Transfer Grammar Rules for Malayalam to English Translation

Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D., Rajeev R. R., M.Sc., M.Phil., Ph.D., Archana S Krishnan, M.A., and Navaneeth S., M.Tech.

Abstract

This paper presents a rule-based approach to machine translation from Malayalam to English. Malayalam, a Dravidian language with rich morphological and syntactic structures, poses significant challenges for machine translation. To address these complexities, we propose a set of transformation rules that map Malayalam expressions onto their English equivalents. The rules are designed to handle various grammatical features, including case markers, tense, and aspect. By focusing on these key linguistic elements, we aim to improve the accuracy and fluency of machine translation between the two languages, contributing to advancements in natural language processing and cross-lingual communication.

Keywords: Malayalam, Dravidian language, Machine translation, morphological analysis, grammar, syntax, semantics.

Introduction

Machine Translation is an important area in computational linguistics that automatically translates text from one language to another. There are so many different methods used for machine translation, which are rule-based methods, transfer grammar-based methods, statistical methods, machine learning, deep learning, etc. In this paper, we attempted to list the transfer grammar rules needed for machine translation from Malayalam to English. Malayalam is a language spoken in southern India. It is the main language in the state of Kerala and also spoken in Lakshadweep and Puducherry. Over 35 million people in India speak Malayalam, making it one of the country's

major languages. Malayalam is officially recognized by Government of India and is considered a classical language of India. Malayalam is a Dravidian language family language and stands out for its fascinating grammar. Unlike English, it relies on an agglutinative structure, where words are built by attaching suffixes to root words. This allows for a rich and expressive vocabulary. The word order in Malayalam is typically subject-object-verb (SOV), but it can be flexible for emphasis. Nouns change their form depending on their role in the sentence (case) and whether they are singular or plural (Lakshmi & Sobha Lalitha Devi, 2014). Verbs take center stage with conjugations that reflect tense, mood, and even whether the action is caused by someone. In older forms of Malayalam, verbs went even further, indicating person, gender, number, and polarity. adjectives, adverbs, prepositions, and conjunctions remain composed by their context, maintaining their original form.

Unlike Malayalam, with its distinct noun and verb inflections, English grammar takes a different approach. English belongs to a category of languages where word classes, like nouns, verbs, and adjectives, aren't readily apparent from a word's form itself. This can be confusing for translation, as the same word can sometimes be used in multiple ways. Take "run" for example, which can function as both a verb ("He runs every morning") and a noun ("We need a good run"). This is because inflectional endings, which modify a word's meaning based on its role in a sentence, are less common in English compared to languages like Malayalam (Nair, 2012). However, English does have its own methods for word formation. Derivational suffixes, like "-er" turning "run" into "runner," can create new words from existing ones. This flexibility makes English a dynamic language, but it can also present a challenge for those trying to untangle its grammatical intricacies.

There are significant variations between Malayalam and English grammar on a number of levels. Word order, the general grammatical structure, the usage of inflections, and even the way meaning is expressed are some examples of these.

Literature Survey

This paper reviews significant research on machine translation transfer rules. By examining these rules, we aim to understand how linguistic knowledge can be applied to improve the accuracy and fluency of machine translation systems.

Design and development of a Malayalam to English Translator - A Transfer based approach, by Latha R Nair.et.al. (2012) [8]. In this paper they proposed a Rule based Translation system for Malayalam to English. Used linguistic components of Malayalam and English for this system. Multiple modules used in this system like compound word splitter, Morphology based parser, Transfer module and Target sentence generator etc. Tested 1000 different kinds of sentences and resulted in 20% exact English versions of exact sentences. This system could give exact Translation of input sentences when adding more Transfer rules.

Transfer grammar components for Malayalam to Hindi and English Machine Translation system, by Stephy Joseph.et.al. (2014) [20]. In this paper they presented a translation system in the Hybrid approach based on rule-based and statistical methods. Tokenization, POS tagging, Chunking, Morphological analysis and Transfer grammar generation are the main modules used in this work. using a rule-based approach for the Data structure for transfer grammar. Nouns exhibit Subject and object properties. In this scenario subject and object identification is very difficult. For that, they used a Translation tagset for identifying nouns which act as a subject or object in a sentence. And used an SVM tool for POS tagging. In each stage they calculated accuracy using functional parameters such as precision and recall. They got the accuracy level of 80% for the Malayalam-Hindi machine translation. The proposed system is only suitable for Malayalam to Hindi Machine translation system. Malayalam to English transfer grammar rules are incomplete in this system.

Relative Study on Malayalam-English Translation Using a Transfer-Based Approach, Shahana.et.al. (2014) [19]. This paper has also discussed Malayalam-English transfer rule constructions using a rule-based methodology. This paper explains the major steps, such as the Malayalam transliteration, the bilingual dictionary-English, the Malayalam morphological generator, and the reordering by transfer rules, etc. and explanations of key computational modules Compound Word Splitter Module, Syntactic Structure Transfer Module and Target Sentence Generator Module. which comprises a preprocessor for splitting the compound words, a morphological parser for context disambiguation and chunking, a syntactic structure transfer module, and a bilingual dictionary.

A Rule Based Approach for Translation of Causative Construction of English and Malayalam for the Development of Prototype for Malayalam to English and English To Malayalam Bilingual Machine Translation System: Bijimol.et.al (2018) [3]. This is a bilingual Malayalam-English and English-Malayalam machine translation system. In this paper they are proposing a Rule-based method for the translation of causative sentences in Malayalam and English and English to Malayalam causative sentences. There are five main modules in the systems: Preprocessing Module, Causative Sentence Processing Module, Malayalam-English Translation Module, English-Malayalam Translation Module and Bilingual Dictionary. The grammar side of both languages Forms of Causative Sentences in English and Malayalam explained and identified changes about the tense marker in both languages. The paper mentioned that the proposed system performs better than that of Google translator for the translation of causative sentences.

Although the above-mentioned studies discuss the various methods of Malayalam and English machine translation, the changes of the suffixes in the Malayalam and English grammatical rules, such as nouns and verbs, are not precisely marked. Our study is making progress by incorporating grammatical changes from Malayalam to English.

Relevance of the study

Transfer grammar rules are meant to connect the syntax of two languages. The use of precise rules regarding the syntax of different languages makes mechanistic translation more accurate. The syntax of Malayalam and English is very different. Therefore, the laws behind them will involve a certain level of complexity. There have been numerous articles and studies related to rule-based machine translation or transfer rules. According to the literature survey on machine translation rule-based methods, none of these articles have explained the syntax rules of Malayalam accurately and completely. It is in this context that this article discusses the transfer grammar rules for translation from Malayalam to English. We are trying to explain how Malayalam grammar connects with English grammar rules.

Neural networking machine translation systems are working only based on the corpus. We found some unsolved error in Neural machine translation especially Malayalam language

translation. especally the grammar rules are also breaking in NMT. There are some simple Malayalam sentences which are breaking in English translation. Here we are proposing the clear rules for translating Malayalam to English. When the system is able to learn the rules, it can give a good output from these rules.

Methodology

The rule-based translation system relies on an extensive repository of linguistic resources to facilitate accurate translation. These resources include JSON files containing Nouns, verbs, Adjectives, Adverbs, irregular verbs, and postpositions, each comprising a collection of Malayalam-English word pairs. Specifically, the noun file contains 1000 pairs, the verb file encompasses 250 pairs, the adjective file comprises 600 pairs, the adverb file consists of 165 pairs, and the irregular verb and postposition files encompass 134 and 42 pairs, respectively. Additionally, pronouns, demonstrative, conjunctions, and other essential data has been added. Malayalam tense markers - uq:u (present), i, tf:u, p:u (past), um (future) - are used to form sentences in each tense. Fifty sentences are crafted for each tense, utilizing the compiled list of words and incorporating pronouns and adjectives. These sentences are manually translated into English, then identified and recorded Malayalam case markers and their corresponding equivalent grammatical elements in English (Nair, 2012). We created as many sentences as possible, incorporating each case marker. Every sentence includes its English translation. We utilized the constructed sentences with case markers for analysis and development of the transfer grammar.

1. Transfer Grammar Rules for Malayalam to English Translation

Typically, in two-word sentences, the word order remains unchanged in most cases. However, if a word ends with 'pa:l' or 'ja:l', swap the positions of the words.

For instance:

- avana: l atits: u becomes "Hit by him".
- vatija:l atits:u becomes "Hit with Stick".

In the case of two-word sentences where the entity is a common noun (N_NN) or a proper noun (N_NNP), and it is not a named entity (NER), several rules are applied. If the word is the subject, the corresponding English word is fetched from the noun dictionary.

Additionally, if the morphological tag indicates a noun or a verb, such as '<n>' or '<v>', we checked if the word is present in the conjunctions list, containing words like 'a:jirun:u' or 'a:n'

For example:

if the sentence is in the plural form. 'were' is added.

- Example: avar kut:ikala:jirun:u becomes "They were children".

 Otherwise 'was' is added in singular form.
- Example: avan kut:ija:jirun:u becomes "He was a child".

if the sentence contains a generic personal pronoun.

• Example: avar kut:ikala:n becomes "They are children".

if it contains a third-person singular pronoun, 'is' is added.

- Example: *aval karajukaja:* η becomes "She is crying". Otherwise, '*am*' is added for first-person pronouns.
- Example: naːn karajukajaːn becomes "I am crying".

If a conjunction is not present in the representation list, we proceed to another rule, which involves adding 's' to the noun to indicate the singular form of the sentence. However, if the noun is present in the noun dictionary, we simply access the corresponding English word. An exception arises with the noun 'child', whose plural form is 'children'. Therefore, we include a condition to check for this exception.

When the word entity is "PR_PRP" (personal pronoun), we implement multiple rules to handle different situations.

2. Rules for Noun Translation

2.1. Case Markers

(a) Nominative

The Nominative case is the Noun or Pronoun that appears as the subject of the verb in a particular sentence.

Example:

lakşmanan e.uti - Lakshmana wrote.

aval in:ale van:u - She came yesterday.

(b) Accusative

Language in India www.languageinindia.com ISSN 1930-2940 Vol. 24:11 November 2024 Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D.,

'e' is the Marker of Accusative case.

The accusative case of a noun is the grammatical case used to receive the direct object of a transitive verb. In English there is no accusative case marker.

In the presence of the morphological tag '<accusative>', we add 'at' before the verb if the noun ends with 'je' and the verb is 'no:k:uka'.

For example: avan ramaje no:k:i becomes "He looked at Rema".

(c) Sociative

Sociative case is very similar to accusative case in that it marks or turns nouns into objects. In English, these sentences can be translated as "to" or "towards". The suffix in this case is 'o:t' in combination with pronouns or nouns, to mark them as object. Basically, you are directing an action, as such this case is only used with verbs that are interactive.

For example,

ivan o:t,pa:n ivano:t paran:u "I told him".

As you can see, in English we don't need to say "to him" because it's already implied there; however, In Malayalam you have to specify. It is a concept that is for the recipient i.e the one receiving the action.

(d) Locative

If the morphological split contains '<perlative>', indicating a perlative case, we add 'through' to the translation. Similarly, for '<ablative>', we add 'from'.

For example:

- si:ta pujajilu:te po:ji becomes "Seetha went through the river".
- si:ta sku:/il nin: van:u becomes "Seetha came from school".

If the word is a noun (object) and it ends with 'le:k:' and the verb is either 'po:kuka' or 'iran:uka', we add 'to' after the verb.

For example:

• *si:ta katajile:k: po:ji* becomes "Seetha went to the shop".

If the noun ends with 't:il' and the verb is 'kajaruka', we add 'to' after the verb.

For example: si:ta marat:il kajari becomes "Seetha climbed to tree".

Language in India www.languageinindia.com ISSN 1930-2940 Vol. 24:11 November 2024

Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D.,

If the noun ends with 'jil' and the verb is 'irik:uka', we add 'on' after the verb.

For example: ra:man kase:rajil irun:u becomes "Raman sat on the chair".

(e) Dative

If the tag is '<dative>', indicating a dative case, we add the corresponding usage word, such as 'to,' 'for,' or 'have.' A dative case often represents the recipient of an action or the beneficiary of a situation.

(f) Instrumental

1. If the word's morphological split contains the tag '<instrumental>', indicating an instrumental case, we append 'by' to the translation. This case typically involves a morphological split of three words.

For example:

• avana: l atits: u becomes "Hit by him".

Before that, we check whether the morph contains the tag '<instrumental>' to determine whether to add 'by' before the person's name.

For example:

• ra:mana:l kol:ap:et:u becomes "killed by Rama".

(g) Genitive

If the tag is '<genitive>', indicating a genitive case, we add corresponding possessive pronouns to the translation.

For example:

• avalute vi:t becomes "Her house".

Otherwise, we add the corresponding singular pronoun.

For example:

• avan o:ti becomes "He ran".

Table 1: Case Markers Rules

Numb		Malayalam case		
er	Case	marker	example	english marker
			paːmpiṇe, maraṯːe :	
1	Accusative	e, <u>n</u> e, <u>t</u> :e	maram	Nil
2	Sociative	o:t	рилајо:t : рила	to
				in, at, on, to, into,
				towards, along,
		aţ:, il, pip:,		through, Among
3	locative	e:k:/o:t:, u:te	ko:ţ:aja <u>ţ</u> ::ko:ţ:ajam	Between
			ka:li <u>n</u> : ka:l	
4	Dative	к:, <u>п</u>	kuţ:ik: : kuţ:i	To, for, have
5	Instrumental	a:l, kont, vetf:, it:	siːt̪ajaːl : siːt̪a	With, by
6	Genitive	nre, ute	siːt̪ajute : siːt̪a	of, 's (belongs to)

3. Rules for Verb Translation

3.1. Tense Markers

When encountering a verb, we implement rules for different tenses such as simple present, simple past, simple future, present continuous, and past perfect.

For the simple present tense, we first check if the last letter of the verb is present in the inflection dictionary.

If the verb ends with the character 'y', we replace it with 'ies'.

For example:

• avan karajun:u becomes "He cries".

However, before applying this rule, we include another condition:

the second-last character of the verb should not be a vowel. This condition is necessary to avoid incorrect verb forms. For instance, the verb 'enjoy' ends with 'y', so if we were to replace 'y' with 'ies', it would create an incorrect verb form.

For the simple past tense,

We check if the verb is irregular. If it is, we replace the word with its irregular past form, such as 'eat' becoming 'ate'. If the verb is not irregular, we simply add the corresponding English verb to the translation.

Example: avan kasits: u becomes "He ate".

• avan tfiritf:u becomes "He smiled".

Additionally, we include an exception where

if the verb ends with 'e', we add 'd' to the verb to form the past tense.

Example:

aval [vasit]:u become "she breathed".

For the simple future tense, we add 'will' before the verb.

Example:

• avan varum becomes "He will come".

For the present continuous tense, we add 'am' after 'I', and 'are' after other pronouns. If no pronoun is present, 'is' is added.

For example:

- naːn karajukajaːη becomes "I am crying".
- avar karajukaja: n becomes "They are crying".
- avan karajukaja:η becomes "He is crying".

For the past perfect tense, we add 'had' before the verb.

For example: avan vi:tunt becomes "He had a house".

In the case of the word entity being 'V_VAUX' (auxiliary verb), which indicates the presence of the word 'unt',

We add 'had' if the pronoun is 'he/she/it', and 'have' if the pronoun is singular.

For example: avan vi:tunt becomes "He has a house".

In three-word sentences, before proceeding to translation, we reorder the representation list to conform to the English sentence structure of subject + verb + object.

Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D.,

In addition to the rules mentioned for two-word sentences, we introduce some additional rules for three-word sentences. Specifically, in these cases, we explicitly mention particular verbs due to potential issues with morph splitting. Adding 'to' or 'on' to the above-mentioned noun ending terms without specifying the verb may lead to errors in translation.

In the case of adverbs, even though the sentence contains three words, after morphological splitting, it may split into four words. Therefore, we introduce an exception case for four-word conditions.

1. If the sentence contains an adverb and the morphological split includes the tag '<adv>', we fetch the corresponding postposition term from the dictionary.

However, since the tag '<adv>' may not always appear in sentences containing adverbs, we implement an additional rule. We check whether the Malayalam word is present in the adverb dictionary. If it is, we add the corresponding adverb from the dictionary.

For example: avan ve:gat:il o:ti becomes "He ran fastly".

2. If the morphological split contains the tag '<post>' or '<cnj>', we fetch the corresponding postposition from the dictionary and add it after the subject.

For example: vati kont atits:u becomes "Hit with stick".

3. We introduce another exception case for sentences like 'si:tajum ra:manum van:u'. In this case, the morphological split contains five words, so we add 'and' after the subject.

For example: ra:muvum ra:dhajum van:u becomes "Ramu and Radha came".

Table 2: Tense markers changes

		Malayalam Verb	
Numbers	Tense	form	English verb form
1	past	parajitʃ:u/parajip:itʃ:u	had/got/made/has/have
		parajik:unːu/parajip:ik	
2	present	ːu <u>ṇ</u> ːu	get/gets/make/makes
3	future	parajik:um	will/have/will/get
4	past perfect	parajip:itf:it:unta:jirun	had/had got/had made

		:u	
	Present		
5	perfect	paraŋːirikːuŋːu	have/has made
	present	parajip:itf:ukontirik:up	am/ is/are
6	Continuous	ïи	Having
	past		
7	Continuous	parajip:itf:ukonticu <u>n</u> :u	was/were having

Conclusion

This paper presents a comprehensive analysis of Malayalam grammar rules, with a particular focus on noun and verb forms, to facilitate accurate translation into English. Malayalam's unique morphophonemic compounding of noun forms poses significant challenges for direct translation. To address this, we propose a set of word-changing rules that effectively identify and transform Malayalam nouns into their English equivalents. These rules encompass both noun and verb forms, considering the various grammatical markers and syntactic structures present in Malayalam. For verb forms, we examine the tense markers un:u (present), i, tf:u, p:u (past), um (future) are used to indicate present, past, and future tenses. Additionally, we explore the case markers pre, ute, e, a:l, il, k: these corresponding English equivalents, recognizing the contextual variations that may influence their translation. By providing a systematic framework of rules for both noun and verb forms, this research contributes to the development of more accurate and efficient machine translation systems for Malayalam to English. These rules can serve as a valuable foundation for future computational linguistics research and applications.

References

- 1. Aasha V C and Ganesh, A. (2015) 'Machine translation from English to Malayalam using transfer approach', 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2, pp. 1565–1570. doi:10.1109/icacci.2015.7275836.
- 2. Asher, R. E., and T. C. Kumari. *Malayalam*. Routledge, London and New York, 1996.
- 3. Bijimol T.K., John Abraham.: A Rule Based Approach for Translation of Causative Construction of English and Malayalam for the Development of Prototype for Malayalam to English and English To Malayalam Bilingual Machine Translation System. International Journal of Engineering & Technology. December 2018.
- 4. Godase, Amruta, and Sharvari Govilkar. "Machine Translation Development for Indian Languages and Its Approaches." *International Journal on Natural Language Computing*, vol. 4, no. 2, 30 Apr. 2015.
- 5. Jayan, Jisha P., R. R. Rajeev, and S. Rajendran. "Morphological analyser for Malayalama comparison of different approaches." *IJCSIT* 2.2 (2009): 155-160.
- 6. Krishnan, Saranya D., Rajeev R R, Sherly Elizabeth, and Mary Priya Sebastian. "Subject and Object Identification in Malayalam Text." *ICACCI'12*, 2012.
- 7. Lakshmi, S., and Sobha Lalitha Devi. "Rule Based Case Transfer in Tamil-Malayalam Machine Translation." *Research in Computing Science*, vol. 84, no. 1, 31 Dec. 2014, pp. 41–52.
- 8. Nair, Latha R., David Peter, and Renjith P Ravindran. "Design and Development of a Malayalam to English Translator- A Transfer Based Approach." *International Journal of Computational Linguistics (IJCL)*, vol. 3, no. 1, 2012.
- 9. Nair, Latha R., David Peter S., and Renjith P Ravindran. "A System for Syntactic Structure Transfer from Malayalam to English." *Proceedings of the International Conference on Recent Advances and Future Trends in Information Technology (iRAFIT2012)*, International Journal of Computer Applications®, vol. 26, 2012.
- 10. Nair, Latha, and S. David. "Machine Translation Systems for Indian Languages." *International Journal of Computer Applications*, vol. 39, no. 2, 2012, pp. 24-31.

- 11. Nair, Ravi Sankar S. "A Grammar of Malayalam." *Language in India*, vol. 12, no. 11, Nov. 2012, pp. 1-135.
- 12. Pralayankar P., Kavitha V., and Sobha L. Case Transfer pattern from Hindi to Tamil MT. In: PIMT Journal of Research. vol. 2. No. 1, pp. 26-31 March-August (2009).
- 13. RR Rajeev, E Sherly., A suffix Stripping based morph analyser for Malayalam language. Proceedings of 20th Kerala Science Congress. 2007.
- 14. Rajan, Remya, et al. *Rule Based Machine Translation from English to Malayalam*. December 2009.
- 15. Saini, Sandeep, and Vineet Sahula. A Survey of Machine Translation Techniques and Systems for Indian Languages. Feb. 2015.
- 16. Shahana. et.al. Relative Study on Malayalam-English Translation using Transfer Based Approach. International Journal of Computing and Technology, Volume 1, Issue 2, March 2014.
- 17. Sobha Lalitha Devi, R. Vijay Sundar Ram, Pravin Pralayankar and T. Bhagyavathi. Syntactic Structure Transfer in a Tamil to Hindi MT System A Hybrid Approach. In: A. Gelbukh (Ed), Computational Linguistics and Intelligent Text Processing, Springer LNCS.Vol. 6008. pp 438 450, (2010a).
- 18. Sobha Lalitha Devi. et.al. Nominal Transfer from Tamil to Hindi. In: International Conference on Asian Language Processing (IALP), Harbin, China. pp. 270 273 (2010b).
- 19. Sobha Lalitha Devi. et.al. Transfer Grammar in Tamil-Hindi MT System. IALP 79-82 (2013).
- 20. Stephy Joseph. et.al. Transfer Grammar Components for Malayalam to Hindi and English Machine Translation System. International Journal of Engineering Sciences, Special Issue I Dravadian, Volume 15, December 2014.

1. Dr. Selvaraj R., M.A., Ph.D.

Senior Linguist

Department of Language Technology

International Center for Free and Open Source Solutions (ICFOSS), Swatantra,

Language in India <u>www.languageinindia.com</u> ISSN 1930-2940 Vol. 24:11 November 2024 Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D.,

Sports Hub, Karyavattom,

Thiruvananthapuram - 695581

selvarajicfoss@gmail.com

Mob: 91+9562526995

2. Jibin Kiran, M.A., M.Phil., Ph.D.

Research Assistant

Department of Language Technology

International Center for Free and Open Source Solutions (ICFOSS), Swatantra,

Sports Hub, Karyavattom,

Thiruvananthapuram - 695581

jibinkiranjk@gmail.com

Mob:7736351541

3. Rajeev R. R., M.Sc., M.Phil., and Ph.D.

Programme Head

Department of Language Technology

International Center for Free and Open-Source Solutions (ICFOSS), Swatantra,

Sports Hub, Karyavattom,

Thiruvananthapuram - 695581

rajeev@icfoss.in

Mob:9447280110

4. Archana S Krishnan, M.A.

Junior Language Editor

International Center for Free and Open-Source Solutions (ICFOSS), Swatantra,

Sports Hub, Karyavattom, Thiruvananthapuram - 695581

archanaskrishnan00@gmail.com

Mob:7902851634

5. Navaneeth S., M.Tech.

Research Associate

International Center for Free and Open-Source Solutions (ICFOSS), Swatantra,

Language in India www.languageinindia.com ISSN 1930-2940 Vol. 24:11 November 2024

Dr. Selvaraj R., M.A., Ph.D., Jibin Kiran, M.A., M.Phil., Ph.D.,

Rajeev R. R., M.Sc., M.Phil., Ph.D., Archana S Krishnan, M.A., and Navaneeth S., M.Tech.

Transfer Grammar Rules for Malayalam to English Translation

Sports Hub, Karyavattom,
Thiruvananthapuram - 695581

navaneethsreedharan@gmail.com
Mob:9497893166