IMPACT: International Journal of Research in Applied, Natural and Social Sciences (IMPACT: IJRANSS)
ISSN (P): 2347-4580; ISSN (E): 2321-8851
Vol. 7, Issue 3, Mar 2019, 21-28

AN AUTOMATED SYSTEM FOR THE TRANSLATION OF ENGLISH ANALYTICAL CAUSATIVE CONSTRUCTIONS INTO MALAYALAM

Bijimol TK[1], John T. Abraham[2] & Jyothi Ratnam D[3]

[1]Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India

[2]Assistant Professor, Bharath Math College Trikkakkara, Cochin, Kerala, India

[3]Research Scholar, Amrita Vishwa VidyaPeetham, Coimbatore, Tamil Nadu, India

Received: 05 Mar 2019 **Accepted:** 11 Mar 2019 **Published:** 21 Mar 2019

ABSTRACT

Machine Translation (MT) is a branch of Natural Language Processing (NLP). Machine translation industry tries to achieve translation excellence. Many translations systems were developed by various agencies for English-Indian language translations. The linguistic features of English and Indian languages are the main hindrance, which affects the quality of various MT systems. The on-line MT providers like Google translator, Bing translator and TDIL developed their system by using the techniques like Statistical MT, Neural MT, and other modern methods. It had been observed that the output of all these systems is not up to the mark, it is simply because of the linguistic phenomena of the natural languages. Here we discuss the linguistic problems related to the translation of the English causative sentence into Malayalam. Then we proposed a Rule-based system to handle the causative sentence in both languages. Malayalam is a highly agglutinative and morphologically rich language these linguistic specialties of Malayalam determine the quality of all kinds of MT systems. Moreover, the scarcity of Malayalam resources and lack of perfect NLP tools are the main obstacles for the development of English–Malayalam, and Malayalam to other Indian language MT systems.

KEYWORDS: *Natural Language Processing, Rule-Based Machine Translation System, Google NMT, Causative Verbs, Word Order Identification, Word Reordering, Impersonal/Interpersonal Causative Verb Identification*

INTRODUCTION

A multi-lingual country like the Republic of India needs MT systems for knowledge transfer from English to Indian languages and from Indian languages to Indian languages. Malayalam is one of the major languages in India, which belongs to the Dravidian language family. It is the official language of Kerala. English also used along with Malayalam for the official correspondence. Officially both Malayalam and English are used in the same stratum at present. The people of Kerala use both English and Malayalam in their day-today life. Google translator provides translation for various Indian languages from English. But the Google translator is not free from flaws. This paper tries to find how to handle the translation of English causative sentence into Malayalam. It is the primary work on the translation of a causative sentence in English to Malayalam MT aspect. Google translator uses the most modern techniques for their online MT systems. But we can find a large number of linguistic misprints in their output. We are proposing a rule-based method for the handling of the English causatives in the context of English to Malayalam MT system.

Google NMT

Neural machine translation (NMT) is an approach to machine translation that uses a large artificial neural network [2]. It requires less memory space compared to traditional statistical machine translation (SMT) models. All parts of the NT model are trained end-to-end to optimize translation quality and performance. NMT systems face the difficulty with rare words and it is computationally expensive both in training and in translation inference.

Google Neural Machine Translation System (GNMTS) is based a neural machine translation (NMT) system developed by Google. It uses an artificial neural network to increase fluency and accuracy. GNMT uses example-based (EBMT) machine translation method to improves the quality of translation [1]. This system learns about the translation from millions of examples.

The Causatives in English and Malayalam

A sentence is a set of words which are grouped together to mean something. It is the basic unit of each language and it expresses a complete thought. A complete sentence contains at least a subject and main verb to declare a complete thought [11]. The main verb is the main part of the sentence which shows what the subject is doing. It signals an action or a state of being or an occurrence. The main verb decides the syntax and semantics of a specific type of sentence. Syntactically verbs are divided into three classes which are transitive verbs, intransitive verbs, and ditransitive verbs and semantically verbs are divided into three categories which are action verbs, process verbs and state verbs [12].

A transitive verb requires transferring its action to someone or something. It has two properties and the first property is, as an action verb it expresses the possible activities like paint, write, eat, kick etc. and second, as a direct object, it receives the action of the verb [12]. A verb that does not take a direct object is the intransitive verb. There is no word in such a sentence to tell who or what received the action. Intransitive verb follows a word or phrase which answer the question ‘how’. There are two characteristics to intransitive verb [13]. First, it is an action verb which shows the activity like the lie, die, arrive, sit etc. Secondly, it has no direct object which receives the action. The third category of the verb is a ditransitive verb which takes a subject and two objects. Theme and recipient are represented by these two objects. They may be called as direct or primary and indirect or secondary object.

A pattern of the verb phrase (VP) in a sentence shows some language-specific features and the causative sentences are one among them. Causation represents the part of the semantics of the verb. The language related features of the causative sentences differ from language to language. Causation is a natural phenomenon. In a causative sentence, the real subject of the sentence caused someone else to do something or being in a certain condition instead of doing byhimself [14]. In other words, one name entity (NP1) makes somebody else do something or cause another named entity (NP2) to be in a certain state. The way of expression of causation varies from one language to another. Most of the Indian languages like Malayalam, Tamil, Urdu, Hindi etc. show morphological causation. English shows morphological, lexical and analytical causation. The causative verbs refer to a causative situation. There are two components which combine the causative situation that are; i) the causing situation or the antecedent and ii) the caused situation or the consequent. Verbs which necessitate or at least imply the presence of three nominals, namely, an Initiator, an Actor or performer, and an Object (patient), may be labeled Causative verbs. Causative verbs always imply an Actor as well as an Initiator of the action performed by the Actor [13].

The causative sentence construction of English and Malayalam is totally different. In English for making of causatives simply use the auxiliary verbs like 'have', 'make', 'get' etc., but in Malayalam, the main verb shows inflections. Other Indian languages like Hindi Malayalam verb also shows two different casual forms named as first casual and second causals. These two causals have two distinct forms. For the making of causative verbs in Malayalam causative suffixes – 'i'(ഇ), 'ppi'(പ്പി), and 'ththu'(ത്തു) are added to the end of the main verb according to the ending vowel of the main verb[15]. Mainly three types of verb ending are found in Malayalam they are:

Verb end with 'a' the suffix 'i' is added with it.

Example:paRaya(പറയ)→transitiveForm→paRayunnu(പറയുന്നു)+'i'→paRayippikkunnu (പറയിക്കുന്നു)

Verb ends with 'ka' the suffix 'ppi' is added with it

Example: eTukka(ഏടുക്ക)→transitive form→ eTukkunnu (ഏടുക്കുന്നു) + 'ppi' →eTuppikkunnu (ഏടുപ്പിക്കുന്നു)

Verbs end with 'la', 'L', 'zha','ra' the suffix is 'ththu' added at the verb end

Example: para (പര) → transitive verb form → parakkunnu (പരക്കുന്നു) → 'ththu'→ (പരത്തുന്നു parathunnu).

Like other transitive verbs all Malayalam causative verbs shows tens inflections also [10].

Example as like 'cheyyippicchu' (ചെയ്യിപ്പിച്ചു), 'cheyyippikkunnu'(ചെയ്യിപ്പികുന്നു) 'cheyyippikkum'(ചെയ്യിപ്പിക്കും) in past-present-future correspondingly.

Example: Malayalam Source Text:

അയാൾഅയാളുടെ കമ്പ്യൂട്ടർന്റെ അറ്റകുറ്റപണികൾ ചെയ്യിപ്പിക്കും

(ayaaL ayaaLuTe kampyuuttarinte attakuttapaNikaL cheyyippikkum)

English Target Text: He will have his computer repaired.

The ending suffix 'um' indicate the future tense of the main verb 'repair'

The Linguistic Problems Found in the Google translator

English Source text: He will have his computer repaired

Google Translation

അയാളുടെ കമ്പ്യൂട്ടർ അറ്റകുറ്റപ്പണികൾ ചെയ്യും

(Ayalude computer attakuttapanikal cheyyum)

Malayalam target text: അയാൾ അയാളുടെ കമ്പ്യൂട്ടറിൻറെ അറ്റകുറ്റപണികൾ ചെയ്യിപ്പിക്കും

(ayaaL ayaaLuTe Computerinte attakuttapanikal cheyyikkum)

The word-order and tense reorganization of Google-translator are correct but the system failed to recognize the causative sense of the English auxiliary verb 'have'/get/make'. We checked the same system with more than 150 English causative sentences; the Google translator did not identify the causal sense of the auxiliary verb 'have'/get/make' in any of the given sentences.

Proposed System

It is a rule-based system and a set of rules are used to implement the translation job. Past, Present and future tense forms of sentences are handled here. The proposed system works in two stages the preprocessing and post-processing stages. In the time of preprocessing state the input sentences were passed through the series of preprocessing stages such as tokenization, POS tagging, causative verb identification, then it transformed source language linguistic units in to target language linguistic units, in the post-processing stage the system translate all the linguistic units and arrange the translated units according to the word-order of the target language. Figure 1 shows the system architecture.

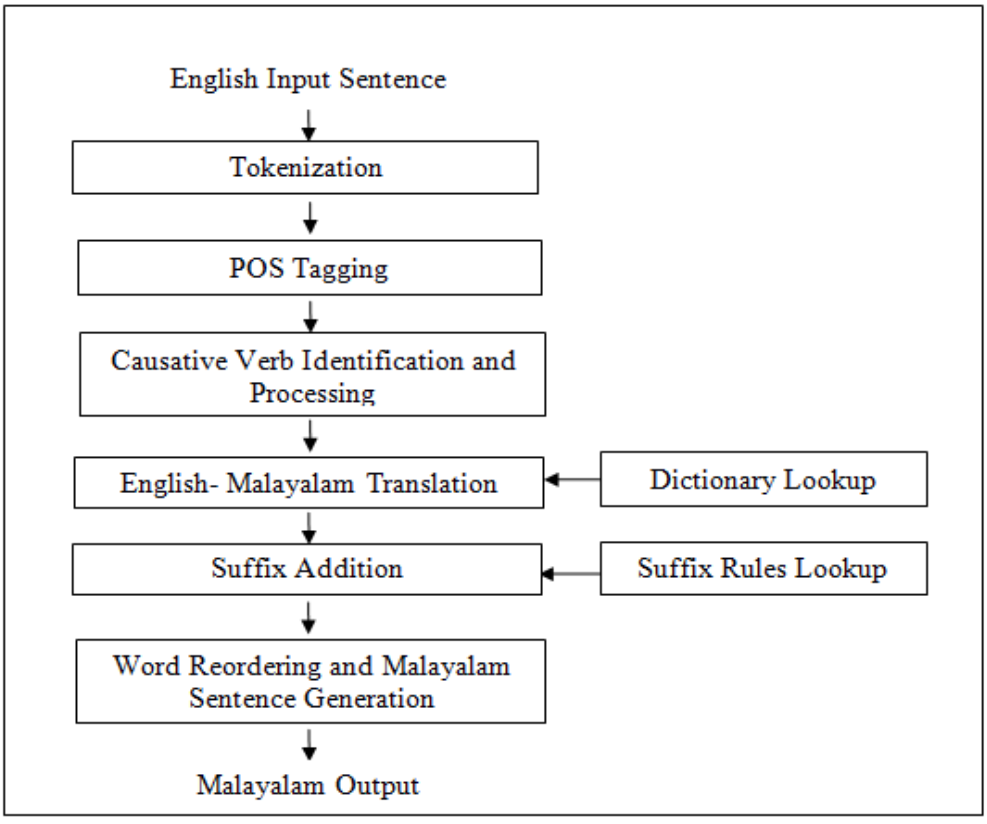

Figure 1: Proposed System Architecture

Tokenization & POS Tagging

It is the basic process in NLP. In tokenization, the input sentence is divided into a part of speech units [1]. Then tags are attached to the part of speech units. Tokenization and POS tagging are done by NLTK module.

Impersonal/Interpersonal Causative Verb Identification

In this module, the proposed system identify the sense and type of causative sentences with help of rules. Figure 2 shows the flowchart for the working of Malayalaminter personal/impersonal causative sentence. Initially, system identifies the position of 'have/get/make'. Then it checks the sentence is in the form of sub+ has/get/make + animate/inanimate object + causative main verb, and select Malayalam 1st causative sentence and tense form if the sentence is impersonal. Otherwise select 2nd causative form. Instructions in the flowchart are executed based on the rules given in section 3.

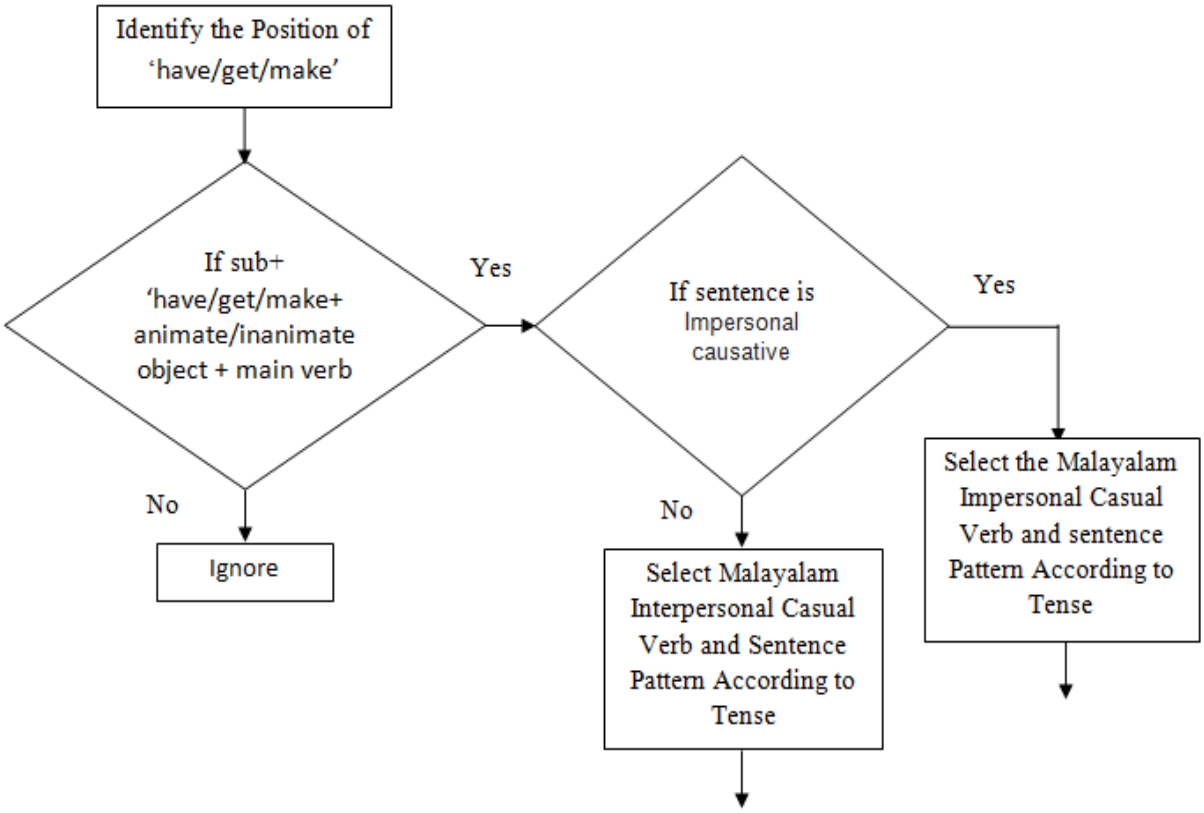

Figure 2: Flowchart for Malayalam Impersonal and Interpersonal Causative Sentence

Dictionary lookup

It is required to keep a dictionary for storing root words and with various inflectional forms of causative verbs of Malayalam and the main verbs and its various tense forms of English verbs and different forms of causative 'have'. In this stage, the system checks into the dictionary of the source and target language and if the verb is available, the process is going into the next step.

Suffix Addition or Morphological Generator

This is a suffix addition step, a list of words in English is translated as Malayalam words for the given input sentence. While its Malayalam output is generated the suffix need to be added with a noun. Sandhi rules are required for generating this suffix. For example, Raman is the word ends with 'in'(ൻ). As per sandhi rules all the words end with 'in' (ൻ) are attached with suffix 'e'(ൎ) and 'raman'+'e' (രാമൻ+ൎ) become Raman (രാമനെ).

Word Reordering Malayalam Sentence Generation

In this module, the system re-arrange the transformed linguistic units according to the word-order pattern of the Malayalam text with the help of rules. Malayalam Sentence Generation is the final step in the proposed system. A list of Malayalam words in prescribed order are combined and get the translation output.

Experiment and Result

Implementation of the proposed system is in Python 2.7.6 which is an interpreted, object-oriented programming language [3]. Morphological processing is done by Rule-Based Machine translation method. Bilingual English Malayalam dictionary is used for finding the translation of root words and causative forms. NLTK tool kit is used to implement tokenization and POS tagging. Figure 3 shows the working method of the proposed system using an example sentence.

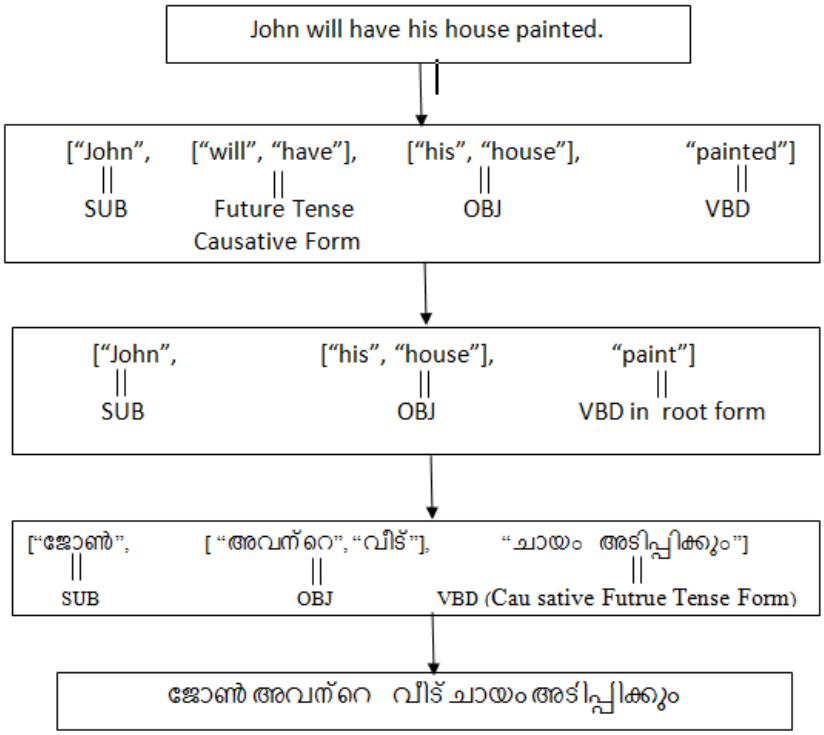

Figure 3: Working of the Proposed System Using an Example Sentence

Sample Output

The proposed system is compared with Google Translator. Table.1 shows the output generated by both Google Translate and the newly developed rule-based system. It shows the difference in the output of both systems. The proposed newly developed rule-based system performs better than the Google translator. English-Malayalam parallel sentences (100) were utilized to train the system and same source sentences gave as the input of Google translator and compared the output of both systems. The Google translator translated all the main verb in their transitive form only it means that Google translator failed to identify the causative sense of the auxiliary-verb 'have'/get/make' in the source English sentence. For testing, we included the impersonal and interpersonal causative sentences. The system gave quality output for simple sentences with nearly an accuracy of 93%. In the context of long sentences, our system failed to recognize the exact

sentence patterns of the source and target languages.

Table 1: Sample Output of Google Translator and Proposed System

Input (English)	Google Translate Output (Malayalam)	System Output (Malayalam)
John will have his house painted	യോഹന്നാൻതന്റെഭവനംനിറച്ചെടുക്കും.	ജോൺ അവന്റെ വീട് ചായം അടിപ്പിക്കും
He has his car washed	അയാളുടെ കാറ് കഴുകിയത്	അവന് അവന്റെ കാർ കഴുകിച്ചു
He made him weep	അവൻ അവനെ കരഞ്ഞു	അവന് അവനെ കരയിച്ചു
I am going to have my task done	എന്റെ ജോലി പൂർത്തിയാക്കാൻ ഞാൻ പോകുന്നു	ഞാന് എന്റെ ജോലി പൂർത്തിയാക്കിപ്പിക്കുവാൻ പോകുന്നു

Conclusion and Future Work

In this paper, we are proposing a rule-based approach for English-Malayalam machine translation system, which translates the causative form of sentences from English to Malayalam. The proposed system utilizes a bilingual dictionary for equivalent Malayalam words. Our studies revels that this proposed system performs better than that of Google translator for the translation of causative sentences. It had been observed that sometimes our system did not correctly identify the correct POS information, this affect the quality of the translation. This proposed system is unable to handle other sentences patterns like complex and long sentences. The proposed system gives quality output for simple sentences. In the future it can be extended to complex and long sentences.

REFERENCES

1. *N. Mariana, "Neural Machine Translation", HPI Uni. Potsdam, BfR, Germany 2017.*
2. *T.Jalaj, Python Natural Language Processing, Packt Publishing, Birmingham, UK, 2017*
3. *B. Steven, L. Edward, and K. Ewan., Natural Language Processing with Python, O'Reilly Media Inc.,USA, 2009*
4. *K.S. Narayanapillai, Adhunika Malayalam vyakaraNam., Kerala Bhasha Institute, Thiruvananthapuram, Kerala, India, 1995.*
5. *S. Tripathi, and J.K. Sarjgek, "Approachesto Machine Translation" Annals of Library and Information Studies, Vol. 57, 2010.*
6. *A.L. Lagarda, V. Alabau, Casacuberta, R. Silva, and E. Díaz-de-Liaño, "Statistical Post-Editing of a Rule-Based Machine Translation System." Proceedings of NAACL HLT, 2009.*
7. *S. Hampshire, C. P Salvia,. "Translation and the Internet: Evaluating the Quality of Free Online Machine Translators", Quaderns. Rev. trad. 17, pp. 197-209, 2010*
8. *R. N. Latha, P. S. David, and P.R. Renjith, "Design and Development of a Malayalam to English Translator-A Transfer Based Approach", International Journal of Computational Linguistics (IJCL), Volume 3, Issue 1, 2012.*

9. *K. S Rajesh, A. K. Veena, and R. Dayakar, "Building a Bilingual Corpus based on Hybrid Approach for Malayalam-English Machine Translation", Special Issue of International Journal of Computer Science & Informatics, Vol.2, Issue.1, 2, 2009.*
10. *[10] V. Ram Kumar, Smapoorna Malayala Vyakaranam, SISO Books, Pattam, Thiruvananthanpuram, 2001.*
11. *Singh Suraj Bhan, English–Hindi translation grammar, Prabhatprakashan: Delhi, 2006.*
12. *Warriner John.E, Warriner's English Grammar ANC composition, Fifth Course: Har court Bracc Jovanovich, Franklin edition, The University of Michigan Press, 1982.*
13. *Michal Auersperger, English Causative Constructions with the Verbs have, get, and make, and their Czech Translation Counterparts.*
14. *Beth Levina and Malka Rappaport Hovavb, A preliminary analysis of causative verbs in English, Department of Linguistics, Northwestern University, Sheridan Road, Evanston, IL 60208-4090, USA,b Department of English, Bar Ilan University, Ramat Gan, Israel, 2016.*
15. *Sunil R, Jayan V, Bhadran V K, Preprocessors in NLP Applications: In the Context of English to Malayalam Machine Translation, Centre for Development of Advanced Computing (C-DAC) Trivandrum, India IEEE 2012.*