



# JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

## A SURVEY ON FORMAL LANGUAGE TRANSLATION (TELUGU - ENGLISH)

Mrs. P. Swaroopa<sup>1</sup>, Kalakonda Vishnu<sup>2</sup>, Pulipati Akshay<sup>3</sup>, S. Siva Sai Sandip<sup>4</sup>,  
Vennam Vivek<sup>5</sup>

<sup>1</sup> Assistant Professor, Department of Computer Science and Engineering, ACE Engineering College, Hyderabad, Telangana, India

<sup>2,3,4,5</sup> IV BTech Students, Department of Computer Science and Engineering, ACE Engineering College, Hyderabad, Telangana, India

### ABSTRACT

Language, as the Information carrier, is most important for humans to communicate and share information. Language barriers occur when people with different language backgrounds communicate. To solve this problem, we need translators who can translate from one language to another quickly and effectively. This is made possible by machine translation. It has been 50 years, since computer scientists have been working on Machine Translation. Machine Translation is not a new term it was in trending from past 50 years, it comes under Machine Learning technology called Natural Language Processing (NLP) [1]. Our paper deals with a Machine translator which converts from Telugu to English, it is a formal translator which aims to translate different shatakas, proverbs, and poems in Telugu to English. We can say that in our system Telugu is the source language and English is the target language, when compared to Telugu, English is morphologically simple language. We have different types of approaches to achieve machine translation namely rule based system, statistical translation system and neural machine translation. This work is a study of pre-processing the corpus and different methodologies used to construct a Machine Translator System.

### INTRODUCTION

India is multilingual country, where there exist more than 18 languages, there are different troops speaking different languages and people in India are not familiar with all languages they can neither understand nor speak all the languages. Most of the population in India speak languages like English and Hindi. Here we can observe a necessity of a translator when ever different people from different regions try to interact with each other, at this time we need a translator which can translate from one language to another quickly and effectively which can be achieved using a Machine Translator.

A Machine Translator is a program which can convert the text from one language to another, so the language in which the input is given is known as source language and the language in which the output is obtained is known as target language. At recent times we have different machine translators in use, which are effective and can solve this problem, but most of the times these existing translators fail when the have to convert poems, proverbs of the source language to target language.

Every language has its own proverbs and poems, which are created based on the ideologies of poets, people and culture. While translating these language specific poems and proverbs a special care must be taken, which is our problem statement, We can say our problem statement is constructing a Machine Translator which can translate different proverbs and shatakas in Telugu to English correctly. So we can say in our case the source language is Telugu where as target language is English. when compared to Telugu, English is morphologically simple language. So our model translates from morphologically rich language to a morphologically simple language.

Machine Translation system can be of mainly three types rule based system, statistical translation system and neural machine translation. Rule based system uses grammar rules and combination of languages with dictionaries for common words. It uses manually created dictionaries. These are robust and cost effective. Statistical translation system are another type of system which do not follow any grammar rules, they can learn by themselves, the accuracy is less when compared to other systems. Neural Machine Translation (NMT) is popular approach which uses neural networks to translate and a network model called Recurrent Neural Network (RNN), the accuracy is high and most of the developers and researches follow NMT for accurate performance in many language pairs when compared with others. We do have different other approaches like Encoder – Decoder architecture, LSTM architecture etc. However, there are different approaches to construct a Machine Translator some of the methods are listed in this paper.

## LITERATURE REVIEW

There have been many researches on machine translator with an aim of improving the accuracy of the system, starting from the data gathering, pre-processing to the type of algorithm used to build a better machine translator many researches have been done, I have read some of the papers to understand the problem in a better approach and to gain knowledge on different techniques their drawbacks used to build a Machine Translator.

[1] The problem statement of Sentence wise translator from Telugu to English is to translate the Vemana sathakam from Telugu to English. They have used neural machine translation (NMT) with long- and short-term Memory (LSTM). Neural machine translation is a new technique for machine translation which uses artificial neural networks (ANN) to increase accuracy and performance. LSTM is very similar to RNN the main difference is the number of layers in the network the RNN uses only one layer whereas LSTM uses 4 neural networks. They have used a bidirectional LSTM to translate poems from Telugu to English. According to the paper NMT with LSTM solves the problem of accuracy and the need of large data set, both for training and evaluation of the results. The drawback of this paper was the idea which is limited to vamana shatakam, this can be elaborated to different shatakas, proverbs and holy books from Telugu to English.

[2] Pre-processing of English to Hindi Corpus for statistical Machine Translation presents that improving the pre-processing technique and giving attention to it would improve the accuracy. The impact of experimenting on the pre-processing of the input are observed on translation quality improvement through BLEU (BILINGUAL EVOLUTION UNDERSTUDY an algorithm for evaluating the quality of the output which has been translated). The pre-processing methodology they have followed includes Casing, Punctuations, Spell Normalization etc. These experiments have proven the improved accuracy in English – Hindi translation. They state that the best combination of pre-processing can be used to improve the accuracy. The drawbacks could be not including the linguistic features like re-ordering source sentences to match the target word order using the source side phrase information.

[3] Statistical Machine Translation is a technique which can be used to solve the problem of Machine Translation. It is a Machine Learning based technique which examines many samples of human-produced translations, using which SMT algorithm learns how to translate automatically. This paper raises an important point involving the morphologically rich languages and other languages, It states that if we translate different morphologically rich languages like German, Arabic into a morphologically simple language like English, can be visualized as movement from higher dimensional space to a lower dimensional space, in which there are less chances of loss of meaning and nuance is harmless. Whereas extra attention must be taken while translating.

[4] This paper explores the Neural Machine Translation Methodology and tries to improve its performance by not using the fixed length vector, It mainly focuses on the translation from English to French. The data set they have used is very large, with more than 275M English-French words, They have taken two different models RNN Encoder-Decoder and Stochastic gradient descent algorithm, they have trained this model for 5 days. The drawbacks could be testing only for English to French and Training model only for 5 days, data set can be improved.

[5] This paper explores the hybrid approach, this approach is designed to translate from Malayalam to English, when compared to English, Malayalam is Morphologically rich language, This approach combines two approaches, the first one is machine translation based on example and the next one is transfer approaches for better efficiency and increased correctness. They have used dataset of Malayalam to English, the drawbacks could be it cannot translate complex sentences.

[6] This paper mainly focuses on translation from English to Telugu with emphasis on prepositions, it mainly focuses on the prepositions and converts the preposition in English to the proportion in Telugu it mainly focuses on different kinds of prepositions that are being used in English which can be translated as post positions in Telugu. time, gender, context and many other features play an important role in selecting the appropriate postposition in Telugu.

[7] This paper talks about the Cross Language Information Retrieval (CLIR) which is a sub topic in IR, it talks about the dictionary based translation approaches, it presents different methodologies like machine translation and corpus based translation. In this process we could face problems like selection of translation for query, selection of the dictionary for possible translation and so-on.

[8] This paper deals with different approaches to construct a machine translator and focuses on machine translation system design based on declension rules. In this paper a Machine translator is built which translates from English to Hindi, the efficiency of this method is high but it cannot deal with complex sentences for translation, It requires a heavy database.

[9] This article explains the procedure to develop a machine translator for five languages. It describes about language components, CRL systems, semantic procedures, pragmatic procedures, The drawbacks could be it cannot handle complex statements, They have taken less vocabulary data, the results obtained are not perfect, they were some semantical errors with lexical failures.

[10] This article analyses some of the systems on the basis of translation of English texts into Hindi. The presented results, systems using statistical approach or hybrid approach are more accurate than those using rule-based approaches. They extend by stating that rule based systems have its own benefits for translation. The hybrid approach which is a combination of both rules based and statistical approach will be seen as a future of machine translation systems.

SNO	PROBLEM STATEMENT	EXISTING SYSTEM	DRAWBACK
1	To translating vemana sathakam using Sentence wise Telugu to English translation.	Neural machine translator with bidirectional LSTM to obtain higher accuracy without large dataset.	1. Cannot handle rare or unknown words 2. It was just limited to vemana sathakas.
2.	Pre-processing of English to Hindi Corpus for statistical Machine Translation	The improvised pre-processing methodology they followed includes Casing, Punctuations, Spell Normalization etc. Their experiments have proven the improved accuracy in English – Hindi translation.	not including the linguistic features like re-ordering source sentences to match the target word order using the source side phrase information

3.	Machine Translator using Statistical machine translation technique.	It is a Machine Learning based technique which examines many samples of human-produced translations, using which SMT algorithm learns how to translate automatically.	Extra attention must be taken while translating from morphologically simple language to morphologically rich language. Cannot handle rare or unknown words.
4.	Building a Machine Translator English to French using Neural Machine Translation approach	It mainly focuses on the translation from English to French. The data set they have used is very large, with more than 275M English-French words, They have taken two different models RNN Encoder-Decoder and Stochastic gradient descent algorithm, the have trained this model for 5 days	be testing only for English to French and Training model only for 5 days, data set can be improved.
5.	English to Malayalam machine translation using hybrid approach	this approach is designed to translate from Malayalam to English, when compared to English ,Malayalam is Morphologically rich language, This approach combines two approaches, the first one is machine translation based on example and the next one is transfer approaches for better efficiency and increased correctness. They have used dataset of Malayalam to English	It cannot handle rare words. Lack of well aligned bitexts Cannot handle complex sentences
6.	Machine translation from English to Telugu using rule-based approach	it mainly focuses on the prepositions and converts the preposition in English to in Telugu it mainly focuses on different kinds of prepositions that are being used in English which can be translated as post positions in Telugu	It cannot handle unknown words.
7	Different approaches to perform better Cross Language Information Retrieval	It discusses about different approaches for CLIR, it deals with different approaches like dictionary based translation, identifying and translating phrases and compound words, out of vocabulary terms etc.	Accuracy of Dictionary based translation is less when compared with other approaches like WSD etc.

8	To construct a Machine translator which translates from English to Hindi	approaches to construct a machine translator and focuses on machine translation system design based on declension rules. In this paper a Machine translator is built which translates from English to Hindi, the efficiency of this method is high	cannot deal with complex sentences for translation, It requires a heavy database.
9.	Constructing a multilingual machine translator	the procedure to develop a machine translator for five languages. It describes about language components, CRL systems, semantic procedures, pragmatic procedures,	cannot handle complex statements, They have taken less vocabulary data, the results obtained are not perfect, they were some semantical errors with lexical failures.
10	To improve the efficiency of machine translator by examining different approaches.	they analysed some of the systems on the basis of translation of English texts into Hindi. the presented results, systems using statistical approach or hybrid approach are more accurate than those using rule-based approaches.	They have only considered translation from English to Hindi

## COCLUSION

The reason for the study is to build up a Machine Translator which translates different shatakas, proverbs, and poems in Telugu to English with out any semantic ad syntactic errors. When any Telugu statement is given as an input, a proper English sentence with same meaning is displayed on the screen. The study include major challenges like Content of poor quality, Technology problems, social Issues etc. This paper talks about different methodologies followed to construct Machine Translator it highlights different approaches using which a Machine Translator can be built, and also explains the comparisons among all the methods.

## ACKNOWLEDGEMENT

We would like to thank our guide Ms. P. Swaroopa and our project coordinator Mrs. Soppari Kavitha for their continuous support and guidance. We are also extremely grateful to Dr. M.V.VIJAYA SARADHI, Head, Department of Computer Science and Engineering, ACE Engineering College for his support and invaluable time.

## REFERENCES

1. P.Sujatha, D. Lalitha Bhaskari “ Sentence Wise Telugu to English Translation of Vemana Sathakam using LSTM “International Journal of Recent Technology and Engineering (IJRTE)
2. Pre-processing English-Hindi Corpus for Statistical Machine Translation ( Karunesh Kumar Arora local Sharma S Agarwal centre for development of advanced computing Noida India, KIIT group of institutions, Sohna road, Bonci Gurugram India go)
3. Adam Lopez, “statistical machine translation”, ACM computing surveys, vol 40, issue no 3, Article 8, August 2008,

Doi: <https://doi.org/10.1145/1380584.1380586>

4. “Neural Machine Translation By Jointly Learning to Align and Translate”, by Dzmitry Bahdanau and KyungHyun Cho Yoshua Bengio, Published as a conference paper at ICLR 2015
5. Rosna P Haroon and Shaharban T A, “Malayalam Machine Translation using Hybrid Approach”, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2016
6. Keerthi Lingam, E. Rama Lakshmi and L Ravi Theja, “Rule-based Machine Translation From English to Telugu With Emphasis on Prepositions”, 2014 First International Conference on Networks & Soft Computing
7. B.N.V Narasimha Raju and M S V S Bhadri Raju,” Statistical Machine Translation System for Indian Languages”, 2016 IEEE 6th International Conference on Advanced Computing
8. Jayashree Nair, Amrutha Krishnan K and Deetha R, “An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism”, 2016 ICACCI, Sept. 21-24, 2016, Jaipur, India
9. ULTRA: A Multilingual Machine Translator David Farwell and Yorick Wilks Computing Research Laboratory New Mexico State University Box 30001, Las Cruces, NM 88003
10. Sandeep Kharb, Hemant Kumar, Manoj Kumar and Dr. Arun Kumar Chaturvedi, “Efficiency of a Machine Translation System”, International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

