Suffixation Rules of Manipuri Verbs in English to Manipuri Machine Translation

MAYANGLAMBAM PREMI DEVI
IRENGBAM TILOKCHAN SINGH
HAOBAM MAMATA DEVI

Abstract

The suffixation of verbs plays an important role in the formation of a meaningful sentence in Manipuri language. All the verbs in Manipuri are bound roots. So, there is always a suffix at the verb in Manipuri sentences. In English language, the tense of a sentence is determined from the verb but in Manipuri language, the tense of a sentence is determined from the suffix(es) associated with the verb. English to Manipuri Machine Translation (EMMT) is a rule-based automatic translation system. In EMMT, the suffixation rules of Manipuri verbs are essential for the derivation of an appropriate suffix(es) of the translated verb in Manipuri from the input English sentence. In this MT, suffixation rules are proposed to use in the post-processing module as a new approach. This paper presents the process of EMMT, different suffixation rules of Manipuri verbs and implementation in EMMT according to the types of tense in English.

Keywords: Verb Suffixation, Machine Translation, Rule-based, Manipuri, Post-processing, Suffixation.

1. Introduction

In computational linguistics, machine translation is the process of translating one language to another language using the computer without human involvement. EMMT system is developed based on the rule-based (RB) transfer approach. MT systems are developed using rule-based approach in the early days (Batra et al. 2010; Singh et al. 2010), but corpus-based methods such as Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) approaches gradually replace this traditional approach. SMT and NMT methods require a sizeable parallel corpus of the pair language. The applications of MT such as parts-of-speech (POS) tagger, parser, morphological analyser and generator of Manipuri language that can be integrated into the proposed MT system are unavailable. Development of such MT applications is under process by many researchers. Due to the non-availability of English and Manipuri (Meitei/Meetei Mayek) parallel corpus and lack of MT applications, the development of English to Manipuri MT system using a corpus-based approach becomes impossible.

In most of the rule-based machine translation (RBMT) systems, the morphological generator (MG) of the target language is used as a post-processing module (Desai et al. 2014; Antony P J 2013; Murthy. K 2002; Nair et al. 2019; Turhan 1997). But Manipuri MG which can be incorporated in the EMMT is not available till now. The development of Manipuri MG as a part of the EMMT system is also a tedious and challenging task. So, suffixation rules of Manipuri verbs are used in the post-processing of the EMMT.

Suffixation is the process of adding a suffix or suffixes to a root word. For example, ꯍꯔꯥꯎꯕ (*haraoba*-happy) + ꯔꯝ ꯏ (*ram i*)=ꯍꯔꯥꯎꯔꯝꯏ (*haraoram i*-was happy). Here, ꯍꯔꯥꯎ (*harao)* is the root word and ꯕ (*ba)* is the suffix of the verb ꯍꯔꯥꯎꯕ (*haraoba).* Since Manipuri verbs are bound root, a suffix is always associated with them. In the example, ꯔꯝ ꯏ (*ram i*) is a suffix that indicates that a Manipuri sentence is in the simple past tense. While writing the verb ꯍꯔꯥꯎꯕ (*haraoba*) and suffix ꯔꯝ ꯏ (*ram i*) together, the suffix ꯕ (*ba*) is removed from the verb ꯍꯔꯥꯎꯕ (haraoba) and added the suffix ꯔꯝ ꯏ (*ram i*) to get the word ꯍꯔꯥꯎꯔꯝꯏ (*haraoram i*-was happy).

This paper is organised as follows: Section 1 describes about MT and suffixation. It also describes the reasons for the non-feasibility of developing corpus-based English to Manipuri

automatic MT system. Section 2 presents the literature review of the works where the suffixation method and rules are employed. Section 3 presents the process of English to Manipuri MT system. Section 4 presents a sample dataset of the suffixation rules of Manipuri verbs regarding simple present tense and a detailed analysis of its application using 12 tense forms. Section 5 presents the implementation of the suffixation rules of Manipuri verbs in EMMT. Section 6 describes the result and discussion, and the conclusion is in section 7.

2. Literature Review

The suffixation method and its rules are used by many researchers in different works like word formation, classification of verbs, POS tagging etc. Some of the works done in relation to the suffixation method are given below. S Rajendran (2001) uses the suffixation method and its rules in the formation of Tamil words to build a vocabulary database. In his work, it is mentioned that suffixation has three kinds: 1. Suffix addition, 2. Phoneme changing and 3. Tense suffix selection. Thoudam Doren Singh and others (2008) use the suffixation method in the formation of possessive pronouns, verbal nouns, possessive adjectives and manner adverbs which are the basic requirements for the development of a Manipuri POS tagger. R. Ravindra Kumar and others (2011) have used the rules of suffixation in the verb classification of Malayalam language. The classification is done to create an English-Malayalam bilingual dictionary for developing a rule-based MT system. Neha Dixit and others (2014) developed an automatic tool for classification of verbs in Hindi based on the syntactic perspective. In their work, suffixation rules are used to get inflected forms of Hindi verbs. Bipul Roy and others (2017) narrate that suffixation is an important process in the Assamese words formation. It has large contribution in the richness of morphosyntax and morphosemantics of the Assamese language. It is used in the new words formation of the language. In the proposed work, the suffixation rules for verbs are used to derive Manipuri verb suffix(es) based on the tense of English sentence. As far as the author's knowledge is concerned, related works on using verb suffixation rules in the post-processing of rule-based English to other Indian languages MT systems are not available. Indian languages like Kannada, Bangla, Telegu, etc. use morphological analyser and generator as the post-processing module in their translation from English language.

Again in this section, we present some of the earlier works in machine translation system. Himangshu Choudhary and others (2018) developed an MT system for English-Tamil called MIDAS translator. It was developed based on neural machine translation. The results showed that the MIDAS translator outperformed Google translate. Sandeep Saini and others (2018) developed a machine translation system for English-Hindi using NMT. The system showed promising results compared to Anusaaraka, AnglaMT and Anglabharati. Kamaljeet Kaur Batra and others (2010) developed a Punjabi to English Noun phrases MT system. They used the transfer approach of RBMT. The accuracy obtained was 85.33%. Thoudam Doren Singh and others (2010) developed Manipuri to English MT system. The system used a rule-based technique. The system achieved a BLEU score of 0.137 and a NIST score of 3.361. Kalyanee Kanchan Baruah and others (2014) developed an Assamese to English bilingual MT system. They followed a statistical phrase-based translation approach. The Assamese-English translation system achieved a BLEU score of 9.72, and the English-Assamese translation system achieved a BLEU score of 5.02. Shivakumar KM and others (2015) developed a Kannada to English MT system. The system was developed based on an SMT approach and achieved a BLEU score of 10.68.

3. English to Manipuri Machine Translation System

EMMT converts an input English sentence to its target Manipuri sentence using rule-based method. The modules involved in EMMT are analysis of the source sentence, re-ordering of input sentence structure to corresponding target sentence structure, retrieving of word's meaning from the English to Manipuri bilingual dictionary, transliteration of non-dictionary words and post-

processing using suffixation of verbs and preposition rules. The process of developing EMMT is shown in figure 1.
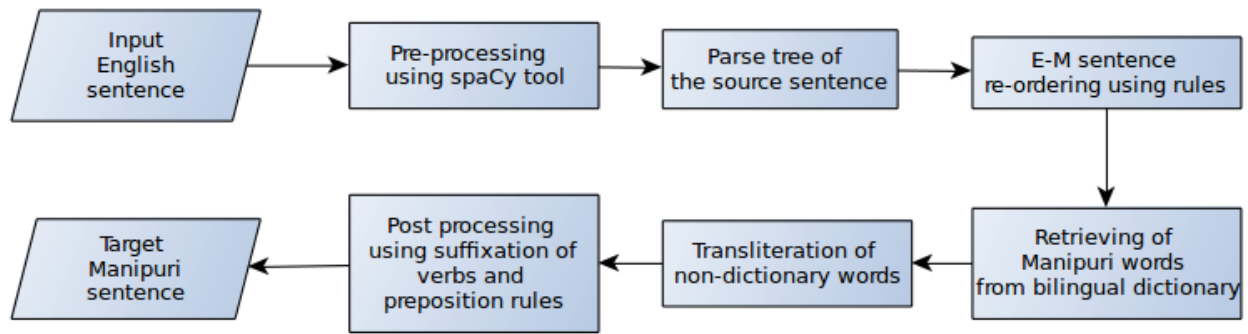

Figure 1: Process flow of English to Manipuri MT

Analysis of source sentence like Parts-of-speech tagging and parsing are performed using spaCy v3.2 tool. Re-ordering of English sentence structure to Manipuri sentence structure is done through phrases. Around 4000 English to Manipuri translated phrases are analysed and obtained 540 unique rules. English to Manipuri bilingual dictionary with 15000 words is designed for retrieving word's meaning. English to Manipuri transliteration system is designed using syllabification method. Around 2215 proper nouns are analysed and obtained 163 transliteration rules. After the process of retrieving word's meaning from the bilingual dictionary and transliteration of proper nouns, suffixation of Manipuri verbs and prepositions rules are applied to get the target sentence. Appendix A presents outputs of the EMMT.

4. Suffixation Rules of Manipuri Verbs

The suffixation rules of Manipuri verbs are constructed upon analysis of Manipuri verbs along with their formation in the English tenses. The verb suffixation rules are different based on (i) the different types of the tense and (ii) different patterns of characters exhibited in the last position and just before the last character of the lemma form of the Manipuri verb. The various nominal and verbal suffixes of the Manipuri language are described in Appendix B. Table 1 shows a dataset of the verb suffixation rules for simple present tense.

Table 1: A Dataset of Suffixation rules for Simple Present Tense

Character before the last character	Last character	Character to be removed	Character to be added for non-negative sentences	Character to be added for negative sentences	Examples	
					Lemma form of the verb	The verb after adding the suffix
ꯀ (kok lonsum)	ꯄ (pa)	ꯄ (pa)	ꯤ (i)	ꯇꯦ (te)	ꯊꯀꯄ(thakpa-drink)	ꯊꯀꯤ (thak i-drink), ꯊꯀꯇꯦ (thakte-do not drink)
�ꯥ (un) / (aatap)/ ` (otnap)/ ꯁ ꯳ (unap) /	ꯕ (ba)	ꯕ (ba)	ꯤ (i)	ꯗꯦ (de)	ꯀꯥꯎꯕ(kaoba-forget)	ꯀꯥꯎꯤ(kao i-forget), ꯀꯥꯎꯗꯦ(kaode-do not forget)

(sounap) ˚						
ꯃ (mit lonsum)	ꯕ (ba)	ꯕ (ba)	ꯅ (i)	ꯗꯦ (de)	ꯃꯤꯅꯝꯕ (minamba-lie)	ꯃꯤꯅꯝꯅ(minam i-lie), ꯃꯤꯅꯝꯗꯦ(minamde-do not lie)
ꯄ (pa lonsum)	ꯄ (pa)	ꯄ (pa)	ꯄꯤ (pi)	ꯇꯦ (te)	ꯀꯞꯄ(kappa-cry)	ꯀꯞꯄꯤ(kappi-cry), ꯀꯞꯇꯦ(kapte-do not cry)
ꯅ (naa lonsum)	ꯕ (ba)	ꯅꯕ/ꯕ	ꯂꯤ (li)	ꯗꯦ (de)	ꯌꯣꯅꯕ(yonba-sell)	ꯌꯣꯜꯂꯤ(yolli-sell), ꯌꯣꯟꯗꯦ(yonde-do not sell)
ꯉ (ngou lonsum)/ ˚ (nung)	ꯕ (ba)	ꯕ (ba)	ꯉꯤ (ngi)	ꯗꯦ (de)	ꯌꯦꯡꯕ(yengba-see)	ꯌꯦꯡꯉꯤ(yengngi-see), ꯌꯦꯡꯗꯦ(yengde-do not see)
ꯂ (til lonsum)	ꯄ (pa)	ꯄ (pa)	ꯂꯤ (li)	ꯇꯦ (te)	ꯆꯠꯄ(chatpa-go)	ꯆꯠꯂꯤ(chatli-go), ꯆꯠꯇꯦ(chatte-do not go)

The first column of table 1 contains the characters just before the last characters of the Manipuri verb, the second column contains the last character of the verb, the third column is the character to be removed from the verb, the fourth column contains the suffix(es) to be added for non-negative sentences, and the fifth column contains the suffix(es) to be added to the verb for negative sentences. The sixth column is the examples containing the lemma form of the verb and the target verb after adding suffix for non-negative and negative sentences.
Let us take an example with the verb *drink*.

       English: He drinks water. (Simple present tense)

       Manipuri: ꯃꯍꯥꯛ (*mahaak*-he) ꯏꯁꯤꯡ (*ishing*-water) ꯊꯛꯅ (*thak i*-drinks) �‖(.)

In this example, 'drinks' is the main verb of the sentence. The lemma form of the verb '*drinks'* is '*drink'*. We get the meaning of drink as ꯊꯛꯄ (*thakpa*) from the bilingual dictionary. The above suffixation rules for simple present tense are applied to the Manipuri verb ꯊꯛꯄ (*thakpa*). Here, the last but one character is ꯛ (*kok lonsum*) and the last character is ꯄ(*pa*). The first rule of the above table is matched and the last character ꯄ(*pa*) is deleted and the suffix ꯅ (*i*) for the non-negative sentence is added to get the target word ꯊꯛꯅ (*thak i*). If the sentence is negative, then the suffix ꯇꯦ (*te*) should be added to get ꯊꯛꯇꯦ (*thakte*-do not drink). Table 2 shows comparative analysis of the suffixation rules of Manipuri verb 'drink' with English tenses. 12 suffixation rules are generated by analysing the suffixes of the verb '*drink*'.

Table 2: Analysis of Suffixes of the verb 'drink' ꯊꯛꯄ(*thakpa*)

Tense	Lemma form of the verb "drink"	Character before the last character	Last character	Character to be removed	Character to be added for non-negative sentences	Character to be added for negative sentences	Resulting verb after suffixation
Simple Present Tense	ꯊꯛꯄ (*thakpa*-drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯅ (i)	ꯇꯦ (te)	ꯊꯛꯅ (thak i-drink), ꯊꯛꯇꯦ (thakte-do not drink)
Present	ꯊꯛꯄ (*thakp*	ꯛ (*kok*	ꯄ (pa)	ꯄ (pa)	ꯂꯤ (li)	ꯇꯔꯦ (tre)	ꯊꯛꯂꯤ(thakli-is drinking), ꯊꯛꯇꯔꯦ(thaktre-is not

Tense							
Continuous Tense	*a-*drink)	*lonsum)*					drinking)
Present Perfect Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯦ (le)	ꯇ꯭ꯔꯤ (tri)	ꯊꯛꯂꯦ(thakle-has drunk), ꯊꯛꯇ꯭ꯔꯤ(thaktri-has not drunk)
Present Perfect Continuous Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯛꯂꯤ (lakli)/ ꯇꯨꯅ ꯂꯩꯔꯤ (tuna leiri)	ꯂꯛꯇ꯭ꯔꯤ (laktri)/ ꯇꯨꯅ ꯂꯩꯇ꯭ꯔꯤ (tuna leitri)	ꯊꯛꯂꯛꯂꯤ(thaklakli-has been drinking), ꯊꯛꯂꯛꯇ꯭ꯔꯤ(thaklaktri-has not been drinking)
Simple Past Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯈꯤ (khi)/ꯂꯝ ꯁ (lam i)	ꯇꯦ (te)/ ꯂꯝꯗꯦ (lamde)	ꯊꯛꯈꯤ(thakkhi-drank), ꯊꯛꯇꯦ(thakte-did not drink)
Past Continuous Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯝꯈꯤ (lamkhi)/ ꯂꯝꯂꯤ (lamli)	ꯂꯝꯗꯦ (lamde)	ꯊꯛꯂꯝꯈꯤ(thaklamkhi-was drinking), ꯊꯛꯂꯝꯗꯦ(thaklamde-was not drinking)
Past Perfect Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯈꯔꯦ (khre)	ꯈꯤꯗ꯭ꯔꯤ (khidri)	ꯊꯛꯈꯔꯦ(thakkhre-had drunk), ꯊꯛꯈꯤꯗ꯭ꯔꯤ(thakkhidri-had not drunk)
Past Perfect Continuous Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯝꯝꯤ (lammi)/ ꯇꯨꯅ ꯂꯩꯔꯝ ꯁ (tuna leiram i)	ꯂꯝꯗꯦ (lamde)/ ꯇꯨꯅ ꯂꯩꯔꯝ ꯗꯦ (tuna leiramde)	ꯊꯛꯂꯝꯝꯤ(yeklammi-had been drinking), ꯊꯛꯂꯝꯗꯦ(yeklamde-had not been drinking)
Simple Future Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯀꯅꯤ (kani)	ꯂꯣꯢ (loi)	ꯊꯛꯀꯅꯤ(thakkani-will drink), ꯊꯛꯂꯣꯢ(thakloi-will not drink)
Future Continuous Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯒꯅꯤ (lagani)	ꯂꯔꯣꯢ (laroi)	ꯊꯛꯂꯒꯅꯤ(thaklagani-will be drinking), ꯊꯛꯂꯔꯣꯢ(thaklaroi-will not be drinking)
Future Perfect Tense	ꯊꯛꯄ (*thakpa-*drink)	ꯛ (*kok lonsum*)	ꯄ (pa)	ꯄ (pa)	ꯂꯝꯒꯅꯤ (lamgani)/ ꯂꯝꯒꯅꯤ	ꯂꯝꯂꯔꯣꯢ (lamlaroi)	ꯊꯛꯂꯝꯒꯅꯤ(thaklamgani-will have drunk), ꯊꯛꯂꯝꯂꯔꯣꯢ(thaklamlaroi-will not have drunk)

						(lamlaga ni)	
Future Perfect Continuou s Tense	ꯊꯛꯄ (*thakp a-* drink)	ꯛ (*kok lonsum*)	ꯝ (pa)	ꯝ (pa)	ꯇꯨꯅ ꯂꯩꯔꯝꯒ ꯅꯤ (tuna leiramga ni)	ꯇꯨꯅ ꯂꯩꯔꯝꯂꯣ ꯏ (tuna leiramlo i)	ꯊꯛꯇꯨꯅ ꯂꯩꯔꯝꯒꯅꯤ(thaktuna leiramgani-will have been drinking), ꯊꯛꯇꯨꯅ ꯂꯩꯔꯝꯂꯣꯏ(thaktuna leiramloi-will have not been drunk)

Examples related to the analysis of suffixes given in table 2 with the verb drink-ꯊꯛꯄ (*thakpa*).

1. (a) English: She drinks water. (Simple present tense)
1. (b) Manipuri: *mahaak(*she) *ishing (*water) *thak i*(drinks).
    ꯃꯍꯥꯛ (*mahaak*) ꯏꯁꯤꯡ (*ishing*) ꯊꯛ ꯏ (*thak i*)‖(.)
    ꯃꯍꯥꯛ ꯏꯁꯤꯡ ꯊꯛ ꯏ ‖

2. (a) English: She does not drink water. (Negative Simple present tense)
2. (b) Manipuri: *mahaak(*she) *ishing (*water) *thakte* (does not drink) .
    ꯃꯍꯥꯛ (*mahaak*) ꯏꯁꯤꯡ (*ishing*) ꯊꯛꯇꯦ (*thakte*)‖(.)
    ꯃꯍꯥꯛ ꯏꯁꯤꯡ ꯊꯛꯇꯦ ‖

3. (a) English: Raam was drinking tea. (Past Continuous tense)
3. (b) Manipuri: *Raam chaa* (tea) *thaklamkhi* (was drinking).
    ꯔꯥꯝ (*Raam*) ꯆꯥ (*chaa*) ꯊꯛꯂꯝꯈꯤ(*thaklamkhi*)‖(.)
    ꯔꯥꯝ ꯆꯥ ꯊꯛꯂꯝꯈꯤ ‖

4. (a) English: Raam was not drinking tea. (Negative Past Continuous tense)
4. (b) Manipuri: *Raam chaa* (tea) *thaklamde* (was not drinking).
    ꯔꯥꯝ (*Raam*) ꯆꯥ (*chaa*) ꯊꯛꯂꯝꯗꯦ(*thaklamde*)‖(.)
    ꯔꯥꯝ ꯆꯥ ꯊꯛꯂꯝꯗꯦ ‖

5. (a) English: They will have drunk coffee. (Future perfect tense)
5. (b) Manipuri: *Makhoi* (they) *coffee* (coffee) *thaklamgani* (will have drunk).
    ꯃꯈꯣꯏ (*makhoi*) ꯀꯣꯐꯤ (*coffee*) ꯊꯛꯂꯝꯒꯅꯤ (*thaklamgani*)‖(.)
    ꯃꯈꯣꯏ ꯀꯣꯐꯤ ꯊꯛꯂꯝꯒꯅꯤ ‖

6. (a) English: They will not have drunk coffee. (Negative Future perfect tense)
6. (b) Manipuri: *Makhoi* (they) *coffee* (coffee) *thaklamlaroi* (will not have drunk).
    ꯃꯈꯣꯏ (*makhoi*) ꯀꯣꯐꯤ (*coffee*) ꯊꯛꯂꯝꯂꯔꯣꯏ (*thaklamlaroi*)‖(.)
    ꯃꯈꯣꯏ ꯀꯣꯐꯤ ꯊꯛꯂꯝꯂꯔꯣꯏ ‖

The '*last character*' ꯝ(*pa*) and '*character before the last character*' ꯛ(*kok lonsum)* of the lemma form of the Manipuri verb '*drink*' (*thakpa*-ꯊꯛꯄ) have similar character as in that of Manipuri verb '*cut*' (*kakpa*-ꯀꯛꯄ), '*suprise*' (*ngakpa*- ꯉꯛꯄ), '*return*' (*hallakpa*-ꯍꯜꯂꯛꯄ), '*come'* (*laakpa*-ꯂꯥꯛꯄ), '*clean'* (*sengdokpa*-ꯁꯦꯡꯗꯣꯛꯄ), '*draw'* (*yekpa*-ꯌꯦꯛꯄ), '*sing'* (*sakpa*-ꯁꯛꯄ), *laugh* (*nokpa*-ꯅꯣꯛꯄ), *smile* (*momon-nokpa*-ꯃꯣꯃꯣꯟ-ꯅꯣꯛꯄ) etc. Such verbs having similar patterns will apply the same suffixation rule. Thus suffixation rules are created on analysing 3274 Manipuri verbs with Present Tenses (Simple Present Tense, Present Continuous Tense, Present Perfect Tense,

and Present Perfect Continuous Tense), Past Tenses (Simple Past Tense, Past Continuous Tense, Past Perfect Tense, and Past Perfect Continuous Tense), and Future Tenses (Simple Future Tense, Future Continuous Tense, Future Perfect Tense, and Future Perfect Continuous Tense). All together 736 suffixation rules are generated for Manipuri verbs.

Figure 2 shows a screenshot of the suffixation rules represented in the verb suffix database. It consists of much information related to the verb- last but one character, last character, character to be removed from the verb, character to be added in the verb and tense information (auxiliary verb and main verb). These rules are implemented in the post-processing of EMMT to retrieve meaningful sentences.

Figure 2: A screenshot of the suffixation rules represented in the verb suffix database

5. Implementation of Verb Suffixation rules in English to Manipuri MT

English to Manipuri MT system translates an input English sentence to its corresponding Manipuri sentence. The steps of translation of EMMT are given as follows: firstly, pre-processing of the input sentence is done. It includes tokenization, tagging and parsing tools. After the process of parsing, English parse sentence is obtained. In English language, the sentence structure is subject-verb-object (SVO) and in Manipuri language, it is subject-object-verb (SOV). Secondly, the reordering of the SVO structure to SOV is done. The third step is the reordering of the subject phrase and object phrase. The fourth step is the lookup of bilingual English to Manipuri dictionary to retrieve translated Manipuri sentences. This translated sentence is not meaningful as the post-processing in Manipuri verbs is not done. Here, verb suffixation rules are applied to the translated Manipuri verb to get the correct form of the tense. The algorithm for the implementation process is given below, and figure 3 shows its graphical illustration.

Algorithm:

1. Read the meaning of the main verb in the input English sentence from the English to Manipuri bilingual dictionary.
2. Identify the last character and last but one character from the Manipuri verb.
3. Identify the auxiliary verb pattern and main verb tag of the input sentence.
4. Database lookup to identify the corresponding verb suffix to be added and the character to be removed from the Manipuri verb.
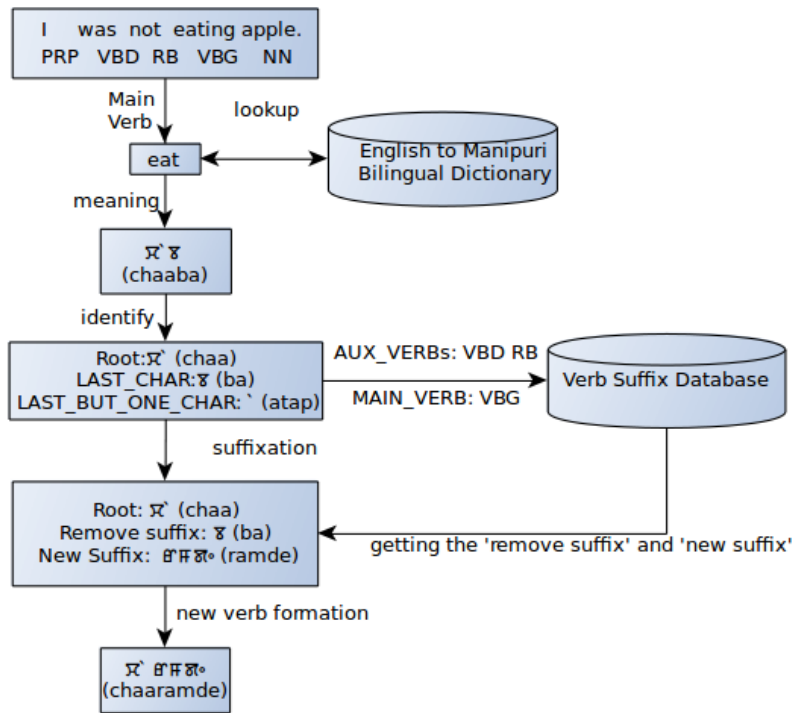5. Remove the old suffix and add the new suffix.

Figure 3: Graphical illustration of the verb suffixation process in English to Manipuri MT using the English sentence "I was not eating apple."

6. Result and Discussion

The performance of EMMT is measured using BLEU score. BLEU (Bilingual Evaluation Understudy) is an evaluation matric for machine translation and the similarity of the machine translated output is measured with reference to a set of human translated sentences. The result of EMMT outperforms the result of Google translate. The BLEU score of the EMMT is 0.70 and that of Google translate is 0.45. The comparison is performed using 1000 sentences of different tense forms. Appendix C presents the comparison of EMMT and Google translate with reference sentence. The analysis and creation of suffixation rules are time-consuming and tedious tasks. 3274 Manipuri verbs are studied, and as a result, 736 unique suffixation rules are developed. The testing result of the proposed verb suffixation module is given in table 3.

Table 3: The testing result of the suffixation of Manipuri verbs in English to Manipuri MT

	No. of input sentences	No. of outputs	Percentage
Correct verb	1000	960	96%
Incorrect verb	1000	40	4%

7. Conclusion

MT systems developed based on rule-based approach is a laborious task. The important tasks included in the EMMT are analysis of English sentence, collection of English to Manipuri re-ordering rules, creation of English to Manipuri bilingual dictionary, development of English to Manipuri transliteration system and post processing of verbs using suffixation rules. Suffixation is

used in various research works like classification of verbs, word formation, POS tagging etc. But in the proposed work, it is used in the post-processing of EMMT to indicate the tense of the translated Manipuri sentence. The verb suffixation rules are developed based on the three tense forms of English language. All the verbs in Manipuri language which have similar last and last but onecharacter works with the same suffix rule. The performance of the EMMT is practicable as compared with others rule-based English to Indian languages MT systems.

APPENDIX A: Outputs of EMMT

Input English sentence	Output Manipuri sentence
My name is Kanchan.	ꯑꯩꯒꯤ ꯃꯤꯡ ꯀꯥꯟꯆꯟ ꯅꯤ॥ *(Eigi ming Kanchan ni//)*
Sunitaa was seeing a dog.	ꯁꯨꯅꯤꯇꯥꯅ ꯍꯨꯏ ꯑꯃ ꯎꯔꯝꯈꯤ॥ *(Sunitaana hui ama uramkhi//)*
I ate an apple.	ꯑꯩ ꯁꯦꯝ ꯑꯃ ꯆꯥꯈꯤ॥ *(Ei sem ama chaakhi//)*
They are going to the park.	ꯃꯈꯣꯏ ꯄꯥꯔꯛ ꯑꯗꯨꯗ ꯆꯠꯂꯤ॥ *(Makhoi park aduda chatli//)*
She plays football.	ꯃꯍꯥꯛ ꯐꯨꯕꯜ ꯁꯥꯟꯅꯩ॥ *(Mahaak fooball saannei//)*
Babita had a red car.	ꯕꯕꯤꯇꯅ ꯑꯉꯥꯡꯕ ꯃꯆꯨꯒꯤ ꯀꯥꯔ ꯑꯃ ꯂꯩ॥ *(Babitana angaangba machugi car ama lei//)*
He is not dancing today.	ꯃꯍꯥꯛ ꯉꯁꯤ ꯖꯒꯣꯏ ꯁꯥꯗ꯭ꯔꯦ॥ *(Mahaak ngasi jagoi saadre//)*
I went to office.	ꯑꯩ ꯊꯕꯛ ꯀꯥꯐꯝꯗ ꯆꯠꯈꯤ॥ *(Ei thabak kaaphamda chatkhi//)*
Ibemhal is a good girl.	ꯏꯕꯦꯝꯍꯜ ꯑꯐꯕ ꯅꯨꯄꯤꯃꯆꯥ ꯑꯃ ꯅꯤ॥ *(Ibemhal aphaba nupimachaa ama ni//)*
I have been working since morning.	ꯑꯩ ꯑꯌꯨꯛ ꯃꯇꯝ ꯑꯗꯨꯗꯒꯤ ꯊꯕꯛ ꯇꯧꯔꯛꯂꯤ॥ *(Ei ayuk matam adudagi thabak tourakli//)*

APPENDIX B: Suffixes in Manipuri language based on the types of English Tense.

Types of tense in English	Complete Suffix in Manipuri	Parts of Suffix in Manipuri
Simple Present Tense	-꯭ꯢ (-i), -ꯃꯤ (-mi), -ꯂꯤ (-li), -ꯉꯤ (-ngi) etc	
Present Continuous Tense	-ꯔꯤ (-ri)/-ꯂꯤ (-li)	
Present Perfect Tense	-ꯔꯦ (-re)/-ꯂꯦ (-le)	
Present Perfect Continuous Tense	-ꯔꯛꯂꯤ (-rakli)/-ꯂꯛꯂꯤ (-lakli), -ꯇꯨꯅ ꯂꯩꯔꯤ (-tuna leiri)/-ꯗꯨꯅ ꯂꯩꯔꯤ (-duna leiri)	-ꯔꯛ (-rak)/-ꯂꯛ (-lak), -ꯂꯤ (-li), -ꯇꯨ (-tu), -ꯗꯨ (-du), -ꯅ (-na), -ꯂꯩ (-lei), -ꯔꯤ (-ri)
Simple Past Tense	-ꯈꯤ (-khi) and -ꯔꯝꯢ (-ram i)/-ꯂꯝꯢ (-lam i)	-ꯔꯝ (-ram)/-ꯂꯝ (-lam), -ꯢ (-i)
Past Continuous Tense	-ꯔꯝꯂꯤ (-ramli)/-ꯂꯝꯂꯤ (-lamli), -ꯔꯝꯈꯤ (-ramkhi)/-ꯂꯝꯈꯤ (-lamkhi)	-ꯔꯝ (-ram)/-ꯂꯝ (-lam), -ꯂꯤ (-li), -ꯈꯤ (-khi)
Past Perfect Tense	-ꯈ꯭ꯔꯦ (-khre)	

Past Perfect Continuous Tense	-ꯔꯝꯃꯤ (-rammi)/-ꯂꯝꯃꯤ (-lammi) , -ꯗꯨꯅ ꯂꯩꯔꯝ ꯏ (-duna leiram i)/-ꯇꯨꯅ ꯂꯩꯔꯝ ꯏ (-tuna leiram i)	-ꯔꯝ (-ram)/-ꯂꯝ (-lam), -ꯃꯤ (-mi), -ꯗꯨ (-du), -ꯇꯨ (-tu), -ꯅ (-na), -ꯂꯩ (-lei), -ꯏ (-i)
Simple Future Tense	-ꯒꯅꯤ (-gani)/-ꯀꯅꯤ (-kani)	-ꯒ (-ga)/-ꯀ (-ka), -ꯅꯤ (-ni)
Future Continuous Tense	-ꯔꯒꯅꯤ (-ragani)/-ꯂꯒꯅꯤ (-lagani)	-ꯔ (-ra)/-ꯂ (-la), -ꯒ (-ga), -ꯅꯤ (-ni)
Future Perfect Tense	-ꯔꯝꯒꯅꯤ (-ramgani)/-ꯂꯝꯒꯅꯤ (-lamgani), -ꯔꯝꯂꯒꯅꯤ (-ramlagani)/-ꯂꯝꯂꯒꯅꯤ (-lamlagani)	-ꯔꯝ (-ram)/-ꯂꯝ (-lam), -ꯒ (-ga), -ꯂ (-la), -ꯅꯤ (-ni)
Future Perfect Continuous Tense	-ꯗꯨꯅ ꯇꯨꯅ ꯂꯩꯔꯝꯒꯅꯤ (-duna leiramgani) /-ꯇꯨꯅ ꯇꯨꯅ ꯂꯩꯔꯝꯒꯅꯤ (-tuna leiramgani)	-ꯗꯨ (-du), -ꯇꯨ (-tu), -ꯅ (-na), -ꯂꯩ (-lei), -ꯔꯝ (-ram), -ꯒ -ga), -ꯅꯤ (-ni)

APPENDIX C: Comparison of the result of EMMT and Google Translate

1	Input Sentence: I learn computer. Reference Sentence: ꯑꯩ ꯀꯝꯄ꯭ꯌꯨꯇꯔ ꯇꯝꯃꯤ॥ *(ei computer tammi)* EMMT Result: ꯑꯩ ꯀꯝꯄ꯭ꯌꯨꯇꯔ ꯇꯝꯃꯤ॥ *(ei computer tammi)* Google Translate Result: ꯑꯩꯅ ꯀꯝꯄ꯭ꯌꯨꯇꯔ ꯇꯝꯃꯤ॥ *(eina NA tammi)*
2	Input Sentence: I like red flowers. Reference Sentence: ꯑꯩꯅ ꯑꯉꯥꯡꯕ ꯃꯆꯨꯒꯤ ꯂꯩꯁꯤꯡ ꯄꯥꯝꯃꯤ॥ *(eina angaangba machugi leising paammi.)* EMMT Result: ꯑꯩꯅ ꯑꯉꯥꯡꯕ ꯃꯆꯨꯒꯤ ꯂꯩꯁꯤꯡ ꯄꯥꯝꯃꯤ॥ *(eina angaangba machugi leising paammi.)* Google Translate Result: ꯑꯩꯅ ꯂꯂꯍꯧꯒꯤ ꯃꯆꯨ ꯄꯥꯝꯃꯤ॥ *(eina laalhougi(meaningless) machu paammi.)*
3	Input Sentence: I am hungry Reference Sentence: ꯑꯩ ꯆꯥꯛ ꯂꯥꯝꯃꯤ॥ *(Ei chaak laammi.)* EMMT Result: ꯑꯩ ꯆꯥꯛ ꯂꯥꯝꯕꯅꯤ॥ *(ei chaak laambani.)* Google Translate Result: ꯑꯩ ꯂꯝꯍꯥꯡꯗ ꯈꯣꯡ ꯍꯥꯝꯗꯔꯤ॥ *(Ei lamhaangda khong haamdari.)*
4	Input Sentence:I was not eating apple. Reference Sentence:ꯑꯩꯅ ꯁꯦꯝ ꯆꯥꯔꯝꯗꯦ॥ *(Einaa sem chaaramde.)* EMMT Result:ꯑꯩꯅ ꯁꯦꯝ ꯆꯥꯔꯝꯗꯦ॥ *(Einaa sem chaaramde.)* Google Translate Result:ꯑꯩꯍꯥꯛ ꯑꯦꯄꯜ ꯆꯥꯗꯔꯤ॥ *(eihaak apple chaadri.)*
5	Input Sentence: He is not dancing today. Reference Sentence: ꯃꯍꯥꯛ ꯉꯁꯤ ꯖꯒꯣꯏ ꯁꯥꯗ꯭ꯔꯦ॥ *(Mahaak ngasi jagoi saadre.)* EMMT Result: ꯃꯍꯥꯛ ꯉꯁꯤ ꯖꯒꯣꯏ ꯁꯥꯗ꯭ꯔꯦ॥ *(Mahaak ngasi jagoi saadre.)* Google Translate Result: ꯃꯍꯥꯛ ꯉꯁꯤꯗꯤ ꯅꯥꯟꯊꯣꯛꯂꯤꯕꯅꯤ॥ *(Mahaak ngasidi naanthoklibani.)*
6	Input Sentence: The dog barks in the night. Reference Sentence: ꯍꯨꯏꯗꯨ ꯅꯨꯃꯤꯗꯥꯡꯗꯨꯗ ꯈꯣꯡꯉꯤ॥ *(Huidu numidaangduda khongngi.)* EMMT Result: ꯍꯨꯏꯗꯨ ꯅꯨꯃꯤꯗꯥꯡꯗꯨꯗ ꯈꯣꯡꯉꯤ॥ *(Huidu numidaangduda khongngi.)*

	Google Translate Result: ꯀꯨꯠꯇ ꯅꯨꯃꯤꯗꯥꯡꯗ ꯍꯨꯔꯥꯜꯂꯤ॥ *(Kutta numidaangda huraalli.)*
7	Input Sentence: Cow gives us milk. Reference Sentence: ꯁꯟꯅ ꯑꯩꯈꯣꯏꯗ ꯁꯪꯒꯣꯝ ꯄꯤ॥ *(Sanna eikhoida sanggom pi.)* EMMT Result: ꯁꯟꯅ ꯑꯩꯈꯣꯏꯕꯨ ꯁꯪꯒꯣꯝ ꯄꯤ॥ *(Sanna eikhoibu sanggom pi.)* Google Translate Result: ꯁꯟ ꯑꯃꯅꯥ ꯑꯩꯈꯣꯏꯗ ꯁꯪꯒꯣꯝ ꯄꯤ॥ *(San amanaa eikhoida sangagom pi.)*

References

ANTONY P J. 2013. Computational Linguistic Tools and Machine Translation System for Kannada Langusge. Ph.D. Thesis. Amrita Vishwa Vidyapeetham: Coimbatore.

BARUAH, KALYANEE KANCHAN; PRANJAL DAS, ABDUL HANNAN. & SHIKHAR KR SARMA. 2014 Assamese-English Bilingual Machine Translation. *International Journal on Natural Language Computing (IJNLC).*

BATRA, KAMALJEET KAUR & G S LEHAL. 2010. Rule Based Machine Translation of Noun Phrases from Punjabi to English. *International Journal of Computer Science Issues.* 409-413.

CHOUDHARY, HIMANSHU; ADITYA KUMAR PATHAK, RAJIV RATN SHAH & PONNURANGAM KUMARAGURU. 2018. Neural Machine Translation for English-Tamil. *In the Proceedings of the Third Conference on Machine Translation (WMT).* Brussels, Belgium. 770–775.

CHUNGKHAM YASHAWANTA SINGH. 2011. Manipuri Grammar. New Delhi: Rajesh Publications.

CIGDEM KEYDER TURHAN. 1997. An English to Turkish Machine Translation System Using Structural Mapping. *In proceedings of Fifth conference on Applied Natural Language Processing of Association for Computational Linguistics.* 320-323.

DESAI, PRATIK; AMIT SANGODKAR AND OM P. DAMANI. 2014. A Domain-Restricted, Rule-Based, English-Hindi Machine Translation System Based on Dependency Parsing. *Proceedings of the 11th International Conference on Natural Language Processing.* 177–185.

DIXIT, NEHA & NARAYAN CHOUDHARY. 2014. Automatic Classification of Hindi Verbs in Syntactic Perspective. *International Journal of Emerging Technology and Advanced Engineering.* 572-579.

JAYASHREE NAIR & VINOD, J. 2019. Design of a Morphological Generator for English to Indian Languages in a Declension Rule-based Machine Translation System. *First International Conference on Advances in Electrical and Computing Technologies 2019.* ICAECT: Coimbatore.

KM, SHIVAKUMAR; NAYANA S & SUPRIYA T. 2015. A Study of Kannada to English Baseline Statistical Machine Translation System. *International Journal of Applied Engineering Research.*

KUMAR, R. RAVINDRA; SULOCHANA KG & JAYAN V. 2011. Computational Aspect of Verb Classification in Malayalam. *International Conference on Information Systems for Indian Languages.* 15-22.

MEITEI, NAOREM JIBIT. 2014. Modern English Grammar cum Manipuri Translation. *Nongmaithem Premlata Devi and Naorem Helenjit (Theba) Meitei.* Naorem Leikai. Manipur.

MURTHY, K. 2002. MAT: A Machine Assisted Translation System. *In Proceedings of Symposium on Translation Support System(STRANS-2002).* IIT Kanpur. 134-139.

ROY, BIPUL & BIPUL SYAM PURKAYASTHA. 2017. A Suffix Based Morphological Analysis of Assamese Word Formation. *International Journal on Recent and Innovation Trends in Computing and Communication.* 445-449.

RAJENDRAN, S. 2001. Word Formation in Tamil. UGC Major Project Report. Department of Linguistics, Tamil University: Thanjavur.

SAINI, SANDEEP & VINEET SAHULA. 2018. Neural Machine Translation for English to Hindi. *Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*.

SINGH, THOUDAM DOREN & SIVAJI BANDYOPADHYAY. 2008. Morphology Driven Manipuri POS Tagger. *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*. 91–98.

SINGH, THOUDAM DOREN & SIVAJI BANDHYOPADHYAY. 2010. Manipuri-English Example Based Machine Translation System. *IJCLA*. 201-216.

SpaCy v3.2. 2021. Part-of-Speech Tagging & Dependency Parsing. Accessed online: https://spacy.io.

About the authors:

Mayanglambam Premi Devi has completed her Master in Computer Application in the Department of Computer Science, Manipur University, Canchipur, Imphal, India. Currently, she is working as Graduate Teacher in the Department of Education (S), Manipur. She has submitted her PhD thesis in the area of English to Manipuri Machine Translation at Manipur University. She is interested in the Natural Language Processing (NLP) and Optical Character Recognition (OCR). She has published several conference papers and International Journal papers in her research area. Email: premimaya007[AT]gmail[DOT]com

Irengbam Tilokchan Singh is a PhD student of Department of Computer Science, Manipur University, India. He was the Former Executive – Software Developer at Shija Hospitals & Research Institute Pvt. Ltd. Imphal, India. He has published several International journals and conference papers. He has keen knowledge of working centralized management information system in corporate & academic. He has interest in Server Load Balancing with Software Defined Networking. Email: tilokchan[AT]gmail[DOT]com

Haobam Mamata Devi is an Associate Professor in the Department of Computer Science, Manipur University, India. She has completed her Ph.D in Computer Science from Manipur University. Her areas of interest are Natural Language Processing, Text Mining and Document Image Processing. She has several publications in various journals and conferences She is a life member of Computer Society of India (SCI) and Institute of Electronic and Telecommunication Engineer (IETE). Email: mamata[_]dh[AT]rediffmail[DOT]com