International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Malayalam Verb Analyzer: An Alternate Computational Model

Abrar K.J[1], Sreenathan M[2]

Department of Linguistics, Malayalam University, Tirur Kerala, India

Abstract: ***The Morphological Analyzer of any language is under increasing scrutiny based on its functional accuracy. In the context of functional inadequacies of Malayalam Morphological analyzers identified in currency, this study proposes a new model incorporating both linguistic and computational aspects for computational verb analysis. This maiden endeavor is a departure from the morphological approaches like suffix striping, paradigm appraoch, Directed Acyclic word diagram; Corpus based approach Finite state Transducer, Two level morphology and Finite State Automata. An alternate model for verb analysis is presented with the expectation of resolving the existing predicaments in computational verb analysis. Interlinking linguistic and computational aspects in establishing a verb analyzer model would be a welcome step to follow in agglutinative language based morphological models.***

1. Introduction

Malayalam is one of the scheduled and classical languages of India. Malayalam, a major literary language with long traditions of literature and scripts, is the main language of Kerala and that of the Lakshadweep .It enjoys the shared official language status along with the English in the homeland Kerala and Lakshadweep. The national status of Malayalam language is tenth in position. Malayalam speakers are mostly inhabited in Kerala. Kerala represents only 1.18 % of nation's physical space but the language covers 2.8 % of total population of the nation. The language technology status of Malayalam is not reached to an appreciable level due to lack of innovative research, policy making and technology development. Most of the efforts followed localization of tools and not paid much attention to develop Malayalam computing.

Despite of having various Academic attempts like M.phil, M.Tec Dissertations and some published papers unlike the other Indian languages Malayalam doesn't have a success story of Morphological Analyzers to claim. IIIT Hyderabad along with the Linguistic Dept. of Tamil University had tried to develop a Morphological Analyzer for Malayalam based on paradigm method using tool Anusaraka(Raajeev R.R.2015). However, it is not available in public.

2. Related Work

Arun Lal (2011) Rani Mole (2011), Reena Ravindran (2013), Rekha M.S. (2013), Anju A.M (2015) and Anitha K. (2016) made some efforts in Processing of Computational Morphology of Malayalam. Rajeev R.R (2006), Shoj Raj (2006), and Sreeja T.D (2011) have tried to contribute in Morphological Generation of Malayalam Language. Besides the above Mphil dissertations submitted in University of Kerala, Awasthy P.V (2007). And Saranya (2008) Submitted M.Tec Dissertations in Amrtha University. Nisha M (2015), Raji Rahmath (2015) Submitted M.tec Dissertations based MBLP Approach in Govt. Engineering College, Palakkad, Kerala. The prominent approach evident in these dissertations is of analyzing Malayalam language through Computer science perspective. Instead of giving any emphasis on the morphological properties of the Malayalam language, the above studies tried to understand the language morphology through computer science lenses by localizing similar attempts surfaced elsewhere. These studies show that the approaches of suffix striping, paradigm approach, Directed Acyclic word diagram; Corpus based approach, Finite state Transducer, Tow level morphology, Finite State Automata are already followed. Ashwani Shaji and Sindhu L (2014) made a comparison of different morphological techniques used for morphological analyzing for Malayalam. In their study both advantages and disadvantages and limitations of the above methods were discussed. A close look on the limitations of this methods outwardly suggest failure of addressing the morphological properties of Language at one side and the non linking of computer science and linguistics on the other. The same is observed in case of papers publish in various journals. Vinod, P.M. (2010), Raveendra Kumar (2011), Rajeev, R.R., (2008), Jisha, Jayan P (2009) Jayan,V (2015) for instants Nirmal J Valat (2014), Rajeev R.R (2008), Vinod P.M. (2012) are some important papers to be mentioned here. The only available online Malayalam morphological analyzer[1] is provided by Swathandra Malayalam Computing (SMC). Under SMC Activity Santhosh Thottingal had made effivetive approach in building Malayalam morphological Analyzer following FST Model[2]. This model is comparatively effective in handling limited morphological features of verb. However, the said product remains unfinished hence complete review cannot be done.

The present study proposes a holistic model for verb analysis by integrating all inflectional features of Malayalam verbs. The model comprises of three steps: classification of verbs, detailing inflectional properties and mapping morpho-phonological alterations.

3. Classification of Verbs

There are varies attempts made as part of traditional grammar and modern linguistics on classification of Malayalam verbs. The structure of stem and morphophonemic forms are considered to classify the verbs

[1] https://morph.scm.org.in
[2] http://thottingal.in/blog/2017/11/26/towards-a-amalayalam-morphological-analysers/

by A.R. Rajaraja Varma (1914/2011). He Classified Malayalam Verbs mainly based on the stem final Phoneme. Shoorand Kunjan Pillai (1965/2000) and V.R. Prabodhachandran (1970) attempted classification based on descriptive linguistic approach. A review on A.R's traditional Manner of classification done as part of his grammar wiring is not found to be effective to adopt to develop computational morphological model. Shooranad Kunjan Pilla's and V.R. Prabodhachnadran's Classifications are not effective in Computational modeling. The inherent limitations are evidently described (Abrar 2019). Besides the above attempts of grammatical classification, Centre for Development of Advanced Computing (CDAC) had developed a Computational verb classification as part of Angla Malayalam machine translation project (Ravindra Kumar et al 2011). This model obviously follows A.R. Rajaraja Varma's and Shooranad Kunjan Pilla's Classification. Considering some of the Morphological Feature but not all inflectional aspects CDAC developed 53 point classification of Malayalam verbs. The complex categorization suggested in this model is found not effective in Morphological analyzer building. The above survey is a self evident state of art of computational Morphology of Malayalam verbs. Considering inadequate platform a new classification is proposed against the 53 point CDAC classification and traditional classification attempts. The proposed classification has identified 11 classes of verb stems based on its past tense forms and each class further inflected based on three Tenses, seven Aspects and fifteen Modality forms. Among the 11 classes, four class (G-J) sub divided into two *kaaritham*(strong verbs) and *akaaritham*(weak verbs) forms.

The Verb classes are shown as:

3.1 Class A

Based on the end phoneme on the verb stem there are two subtypes identified in this class – jə ending and –ɻu ending both take -tu as the past tense marker. Total of eleven verb stems are found in this class.

3.2 Class B

Stem ended with -n̲, -l̲ and both take -ɳɻu as past tense marker. Twenty eight verb stems are grouped under this class.

3.3 Class C

C type Contains of Vowel ending forms. (-a, -ā, -i, -ī, -ū, -e). This group takes -ɲɲu as the past tense marker. One hundred thirty four verb stems are found in this class.

3.4 Class D

Contain two stems of -o (no) ending and -e (ve) ending and -ntu is the past tense marker. Only two verb stems cover this class.

3.5 Class E

This based on -ɻu ending stems. -ɳu, -ɳɳu is the past tense marker. Nineteen verb stems shows Class E characters.

3.6 Class F

The F class contain Stem with -a, -ā, -i, -e ending forms. -cu is its past tense marker. Nine hundred eighty seven verb stems are grouped under this class. This is most productive class.

3.7 Class G

G contains both kaaritha and akaaritha stems. Akaaritha has only one form and its end phoneme is -ā. kaaritha forms ended in -a, -ā, -i, -ī, -u, -ū, -o, -ṛ. Both kaaritha and akaaritha take -ttu asthe past marker. This class has two sub types based on kaarita akaarita dichotomy. Akarita subclass G1 contains only one verb stem and in G2 has one hundred and seventeen (117).

3.8 Class H

H type contain verb stem ending -ʈə and l̲. -ʈə ending are akaaritha form which takes -ʈʈu as the past tense marker. The kaaritha forms normally end with l̲. This group also has two sub types. H1 akaarita form of elevn number and H2 has three stems.

3.9 Class I

I type is based on -r̲ə, -l stem ending. And both take -rru has past tense marker. The -r̲ə forms are akaarithas and -l forms are
kaarithas. In subtype I1 has three and I2 has four stems.

3.10 Class J

J Contain -a, -u, -l, -r̲, -n ending stems and nnu is taken as the past tense marker. Among this group the stems ending with, -r̲, -l, -n are akaarithas and -a, -u, -l are kaarithas. -l forms are not geminating unlike the akaaritha forms. J1 contains 96 and J2 has 30 verb stems.

3.11 Class K

K class stems are ending in consonants -k, -c, -t, -p, -ŋ, -ɳ, -ɲ, -n, -m, -v, -ɻ, -ṛ, -l, -l̲, -s, -z . Along with these forms, ending in geminated forms kk, cc, pp, tt, ll, l̲l̲, nn and Consonants Clusters like mpa, ɳca, ŋka are also found. All these forms take i as the past tense marker. Six hundred and seven (607) are grouped under this class.

Total of 2053 verb stems of Malayalam are classified and clustered as described above. This scientifically classified verb stems are treated separately in the environment of its inflections. It means each verb stem reflexes three forms in tense context and eight Aspects and nineteen modality contexts and altogether 30 inflected forms to each verb stem is generated. For detailing the Tense Aspect modality forms of one class is shown below as an example.

3.1. Class A

Two verb stems under this class is tested based on the three tenses, eight Aspects and nineteen Modality inflections as follows.

3.1.1 Table 1: Tense Forms

	V1	V2
Stem end	pejə	uɻu
Past	pejtu	uɻutu
Present	pejjunnu	uɻunnu
Future	Pejjum	uɻum

3.1.2. Table 2: Aspect Forms

	V1	V2
Progressives 1, Present tens + unʈə 2, Infinitive + āɳə	pejjununʈə pejjukajāɳə	uɻununʈə uɻukajāɳə
Iterative 1, koɳʈə + irikkə + Tense markers 2, koɳʈə + ē + irikkə + Tense markers	pejtukoɳʈirikkunnu pejtukoɳʈēirikkunnu	uɻutukoɳʈirikkunnu uɻutukoɳʈēirikkunnu
Perfect 1, Simple- Past + irunnu 2, Contemporaneous- Past + irikkunnu 3, Remote- Past + iṯṯə + unʈə	pejtirunnu pejtirikkunnu pejtiṯṯunʈə	uɻutirunnu uɻutirikkunnu uɻutiṯṯunʈə
Habitual 1, ārə + unʈə + Tense markers	pejjāṛunʈājirunnu	uɻāṛunʈājirunnu

3.1.3. Table 3: Modality Forms

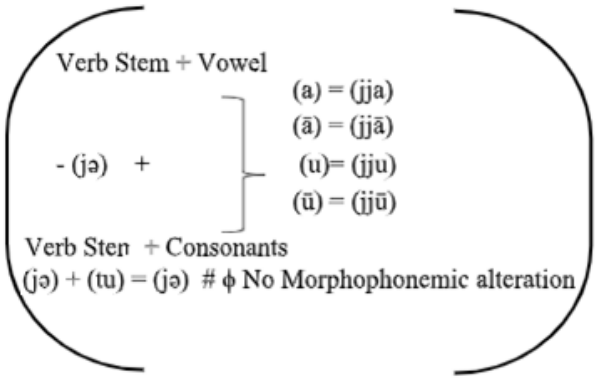
	V1	V2
Imperative	Pejjū, pejjaɳam	uɻū, uɻaɳam
Compulsive	Pejtē prrū	uɻutē prrū
Negative Compulsive	Pejjāte prrilla	uɻāte prrilla
Promissive	Pejjām	uɻām
Permissive	pejtukollu	uɻutukollu
Optative	pejjaṯṯe	uɻaṯṯe
Precative	pejjaɳē	uɻaɳē
Negative Precative	Pejjarutē	uɻarutē
Desiderative	pejjaɳamājirunnu	uɻaɳamājirunnu
Abilitative	pejjāvunnatējullu	uɻāvunnatējullu
Irrealis	Pejtēne	uɻātēne
Purposive	Pejjān	uɻān
Conditional	Pejtāl	uɻutāl
Satisfactive	Pejtallo	uɻutallo
Monitory	Pejjumē	uɻumē
Epistemic Modality		
Evidential Modality	Pejjām	uɻām
Judgmental Modality		
Alethic	pejjapeṯunnu	uɻapeṯunnu
Quotative	pejtatre	uɻutatre
Reportive	pejtupōlum	uɻutupōlum

Like the above, reaming ten classes can also be tested. In this way the verb stems and its inflections in currency can be fully verified and eventually a data base can be generated.

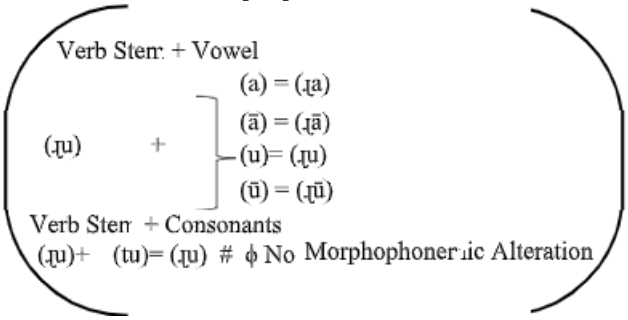
4. Morphophonemic Alterations

The inflections discussed in 3.1.1., 3.1.2, and 3.1.3 opens a new vista of morphophonemic alterations resulted after the verb stem conjoins with the Tense, Aspect and Modality inflections. All the eleven types conjoined with tense Aspect modality inflections can be tested class wise and rule of Morphophonemic alteration can be generated. All Categories of verb stem undergoes following changes when it encounter with Tense, Aspect, Modality forms is shown in picture 1-16.. .

4.1 Picture 1: V1 Morphophonemic Alteration

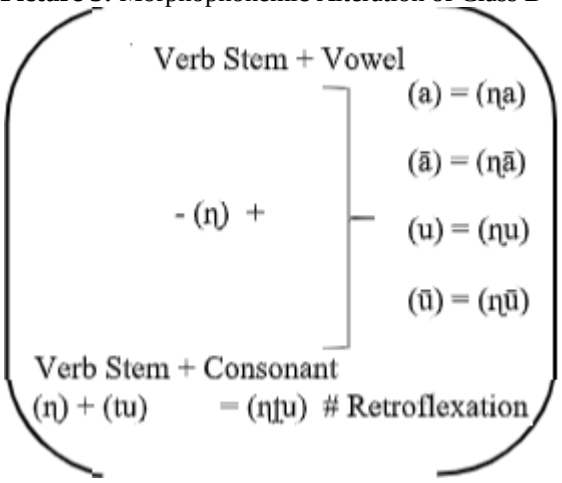


4.1 Picture 2: V2 Morphophonemic Alteration



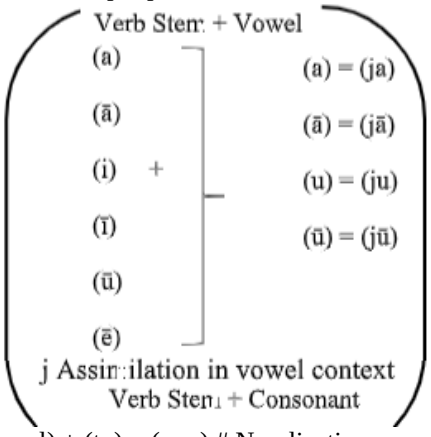
4.2 Class B

Picture 3: Morphophonemic Alteration of Class B



4.3 Class C

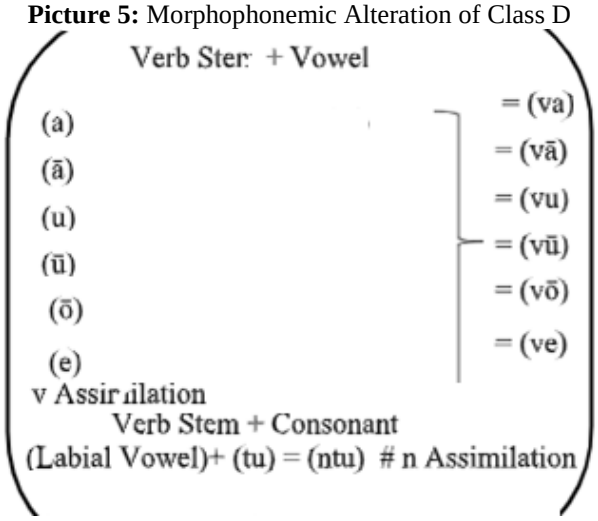
Picture 4: Morphophonemic Alteration of Class C



(Palatal Vowel) + (tu) = (ɲɲu) # Nasalization

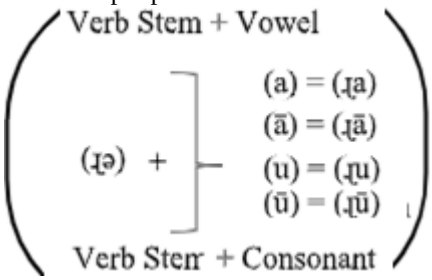
4.4 Class D

Picture 5: Morphophonemic Alteration of Class D



4.5 Class E

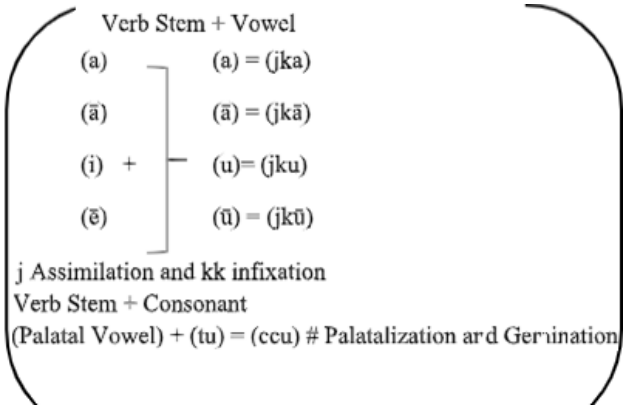
Picture 6: Morphophonemic Alteration of Class E



(ɻə)+ (tu) = (ɳu) # Rertoflexation

4.6 Class F

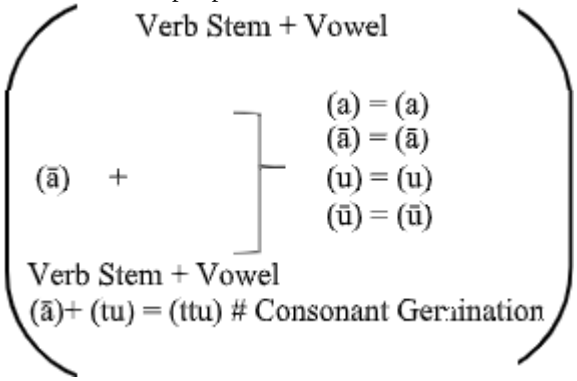
Picture 7: Morphophonemic Alteration of Class F



4.7 Class G

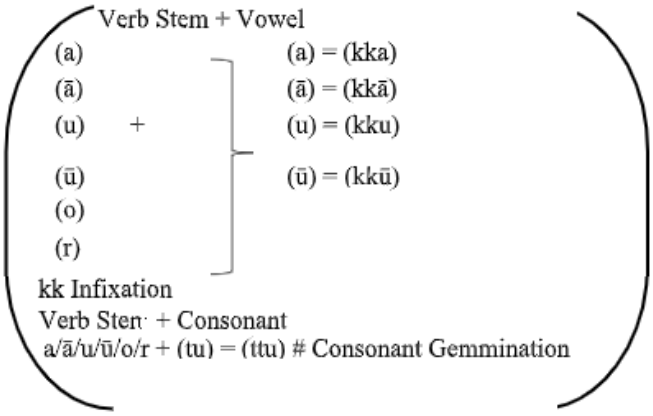
4.7.1Subclass G_1

Picture 8: Morphophonemic Alteration of Class G_1



4.7.2 Subclass G_2

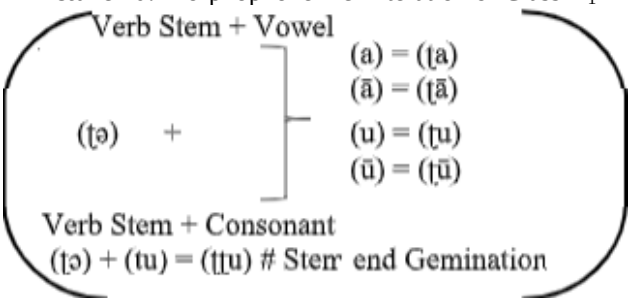
Picture 9: Morphophonemic Alteration of Class G_2



4.8. Class H

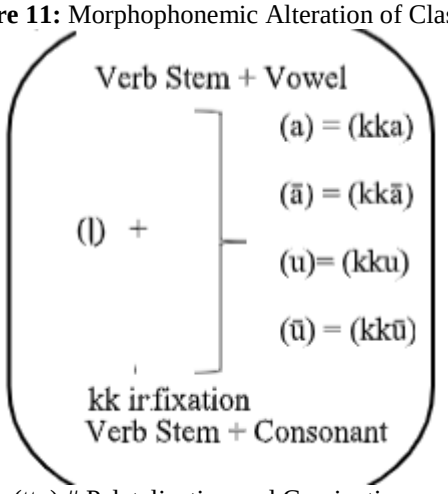
4.8.1Subclass H_1

Picture 10: Morphophonemic Alteration of Class H_1



4.8.2 Subclass H_2

Picture 11: Morphophonemic Alteration of Class H_2

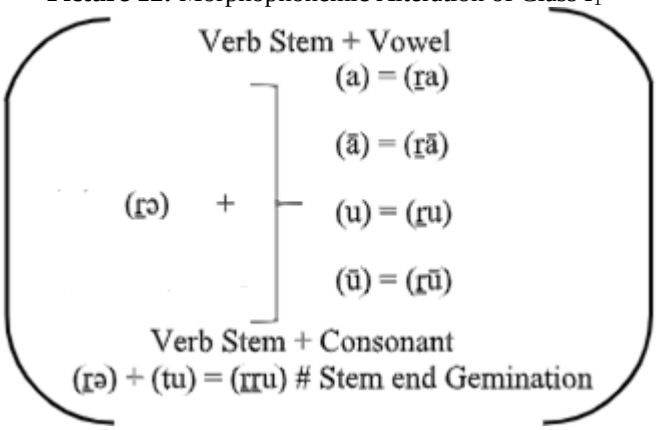


(l) + (tu) = (ṯṯu) # Palatalization and Gemination

4.9 Class I

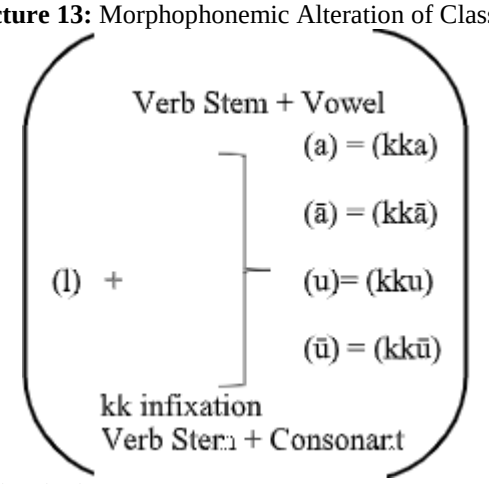
4.9.1 Subclass I_1

Picture 12: Morphophonemic Alteration of Class I_1



4.9.2 Subclass I_2

Picture 13: Morphophonemic Alteration of Class I_2

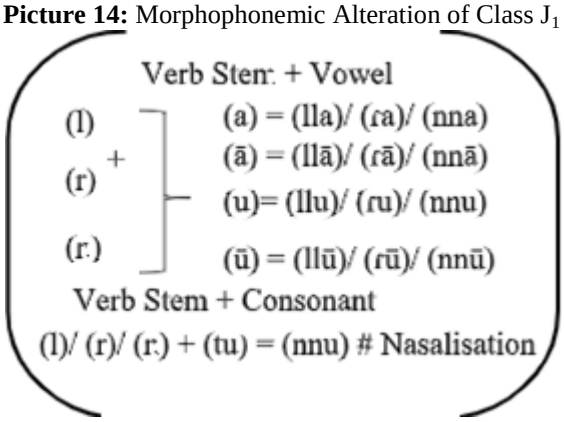


(l) + (tu) = (ṟṟu) # ṟ Gemination

4.10 Class J

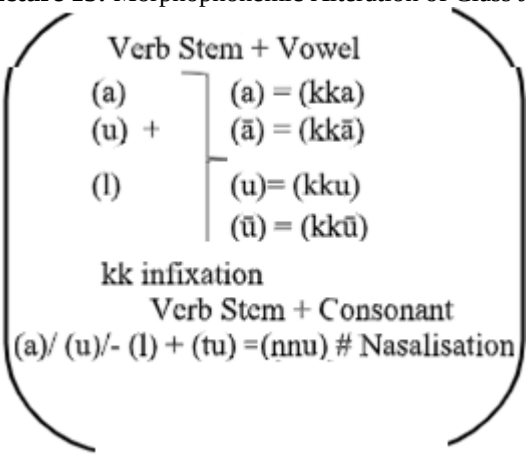
4.10.1.Subclass J_1

Picture 14: Morphophonemic Alteration of Class J_1



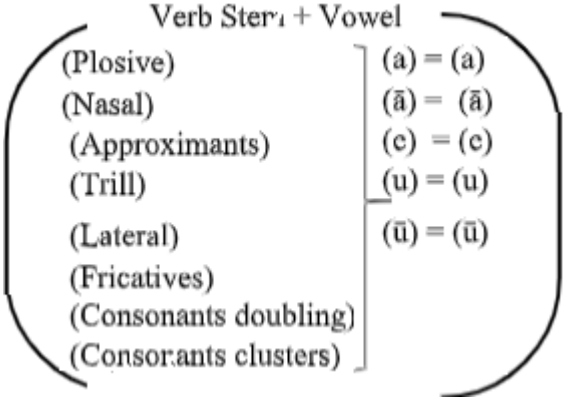
4.10.2 Subclass J_2

Picture 15: Morphophonemic Alteration of Class J_2



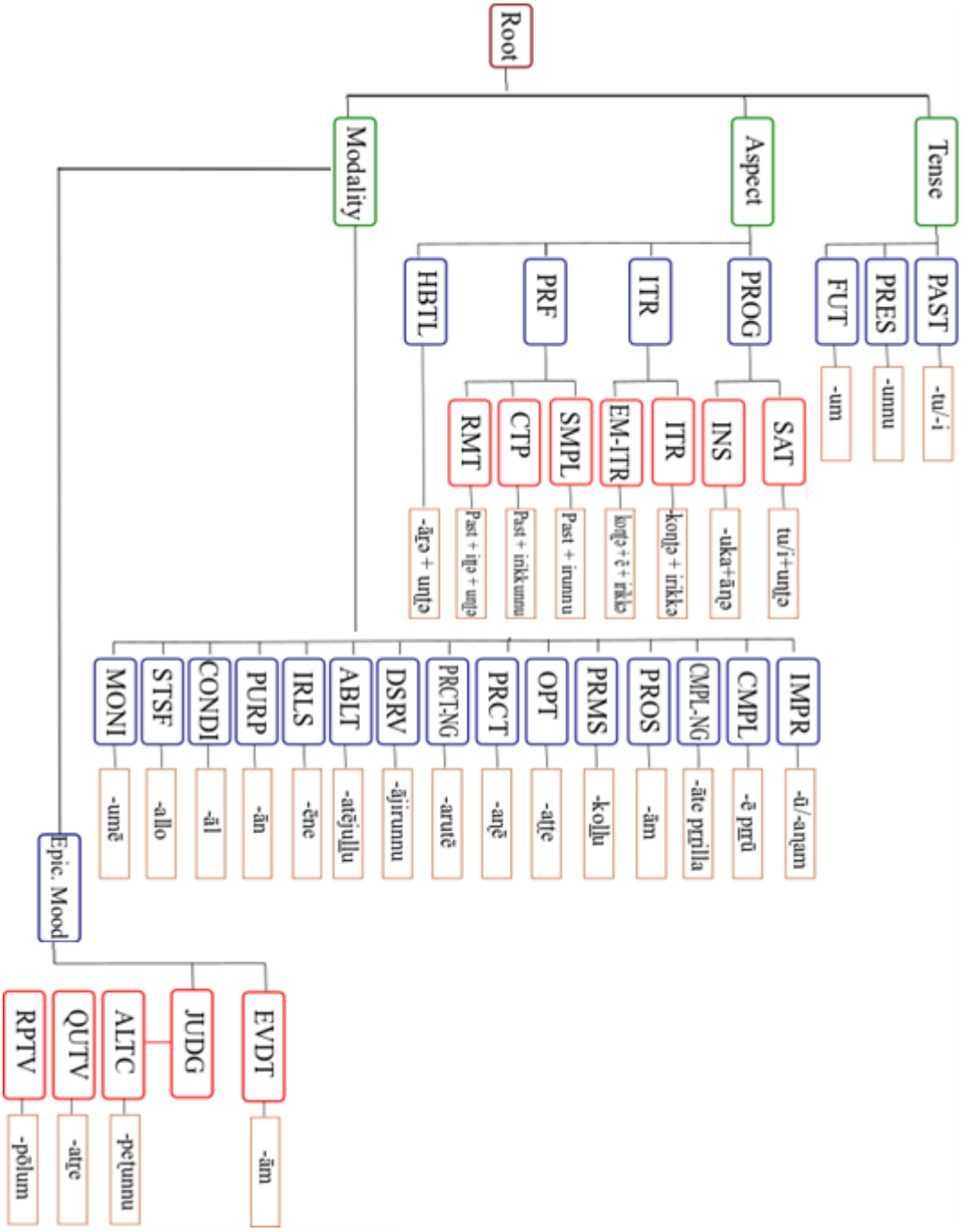
4.11. Class K

Picture16: Morphophonemic Alteration of Class K



5. Model Architecture

Like the above, eleven categories of verb stem in the inflected environments are tested and undisputable result can be formulated. This suggests a holistic model of verb analyzer for Malayalam shown in the flowchart below.



6. Conclusion and Future Work

In this article, we have presented a new model for processing verbs as part of Morphological analyzer. The classified verb stems along with the 30 inflections of each verb would demonstrate the power of the empirical data base and indexing of morphophonemic alterations would further predict the environments easily. This model has the Advantage of utilizing Computational Linguistics base for designing Malayalam Morphological analyzer. Based on the architecture proposed here software can be designed to generate morphological out comes as future work. The same model can be extended to other grammatical categories in order to generate a functional morphological analyzer for Malayalam language.

References

[1] Abrar, K.J. 2018. Computational Analysis of Verbs in Malayalam. Vol18:11 November. Language iIndia. www.languageinindia.com
[2] Abrar, K.J. 2019. Malayalam Verb Morphological Analysis; A Computational Linguistics Approach. PhD Thesis (Unpublished). Tirur: Malayalam University.
[3] Anitha, K. 2016. Identification of Verb from Malayalam Corpus: A Computational Approach (Mphil Thesis). Thiruvananthapuram: Kerala University
[4] Anju, A. M. 2015. Identification of Verb from a Malayalam Corpus (MPhil Thesis). Thiruvananthapuram: Kerala University
[5] Arunlal, N. V. 2011. Statistical Morph Analysis for Malayalam (MPhil Thesis). Thiruvananthapuram: Kerala University.

[6] Asher, R.E., Kumari, T.C. 1997(2012). *Malayalam*. Oxen: Routledge.
[7] Jayan,V, Bhadran,V.K. 2015. *Difficulties in Processing Malayalam verbs for Statistical Machine Translation*. Trivandrum; CDAC.
[8] Jisha, Jayan P. Rajeev, R.R., Rajendran, N. 2009. *Morphological Analyser for Malayalam- A Comparison of different Approaches*. Vol.2, No.2. IJCSIT.
[9] Kumar, K.K., Sulochana, K.G., Jayan, V. 2011. Computational aspect of verb classification in Malayalam. Information Systems for Indian Languages, 15-22.
[10] Kunjan Pillai, Shuranad. 1965/2001. Malayalam Lexicon Vol.I. Trivandrum: University of Kerala.
[11] Nisha, M. 2015. Malayalam Morphological Analyzer Using MBLP Approach (M.tec Thesis). Palakkad: Govt. Engineering College.
[12] Prabhodhachandran nair, V.R. 1970. Malayalam Descriptive Grammar. Trivandrum: DLA.
[13] Rajaraja Varam, A.R. 2011. Keralapāṇinīǰam. Kottayam: DC books.
[14] Rajeev, R. R. 2006. Morphological Generator for Malayalam nouns (Mphil Thesis). Thiruvananthapuram: Kerala University
[15] Rajeev, R.R., Rajendran N., Elizabeth, Sherly. 2008. *A suffix striping morph analyser for Malayalam language*. Proceedings of 20th Kerala Science Conference.
[16] Raajeev R.R.2015 Natural language processing works in Malayalam in The Languages of Kerala and Lakshadweep (eds) Sreenathan M and Joseph Koyippally, New DelhiOrient Black swan
[17] Raveendra, Kumar, Sulochana, K.G., Jayan, V. 2011. *Computational Aspect of Verb Classification in Malayalam*. ICISIL.
[18] Ravi Shankar, Nair. 2012. *A Grammar of Malayalam*. Language in India.
[19] Ranimol, R. 2011. Malayalam Verb Morphological Analyser- A Hybrid Approach (MPhil Thesis). Thiruvananthapuram: Kerala University
[20] Reena Ravindran. 2013. Morphological Analyser for Malayalam Nouns- program Based (MPhil Thesis). Thiruvananthapuram: Kerala University
[21] Reji Rahmath, K. 2015. Morphological Generator for Malayalam Using MBLP Approach (M.tec Thesis). Palakkad: Govt. Engineering College.
[22] Rekha, M. S. 2013. Morphological Analyser for Malayalam Verb program based (MPhil Thesis). Thiruvananthapuram: Kerala University
[23] Saranya, S.K. 2008. Morphological Analyser for Malayalam verb. M.tec. Thesis. Coimbatore: Amrtha University.
[24] Shoja Raj. 2006. Morphological Generator for Malayalam Verbs (Mphil Thesis). Thiruvananthapuram: Kerala University
[25] Shuranad Kunjan pillai (Ed.). 2000. Malayalam Lexicon, Index1. Trivandrum: University of Kerala.
[26] Sreeja, T. D. 2011. Computational Analysis of Malayalam Suffixes (Mphil Thesis). Thiruvananthapuram: Kerala University.
[27] Sukumara Pillai, K. 1980. Kairaḷishabdhanushāsanam. Trivandrum: Kerala Language Institute.
[28] Sunil,R. Nimtha, Manohar. Jayan, V. Sulochana, K.G. 2012. *Morphological Analysis and Synthesis of Verbs in Malayalam*. ICTAM.
[29] Vinod, P.M., Jayan, V., Bhadran, V.K. 2010. *Implementation of Malayalam Morphological Analyser Based on Hybrid Approach*. ROCLING.