



**THE EFFECT OF AUXILIARY VERBS IN MALAYALAM ON MACHINE
TRANSLATION**

Aaruni Prasad,

Postgraduate student,

Department of Linguistics,

University of Kerala,

Thiruvananthapuram, India,

aaruniprasad352@gmail.com,

8301958352.

Dr. Darwin L,

Head of the Department,

Department of Linguistics,

University of Kerala,

Thiruvananthapuram, India,

darwin@keralauniversity.ac.in,

9446904488

APA Citation:

Prasad, A., & Darwin, L. (2025). The Effects of Auxiliary Verbs in Malayalam on Machine Translation, *Journal of Indian Languages and Indian literature in English*, 02(04), 176-185; 2025.

Submission Date: 16/02/2025

Acceptance Date: 24/03/2025

Abstract

Auxiliary verbs (AUX) are grammatical elements that modify the tense, aspect, modality, or voice of a verb, playing a crucial role in sentence structure and meaning. In Malayalam, auxiliary

verbs exhibit semantic variability depending on their occurrence and collocation, making their

interpretation highly domain-dependent. This variability presents significant challenges for machine translation (MT) systems, which often struggle to accurately capture the intricate meanings of auxiliaries across different semantic domains. This study highlights how auxiliary verbs in Malayalam, such as *irikunnu* (expressing continuous or progressive form), *untu* (denoting completive form), *a:nə* (denoting progressive form), *no:kk* (indicating an attempt or consideration), and *po:* (suggesting completion or movement), contribute to translation inaccuracies. The paper emphasizes the need to analyze each auxiliary verb individually, mapping their multiple semantic interpretations across different domains. Misinterpretation of these auxiliaries leads to translation errors, affecting the overall accuracy of Malayalam MT systems. A key focus of this study is the auxiliary verb *no:kk*, which exhibits semantic variability based on its semantic domain. For instance, *no:kk* can indicate an attempt (as in "try to do something") or function as a softening particle in certain imperative expressions. MT models frequently fail to differentiate these nuances, leading to incorrect or unnatural translations. By analyzing real translation errors involving *no:kk* using a morphological analysis-based methodology, this paper identifies patterns of misinterpretation and proposes improvements to better handle auxiliary verbs in machine translation.

The study underscores the necessity of enhancing MT models by incorporating a deeper understanding of semantic domains and variability in auxiliary verbs. A linguistic analysis of Malayalam auxiliaries and their multiple semantic roles can help improve translation accuracy.

Keywords: Auxiliary Verbs, Machine Translation, Semantic Variability, Semantic domains, Collocations, Co-occurrences.

1. Introduction

Language is an elaborate and complicated system of communication made up of rules and grammatical structures. One of these is the auxiliary verb (AUX) it is widely used to change the tense, aspect, modality, or voice of a verb. These also appear in tandem with main verbs and include aspects in

meaning without which a correct interpretation is impossible. Auxiliary verbs have wide distribution across the world's languages, but the precise function of auxiliary verbs and the degree of semantic variability they express can vary widely from another.

Malayalam, a Dravidian language spoken primarily in the state of Kerala, India, offers a classic example of such semantic variability, whereby an auxiliary verb may exhibit allomorphic behaviour, depending upon its occurrence or collocation. The auxiliary verb *irikkunnu* (It is a continuous or progressive form of this auxiliary) *untə* also indicates this is in a completive form of this auxiliary verb) *a:nə* also indicates this is in a progressive form of this auxiliary verb), *no:kk* (indicating to attempt or to consider), *po:* (indicates this is in a completing form of this auxiliary verb) etc. In other domains, the auxiliary verbs are main verbs, while in others, the main verbs are auxiliaries. This bimodality leads to translation inconsistencies during translation through a machine translation system. Machine Translation (MT) techniques are used to automatically translate text from one language into other language using computational models, that is from Source language to Target language. While machine translation (MT) systems have improved tremendously over the years, translating Malayalam is still a challenging task due to its rich morphology, complex syntactic structures, and semantically rich vocabulary. MT potentially suffers from semantic instability across the various semantic domains corresponding to the semantics of auxiliary verbs, leading MT models to fail to capture meanings properly and, hence, generate incorrect or unnatural translations. Specifically, it illustrates how the complications in translating Malayalam auxiliary verbs are symptomatic of the larger hurdle of making robust MT systems for a low-resource language like Malayalam, ultimately part of the Dravidian family of languages. Malayalam Machine Translation (MT) is an evolving field as ample parallel corpora are now accessible along with high resource machines to help predict translations, where many translation errors usually boil down to the fact that auxiliary verbs cannot be correctly interpreted.

1.1. Literature review

Nair (1990) conducted a comprehensive study on auxiliary verbs in Malayalam, examining their structural and semantic properties. The study highlighted the functional versatility of auxiliaries in Malayalam, noting their role in expressing various tenses, aspects, and modalities. However, the work did not focus on how these auxiliaries are interpreted by computational models, leaving a gap in understanding their impact on machine translation.

Kiranmayee et al. (2023) examine the challenges in developing MT systems for Dravidian languages,

including Malayalam. They emphasize that lexical divergence, ambiguity, and syntactic and semantic issues contribute to translation inaccuracies.

The study underscores the need for enhancing MT systems to address these challenges effectively.

Jayan and Bhadran (2015) discuss the difficulties in processing Malayalam verbs for Statistical Machine Translation (SMT). They highlight that the agglutinative nature of Malayalam leads to a vast number of verb forms, posing challenges in alignment and translation within SMT frameworks.

1.2. Research questions

This study seeks to address the following research question:

How do auxiliary verbs in Malayalam, particularly those exhibiting semantic variability, affect the accuracy of machine translation systems, and what improvements can be proposed to enhance translation quality?

2. Role of Malayalam Auxiliary verbs in MT

Auxiliary verbs in Malayalam (saha:yakakRiya) plays as crucial elements in expressing tense, aspect, mood, and voice within sentences. Their semantic variability, influenced by context and collocation, brings significant challenges for machine translation (MT) systems. This chapter examines the complexities of Malayalam auxiliary verbs and examines their implications for MT accuracy. In Malayalam, auxiliary verbs function to modify the main verb, providing additional information about the action or state described. They can indicate or show continuous actions, completive aspects, progressive forms, attempts, or movements. Like, the auxiliary verb "irikunnu" expresses a continuous or progressive aspect, while "po:" suggests completion or movement. Most importantly, these auxiliary verbs are not confined to auxiliary functions alone; they can also act as main verbs, depending on the context, even more complicating their interpretation. The dual role of auxiliary verbs in Malayalam serving as both auxiliary and main verbs introduces semantic variability that MT systems often struggle to handle. This variability is highly domain-dependent, meaning that the interpretation of an auxiliary verb can change based on the specific context or field in which it is used. As a result, MT systems may misinterpret these verbs, leading to translation inaccuracies.

According to Sree Veeramanikantan Pillai (1974), Malayalam has a significant number of vocabulary verbs, with some of the most commonly used ones being *kol*, *po:k*, *no:kk*, etc. In a well-researched article, Prabhodhachandran Nair (1979) provides a comprehensive analysis of Malayalam verbs. He examines the distinction between auxiliary verbs and the semantics of acceptual verbs. According to him,

Malayalam has fewer than twenty auxiliary verbs, which convey various semantic notions such as completion, continuance, probability, manner, benefit, and attempts toward an action.

Additionally, he briefly discusses moods in Malayalam, illustrating different types, including permissive, imperative, compulsive, optative, prohibitive, and perfective moods.

The following are some of the examples in which auxiliary verbs are present in the sentence:

1. avan kaḷiccə koṇṭirikkuṇṇu

“He is playing” (progressive aspect)

2. avaḷ pa:dino:kki

“She tried to sing” (notion of trying)

3. avan eḷutikka:niccu

“He demonstrated how to write” (notion of demonstration)

4. kaṇṇa:ṭi uṭaṇṇupo:yi

“The mirror was broken off accidentally” (accidental aspect)

5. avaḷ paraṇṇə kaḷayum

“She will say intentionally” (intentional aspect)

The above mentioned are some of the commonly used auxiliary and the following are their machine translation outputs:

6. avan kaḷiccə koṇṭirikkuṇṇu

“He is playing”

7. avaḷ pa:dino:kki

“She sang and looked”

8. avan eḷutikka:niccu

“He wrote and showed”

9. kaṇṇa:ṭi uṭaṇṇupo:yi

“The mirror is broken”

10. *avaḷ paraṇṇa kaḷayum*

“She will say”

The examples provided showcase the difficulties machine translation (MT) systems encounter when trying to accurately interpret Malayalam auxiliary verbs, often resulting in translation errors.

- Original Sentence: *avaḷ pa:dino:kki*
- Intended Meaning: “She tried to sing” (implying an attempt to do something)
- MT Output: “She sang and looked.”

Here, the auxiliary verb *no:kki* is meant to express an attempt or effort. But the MT system lacks to find its intended meaning and provided the verbal meaning of the words, “sang” and “looked,” which completely misses the intended meaning of trying to sing.

- Original Sentence: *avan eḷutikka:niccu*
- Intended Meaning: “He demonstrated how to write” (indicating a demonstration or explanation)
- MT Output: “He wrote and showed.”

The auxiliary verb *ka:niccu* is supposed to convey the act of demonstration, suggesting that the person is showing or teaching how to write. However, the MT system renders it as two distinct actions, “wrote” and “showed,” failing to capture the intended instructional meaning.

This is because, the MT model uses word-by-word or morpheme-by-morpheme translation, which aligns with the traditional approach to morpheme identification as outlined by Eugene Nida,

Principle 6 (pp 58-59):

“A morpheme is isolatable if it occurs under the following conditions:

1. In isolation.
2. In multiple combinations in at least one of which the unit with which it is combined occurs in isolation or in other combinations.
3. In a single combination provided the element with which it is combined occurs in isolation or in other combinations with no unique constituents.”

When following a morpheme-by-morpheme translation approach, machine translation (MT) models struggle to interpret the meaning of auxiliary verbs in specific contexts. Instead of analyzing their

functional role, the model focuses solely on the semantic meaning of the morpheme, which typically corresponds to the verbal form of the auxiliary. This happens because auxiliary verbs in Malayalam can function both as main verbs and auxiliaries. As a result, when relying on morpheme-by-morpheme translation, the MT model tends to capture only the meaning of the main verb, overlooking the auxiliary's intended function.

The case of Auxiliary – *no:kk* ;

In Malayalam, *no:kk* is a commonly used auxiliary verb. While its main verb form, *no:kk*, means "to look," it functions as an auxiliary in various semantic domains, serving multiple roles depending on the context. The same auxiliary verb can convey meanings such as an attempt, a try, or a look. It can also indicate an inceptive aspect or function as a request, imperative, and more, depending on the context.

11. avan enne no:kki
12. avan o:ṭa:n no:kki
13. avaḷ pa:di no:kki
14. ra:man marattil kayaRi no:kki
15. ceytu no:kku

Even though examples (11) to (15) contain the morpheme *no:kk*, its meaning varies across the utterances. In the 11th example, it functions as the main verb, meaning "to look." However, in the other four examples, the morpheme appears as auxiliary, it conveys meanings such as "try," "attempt," or an imperative statement. Examining how *no:kk* acquires these multiple meanings reveals how this variation can lead to translation errors.

MT outputs of (11) to (15):

- Input: avan enne no:kki
- MT Output: "He looked at me"
- Intended Meaning: "He looked at me"
- Input: avan o:ṭa:n no:kki
- MT Output: "He looked at running"
- Intended Meaning: "he attempted to run"
- Input: avaḷ pa:di no:kki
- MT Output: "She looked at singing"
- Intended Meaning: "she attempted to sing"
- Input: ra:man marattil kayaRi no:kki

- MT Output: “Rama climbed the tree and looked”
- Intended Meaning: “Rama tried to climb the tree”

- Input: ceytu no:kku
- MT Output: “do it”
- Intended Meaning: “do it”

The output reveals that, in most cases, the machine translation system interprets only the base meaning of the morpheme. It fails to consider the various semantic interpretations associated with different semantic domains of the auxiliary verb. As a result, the translation produced is often inaccurate and error. This issue is also evident in the case of complex sentences. The machine translation system continues to overlook the diverse semantic interpretations of auxiliary verbs, resulting in incorrect translations.

16. Input: “manu itatuvaṣattekkə no:kkikkoṇṭirikukaya:yirunnu , appo:ḷ si:ta avano:ṭə ca:ṭi no:kku ennu paraṇu avan ca:ṭi no:kki”

MT Output: “Manu was looking to the left side when Sita told him to jump and look and he jumped and looked”

Intended Meaning: “Manu was looking to the left when Sita asked him to try jumping, and he did”

This example also demonstrates the problem. The Machine Translation System incorrectly translates the auxiliary verb “no:kk” as “look” even in a complex sentence. The intended meaning, where “no:kk” conveys the sense of “trying” (i.e., “try jumping”), is completely missed by the system. Instead, the output presents a literal and inaccurate interpretation.

3. Conclusions

Machine translation software uses morpheme-by-morpheme translation primarily as its main approach whereas semantic word evaluation through contextual analysis is excluded. The translation method

generates proper results in specific cases yet it makes critical translation mistakes when processing auxiliary verbs in Malayalam. The semantic understanding of auxiliaries constitutes the sole analysis range of machine translation systems which applies to both main verb scenarios and their listed semantic interpretations. Each auxiliary verb in Malayalam functions differently according to semantic areas and specific circumstances due to its complex morphological structure. Complexities regarding

auxiliary verb interpretation exist for machine translation systems because their meanings require full understanding of linguistic context and surrounding words and sentence construction. A sufficient number of contexts must be studied for each auxiliary verb before incorporating its meanings into translation models properly. The incorrect operation of machine translation systems occurs during processing of dialogues written in conversation style and negative construction types including both complex and simple sentences. The evaluation approach for Malayalam auxiliary verbs needs development to shift from form-based methodologies toward better identification of various lexical meanings contained within these words.

An expanded translation model which performs contextual analysis and collocational analysis stands essential for accomplishing better machine translation of Malayalam auxiliary verbs. The analysis of semantic variations demands testing auxiliary verbs across various co-occurring word combinations and associating groups. A wide range of Malayalam texts that include various sentence patterns and conversational exchanges and complex linguistic structures should be gathered to build a complete database of auxiliary verbs. When MT systems properly document semantic domain variations they become more able to understand meanings which extend beyond standard meanings of auxiliary verbs. A framework which detects auxiliary verbs based on context should be developed by implementing semantic role labeling (SRL) together with contextual embeddings or multi-task learning techniques to boost translation accuracy. Researchers must create algorithms which identify complementary verb functions through context detection before integrating the information into MT systems. The translation quality improves when researchers move from inconsistent use of morpheme rules toward alternative methods that account for the semantic versatility of Malayalam auxiliary verbs.

4. Reference

Asher, R. E., & Kumari, T. C. (1997). *Malayalam*. Routledge.

Ejabena, H. (n.d.). *A critique of Nida's (1949) The identification of morphemes*.

Jayan, V., & Bhadrar, V. K. (2015). Difficulties in processing Malayalam verbs for statistical machine translation.

International Journal of Artificial Intelligence & Applications, 6(3), 13– 24.
<https://doi.org/10.5121/ijaia.2015.6302>

Kiranmayee, B. V., Sai Priya, R., Vijaya, R., Suresh, P., & Goutham, R. V. (2023). Machine translation on Dravidian languages. *International Journal of Recent Technology and Engineering*, 12(1).

<https://www.ijrte.org/wp-content/uploads/papers/v12i1/A75380512123.pdf>

Nair, N. R. (1990). *Auxiliary verbs in Malayalam*. Annamalai University.

Nair, V. R. P. (1972). *Malayalam verbal forms*. Dravidian Linguistics Association.

Nida, E. A. (1976). *Morphology: The descriptive analysis of words*. University of Michigan Press.

UpToWord. (n.d.). *Auxiliary verb meaning in Malayalam - മലയാളം അർത്ഥം*. Retrieved March 15, 2025, from <https://www.uptoword.com/en/auxiliary-verb-meaning-in-malayalam>