Advanced technology in kannada to Telugu Translation by Using Transfer Based Method:
An Accurate Approach

[#1] Sreenivasulu Madichetty,   [*2] Dr.A.Ananda Rao,   [#3] Radhika Raju.P,

[1] PG scholar, CSE,JNTUA, Ananthapuramu. India
[2] Professor&DAP, CSE,JNTUA, Ananthapuramu. India
[3] Lecturer, CSE,JNTUA, Ananthapuramu. India
[1] Sreea568@gmail.com
[2] akepogu@gmail.com
[3] radhikaraju.p@gmail.com

Abstract: **The term Machine Translation can be defined as Translation of sentences or words from one language to another language automatically with or without any human involvement. Today Machine Translation Systems plays an important role for sharing the information from one language to another language like Sanskrit to Hindi, Devanagari to English etc., which are life transforming stories available in India. In this work, translation of Kannada to Telugu languages has been considered which is mainly used in southern part of India (Karnataka, Andhra Pradesh, and Telangana). The basic activity of any machine translation application is to manage the vocabulary of words .The existing literature has many machine translation systems like Directed Machine Translation, Interlingual Machine Translation system, Statistical Machine Translation, Hybrid Machine translated System, Transfer Based approach and Corpus Based Machine Translated System etc. In this work, Transfer Based Machine Translation has been considered for translating from Kannada language as input language to Telugu Language as a output language, which is predicted to provide better results when compared to the other approaches.**

1. INTRODUCTION

Machine translation is the task of translating the text in input language to output language, automatically. Machine translation can be considered as an area of applied research that draws ideas and techniques from linguistics, computer science, artificial intelligence, translation theory, and statistics. Even though machine translation was envisioned as a computer application in the 1950's and research has been made for 60 years, machine translation is still considered to be an open problem [12].

India is a linguistically rich country. In India, mainly two large language families are there.1. Indo- Aryan language family and 2. Dravidian language family. Majority of people belongs to Indo-Aryan language family and secondly most people belongs to the Dravidian language family. Kannada and Telugu belongs to the same language family which is nothing but Dravidian language family. In order to provide the communication between the different language families there is a need for the machine translation. India has eighteen official languages, which were written in ten kinds of scripts [1], [16]. Hindi is the common language which is used in India. Kannada is the language which is most widely used in the southern part of India. More number of states have their own local language, which is either Hindi or one of the other official languages. Only about 7% of the population speaks English. Currently, the translation is done manually. Automation is used for strictly restricted to word processing. There are two specific examples for large volume manual translation are –(i).Sports news can be translated from Kannada into local languages. (ii).Government department's annual reports and public sector units can be translated among Hindi, English and the local language. Many resources such as employee details, weather reports, books, etc., in Kannada are being manually translated to local language. The main disadvantage of human translation is it requires more time and cost. Machine translation has the advantage is it is faster, cheaper and it is better compared to the human translation. The main goal of the machine translation is to improve the accuracy and speed of the translation. It has different approaches for machine translation 1.Linguistic approach 2.Non-Linguistic Approach 3.Hybrid Approach [2].

1.1 LINGUISTIC APPROACH:

Linguistic Approach is also known as Rule Based Approach. In India many translations can be done using Rule Based Approach only. Rule Based Approach can be classified into three types.

1. Directed Machine Translation

2. Interlingual Machine Translation

3. Transfer Based Approach.

a). DIRECTED MACHINE TRANSLATION:

According to the name, it uses the direct translation using bilingual dictionary by word to word. It doesn't use any intermediate representation but it follows the some syntactic rules [7]. The following procedure shown below:

1. Removing the suffixes from the input language and identify the root words.
2. Looking up the dictionary for translating to the output language.
3. There is a need for changing the position of the words in a sentence for some languages in which the structure of both languages are different. But Kannada to Telugu, it can be no need for changing the position of words in a sentence because structure of the both languages are similar.

b). INTERLINGUAL MACHINE TRANSLATION:

In Interlingual machine translation the depth of analysis is more when compared to the other rule based translation approaches. The main aim of   Interlingual machine translation is transforming the texts in the input language to a unique representation and which is helpful to many languages, and using the unique representation translating the text into output language. Interlingual approach knows machine translation as a two stage process:

1. Analyzing and transforming the text from input language to unique representation.

2. With the help of unique representation, text can be generated in the output language.

c). TRANSFER BASED APPROACH:

Transfer based approach can be used mostly when the structure of both the input and output language are dissimilar. In this approach consists of three phases. They are analysis phase, transfer phase, and generation phase. In the first phase, the input language sentence or word is parsed, the sentence or word structure can be generated as parse tree form. In the transfer phase, grammar rules are applied to the parse tree which is generated from input language to be converted into the structure of the output language. The generation phase words can be generated from the parse tree.

1.2 NON RULE BASED MACHINE TRANSLATION:

Non Rule based machine translations doesn't require any linguistic knowledge. It requires more number of resources which are not available in all languages. Therefore it is difficult to implement Non Linguistic machine translation like Example based machine translation, Hybrid based machine translation etc.

a*) HYBRID BASED MACHINE TRANSLATION:*

Hybrid based machine translation is nothing but which is combination of any two machine translation approaches either Rule based machine translation or Non Rule based machine translation or both.

b) EXAMPLE BASED MACHINE TRANSLATION:

Example based machine translation is a Non-Rule based machine translation which requires bilingual parallel corpora which is having the sentences in both languages. In this it requires more depth of analysis when compared to other machine translation methods which is one of the main draw back in the Example based machine translation.

2. PREVIOUS WORK

The methods which are used in machine translation systems which mainly depends upon the structure of both the input language and output language. If the structure of the input and output language are similar, it can use Direct Machine Translation System, else it can use Transfer based approach.

In the past it are having the different types of machine translation systems which are using the Transfer based approach. The Machine Translation System is MANTRA system which was developed in the year 1997 in which the languages used for translating are English and Hindi [17]. It is mainly applicable for office administration documents and which was further developed in the year 1999 for the application proceeding Rajyasabaha. An English to Hindi Machine Translation System which was developed in the year 2002 and it is mainly applicable for weather narration. An English to Kannada Machine Translation System which was developed in the year 2002 and which was named as MAT system. This system was tested for government
circular. Shakti Machine Translation System which was developed in the year 2003 which is used to translate English to Indian Languages. An English to Telugu Machine Translation System which was developed in the year 2004 and it was tested simple sentences. It are also having the Machine Translation Systems which are using the Direct based approach.
Anusaaraka System which was developed in the year 1995 among Indian languages which are Telugu, Kannada, Bengali,

Punjabi, and Marathi to Hindi. It is applicable for translating children stories. Punjabi to Hindi Machine Translation System which was developed in the years 2007, 2008 and which can be applicable for general purpose. Hindi to Punjabi Machine Translation System which was developed in the year 2010 and it can be used for translating itb pages, emails. Hindi to Punjabi Machine Translation System which was developed in the years 2009 and 2010 and it can be used for general purpose.

3. DEVELOPMENT AND IMPLEMENTATION OF KANNADA TO TELUGU MACHINE TRANSLATION USING TRANSFER BASED APPROACH
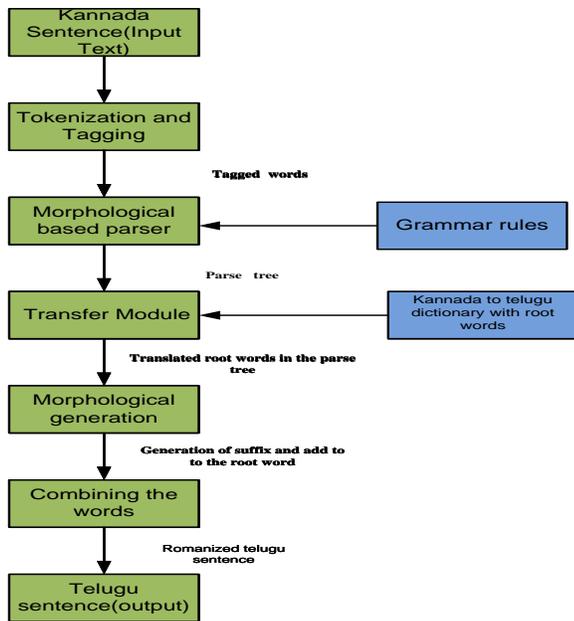

Fig: 1 Block diagram from Kannada to Telugu Translation using Transfer Based Approach.

As it have seen in the existing literature, if the structure of the both input and output languages of Machine Translation Systems are similar, then Direct based approach [1], [9] is used. If the structure of the both input and output languages of Machine Translation Systems are dissimilar, then Transfer based approach [1], [5] is used. In this work even though both input and output languages are similar,
it have not used the Direct based approach because if it use the Transfer based approach performance will
be increased. Therefore in this paper it used the Transfer based approach

a) TOKENIZATION AND TAGGING:

In this tokenization and tagging phase, Kannada sentence or paragraph can be taken from the input file and it can be tokenized into words or sentences. If sentences again it can be tokenized into words or if words that can be tagged for each

sentence [3]. In this phase the output gives Kannada words with tagging.

FOR VERBS

1. hoodanu||hoogu||V-PAST-P3.M.SL

FOR NOUNS

2. raamanu||raama||N-PRP-PER-M.SL-NOM

As shown in the above sentence 1, from the word hoodanu the root word hoogu is generated and its tag V-IN-ABS-PAST-P3.M.SL is generated .In most number of words V-IN-ABS is common so it did not classified that tag in parser. And from sentence 2, from the word raamanu, raama is generated as root word and its tag N-PRP-PER-M.SL-NOM is generated.

Some tags are:

N-Noun	PER-Person
PRP-Proper	NOM-Nominative
M-Male	
SL-Singular	

b) MORPHOLOGICAL BASED PARSER:

In this morphological based parser, tagged output taken from the tokenization and tagging phase. In this phase generate a parse tree for each tagged word using Brute force Parsing Mechanism from the grammar rules. And gives the output parse tree from each tagged word structure [7] [6].
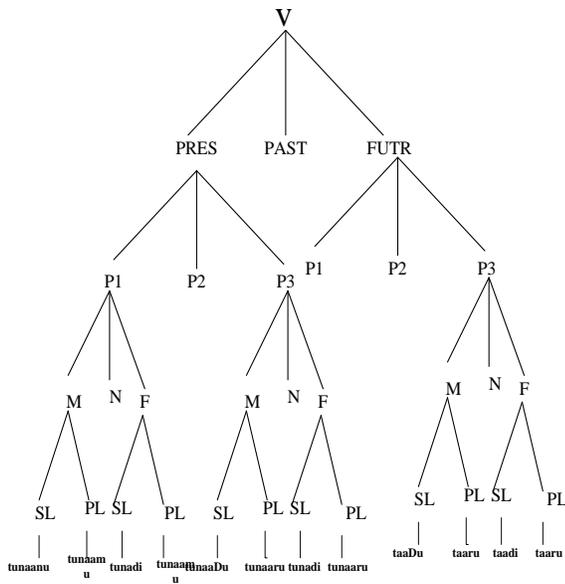
Fig: 2 Parse tree for verb structure

In this phase select the path from the parse tree according to the given tag. If the path is matched then go to the next module. As shown in the above parser some of the symbols are

V- Verb

PRES-Present tense

PAST-Past tense

FUTR-Future tense

P1-First person

P2-Second person

P3-Third person

M-Male

F-Female

N-Neutral

SL-Singular

PL-plural

c) CROSS LINGUAL DICTIONARY:

In this cross lingual dictionary contains the Kannada to Telugu meanings of root words only. In this dictionary contains most occurring root words of Nouns, Pronouns, and Verbs and so on. Each entry has two fields.one is Kannada root word another field has equivalent Telugu root word in Romanized form for most common Verbs, Nouns, Pronouns and so on[9].

Kannada root word	Translation of Telugu root word
niiDu	Iccu
negu	geMtu

TABLE: Cross lingual Dictionary

As shown in the above table containing root words from Kannada to Telugu .In this dictionary all words can be stored in romanized form only. It can converted exact scripting language after the completion of translation from Kannada to Telugu. Here **"niiDu"** is the Kannada word which can be translated as a Telugu word as **"iccu".** In the same way many words can be translated**.**

d) TRANSFER MODULE:

In this transfer module, root words can be translated from Kannada to Telugu which is taking from the Kannada to Telugu Dictionary. And also gives the output parse tree**.**

e) MORPHOLOGICAL GENERATION:

In this Morphological generation, it can generate the suffix according to the given tag for each word from parse tree and added to the root word. In this phase the output gives as the Telugu word. In the generation of suffix it use depth first search in the parse tree for getting the suffix [6]. And that suffix can be added to the root word.

In the next step combining all the words which can be generated from the morphological generation (Romanized words in Telugu).

In this last phase Romanized Telugu sentence can be taken as input and gives the output as exact Telugu sentence using Telugu Saara system [4].

4. IMPLEMENTATION AND TESTING OF A SYSTEM

This system can be implemented using Python programming language [10,11] and with the help of saara system for getting tagged words[12] and which can be tested the text file as input file which contains 1000 Kannada sentences can be translated into Telugu language. Broadly the system has been divided into modules.

1. Conversion of Roman Kannada (Transliteration)
2. Divide the Sentences into words and words can be tagged
3. From the tagged words generate the parse tree
4. Root word can be translated from Kannada to Telugu using Bilingual Dictionary
5. From the parse tree generate the suffix which can be added to the root word
6. Combining the all the words and convert from Roman Telugu to Telugu (Transliteration).

In our system input sentences are taken from a text file and output can be stored in another text file.

5. RESULTS AND DISCUSSION

Various Machine translation approaches are there for Machine Translation. In that Transfer based approach is more flexible and it can be extended to multi lingual environment and which gives better performance. In this system simple Kannada to Telugu Sentences can be tested for Machine translation.

REFERENCES:

[1] G V Gharaje and G K kharate "Survey of Machine Translation System in India" International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, October2013.
[2] Latha R. Nair and David Peter S "Machine Translation Systems for Indian Languages" International Journal of Computer Applications (0975 – 8887) Volume 39– No.1, February 2012.
[3] Kavi Narayana Murthy and Srinivasan Badugu "A New Approach to Tagging in Indian Languages" Research in computing 2013.
[4] Kavi Narayana Murthy and Srinivasan Badugu developed a paper on "Roman Transliteration for Indian scripts"
[5] Latha R Nair, David Peter & Renjith P Ravindran "Design and Development of a Malayalam to English Translator- A Transfer Based Approach" International Journal of Computational Linguistics (IJCL), Volume (3): Issue (1): 2012
[6] T.Suryakanthi Research Scholar, and Dr. S.V.A.V. Prasad Translation of Pronominal Anaphora from English to Telugu Language (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.
[7] T. Venkateswara Prasad1, G. Mayil Muthukumaran2 Telugu to English Translation using Direct Machine Translation Approach International Journal of Science and Engineering Investigations vol. 2, issue 12, January 2013
[8] David Peter S. School of Engineering Cochin University of Science and Technology Machine Translation Systems for Indian Languages International Journal of Computer Applications (0975 – 8887) Volume 39– No.1, February 2012.
[9] A Punjabi To Hindi Machine Translation System Gurpreet Singh Lehal by Professor, Dept. of Comp. Sci.,Punjabi University Patiala.
[10] Saara System is an integrated system that includes monolingual and bilingual dictionaries, stemmer, morphological analyzers and generators, etc., developed by Dr. Kavi Narayana Murthy at University of Hyderabad in Natural Language Engineering Lab.
[11] Machine Translation, Doug Arnold, University of Essex,doug@essex.ac.uk
[12] Natural language processing with python
[13] https://www.nltk.org
[14] https://www.google.co.in/
[15] http://en.wikipedia.org/wiki/
[16] Mallama V Reddy, DR. M. Hanumathappa "NLP challenges for Machine Translation from English to Indian Languages" International Journal of Computer Science and Informatics, ISSN (PRINT): 2231–5292, Volume-3, Issue-1, 2013
[17] Mantra Machine Translation System from English to Hindi which was developed by C-DAC

Selected Paper from International Conference on Computing (NECICC-2k15)