text, or a webpage, into another language. It does not always deliver accurate translations and does not apply grammatical rules, since its algorithms are based on statistical analysis rather than traditional rule-based analysis.

In the Indian region, Udupa and Faruquie have developed an English-Hindi Statistical Machine Translation System [47]. This machine translation system is based on IBM Models 1, 2, and 3. The system has been tested through the English-Hindi parallel corpus consist of 150,000 sentence pairs.

Singh and Bandyopadhyay have been developed Manipuri-English bidirectional statistical machine translation system [19]. The system uses four useful translation factors namely case markers and POS tags information at the source side and suffixes and dependency relations at the target side. This translation system has been evaluated through the BLEU score.

E Example-based Machine Translation

The example-based machine translation system uses bilingual corpus with the parcel text for the machine translation. These systems are trained through the bilingual parallel copra, which contain sentence pairs. The example based approach is more useful for detecting the context from the text. Also this approach uses translation memories [48]. Using this approach number of machine translation systems have been developed all over the world. Among others, OpenMaTrExis one of the open source Example-based machine translation systems which is freely available on the OpenMaTrEx web site [49].

OpenMaTrEx has been developed through the marker hypothesis, which is compressed on marker-driven chunker, a collection of chunk aligners and two engines.

Kyoto-U is a successful Example based machine translation system that translates English-Japanese [50]. This system uses a morphological analyzer and dependency analyzer to detect Japanese sentence structures and converted into dependency structures. In addition, Japanese and English parsers and bilingual dictionary were used as external resources.

At present many researchers are researching to develop example-based machine translation systems by using World Wide Web as parallel corpora [51]. The wEBMT is an example-based machine translation (EBMT) system that uses the World Wide Web as the parallel corpus.

F Knowledge-based Machine Translation

Knowledge-based machine translation approach uses knowledge for machine translation. This is an extended idea of the example-based machine translation. This approach uses linguistic and computational instructions, which are supplied by a human. Numbers of commercial quality Machine Translation systems have used this knowledge-based approach. Among others EDR[52] and KANT [53] are the major knowledge-based machine translation systems.

EDR (Electronic Dictionary Research), 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 machine translation. As compared with other machine translation systems,

EDR is more than a mere translation system but provides lots of related information.

KANT (Knowledge-based Accurate Natural-language Translation) is a knowledge based machine translation system for specific domain. Prototype of the KANT architecture translates French, German, and Japanese successfully. KANT is currently being extended in a large-scale commercial application [118]. The KANT prototype has been implemented in the domain of technical electronics manuals, and translates from English to Japanese, French and German.

G Hybrid Machine Translation

The Hybrid machine translation system uses combine method in rule-based and Statistical machine translation approaches. This hybrid approach has several advantages.

Among others, SYSTRAN is the market leading provider of language translation software products and solutions for the desktop, enterprise and Internet that facilitate communication in 52 language combinations and in 20 vertical domains [54]. Introducing combination of self-learning and linguistic technologies SYSTRANS has been developed hybrid machine translation system [44] named as a SYSTEMS Enterprise server 7.

H Agent-based Machine Translation

Agent technology, more specifically multi-agent systems, have also been used to handle machine translations. This Multi-agent system provides tools for building artificial Complex Adaptive Systems [56].

In general any multi agent system contains four key components, namely Multi-Agent Engine, Virtual world, Ontology and Interfaces [57][58]. The multi agent engine provides a run time support for agents. The engine starts as the first step of the system. Virtual world is the environment of the multi agent systems. Using this Virtual world, agents are cooperated and competed with each other as they construct and modify the current scene. The Ontology contains conceptual problem domain knowledge of each agent. There are a number of NLP systems that have been developed using multi agent system technology. Most of these systems use agents to handle semantics in the translation.

Minakow and others [57] have developed a Multi Agent-based text understanding system for car insurance domain. This system uses Multi agent system based approach to understand a given text. The system uses four steps to text understanding namely morphological analysis, Syntax analysis, semantic analysis and pragmatic analysis. To analyze the whole text is divided into sentences. Then first three stages are applied to each sentence. After analyzing each paragraph text is passed to pragmatic analysis.

Stefanini and others have developed a Multi-agent based general Natural language processing system named Talisman [58]. Talisman agents can communicate with each other without the central control. These agents are able to directly exchange information using an interaction language. Linguistic agents are governed by a set of local rules. The TALISMAN deals with ambiguities and provides a distributed algorithm for conflict resolutions arising from uncertain information.

4. Existing Approaches for English to Sinhala Machine Translation

During the past few years many Sri Lankan researchers contributed to develop Machine Translation systems for local languages. Among others University of Colombo has recorded a significant research to develop English to Sinhala and Sinhala-Tamil machine translation system with several Local language resources such as Sinhala corpus [60][99], Sinhala text to Speech system [61], Parts of Speech Tagger[62] and OCR system for Sinhala language [63]. As a first attempt Weersinghe and others have been researching to develop Sinhala to Tamil machine translation system through the corpus based approach [64]. This translation system evaluates through the BLUE score matrix and reasonable result were achieved. At present they are researching to develop English to Sinhala machine translation system through the translation memories[66]. They have designed translation tool named OpenTM, which is based on the translation memories. They have mentioned that this OpenTM is suitable for any language pairs around the world, where at least one language requires complex script support. 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 [66]. Vithanage's translation system can translate simple sentences and works on the limited set of words and the limited sentence patterns. This translation system is fundamental rule-based and it has used Paragraphs and sentence tokenization, simple parsers (English and Sinhala), translators and Sinhala sentence generators for English to Sinhala translation.

In 2008, Fernando and others have developed English to Sinhala machine translation system using Artificial Neural Networks [67]. A Probabilistic Neural Network is used to identify the English grammar and it is based on Bayesian classifiers. This system has been achieved 50% accuracy in the grammatical translation. It has been tested through 84 test cases including 12 tenses and it only capable to translate only the simple sentences. In addition to above, some people all over the world have attempted to develop machine translation system for Sinhala. Among others, Hearth and others have attempted to develop translation system for Japanese to modern Sinhalese [68]. The system has a limited vocabulary and it handles translations only within its domain

BEES, an acronym for Bilingual Expert for English to Sinhala machine translation. The concept of Varanegeema (conjugation) in Sinhala language has been considered as the philosophical basis of this approach to the development of BEES. The Varanegeema in Sinhala language is able to handle large number of language primitives associated with nouns and verbs. For instance, Varanegeema handles the language primitives such as person, gender, tense, number, preposition and subjectivity/objectivity. More importantly, Varanegeema allows deriving all associated word forms from a given base word. This enables to drastically reduce the size of the Sinhala dictionary. Since the concept of Varanegeema can be expressed by a set of rules, it nicely goes with rule-based implementation of machine translation systems. BEES implements 85 grammar rules for Sinhala nouns and 18 rules for Sinhala verbs.

The English to Sinhala Machine Translation system has been designed and developed as a rule-based System. It 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. Figure 1 shows top-level design of the English to Sinhala machine translation system.

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

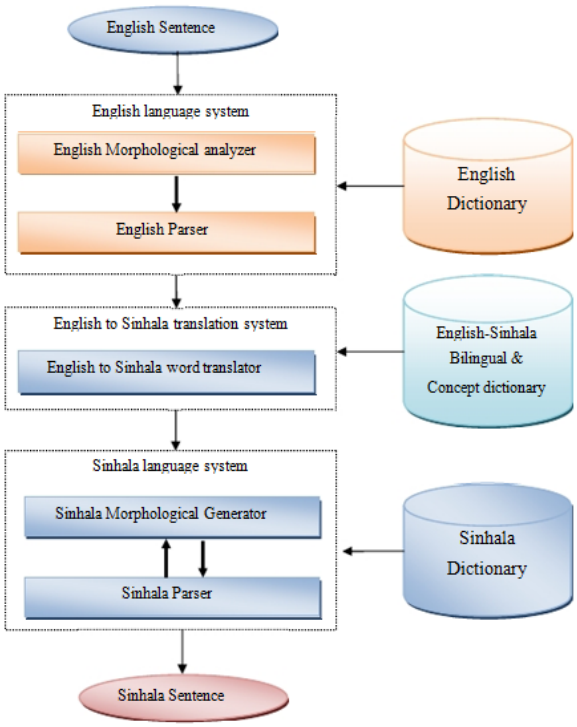

Figure 1: op-level design of the English to Sinhala machine translation system

5. Discussion

This paper has reviewed on existing approaches for machine translation including human-assisted, rule-based, statistical, example-based, hybrid and agent based. Compared with other approaches, Statistical machine translation approach is by far the most widely-studied machine translation approach in the field of natural language processing. However, this approach is more suitable for related language pairs and need to large parallel corpus. Further, many of these structural difference language pairs such as, English-Hindi, English-Japanese are used different approaches namely human-assisted rule-based and knowledge based. This is because, these types translation systems need more syntax and semantics information for the successful translation.

English and Sinhala languages are different language pairs and Sinhala is one of the related languages of the Hindi. Considering the existing English- to Hindi Translators (Anusaaraka, Angala Hindi) human assisted and rule-based approaches have been used. The Agent based approach is a modern approach for machine translation and it gives more fixable way to join two more approach for the translation process. Therefore, multi-agent approach is more suitable for English to Sinhala machine translation.

References

[1] B. Manaris, "Natural Language Processing: A Human-Computer Interaction Perspective", Appears in Advances in Computers (Marvin V. Zelkowitz, ed.), vol.47, pp. 1-66, Academic Press, New York, 1998.

[2] A. M. Turing, “Computing Machinery and IntelligenceMind”, New Series, Vol. 59, No. 236, pp 433-460.

[3] N. Chomsky "Syntactic Structures", Mouton, 1957.

[4] D. Jurafsky, J. H. Martin, "Speech and Language Processing", Boulder: University of Colorado, 2005.

[5] Wikipedia, Available: http://en.wikipedia.org

[6] J. Hutchins, "Machine Translation: past, present, future", New York : Halsted Press, 1986

[7] J. Hutchins, "Machine translation over fifty years", Published in: Histoire, Epistemologie, Langage, Tome XXII. 2001, pp. 7-31

[8] J. Hutchins, "Machine Translation: A Brief History", Oxford Pergamon Press,1995,pp 431-445.

[9] SYSTRAN, Available: http://www.systransoft.com

[10] S. Kang and et al, "An English to Korean System for Human Assisted Language Translation", TENCON 87, IEEE, 1987, pp 509-515.

[11] B. Akshar, C. Vineet, P. A. Kulkarni,R. Sangal,"Anusaaraka: Overcoming language barrier in India", New Delhi, India, 2001.

[12] S. Chaudhury, A. Rao, D. Sharma "Anusaaraka: An Expert system based MT System", In the proceedings of IEEE conference on Natural language processing and knowledge management. (IEEE-NLPKE), China, 2010.

[13] Anusaaraka, URL :http://anusaaraka.iiit.ac.in/

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

[15] MaTra, URL: http://www.cdacmumbai.in/matra/

[16] R. Ananthakrishnan, et al., "MaTra: A Practical Approach to Fully-Automatic Indicative English-Hindi Machine Translation", Inthe proceedings of MSPIL-06., 2006.

[17] ManTra, URL :http://mantra-rajbhasha.cdac.in/mantrarajbhasha

[18] R. Mahesh, K. Sinha, "Integrating CAT and MT in AnglaBhart-II architecture", 10th EAMT conference "Practical applications of machine translation". - 2005. pp. 235-244.

[19] K. D. Sanjay, P. S. Pramod, "Machine Translation System in Indian Perspectives", Journal of Computer Science, 1082-1087, 2010, pp 1082-1087.

[20] AU-KBC, Available: http://www.au-kbc.org/.

[21] S. Kang and et al, "An English to Korean System for Human Assisted Language, Translation", TENCON 87, IEEE, 1987, pp 509-515

[22] Apertium, URL: http://www.apertium.org/.

[23] I. Tatsuya, K. Akira, K. Yuka, "Toshiba Rule-Based Machine Translation System", NTCIR-7 PAT MT, Proceedings of NTCIR-7 Workshop Meeting, Japan, 2008, pp.430-434.

[24] I. Alegria, et al, "An Open Architecture for Transfer-based Machine Translation between Spanish and Basque", MT Summit, A workshop at Machine Translation Summit X, Thailand 2005.

[25] A. Lavie and et al, "Experiments with a Hindi-to-English Transfer-based MT System under a Miserly Data Scenario", ACM Transactions on Computational Logic, Vol.V, 2004.

[26] IIIT Available: http://www.iiit.net.

[27] S. Omar and et al., "Machine Translation of Noun Phrases from Arabic to English Using Transfer-Based Approach", Journal of Computer Science 6, pp 350-356.

[28] J. A. Perez-Ortiz, F. Sanchez-Martne and F. M. Tyers, "Shallow-transfer rule-based machine translation for Swedish to Danish", Proceedings of the First International Workshop on Free Open-Source Rule-Based Machine Translation. - Spain , 2009. -pp. 27-33

[29] J. A. Yara, "A Tagalog-to-Cebuano Affix-Transfer-Based Machine Translator", Proceedings of the 4th NNLPRS, 2007.

[30] I. Aizpurua, G. Ramirez, J. Pichel, J. Waliño, "Opentrad: bringing to the market open source based Machine Translators",Langtech, Rome, 2008.

[31] Matxin: an open-source transfer machine translation engine, URL: http://matxin.sourceforge.net/

[32] OpenLogos Machine Translation, Available: http://logos-os.dfki.de/

[33] B. Scott, A. Barreiro, "OpenLogos MT and the SAL representation language", In Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation, 2009, pp. 19-26.

[34] K. Shin-ichiro, M. Kazunori, "Interlingua Developed and Utilized inReal Multilingual MT Product Systems", AMTA/SIG-IL First Workshop on Interlinguas. – 1997

[35] S. Abdelhadi, C. Violetta, J. Abderrahim, "A prototype English-to-Arabic interlingua-based MT system", Third International Conference on Language Resources and Evaluation, Workshop Arabic language resources and evaluation: status and prospects, 2002.

[36] Q. H. Z. Xuan, C. Huowang, "An interlingua-based Chinese-English MT system", Journal of Computer Science and Technology Volume 17, 2002, pp 464-472.

[37] T. Chimsuk, S. Auwatanamongkol, "A Thai to English Machine Translation System using Thai LFG tree structure as Interlingua", World Academy of Science, Engineering and Technology. - 2009. pp 690-695.

[38] D. Hull, G. Grefenstette, "Querying across languages: A dictionary-based approach to multilingual information retrieval", In Proceedings of the 19th Annual Interna- tional ACM SIGIR Conference on Research and Development in Information Re- trieval, Zurich, Switzerland, 49–57, 1996

[39] D. Mandal, M. Gupta, S. Dandapat, P. Banerjee, S. Sarkar, "Bengali and Hindi to English CLIR Evaluation", Advances in Multilingual and Multimodal Information Retrieval, Lecture Notes in Computer Science, 2008, Volume 5152/2008, 95-102, 2008

[40] D. Thenmozhi, C. Aravindan, "Tamil-English Cross Lingual Information Retrieval System for Agriculture Society", International forum for Information Technology for Tamil (INFITT), Tamil International conference 2009.

[41] D. Jurafsky, J. H. Martin, "Speech and Language Processing", Boulder: University of Colorado, 2005.

[42] Moses: Available: http://www.statmt.org/moses/

[43] Yahoo Babel fish, 2008, Available: http://babelfish.yahoo.com/

[44] SYSTRAN, Available: http://www.systransoft.com.

[45] Microsoft Translator, 2010, Available: http://www.microsofttranslator.com

[46] Google Translator, Available: http://translate.google.com

[47] R. Udupa, A. Faruquie, "An English-Hindi Statistical Machine Translation System", Natural Language Processing, IJCNLP 2004.

[48] W. Andy, G. Nano, "wEBMT: Developing and Validating an Example-Based Machine Translation System Using the World Wide Web", Computational Linguistics Volume 29, Number 3, 2003, pp 421-457.

[49] OpenMaTrEx, Available: http://www.openmatrex.org

[50] T. Nakazawa, S. Kurohashi, "Kyoto-U: Syntactical EBMT System for NTCIR-7 Patent Translation Task", Proceedings of NTCIR-7 Workshop Meeting, Japan, 2008.

[51] G. Gregory, "The World Wide Web as a resource for example-based machine translation tasks", In Proceedings of the ASLIB Conference on Translating and the Computer, volume 21, London 1999.

[52] Y. Toshio, "The EDR electronic dictionary", Communications of the ACM, Volume 38, Issue 11, 1995, pp. 42-44

[53] KANT, Available: http://www.lti.cs.cmu.edu/Research/Kant

[54] F. Paris, D. San, "SYSTRAN Introduces hybrid machine translation solution for enterprises", SYSTRAN, 2009, Available: http://www.systran.co.uk/systran/news-and-events/press-release/hybrid-machine-translation-solution-for-enterprises

[55] G. Rzevski Home page: http://www.rzevski.net/

[56] G. Rzevski, J. Himoff, P. Skobelev, "MAGENTA Technology: A Family of Multi-Agent Intelligent Schedulers", conference on multi-agent systems in Karlsruhe,February 2006

[57] G. Rzevski, "A new direction of research into Artificial Intelligence", Sri Lanka Association for Artificial Intelligence 5th Annual Sessions. - 2008.

[58] M. H. Stefanini, Y. Demazeau, "TALISMAN: A multi-agent system for natural language processing", In Proceedings of SBIA'95. - Springer Verlag:, 1995, pp. 312-322

[59] LTRL, Language Technology Research Laboratory, Available:http://ucsc.cmb.ac.lk/ltrl/.

[60] A. R. Weerasinghe, D. Herath, V. Welgama, "Corpus-based Sinhala Lexicon", Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP, Singapore : ACL-IJCNLP, 2009 pp. 17-23.

[61] A. R. Weerasinghe., A. Wasala, K. Gamage, "A Rule Based Syllabification Algorithm for Sinhala", Proceedings of 2nd International Joint Conference on Natural Language Processing (IJCNLP-05), Korea, 2005, pp. 438-449.

[62] H. Dulip, R. Weerasinghe, "A Stochastic Part of Speech Tagger for Sinhala", Proceedings of 6th International Information Technology Conference. Colombo, Sri Lanka, 2004

[63] A. R. Weerasinghe, D.Herath, N. P. K.Medagoda, "A KNN based Algorithm for Printed Sinhala Character Recognition", Proceedings of 8th International Information Technology Conference. - Colombo, 2006.

[64] A. R. Weerasinghe, "A Statistical Machine Translation approach to Sinhala-Tamil language translation", SCALLA. - 2004.

[65] A.R.Weerasinghe and et al., "OpenTM: A Translation Memory System for Complex Script Languages", conference on Localized Systems and Applications CLSA2010, Kalutara, 2010. - pp. 72-73.

[66] N. V. C. Vithanage, "English to Sinhala Intelligent Translator for Weather forecasting domain", Colombo: Thesis submitted BIT degree, University of Colombo, Sri Lanka, 2003.

[67] B. T. L. Fernando, et al., "English to Sinhala language Translator using Artificial Neural Networks", PSLIIT Vol2., SLIIT, 2008, pp. 42-45

[68] A. Herath and et al., "A Practical Machine Translation, System from Japanese to Modern Sinhalese", The Logico-Linguistic Society of Japan.,1995.

[69] B. Hettige, A. S. Karunananda, "A Parser for Sinhala Language - First Step Towards English to Sihala Machine Translation", proceedings of International Conference on Industrial and Information Systems ICIIS, Colombo : IEEE, 2006.pp 583-587.

[70] B. Hettige, Bilingual Expert for English to Sinhala, URL: http://dscs.sjp.ac.lk/~budditha/bees.htm.

[71] B. Hettige, A. S. Karunananda, "Developing Lexicon Databases for English to Sinhala Machine Translation", proceedings of second International Conference on Industrial and Information Systems (ICIIS2007), Colombo, IEEE, 2007.

[72] B. Hettige, A. S. Karunananda, "Transliteration System for English to Sinhala Machine Translation", Proceedings of second International Conference on Industrial and Information Systems (ICIIS2007), Colombo: IEEE, 2007.

[73] B. Hettige, A. S. Karunnanda, "An Evaluation methodology for English to Sinhala machine translation", Accepted to present 6th International conference on Information and Automation foe Sustainability (ICIAfS 2010), IEEE., 2010

[74] B. Hettige, A. S. Karunananda, "Varanageema: A Theoretical basics for English to Sinhala", Accepted to present, 7th Annual Sessions of Sri Lanka Association for Artificial Intelligence (SLAAI), Kelaniya, 2010.

[75] B. Hettige, A. S. Karunananda, "Context-based approach to semantics handling in English to Sinhala Machine Translation", Poster presentation of the 26th National IT conference (NITC), Sri Lanka. - Colombo, 2009.