==========
Language in India www.languageinindia.com **ISSN 1930-2940 Vol. 18:11 November 2018**
India's Higher Education Authority UGC Approved List of Journals Serial Number 49042
==========

Computational Analysis of Verbs in Malayalam

Abrar K.J., Ph.D. Research Scholar

==========

Abstract

Verb analysis is an important task in Computational Morphology. Its status in Malayalam is not appreciable when we compare it with the other Indian languages like Hindi, Tamil, Telugu, etc. In this paper, I would propose a model for Malayalam verb Morphological Analyser. This model is proposed by making use of the analysis of Malayalam verb morphology attempted by traditional grammarians and computational linguists. Grammarians have concentrated mainly on describing the nature, structure and categorisation of verbs. Computational linguist's approaches it in a way to suit it for machine learning environment. Both the approaches were taken into account for the framework proposed in the present study. This study is intended to streamline the limitations of verb analysis practiced by the Malayalam computational linguists. It suggests the possibility of a new model of verb analyser for computational grammar of Malayalam verbs. The scope of this study is to avoid all existing grammatically linked computational issues of verbs in Malayalam language for achieving the highest performance accuracy in Machine Translation (MT) and allied areas.

Keywords: Morphological analyser, Machine Translation, Malayalam verb structure analysis, Computational grammar.

1, Introduction

Morphological Segmentation is an important preliminary task in computer assist text analysis or Natural Language Processing (NLP). In NLP, the morphological processing is to recognize and split each morpheme of word or sentence as free or bound morphemes based on their semantic status. Morphological analyser in NLP has dual steps to follow. First, to identify each morpheme of word and tagging or labelling with its grammatical functions is the second step.

Word = stem/root or stem/root + suffix/suffixes
Birds = bird (N) + s (PL)

This process is depending upon the morphological and morpho-phonological behaviour of a language. Morphological analyser is using varies NLP tools like parser, stemmer, POS tagging etc. and has high role in MT, Automatic Text Summarization, Spell checker, Grammar checker, Information retrieval, Sentimental analysis etc.

The NLP and its Applications status of Malayalam Language is comparatively underdeveloped against other Indian languages despite of having number of on-going projects. In Malayalam there is wider possibility of suffixing using diverse morphemes after stem/root due to its agglutinative nature.

Identifying morpheme in Malayalam is a challenging task before machine. Verb morphology is more complex, compared to other Parts of Speech categories like noun and adjectives. Analysing verbs stem and recognizing inflectional pattern are the basic processing in Malayalam morphological analyser. But its accuracy rate is not appreciable due to the limitations in the adopted approaches in handling Malayalam verbs. In this paper I would suggest a different method of morphological segmentation and tagging of Malayalam verbs with better accuracy in the processing of inflectional and derivational patterns of Malayalam verb morphology

2. Malayalam Morphological Analyser

In the field of NLP several methods are being used for morphological recognition. Malayalam morphological analyser so far used three methods- Suffix stripping method (Rajeev, Sherly 2008), Paradigm approach (Saranya 2008), and Hybrid approach (Vinod, Jayan, Bhadran. 2012). In suffix stripping method, system recognizes suffixes attached with stem/root and segments each suffix. The programme is designed in a way to identify the morphemes form left to right. The part after removing suffixes is considered as the stem/root of the word. Working of this system is based on a suffix dictionary and morpho-phonemic rules. Paradigm based approach is another method for morphological recognition, 'Paradigm is the complete set of related word-forms associated with a given lexeme' (Jisha, Rajeev, Rajendran. 2009:156). Each grammatical category contains different paradigm depending on the phonotactic structure of a word like word ending (vowel ending, consonant ending, and Chillu[1] ending). pālam 'Bridge', makan 'Son' both are Noun words but with different paradigm. Hybrid approach is the combination of suffix stripping and paradigm-based approach. It needs both paradigm and suffix dictionaries. This type of analysis claims 83.67 % of accuracy (Vinod, Jayan, Bhadran. 2012:316). The working accuracy is counted based on the paradigm inflection list. Hundred percentage of accuracy can be achieved only when the complete inflection list of each paradigm categorises is made available. Malayalam takes free inflection order (discuss 3) and many affixations are possible in a word root/stem. Other methods like Finite State Automata (FSA), Finite State Transducer (FST) Memory Based approach, Corpus based approach etc. are proposed by different computer programmers. Using FST method Santosh Thottingal has prepared a Morphological analyser[2]. In this analyser there is a no segmentation of root/stem and each affix. In this, infinite form of verbs (conjugation of uka) are considered as a base form of the verb.

3. Verb Morphology of Malayalam

In the field of morphology stem/root structure, inflectional and derivational forms are the basic considerations of a word formation system. In Malayalam verbs, most of the verb stems are expressed as Intransitive, Transitive, Causative, Negative forms and each form contain TAM (Tense, Aspect, and Modality) inflection. This is the frame of basic inflectional morphology of a Malayalam verb. In nominalising (verb change into noun) and Adjectivising (Verb change into adjectives) this acts as the basic form of derivational morphology of verb. This study follows the above morphological pattern of Malayalam verbs for computer based analysis.

[1]Pure consonants or vowel less sound.
[2] httm/smc.morph.in

3.1. Root/Stem Structure

Most of the Malayalam verbs root/stems are of Dravidian and Sanskrit origin. Some of verbs are foreign origin borrowed from languages like Arabic, Persian, and Syrian etc. Adding *-i* in the ending of the verb root, it become a Malayalam base stem. Most of verb stem ends in vowel, either short or long, and comparatively less verbs are consonants ending, and with *chillu* ending. General Phonotatic pattern of Malayalam VCV, V(C)V (short vowel) VCVV, CVV (long vowel), VCVC, CVCVC (Consonants ending) VVC, VCVC, CVCVC (Chillu ending). Trill [ɾ], Alveolar Lateral [l], Retroflex Lateral [ɭ], Retroflex Approximant [ɻ] are verb root ending Chillus in Malayalam. verb stem classified into *kaaritha* and *akaaritha* based on the link morpheme *–kk* before adding present and future tense marker, *-kk* **infixed** verbs are *kaaritha* verbs and other verbs are *akaaritha*. There is no **infixation** in past tense form of *kaaritha* verbs. Only vowel and chillu ending possess this feature. A.R. Rajaraja Varma (2011:304-312) has listed verbs on the basis of root ending. And he also classifies kaaritha and Akaaritha verbs.

Kaaritha verb	
kalikkunnu	kalicu
kali-kk-unnu	Kali-cu
Play-kk (infix)-PRES	Play-PAST
‘Playing’	‘Played’
Akaaritha verb	
Parajunnu	Paraɲɲu
Para-(y) unnu	Para-ɲɲu
Say-PRES	Say-PAST
‘Says’	‘said’

Other classification is Transitive (*sakarmaka*) and Intransitive (*akarmaka*). This is universal classification of verbs. Verbs contain Object in sentence that verbs are Transitive verbs. Without Object are Intransitive verbs.

Niraccu (Transitive)	Niraɲɲu (Intransitive)
‘to fill’ (TRAN-PRESE)	‘to fill (INTR-PAST)

Transitive-Intransitive classification is depending on the semantic feature of a verb and Kaaritha-Akaaritha classification is depending on the structural features of a verb.

3.2 Inflection

A verb expresses three types of inflections, i.e. Tense, Aspect, and Mood. It has inflections both in predication and in its negation. Tense appears as Present, Past and Future forms and -unnu, -u and -i, -um are the tense markers respectively. Past tense morphology is very complex in Malayalam. -u marker has different allomorphs depending on the phoneme in the verb ending. Sooranad Kunnjan Pillai has indexed Verbs in the work ‘Malayalam Lexicon’ (2000:1-105). He has classified 12 classes of -u Past tense markers as given in the following table.

Table 1

Present Tense	Past Tense	Future Tense
-unnu parajunnu ‘talk’ pejjunnu ‘rain ’ ʋiɭikkunnu ‘call’ pularunnu ‘raising ’	i- muŋi ‘drowned’ āṭṭi ‘’ minni ‘glitter’ nakki ‘lick’ u- tu- ceytu ‘done’ cu- ʋiɭicu ‘called’ nnu- parannu ‘fly’ nʈu- kanʈu ‘saw’ ntu- nontu ‘pained’ ɲɲu- karaɲɲu ‘cry’ ʈʈu- keʈʈu ‘listen’ ttu- cattu ‘died’ rru- perru ‘delivered’	um- parajum ‘will talk’ pejjum ‘will ’ ʋiɭikkum ‘will call’ pularum ‘will raise’

Table 1 shows the different pattern of three tenses in Malayalam. The change of phonological behaviour of past tense depending on the Morpho-phonemic rules of the Language.

Modality is another inflection of verbs. Mood is defined on the basis of how it expresses or present in a verb. Tense forms are also a mood. Traditional grammarians considered three or four moods in language as Imperative, Optative etc. The morpho-syntactic description of Malayalam done by Asher (2012:304-314) and Ravi Shankar (2012:58-66) have list out ten plus modality forms of Malayalam. Here I presents Ravi Shankar’s (Ravi Shankar: 2012:58-66) description of Moods.

Table 2

Mood	Example
Imperative	nirakkū, nirakkaɳam
Promissive	niracekkam, niracōɭam
Permissive	nirakkam, nirakkaʋunnatāɳ
Optative	nirakkaʈʈe
Precative	nirakkaɳē
Negative Precative	Nirakkarut
Desidarative	nirakkaɳamajirunnu
Abilitative	nirakkaʋunnatejuḷḷu
Irrealis	Niracene
Dubitative	nirakkunnunʈo
Purposive	nirakkan pokunnu
Conditional	Niracāl
Satisfactive	Niracallo

Monitory	Nirakkume

This list contains some interrogational forms (Dubitative), Infinitive forms (Purposive) and Negative forms (Negative Precative can be removed from the list as they are considering interrogative, negative and infinitive forms). Remaining forms in daily communication can be treated as moods.

In Malayalam, aspectual system is also expressed using inflection with verb. Traditional grammarians discussed this type of alteration in Malayalam. But they have not used the term aspect or any equivalents in their grammar texts. Someone described it as part of the auxiliary verb and others considered it as the expansion of tense forms. Following are the main aspect forms of Malayalam.

Table 3

Aspect	Example
Progressives 1, Present tens + unʈə 2, Infinitive + āɳə	 nirajunnunʈə nirajukajāɳə
Iterative 1, konʈə + irikkə + Tense markers 2, konʈə + ē + irikkə + Tense markers	 niraɲɲukonʈirikkunnu niraɲɲukonʈējirikkunnu
Perfect 1, Simple- Past + irunnu 2, Contemporaneous- Past + irikkunnu 3, Remote- Past + iṯṯə + unʈə	 niraɲɲirunnu niraɲɲirikkunnu niraɲɲiṯṯunʈə
Habitual 1, ārə + unʈə + Tense markers	 nirajarunnʈə, nirajarunʈājirunnu

Above is the linguistic structure of the aspectual system of Malayalam. All forms of Tense, Aspect, and Mood can be inflected with negative markers -illa, -alla, -arutə
nirajunnu ‘is filling’- nirajunnilla ‘is not filling’, nirakkaɳam ‘must fill’ nirakkarutə ‘don’t fill’
Only transitive verbs take the causative forms. i, ppə, ccu is the main causative markers. Some verb takes more than one marker.
āṯṯi- ‘’ āṯṯicu- āṯṯippiccu
cejjiccu- cejjippicu-

The above description gives only primary inflections of a Malayalam Verb. Each inflection takes many more combinations with other conjugations like be verbs ākə, unʈə. Some example for this type of conjugations are listed out below.

Table 4

nirakkunnu ‘Fill_PRES’	Niracu ‘Fill_PAST’
nirakkunnunʈə	niraciṯṯunʈə

‘Fill_PROG’	‘Fill_PERF-REMO’
nirakkunnunʈājirunnu ‘Fill_PROG_be PAST’	niraciʈʈunʈājirunnu ‘Fill_PERF-REMO _be PAST’
nirakkunnunʈājirunneŋil ‘Fill_PROG_be PAST_CONDI’	niraciʈʈunʈājirunneŋil ‘Fill_PERF-REMO _be PAST_CONDI’
nirakkunnunʈājirunneŋilo ‘Fill_PROG_be PAST_CONDI_IP’	niraciʈʈunʈājirunneŋilo ‘Fill_PERF-REMO _be PAST_CONDI_IP’

Here we can see the structure of Present and Past tense inflection. The past tense take Perfect-Remote aspect, and Present tense take Progressive aspect form in the same pattern with difference in meaning. Such kinds of pattern recognition would streamline the rules of inflection with respect to the type of aspects or mood forms appear with tense. These generalised rules help the computer based morphological analysis easy. Following is some examples for inflection in causative.

Table 5

nirappikkunnu ‘Fill_CAUS_PRES’	Nirappicu ‘Fill_CAUS_PAST’
nirappikkunnunʈə ‘Fill_CAUS_PROG’	nirappiciʈʈunʈə ‘Fill_CAUS_PERF-REMO’
nirappikkunnunʈājirunnu ‘Fill_CAUS_PROG_be PAST’	nirappiciʈʈunʈājirunnu ‘Fill_CAUS_PERF-REMO _be PAST’
nirakkunnunʈājirunneŋil ‘Fill_CAUS_PROG_be PAST_CONDI’	nirappiciʈʈunʈājirunneŋil ‘Fill_CAUS_PERF-REMO _be PAST_CONDI’

3.3. Derivation

Derivations are adding affixes into root/stem that makes change in the basic grammatical categories i.e., Verbs change to Noun and Adjectives. These affixes can be termed as derivational affixes. In Malayalam, grammatical condition of the verb stem can be changed in to a noun or adjectives by adding some suffixes. Most of the verb becomes verbal noun by suffixing nominal marker like -al (nirakkal Fill_NOML), -atə (niracatə Fill_NOML) and gender markers -an (niracavan -masculine), -aḷ (niracavaḷ -feminine), and-ar (niracavar -neuter). Adding –a, -āya and -uḷḷa the verb becomes Adjectives. ‘Most of Malayalam Adjectivized forms are relative participle’ (Ravi Shankar: 2012:82). –āya conjugations is most probably seen after nominalised verb with -atə as a Relative Participle form; niraɲɲatāya‘Fill_PAST_NOML_ADJV’. Adding –āyi after nominalised form, it become Past Participle form; niraɲɲatāyi ‘Fill_PAST_NOML_ADJV’. And -uḷḷa form combine after Past tense forms, Past Participle -iʈʈə ending forms and Iterative aspect koɳʈə ending forms as a Relative participle; niraculḷa ‘Fill_PAST_ADJV’, niraciʈʈuḷḷa ‘Fill_PP_ADJV’, niracukoɳʈuḷḷa. ‘Fill_ITR_ADJV’. Following are some examples of lengthy derivational (both Nominal and adjective forms) string in Malayalam.

Table 6

Nominal	**Adjective**
nirakkunnatə 'Fill_PRES_NOML'	niracuḷḷa 'Fill_PAST_ADJV'
nirakkunnatukoɳʈə 'Fill_PRES _NOML_INTS'	niracukoɳʈuḷḷa 'Fill_ITR_ADJV'
nirakkunnatukoɳʈāɳə 'Fill_PRES_NOML_INTS_be PRES'	niracukoɳʈiruniʈʈuḷḷa 'Fill_ITR_PP_ADJV'
nirakkunnatukoɳʈāɳeɳkil 'Fill_PRES_NOML_INTS_bePRES_CONDI'	niracukoɳʈirikkunnatājuḷḷa 'Fill_ITR_NOML_ADJV'

Above discussed morphological structures are only basic inflectional and derivational patterns of Malayalam. A Morphological analyser is expected to be addressed all possible inflections and derivations of a language.

4. Model

In this session I am proposing a model of Malayalam morphological analyser useful both in segmentation and tagging. We also see the inflectional structure and conjugative form of Malayalam verbs. Each conjugative form is morphologically and semantically relevant. So each form is to be listed out separately in root/stem of a word. For grammatical tagging the glossing method are found useful. It will cover all conjugative form by adding root/stem of a word (see Table 4 and Table 5). Some grammatical forms may have one or more grammatical functions. Example; pōji 'go', vannu 'came', tannu 'gave' this type of verbs are past tense form in some context like after past participle form this verb forms are express Auxiliary function. parɲɲupōji, cejtupōji, parɲɲutannu, cejtuvannu etc. in Malayalam morphological situation all grammatical functional forms are must be tagged. Figure 1. Show the model of Malayalam morphological analyser interface.

Figure 1.

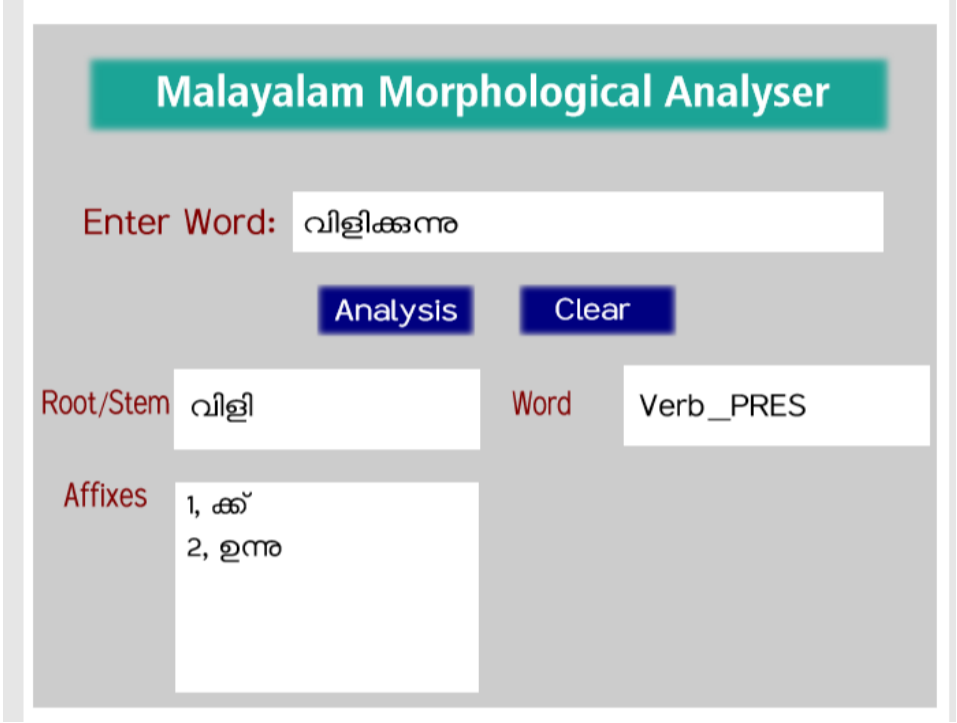

Figure 1. Analysed present tense form of *kaarita* (see 3.1) verb segmentation. Here first Segmented in root/stem of a word, its affixation listed one by one. And also showing the grammatical categorises of a given word using glossing method. This type of analysis is useful for computational purpose as it would help in developing a verb analyser.

5. Discussion and Future Work

Computational linguistics or NLP work is an inter-disciplinary field. For achieving hundred percentage of accuracy in computer-based products there must be collaborations of Computer scientists and linguists. The consortium type of working pattern is useful method in this field. In the Malayalam NLP, there is no attempt for collaboration. Most of the NLP tools making attempts are part of M.Tech. Project works. The available attempts of morphological analyser have not succeeded to achieve its objective. From the review of earlier works, this study suggests that the affixes segmentation and grammatical glossing are useful method for POS tagging and semantics analysis of a word.

Abbreviations;

ADJV	: Adjective	PAST	: Past Tense
C	: Consonants	PERF-REMO	: Perfect Remote aspect
CAUS	: Causative	PP	: Past Participle
COND	: Conditional	PRES	: Present tense
ITR	: Iterative aspect	PROG	: Progressive aspect
NOML	: Nominal	V	: Vowel

==========

References

Asher, R.E., Kumari, T.C. 2012 (1997). *Malayalam*. Oxen: Routledge.

Jayan, V, Bhadran,V.K. 2015. *Difficulties in Processing Malayalam verbs for Statistical Machine Translation*. Trivandrum; CDAC.

Jisha, Jayan P. Rajeev, R.R., Rajendran, N. 2009. *Morphological Analyser for Malayalam- A Comparison of different Approaches*. Vol.2, No.2. IJCSIT.

Prabhodhachandran Nair, V.R. 1970. *Malayalam Descriptive Grammar*. Trivandrum: DLA.

Rajaraja Varam, A.R. 2011. *Keralapāṇinīyam*. Kottayam: DC books.

Rajeev, R.R., Rajendran N., Elizabeth, Sherly. 2008. *A suffix striping morph analyser for Malayalam language*. Proceedings of 20th Kerala Science Conference.

Raveendra, Kumar, Sulochana, K.G., Jayan, V. 2011. *Computational Aspect of Verb Classification in Malayalam*. ICISIL.

Ravi Shankar, Nair. 2012. *A Grammar of Malayalam*. *Language in India* www.languageinindia.com.

Ravi Shankar, Nair. 2012. *Vākyadharshanam*. Trivandrum: Kerala Language Institute.

Saranya, S.K. 2008. *Morphological Analyser for Malayalam verb*. M.tec. Thesis. Coimbatore: Amrutha University.

Sunil, R. Nimtha, Manohar. Jayan, V. Sulochana, K.G. 2012. *Morphological Analysis and Synthesis of Verbs in Malayalam*. ICTAM.

Shuranad Kunjan pillai (Ed.). 2000. *Malayalam Lexicon, Index1*. Trivandrum: University of Kerala.

Vinod, P.M., Jayan, V., Bhadran, V.K. 2010. *Implementation of Malayalam Morphological Analyser Based on Hybrid Approach*. ROCLING.

==========

ABRAR K.J.
Research Scholar, Dpt. Of Linguistics
Thunchath Ezhuthachan Malayalam University
Tirur, Malappuram,
Vakkad (PO) 676502
Kerala
abrarkj@gmail.com