Machine Translation from English to Tamil using Hybrid Technique

Lakshmana Pandian S
Department of Computer Science & Engineering
Pondicherry Engineering College
Puducherry, India

Kumanan Kadhirvelu
Department of Computer Science & Engineering
Pondicherry Engineering College
Puducherry, India

ABSTRACT
The corpus based techniques in Machine Translation involves parallel corpora, but it is not applicable for the languages for which there are less or no parallel corpora available. In such case the Rule based machine Translation suits best. The main objective of our work is to build a translation system that translates English sentences to Tamil Sentences. Due to the less availability of parallel corpora for English to Tamil the system is implemented using a Hybrid Technique (the combination of both Rule Based Technique and Statistical Technique). The system is first implemented in a Rule Based approach which involves segmentation and tagging, Rule Based Reordering, Morphological Analyzing, and dictionary based translation to the Target language. Then the errors in the translated sentences are corrected by applying Statistical technique.

Keywords
Segmentation and tagging, Rule based reordering, chunking, Morphological analyzer.

1. INTRODUCTION
Machine Translation is a process of translating the sentences from one language to the other. In earlier days the machine translation is done only at the word level i.e. word by word translation. Later as the technology grows many researches have been carried out in this field and thus the translation at the phrase and sentence level has been implemented. The three major techniques involved machine translations are [1] Rule Based, Statistical, and Example Based Technique. The Statistical and Example Based Techniques needs parallel corpora for translation. But for south Indian languages such as Tamil, Telugu, Malayalam has only few parallel corpora available. In such cases adopting only the statistical technique will not result in proper translation to the target language.

In this paper we propose a hybrid technique (combination of Rule Based Technique and the Statistical Technique) to translate the sentences from English to Tamil language.The various challenges in this work are 1. Tamil language is morphologically quite rich and mostly follows SOV order but the order is flexible according to the sentences, 2. It is possible to construct meaningful sentences which are grammatically correct with one word, for example “odinaan” shows the tense, action performed, and the gender of the person performing that action. So implementing the gender ending is a difficult task, 3. Ambiguity problem [2] arise i.e. words that have more than one meaning when they occur in parts of speech cannot be translated accurately 4. No huge parallel corpora are available so that implementing it in the combination of both rule based technique and statistical technique gives more accuracy in translation.

2. RELATED WORK
Only few researches have been done in the case of English to Tamil translation. RemyaRajan, et.al [3] proposed English to Malayalam Translation system based on Rule Based technique. They used five key factors namely word dictionary file, Roman to Unicode file, Unicode to Roman file, Morph-dictionary file and Transfer link rule for translation. The above system works well for a sentence having not more than 6 words.

Poornima C, Dhanalakshmi V, et.al [4] proposed a preprocessing tool for machine translation that simplifies the complex sentences into simple sentences. This system uses rule based technique for sentence simplification and uses characters such as (‘,’, ‘?’) as delimiters for sentence separation. They have designed this system as a preprocessing tool for English to Tamil translation.

R. Harshawardhan, et.al [5] presented a novel framework for English to Tamil translation system. This is a phrase based translation using translation memory and concept labeling. The given input text are labeled and converted into phrases. This phrases are searched in the parallel corpus and the related phrases are extracted from it. Among the related phrases the best one is chosen as the target output sentence.

Rahul.C, et. al [6] proposed a system for translating from English to Malayalam which gives better performance than Phrase based system by avoiding parsing of the target language. First it reorders the source language to Malayalam language by using simplemodification in the English parse tree and then uses the morph analyzer. The result of the system shows that improvements could be possible by incorporating syntactic and morphological information.

S. Sripirakas, et.al [7] proposed a statistical based translation system from Sinhala –Tamil. They prepared parallel corpora from the parliament order papers. The decoder is tuned with the MERT technique in order to increase the quality of the output. Two other automatic techniques such as TER and NIST are used for evaluation instead of depending on BLEU.

3. SYSTEM OVERVIEW

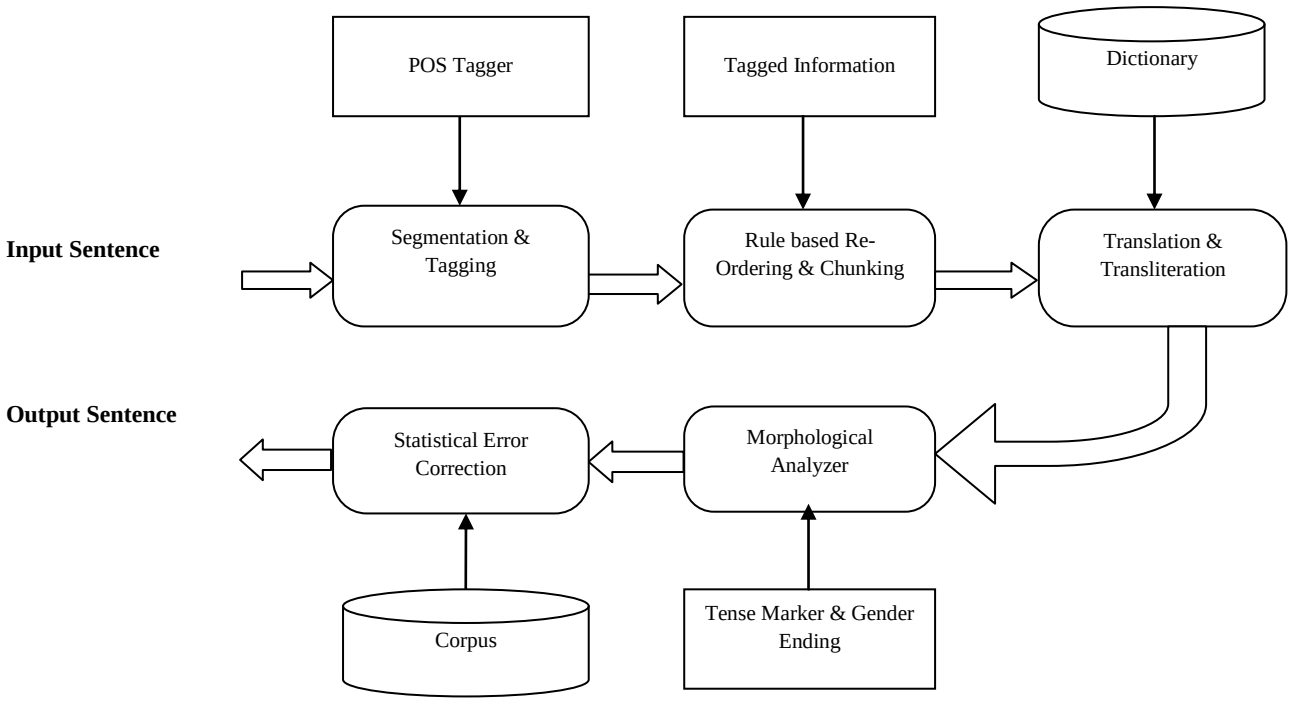

Fig 1: overview of proposed system

4. KEY FACTORS OF HYBRID TECHNIQUE

4.1 Segmentation and Tagging

The segmentation and tagging [8] of the source sentence is done using the PartsOf speech tagger. In our work we use Stanford POStaggerfor the tagging purpose. The English sentences are taken as a input to the parts of speech tagger. The tagger tokenizes each word in a sentence and identifies the parts of speech information such as verb, noun, adjective etc. of that word. Then the words and their tagged information are stored in a separate file which is used for reordering of sentences.

Tags and Their Abbreviation

TAG DEFINITIONS	examples
UH Interjection	oh, oops
PP Personal pronoun	I, you, she
WP Wh pronoun	who, what
WRB Wh-adverb	how, where
WDT Wh-determiner	which, that
NNP Proper noun	Edinburgh
PRP Pronoun	he, she
RB Adverb	quickly
DT Determiner	the, some
CC Conjunction	and, or
JJ Adjective	Big
JJS Adj- Superlative	Biggest
JJR Adj- Comparative	Bigger
WP Wh Pronoun	Who
IN Preposition	of, in
NN Noun	Apple, dog
NNS Noun Plural	Apples
TO	To
RP Particle	up, off
VB Verb	eat, play
VBP Verb-Present	eat, is, are
VBD Verb-Past	Ate, was, were
VBG Verb-gerund	eating
VBZ Verb- Present	eats
VBN Verb-Past part	eaten
MD modal	can, should

4.2 Rule Based Reordering

The tagged words are stored separately for the purpose of reordering according to the morphological structure of the Tamil language. The tagged words are arranged according to the order which is mentioned below.

UH/ PP/ WP/WRB/WDT/ NNP/ PRP/ RB/ DT /CC/ JJ/ PP$/ WP/JJR/ JJS/ IN/ NN/ NNS/TO/ VB/ VBD/ VBG/ VBN/ VBP/ VBZ/ MD

The above mentioned order suits all most all the types of simple sentence when reordering it from English to Tamil language.

Input sentence: he is going to college
Tagged sentence: he/PRP is/VBP going/VBG to/TO college/NN =>English pattern
Sentence reordered: he /PRP college/NN to/TO going/VBG is/VBP=> Tamil pattern
Output sentence:he college to going is

4.3 Chunking

The words in one language will not have equivalent words in another language. In some cases a word in one language is to be expressed by group of words in another. In such cases chunking of the source language has to be done. Chunking is

nothing but grouping of words into a phrase. A noun phrase contains an adjective and a noun, for example "sweetest apple". Here we chunk the words using bi-gram model [9]. Using this model possible bi-grams in a sentence are found and are translated into target language using the dictionary file.Consider an example "he bought a **wrist watch** yesterday". In this example the words "**wrist watch**" is chunked and grouped together as a single entity and then it is translated as **"கைக்கடிகாரம்"**.So for the above sentence the generated output sentence will be "அவன் நேற்று ஒரு **கைக்கடிகாரம்** வாங்கினான்".

4.4 Translation and Transliteration

The Translation is done with the help of word dictionary file. Our word dictionary file contains around 5000 words. . If a word in an input sentence does not exist in the dictionary file then it will be considered as a proper noun. So those words are transliterated to the target language. Transliteration is implemented using the rule based technique. Identifying the gender for the transliterated proper nouns is an impossible task. So for any proper noun that is transliterated the common gender ending in Tamil language for both gender "கிறார்"is applied. Let us consider an example sentence **"arun** is driving a car". For this the translated output sentence is **"அருன்** கார் ஓட்டி கொண்டிஇருக்**கிறார்**" which will be common for both genders.

4.5Morphological Analyzer

Once the sentences are translated into the target language the morphological analyzing [10] has to be done. Because in Tamil language a single word for example "ஒடுகிறான்" shows the tenses, gender, and the action performed which appears like a complete sentence. So the gender ending with the proper tenses has to be added with the verbs.

Consider the above example, the output after the reordering process will be "he college to going is". Here the "going" is the verb which is in the continuous tense. Here the root word for that word "going" has to be found which is "go". After finding the root word it will be easy to find the meaning in the dictionary.

With the help of the tags the gender ending and tense marker process is done.
Example of Gender Ending Rules

Rules Simple Continuous Tense

If verb=> VBG (verb gerund) then add "கொண்டுஇரு" after the verb.

If pronoun (PRP) = "he" and VBP= "is" then add "கிறான்" after it.

If pronoun (PRP) = "she" and VBP= "is" then add "கிறாள்" after it.

If pronoun (PRP) = "they" and VBP= "are" then add "கிறார்கள்" after it.
If pronoun (PRP) = "he" and VBD= "was" then add "ந்தான்"after it.

If pronoun (PRP) = "she" and VBD= "was" then add "ந்தாள்" after it.

If pronoun (PRP) = "it" and VBD= "is" then add "கிறது" after it.

If pronoun (PRP) = "it" and VBD= "was" then add "ந்தது" after it

If pronoun (PRP) = "they" and VBD= "were" then add "ந்தார்கள்" after it.

If personal pronoun (PP) = "I" and VBP= "am" then add "கிறேன்" after it.

If personal pronoun (PP) = "we" and VBP= "are" then add "கிறோம்" after it.

If personal pronoun (PP) = "I" andVBP= "was" then add "ந்தேன்"after it.

If personal pronoun (PP) = "we" and VBP= "were" then add "ந்தோம்" after it.

If To= "to" then append "க்கு" after the noun.

Rules for Simple Present Tense

If pronoun (PRP) = "he" and VBP= "is" then add "கிறான்" after it.
If pronoun (PRP) = "it" and VBD= "is" then add "கிறது" after it.
If pronoun (PRP) = "she" and VBP= "is" then add "கிறாள்" after it.

If pronoun (PRP) = "they" and VBP= "are" then add "கிறார்கள்" after it.

If personal pronoun (PP) = "I" and VBP= "am" then add "கிறேன்"after it.

If personal pronoun (PP) = "we" and VBP= "are" then add "கிறோம்" after it.

If To= "to" then append "க்கு" after the noun.

Rules for Simple Past Tense

If pronoun (PRP) = "he" and VBD= "was" then add "ந்தான்"after it.

If pronoun (PRP) = "it" and VBD= "was" then add "ந்தது" after it

If pronoun (PRP) = "she" and VBD= "was" then add "ந்தாள்" after it.

If pronoun (PRP) = "they" and VBD= "were" then add "ந்தார்கள்" after it.

If personal pronoun (PP) = "I" and VBP= "was" then add "ந்தேன்"after it.

If personal pronoun (PP) = "we" and VBP= "were" then add "ந்தோம்" after it.

If To= "to" then append "க்கு" after the noun.

Rules for Simple Future Tense

If Modal (MD) = "Will" or "Shall" or "could" or "should" Then

If pronoun (PRP) = "he" then add "வான்"after it.

If pronoun (PRP) = "she" then add "வாள்" after it.

If pronoun (PRP) = "they" add "வார்கள்" after it.

If personal pronoun (PP) = "I" then add "வேன்"after it.

If personal pronoun (PP) = "we" then add "வோம்" after

it
If pronoun (PRP) = "it" then add "உம்" after it

If To= "to" then append "க்கு" after the noun.

The above mentioned are the possible cases and their corresponding rules that are to be applied to the sentences in the cases of a simple sentence.

4.6 Error Correction Using Statistical Technique

Finally we get the target sentence for the corresponding source sentence. Even though we write gender ending rules, in some cases accurate verb with proper gender ending cannot be obtained for Tamil language. Particularly when writing rules for past tense sentences many contradictions arises. In such cases there occurs an error in the target sentence. For example consider a source sentence "he played cricket" for that we may get the target sentence as "அவன் கிரிகெட் விளையாடிதான்". Here we get the wrong gender ending verb "விளையாடிதான்" instead of correct one as "விளையாடினான்". So here we correct the errors statistically by maintaining a separate file that contains gender ending verbs in Tamil. We have around 70 base verbs in Tamil language with all possible gender ending and tenses. The final output is compared with that file and the word with maximum similarity in that file will be replaced in the target sentence. So the wrong word "விளையாடிதான்" will be replaced by the correct word "விளையாடினான்".so the final output will be "அவன் கிரிகெட் விளையாடினான்" which is the correct target sentence.

5. HOW THE SYSTEM WORKS STEPS IN TRANSLATION

Step 1: The given source sentence is parsed and tagged using POS tagger, the tagged information is stored in a separate file.
Step 2: The rule based reordering of the sentence has to be done in the above formulated order using the tagged information.
Step 3: Chunking of the source sentence has to be done using the bi-gram model, and the bi-grams are translated into Tamil language by means of a word dictionary file.

Step 4: Then the word by word translation has to be done with the bilingual dictionary and if a word does not exist in dictionary, it may be a proper noun which is to be transliterated to Tamil language.
Step 5: Then apply gender ending rules to get the target output sentence.
Step 6: The error in the target sentence is corrected statistically using the file which contains collection of Tamil verbs with proper tense and gender endings. Finally the Tamil sentence for the corresponding English sentence is generated.

6. RESULTS

We have used Stanford POS tagger for getting the Parts Of Speech information of the input sentences. So far we have created the word dictionary file which contains around 4500 words from English to Tamil. Gender ending rules for simple sentences such as simple present, simple past, simple future were created. Gender ending verbs for all possible tenses were created for the purpose of statistically correcting the errors in the output sentences. We have used Word-Net for stemming of the words. The system is implemented using java codes. Simple sentences with all possible tenses were worked correctly. Figure 2 shows the comparison of proposed system with the existing google translator and our system gives the better results than the existing translator.

Table 1: Sample output Comparison with Existing System

Input sentence	Google translation	My translation
I swim daily	நான் தினமும் நீந்த.	நான் தினமும் நீந்துகிறேன்
I am eating a sweet apple	நான் ஒரு இனிப்பு ஆப்பிள் சாப்பிட்டு கொண்டிருக்கிறேன்	நான் ஒரு இனிப்பான ஆப்பிள் சாப்பிடுகொண்டு இருகின்றேன்
I am the tallest boy	நான் உயரமான பையன் இருக்கிறேன்	நான் உயரமான பையன்
I was playing cricket yesterday	நான் நேற்று கிரிக்கெட் விளையாட இருந்தது	நான் கிரிக்கட் நேற்று விளையாடிகொண்டு இருந்தேன்
I give him a pen	நான் அவரை ஒரு பேனா கொடுக்க	நான் ஒரு அவனுக்கு பேனா கொடுகிறேன்

I am driving a car	நான் ஒரு காரில் நான்	நான் ஒரு கார் ஓட்டிகொண்டு இருகின்றேன்

7. EVALUATION METRICS

The automatic Evaluation for Machine Translation are BLEU (Bilingual Evaluation Under-Study), NIST (National Institute of Standards and Technology) WER (Word Error Rate) and METEOR.

BLEU[11] metric gives high correlation against human judgment and it is the one of the popular metrics in Machine Translation since it gives maximum correlation with the human translation. BLEU calculates the average score of the individual sentences to get the final score for the whole corpus. To compare the candidate translation and machine translation BLEU uses modified form of precision because Machine Translation system generates more words than they actually appear in the reference text. BLEU is not best suited to evaluate individual sentences and increase in BLEU score does not mean improved quality of translation.

NIST [12] is based on BLEU metric with some alterations. It calculates information content in the particular n-gram by giving high height to rarely appearing n-gram whereas BLEU gives equal weight to each n-gram precision and also calculation of brevity penalty differs from BLEU. The overall score of the translation in NIST is not affected by small variations in translation length.

WER [13] works at the word level instead of phoneme level which is derived from Levenshtein distance. It evaluates by arranging the candidate word sequence in accordance to the reference sequence using dynamic string alignment. The final score is based on number of insertions, number of deletions, number of substitutions, number of corrects, number of words in reference.

METEOR [14] is based on the harmonic mean of unigram precision and recall with precision weighted less than recall. Unlike other metrics it has features like stemming and synonym matching along with the standard word matching. It has been designed to overcome the problems in BLEU which means it gives high correlation with the human judgment at the sentence level whereas the BLEU works with corpus level.

Any metric should assign a quality score for a translation, so that the score would highly correlate with the human judgment. Correlation is found between the scores of automatic evaluation metrics with the scores of the human judgment as the humans are the end users of the translation. In our work BLEU metric does not suits well as it works with multiple reference translations. Since Tamil is a free order language the words can be arranged in more than one order to give the same meaning. So WER does not suit well as it mainly works on rearranging the sentences according to the reference translation. So in order to measure the quality of our target language we have considered the three main aspects namely meaning of the sentence, alignment and the correctness of the words. For each sentence the scores of the above aspects is calculated manually. Finally the average score is calculated for each sentence. The score ranges between 0 -1 where the score nearer to the value 1 shows the better translation quality.

We have taken six types of sentences namely simple present, simple past, simple future, present continuous, past continuous and noun phrase for the purpose of evaluation the translation quality. In each sentence type we have taken minimum of twenty five sentences to manually calculate the average score based on the above three aspects. We have compared our system with the existing Google translator. The tables are listed below.

Table 2 .Scores of existing system (Google Translator)

Sentence type	Average Alignment score	Average word score	Average meaning score	Total average score
Simple present	0.941	0.708	0.625	0.758
Simple past	0.941	0.801	0.717	0.819
Simple future	1.000	0.910	0.910	0.940
Present continuous	1.000	0.748	0.748	0.832
Past continuous	1.000	0.742	0.808	0.850
Noun	0.960	0.900	0.850	0.903

Table 3. Scores of proposed work (hybrid technique)

Sentence type	Average Alignment score	Average word score	Average meaning score	Total average score
Simple present	1.000	0.915	0.915	0.943
Simple past	1.000	0.915	0.915	0.943
Simple future	1.000	0.830	0.830	0.886
Present continuous	1.000	0.910	0.830	0.913
Past continuous	1.000	0.910	0.880	0.930
Noun	0.800	0.900	0.800	0.833

The above tables represent the average score of the sentence types for both Google translator and our work. Then the correlation coefficient for the existing and proposed work is calculated with the human judgment.

Here we use Spearman's rank Correlation [15][16] Coefficient tocalculate the correlation values for the above two systems. The formula for calculating Spearman's rank correlation coefficient is

$$(R) = 1 - \frac{6\sum d^2}{n^3 - n}$$

Where $d_i = x_i - y_i$, differences between the ranks of each observation on the two variables, n is the number of observations.

7.1 Spearman's rank Correlation Coefficient

The Spearman's rank correlation coefficient is calculated for both existing work and our work with the human judgment.

The below table shows the correlation coefficient values for Google translator and our Proposed Hybrid technique.

Table 4.Spearman's rank correlation coefficient

Sentence Type	Google Translator	Proposed Work
Simple Present	0.515	0.618
Simple Past	0.512	0.618
Present Continuous	0.539	0.609
Past Continuous	0.515	0.609
Noun Phrase	0.548	0.615
Future	0.612	0.612

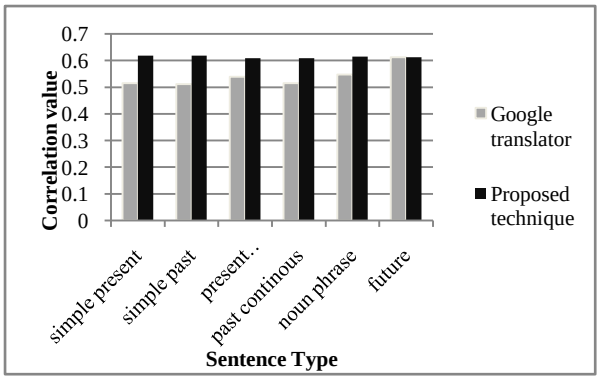

Figure 2 Comparison of proposed technique with existing work

A graph is plotted for the above values taking data sets that is sentence type such as simple present, simple past, present continuous , past continuous noun phrase and future sentences as the X axis and corresponding correlation values in the Y axis. The graph shows that our proposed technique for translation hasthe higher correlation with the human judgment than the existing Google translator.

7.2 ANOVA CALCULATION

Analysis of Variance calculation [17] is done between the correlation values of existing and proposed work to find whether the differences between these groups are significant are not.

Table 5.Correlation values

	Google Translator	Proposed
1	0.515	0.618
2	0.512	0.618
3	0.539	0.609
4	0.515	0.609
6	0.612	0.612
N	6	6
X	0.540	0.614
S	0.038	0.004
1	0.515	0.618

Where
nis the size of the group
S is the standard deviation
Xis the mean value of the group
Dfis the degrees of freedom
SS is the squares of the sum
MS square of the means

F test statistic is the ratio of two sampleVariances.

$$F = \frac{\text{Squares of mean (MS)of Between the group}}{\text{Squares of mean (MS)of within the group}}$$

F= 0.016 / 0.001

F(observed value) = 21.8955

F test is a right tail test. The F test statistic has an F distribution with df(B) numerator df and df(W) denominator df. F (critical value) is found using the F table.

F(1,10) (critical value) = 4.9646 at α = 0.05 (1/3)

Source	**Df**	**SS**	**MS**	**F**
Between the group	1	0.016	0.016	21.8955
Within the group	10	0.007	0.001	
Total	11	0.024		

Since the critical value of F is less than the observed value we reject the null hypothesis. The null hypothesis is that the means of the two groups in class were the same, but we reject

that, so at least one row has a different mean. That is means of the correlation values of Google translator and our proposed work are significantly different and our propose work has the higher correlation value with the human judgment than the existing system.

8. CONCLUSION

This paper presents an effective methodology for English to Tamil translation. Up to our knowledge the approach that we have adopted is a quite novel one. The work done based on this till now limits to translation of simple sentences from English to Tamil. The system can be further enhanced to translate complex sentences by creating new morphological reordering rules. The complex sentences can be split into two more simple sentences by using delimiters such a (',', '?') then it will be easy to translate those sentences based on our approach. Since a word in English has multiple meaning in Tamil, an effective word dictionary file is needed in order to achieve better results in translation**.**

9. REFERENCES

[1] Hutchins, W. John; and Harold L. Somers "An Introduction to Machine Translation" London: Academic Press, *pp. ISBN: 0-12-362830-x*. March 1992.

[2] Hui Yang, Anne de Roeck, Vincenzo Gervasi, Alistair Willis , and Bashar Nuseibeh, "Analysing anaphoric ambiguity in natural language requirements", © *Springer-Verlag London Limited*, 2011.

[3]RemyaRajan, Remya Sivan, RemyaRavindran and K.P Soman, "Rule Based Machine Translation from English to Malayalam", *in the proceedings of the International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp.439-441*, 2009.

[4] Poornima C , Dhanalakshmi V, Anand Kumar M and Soman K P, "Rule based Sentence Simplification for English to Tamil Machine Translation System", *International Journal of Computer Applications (0975 – 8887),Volume 25– No.8,* July 2011.

[5] R. Harshawardhan, Mridula Sara Augustine and K.P. Soman, "Phrase based English – Tamil Translation System by Concept Labeling using Translation Memory", *International Journal of Computer Applications (0975 – 8887), Volume 20– No.3*, April 2011.

[6] Rahul.C, Dinunath.K, RemyaRavindran, K.P.Soman, Department of Computational Engineering &NetworkingAmritaVishwaVidyapeetham, "Rule Based Reordering and Morphological Processing For English-Malayalam Statistical Machine Translation", *in the proceedings of the International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp.458-460*, 2009.

[7] S. Sripirakas, A. R. Weerasinghe, D. L. Herath, University of Colombo School of Computing, "Statistical Machine Translation of Systems for Sinhala - Tamil*", In the proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer), pp.62-68,* 2010.

[8]Felipe Sánchez-Martínez ,Juan Antonio Pérez-Ortiz and Mikel L . Forcada, "Using target-language information to train part-of-speech taggers for machine translation", *Springer Science+Business Media B.V,*. 2008.

[9] Fu-Hua Liu, Liang Gu, YuqingGao and Michael Picheny, IBM T.J. Watson Research Center Yorktown Heights, NY 10598, U.S.A, "Use of Statistical N-Gram Models in Natural Language Generation for Machine Translation", *In the proceedings of the International Conference on Computer Science and Engineering, pp.636-639,* 2003.

[10] ArwaHatem, Nazlia Omar, Khalid Shaker, School of Computer Science, Faculty of Information Science and Technology, "Morphological Analysis for Rule based Machine Translation", *in the proceedings of the International Conference on Semantic Technology and Information Retrieval, Putrajaya, Malaysia, pp.260-263,* june,2011.

[11] MuyunYang Sch. of Comput. Sci. & Technol., Harbin Eng. Univ. ,Harbin Junguo Zhu, Jufeng Li, Lixin Wang Haoliang Qi , Sheng Li, Liu Daxin , "Extending BLEU Evaluation Method with Linguistic Weight", *in the proceedings of the 9th International Conference on young computer scientist, ICYCS 2008, pp.1683 - 1688 ,2008.*

[12] Yang Liu, Dept. of Comput. Sci., Texas Univ., Richardson, TX, *Shriberg, E. "Comparing Evaluation Metrics for Sentence Boundary Detection", IEEE International Conference onAcoustics, Speech and Signal Processing, 2007. ICASSP 2007.*

[13]Wang, Y., Acero, A., and Chelba, C**, "**Is Word Error Rate a Good Indicator for Spoken Language Understanding Accuracy"**,** *IEEE Workshop on Automatic Speech Recognition and Understanding, St. Thomas, US Virgin Islands, 2003.*

[14] Subramanian, K. BBN Technol., Cambridge, Stallard, D. ,Prasad, R. , Saleem, S. , Natarajan, P., "Semantic translation error rate for evaluating translation systems", *IEEE Workshop on Automatic Speech Recognition and Understanding, 2007.*

[15]Myers, Jerome L., Well, Arnold D, "Research Design and Statistical Analysis",*Second Edition*, Lawrence Erlbaum,*pp. 508, ISBN 0-8058-4037-0,*2003.

[16]Maritz. J.S. "Distribution-Free Statistical Methods", Chapman & Hall. *ISBN 0-412-15940-6, page 217*, 1981.

[17]http://turner.faculty.swau.edu/mathematics/math241/materials/anova/