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, ꯍꯥꯡꯕꯥ (*haraoba*-happy) + ꯔꯥꯏ (*ram i*)=ꯍꯥꯡꯕꯔꯥꯏ (*haraoram i*-was happy). Here, ꯍꯥꯡꯕ (*harao)* is the root word and ꯥ (*ba)* is the suffix of the verb ꯍꯥꯡꯕꯥ (*haraoba).* Since Manipuri verbs are bound root, a suffix is always associated with them. In the example, ꯔꯥꯏ (*ram i*) is a suffix that indicates that a Manipuri sentence is in the simple past tense. While writing the verb ꯍꯥꯡꯕꯥ (*haraoba*) and suffix ꯔꯥꯏ (*ram i*) together, the suffix ꯥ (*ba*) is removed from the verb ꯍꯥꯡꯕꯥ (haraoba) and added the suffix ꯔꯥꯏ (*ram i*) to get the word ꯍꯥꯡꯕꯔꯥꯏ (*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 a corpus-based English to Manipuri

automatic MT system. Section 2 presents the literature review of the works wherein 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.

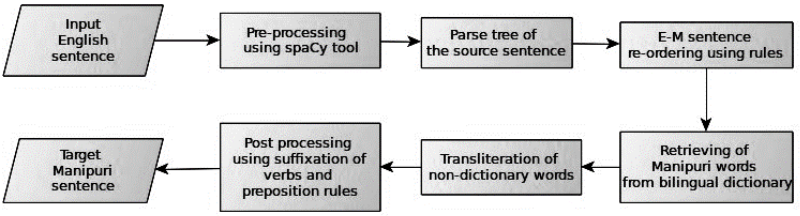

Figure 1: Process flow of English to Manipuri MT

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.

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)	ꯍ (i)	ꯇꯣ (te)	ꯊꯀꯄ(thakpa-drink)	ꯊꯀꯍ (thak i-drink), ꯊꯀꯇꯣ (thakte-do not drink)
ꯑ (un) / ꯔꯥ (aatap)/ ꯔꯣ (otnap)/ ꯔꯨ (unap) / ꯔꯪ(sounap)	ꯕ (ba)	ꯕ (ba)	ꯍ (i)	ꯗꯤ (de)	ꯑꯔꯕ(kaoba-forget)	ꯑꯔꯍ(kao i-forget), ꯑꯔꯗꯤ(kaode-do not forget)
ꯃ (mit lonsum)	ꯕ (ba)	ꯕ (ba)	ꯍ (i)	ꯗꯤ (de)	ꯃꯤꯁꯤꯕ(minamba-lie)	ꯃꯤꯁꯤꯍ(minami-lie), ꯃꯤꯁꯤꯗꯤ(minamde-do not lie)

ꯄ (pa lonsum)	ꯄ (pa)	ꯄ (pa)	ꯄꯤ (pi)	ꯇꯣ (te)	ꯀꯄꯄ(kappa-cry)	ꯀꯄꯄꯤ(kappi-cry), ꯀꯄꯇꯣ(kapte-do not cry)
ꯅ (naa lonsum)	ꯕ (ba)	ꯅꯕ/ ꯕ	ꯂꯤ (li)	ꯗꯣ (de)	ꯌꯣꯅꯕ(yonba-sell)	ꯌꯣꯂꯤ(yolli-sell), ꯌꯣꯅꯗꯣ(yonde-do not sell)
ꯉ (ngou lonsum)/ ꯭ (nung)	ꯕ (ba)	ꯕ (ba)	ꯉꯤ (ngi)	ꯗꯣ (de)	ꯌꯦꯡꯕ(yengba-see)	ꯌꯦꯡꯉꯤ(yengngi-see), ꯌꯦꯡꯗꯣ(yengde-do not see)
ꯎ (til lonsum)	ꯄ (pa)	ꯄ (pa)	ꯆꯤ (li)	ꯇꯣ (te)	ꯆꯎꯄ(chatpa-go)	ꯆꯎꯆꯤ(chatli-go), ꯆꯎꯇꯣ(chatte-do not go)

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

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: ꯍꯥꯡ꯭ꯄ (*mahaak*-he) ꯏ꯴꯭ꯤꯡ (*ishing*-water) ꯊꯄꯏ (*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 ꯏ (*i*) for the non-negative sentence is added to get the target word ꯊꯛꯏ (*thak i*). If the sentence is negative, then the suffix ꯞ꯭ꯠ (*te*) should be added to get ꯊꯛꯞ꯭ꯠ (*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*'.

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)	ꯏ (i)	ꯞ꯭ꯠ (te)	ꯊꯛꯏ (thak i-drink), ꯊꯛꯞ꯭ꯠ (thakte-do not drink)
Present Continuous Tense	ꯊꯛꯄ (*thakpa*-drink)	ꯛ (*kok lonsum)*	ꯄ (pa)	ꯄ (pa)	ꯊꯤ (li)	ꯞ꯭ꯠ꯭ꯔ (tre)	ꯊꯛꯊꯤ(thakli-is drinking), ꯊꯛꯞ꯭ꯠ꯭ꯔ꯫(thaktre-is not drinking)

Present Perfect Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂ꯭ (le)	ꯇ꯭ꯔꯤ (tri)	ꯊꯕꯂ꯭(thakle-has drunk), ꯊꯕꯇ꯭ꯔꯤ(thaktri-has not drunk)
Present Perfect Continuous Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂꯕꯂꯤ (lakli)/ ꯇꯨꯅ ꯂ꯭ꯦꯔꯤ (tuna leiri)	ꯂꯕ ꯇ꯭ꯔꯤ (laktri) / ꯇꯨꯅ ꯂ꯭ꯦꯇ꯭ꯔꯤ (tuna leitri)	ꯊꯕꯂꯕꯂꯤ(thaklakli-has been drinking), ꯊꯕꯂꯕꯇ꯭ꯔꯤ(thaklaktri-has not been drinking)
Simple Past Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯈꯤ (khi)/ ꯂꯝ ꯏ (lam i)	ꯇ꯭ꯦ(te) / ꯂꯝ ꯗꯦ꯭ (lamde)	ꯊꯕꯈꯤ(thakkhi-drank), ꯊꯕꯇ꯭ꯦ(thakte-did not drink)
Past Continuous Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂꯝꯈꯤ (lamkhi)/ ꯂꯝꯂꯤ (lamli)	ꯂꯝ ꯗꯦ꯭ (lamde)	ꯊꯕꯂꯝꯈꯤ(thaklamkhi-was drinking), ꯊꯕꯂꯝ ꯗꯦ꯭(thaklamde-was not drinking)
Past Perfect Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯈ꯭ꯔꯦ (khre)	ꯈꯤ ꯗ꯭ꯔꯤ (khidri)	ꯊꯕꯈ꯭ꯔꯦ(thakkhre-had drunk), ꯊꯕꯈꯤꯗ꯭ꯔꯤ(thakkhidri-had not drunk)
Past Perfect Continuous Tense	ꯊꯕ ꯄ (*thak pa-* drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂꯝ ꯍꯤ (lammi)/ ꯇꯨꯅ ꯂ꯭ꯦꯝ ꯍ (tuna	ꯂꯝ ꯗꯦ꯭ (lamde)/ ꯇꯨꯅ ꯂ꯭ꯦꯝ ꯗꯦ꯭ (tuna	ꯊꯕꯂꯝ ꯍꯤ(thaklammi-had been drinking), ꯊꯕꯂꯝ ꯗꯦ꯭(thaklamde-had not been drinking)

					leiram i)	leiram de)	
Simple Future Tense	ꯊꯕ ꯄ (*thak pa*-drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯀꯟꯍꯤ (kani)	ꯂꯣꯏ (loi)	ꯊꯕꯀꯟꯍꯤ(thakkani-will drink), ꯊꯕꯂꯣꯏ(thakloi-will not drink)
Future Continuous Tense	ꯊꯕ ꯄ (*thak pa*-drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂꯥꯡꯍꯤ (lagani)	ꯂꯥꯣꯏ (laroi)	ꯊꯕꯂꯥꯡꯍꯤ(thaklagani-will be drinking), ꯊꯕꯂꯥꯣꯏ(thaklaroi-will not be drinking)
Future Perfect Tense	ꯊꯕ ꯄ (*thak pa*-drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯂꯝꯃꯍꯤ (lamgani)/ ꯂꯝꯂꯥꯡꯍꯤ (lamlagani)	ꯂꯝꯂꯣꯏ (lamlaroi)	ꯊꯕꯂꯝꯃꯍꯤ(thaklamgani-will have drunk), ꯊꯕꯂꯝꯂꯣꯏ(thaklamlaroi-will not have drunk)
Future Perfect Continuous Tense	ꯊꯕ ꯄ (*thak pa*-drink)	ꯀ (*kok lons um)*	ꯄ (pa)	ꯄ (pa)	ꯇꯨꯅ ꯂꯩꯔꯝ ꯍꯤ (tuna leiramgani)	ꯇꯨꯅ ꯂꯩꯔꯝ ꯂꯣꯏ (tuna leiramloi)	ꯊꯕꯇꯨꯅ ꯂꯩꯔꯝꯍꯤ(thaktuna leiramgani-will have been drinking), ꯊꯕꯇꯨꯅ ꯂꯩꯔꯝꯂꯣꯏ(thaktuna leiramloi-will have not been drinking)

Table 2: Analysis of Suffixes of the Verb ‘drink’ ꯊꯕꯄ(*thakpa*)

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

1. (a) English: She drinks water. (Simple present tense)
1. (b) Manipuri: *mahaak(*she) *ishing (*water) *thak i(*drinks).
ꯍꯥꯡ ꯃ (*mahaak*) ꯏꯁꯤꯡ (*ishing*) ꯊꯕꯛ (*thak i*)॥(.)

ꯍꯥꯡ ꯃ ꯏꯁꯤꯡ ꯊꯕꯛ ॥

2. (a) English: She does not drink water. (Negative Simple present tense)
2. (b) Manipuri: *mahaak(*she) *ishing (*water) *thakte (*does not drink).
ꯍꯥꯡ ꯃ (*mahaak*) ꯏꯁꯤꯡ (*ishing*) ꯊꯕꯗꯦ (*thakte*)॥(.)
ꯍꯥꯡ ꯃ ꯏꯁꯤꯡ ꯊꯕꯗꯦ ॥

3. (a) English: Raam was drinking tea. (Past Continuous tense)
3. (b) Manipuri: *Raam chaa* (tea) *thaklamkhi* (was drinking).
ꯔꯥ (*Raam*) ꯆꯥ (*chaa*) ꯊꯕꯀꯝꯍꯤ(*thaklamkhi*)॥(.)
ꯔꯥ ꯆꯥ ꯊꯕꯀꯝꯍꯤ ॥

4. (a) English: Raam was not drinking tea. (Negative Past Continuous tense)
4. (b) Manipuri: *Raam chaa* (tea) *thaklamde* (was not drinking).
ꯔꯥ (*Raam*) ꯆꯥ (*chaa*) ꯊꯕꯀꯝꯗꯤ(*thaklamde*)॥(.)
ꯔꯥ ꯆꯥ ꯊꯕꯀꯝꯗꯤ ॥

5. (a) English: They will have drunk coffee. (Future perfect tense)
5. (b) Manipuri: *Makhoi* (they) *coffee* (coffee) *thaklamgani* (will have drunk).
ꯍꯛꯍꯥꯡ (*makhoi*) ꯀꯣꯐꯤ (*coffee*) ꯊꯕꯀꯝꯄꯤ (*thaklamgani*) ॥(.)
ꯍꯛꯍꯥꯡ ꯀꯣꯐꯤ ꯊꯕꯀꯝꯄꯤ ॥

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).
ꯍꯛꯍꯥꯡ (*makhoi*) ꯀꯣꯐꯤ (*coffee*) ꯊꯕꯀꯝꯔꯣꯍꯥꯡ (*thaklamlaroi*) ॥(.)
ꯍꯛꯍꯥꯡ ꯀꯣꯐꯤ ꯊꯕꯀꯝꯔꯣꯍꯥꯡ ॥

The ‘*last character*’ ꯄ(*pa*) and ‘*character before the last character*’ ꯀ(*kok lonsum)* of the lemma form of the Manipuri verb ‘*drink*’ (*thakpa*-ꯊꯕꯄ) have similar character as in that of Manipuri verb

‘*cut*’ (*kakpa-*ꯀꯛꯄꯥ), ‘*surprise*’ (*ngakpa-* ꯉꯛꯄꯥ), ‘*return*’ (*hallakpa-*ꯍꯜꯂꯛꯄꯥ), ‘*come’* (*laakpa-*ꯂꯥꯛꯄꯥ), *‘clean’* (*sengdokpa-*ꯁꯩꯒꯣꯛꯄꯥ), *‘draw’* (*yekpa-*ꯌꯦꯛꯄꯥ), *‘sing’* (*sakpa-*ꯁꯛꯄꯥ), *laugh* (*nokpa-*ꯅꯣꯛꯄꯥ), *smile* (*momon-nokpa-*ꯃꯣꯃꯣꯟ-ꯅꯣꯛꯄꯥ) 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.

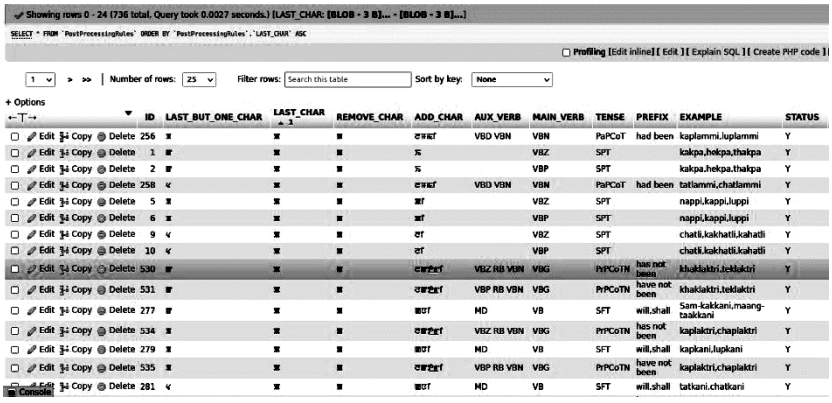

Figure 2: A Screenshot of the Suffixation Rules Represented in the Verb Suffix Database

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 the 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.

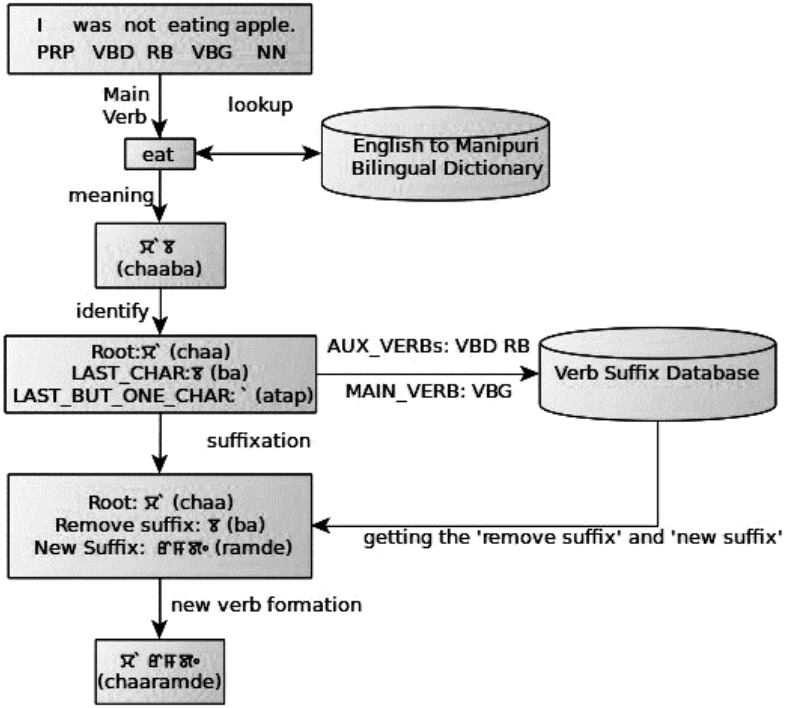

Figure 3: Graphical illustration of the verb suffixation process in English to Manipuri MT using the English sentence "I was not eating apple."

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.

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

Table 3: The Testing Result of the Suffixation of Manipuri Verbs in English to Manipuri MT

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 one-character works with the same suffix rule. The performance of the EMMT is practicable as compared with other 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 ꯑꯣꯏꯔꯤ꯫ *(Eigi ming Kanchan ni||)*
Sunitaa was seeing a dog.	ꯁꯨꯅꯤꯇꯥ ꯑꯣꯃ ꯑꯃ ꯀꯨꯔꯥꯃ꯫ *(Sunitaana hui ama uramkhi||)*
I ate an apple.	ꯑꯩ ꯁꯦꯃ ꯑꯃ ꯆꯥꯀꯤ꯫ *(Ei sem ama chaakhi||)*
They are going to the park.	ꯃꯈꯣꯏ ꯄꯥꯔ ꯑꯇꯨꯞ ꯆꯜꯂꯤ꯫ *(Makhoi park aduda chatli||)*
She plays football.	ꯃꯍꯥꯀ ꯐꯨꯂꯣꯇ ꯁꯥꯟꯅꯩ꯫ *(Mahaak fooball saannei||)*
Babita had a red car.	ꯕꯕꯤꯇꯥ ꯑꯉꯥꯡꯕꯥ ꯃꯍꯧꯁꯤ ꯄꯥꯞ ꯑꯃ ꯀꯥꯛ꯫ *(Babitana angaangba machugi car ama lei||)*
He is not dancing today.	ꯃꯍꯥꯀ ꯉꯁꯤ ꯆꯦꯏꯅ ꯁꯥꯞ꯫ *(Mahaak ngasi jagoi saadre||)*
I went to office.	ꯑꯩ ꯊꯕꯛ ꯄꯥꯡꯐꯝ ꯆꯜꯀꯤ꯫ *(Ei thabak kaaphamda chatkhi||)*
Ibemhal is a good girl.	ꯏꯕꯦꯝꯍꯥꯜ ꯑꯐꯕ ꯅꯨꯄꯤꯃꯆꯥ ꯑꯃꯀꯤ꯫ *(Ibemhal aphaba nupimachaa ama ni||)*
I have been working since morning.	ꯑꯩ ꯑꯇꯨꯝ ꯃꯁꯤ ꯑꯃꯁꯨꯄꯤ ꯊꯕꯛ ꯁꯤꯔꯕꯥꯀꯤ꯫ *(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	-ꯏ (-i), −ꯃꯤ (-mi), -ꯂꯤ (-li), −ꯉꯤ (-ngi) etc	
Present Continuous Tense	-ꯔꯤ (−ri)/-ꯂꯤ (-li)	
Present Perfect Tense	-ꯔꯣ (-re)/-ꯂꯣ (-le)	
Present Perfect Continuous Tense	-ꯔꯗꯂꯤ (-rakli)/-ꯂꯗꯂꯤ (-lakli), -ꯇꯨꯅ ꯂꯥꯔꯤ (-tuna leiri)/-ꯗꯨꯅ ꯂꯥꯔꯤ (-duna leiri)	-ꯔꯗ (-rak)/-ꯂꯗ (-lak), -ꯂꯤ (-li), -ꯇꯨ (-tu), -ꯗꯨ (-du), -ꯅ (-na), -ꯂꯥ (-lei), -ꯔꯤ (-ri)
Simple Past Tense	-ꯈꯤ (-khi) and -ꯔꯃꯏ (-ram i)/-ꯂꯃꯏ (-lam i)	-ꯔꯃ (-ram)/-ꯂꯃ (-lam), -ꯏ (-i)
Past Continuous Tense	-ꯔꯃꯂꯤ (-ramli)/-ꯂꯃꯂꯤ (-lamli), -ꯔꯃꯈꯤ (-ramkhi)/-ꯂꯃꯈꯤ (-lamkhi)	-ꯔꯃ (-ram)/-ꯂꯃ (-lam), -ꯂꯤ (-li), -ꯈꯤ (-khi)
Past Perfect Tense	-ꯈꯔꯣ (-khre)	
Past Perfect Continuous Tense	-ꯔꯃꯃꯤ (-rammi)/-ꯂꯃꯃꯤ (-lammi) , -ꯗꯨꯅ ꯂꯥꯔꯃꯏ (-duna leiram i)/-ꯇꯨꯅ ꯂꯥꯔꯃꯏ (-tuna leiram i)	-ꯔꯃ (-ram)/-ꯂꯃ (-lam), −ꯃꯤ (−mi), -ꯗꯨ (-du), −ꯇꯨ(-tu), −ꯅ (-na), −ꯂꯥ (-lei), -ꯏ (-i)
Simple Future Tense	-ꯒꯤ (-gani)/-ꯀꯤ (-kani)	-ꯒ (-ga)/-ꯀ (-ka), -ꯤ (-ni)
Future Continuous Tense	-ꯔꯒꯤ (-ragani)/-ꯂꯒꯤ (-lagani)	-ꯔ (-ra)/-ꯂ (-la), -ꯒ (-ga), -ꯤ (-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 ꯂꯩꯔꯝꯒꯅꯤ (-duna leiramgani) /-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 ꯀꯝꯄꯨꯇꯔ ꯇꯝꯃꯤ꯫ *(ei computer tammi)* EMMT Result: ꯑꯩ ꯀꯝꯄꯨꯇꯔ ꯇꯝꯃꯤ꯫ *(ei computer tammi)* Google Translate Result: ꯑꯩꯅꯥ ꯀꯝꯄꯨꯇꯔ ꯇꯝꯃꯤ꯫ *(eina NA tammi)*
2	Input Sentence: I like red flowers. Reference Sentence: ꯑꯩꯅꯥ ꯑꯉꯥꯡꯕ ꯃꯆꯨꯒꯤ ꯂꯩꯁꯤꯡ ꯄꯥꯝꯃꯤ꯫ *(eina angaangba machugi leising paammi.)* EMMT Result: ꯑꯩꯅꯥ ꯑꯉꯥꯡꯕ ꯃꯆꯨꯒꯤ ꯂꯩꯁꯤꯡ ꯄꯥꯝꯃꯤ꯫ *(eina angaangba machugi leising paammi.)* Google Translate Result: ꯑꯩꯅꯥ ꯂꯥꯜꯍꯧꯒꯤ 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 ꯂꯝꯍꯥꯡꯗꯥ ꯈꯣꯡꯍꯥꯝꯗꯔꯤ꯫ *(Ei lamhaangda khong haamdari.)*
4	Input Sentence:I was not eating apple. Reference Sentence:ꯑꯩꯅꯥ ꯁꯦꯝ ꯆꯥꯔꯝꯗꯦ꯫ *(Einaa sem chaaramde.)* EMMT Result:ꯑꯩꯅꯥ 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 ꯖꯒꯣꯢ ꯁꯥꯗ꯭ꯔꯦ꯫ *(Mahaak ngasi jagoi saadre.)* EMMT Result: ꯃꯍꯥꯛ ꯉꯁꯤ ꯖꯒꯣꯢ 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: ꯍꯨꯢꯗꯨ ꯅꯨꯃꯤꯗꯥꯡꯗꯨꯗ ꯈꯣꯡꯉꯤ꯫ *(Huidu numidaangduda khongngi.)* EMMT Result: ꯍꯨꯢꯗꯨ ꯅꯨꯃꯤꯗꯥꯡꯗꯨꯗ ꯈꯣꯡꯉꯤ꯫ *(Huidu numidaangduda khongngi.)* Google Translate Result: ꯀꯨꯇ꯭ꯇꯥ ꯅꯨꯃꯤꯗꯥꯡꯗ ꯍꯨꯔꯥꯜꯂꯤ꯫ *(Kutta numidaangda huraalli.)*
7	Input Sentence: Cow gives us milk. Reference Sentence: ꯁꯅ꯭ꯅ ꯑꯩꯈꯣꯢꯗ ꯁꯪꯒꯣꯝ ꯄꯤ꯫ *(Sanna eikhoida sanggom pi.)* EMMT Result: ꯁꯅ꯭ꯅ ꯑꯩꯈꯣꯢꯕꯨ ꯁꯪꯒꯣꯝ ꯄꯤ꯫ *(Sanna eikhoibu sanggom pi.)* Google Translate Result: ꯁꯟ ꯑꯃꯅꯥ ꯑꯩꯈꯣꯢꯗ ꯁꯉꯒꯣꯝ 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. *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. *The 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. *The 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.* Manipur: Naorem Leikai.

MURTHY, K. 2002. MAT: A Machine Assisted Translation System. *The 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. The *Proceedings of 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 Author

Mayanglambam Premi Devi 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 has published several research papers in international journals and conferences. 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. Her areas of interest are Natural Language Processing and 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

Cite this Work:

Premi Devi, Mayanglambam, Tilokchan Singh, Irengbam & Mamata Devi, Haobam. 2022. Suffixation Rules of Manipuri Verbs in English to Manipuri Machine Translation. *Translation Today*, Vol. 16(2). 23-44. DOI: 10.46623/tt/2022.16.2.ar2