Malayalam Verb Analyzer: An Alternate Computational Model

Abrar K.J1, Sreenathan M2

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 stripping, paradigm approach, 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(Rajeev 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 stripping, 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 analyzer1 is provided by Swathandrap Malayalam Computing (SMC). Under SMC Activity Santhosh Thottingal had made effective approach in building Malayalam morphological Analyzer following PST Model2. 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 morphophonological 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-malayalam-morphological-analysers/
3.3 Class B

Stem ended with -ŋ, -l and both take -ŋu as past tense marker. Twenty eight verb stems are grouped under this class.

3.4 Class C

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

3.5 Class E

This based on -u ending stems. -ŋu, -ŋũu is the past tense marker. Nineteen verb stems shows Class E character.

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 akāraitha stems. Akāraitha has only one form and its end phoneme is -ā. Kaaritha forms ended in -a, -ā, -l, -ī, -u, -ū, -o, -ē. Both kaaritha and akāraitha take -tu as the past tense marker. This class has two sub types based on kaaritha akāraitha 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 -ṭā and ḫ. -ṭā ending are akāraitha form which takes -ṭu as the past tense marker. The kaaritha forms normally end with ḫ. This group also has two sub types. H1 akāraitha form of elevn number and H2 has three stems.

3.9 Class I

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

3.10 Class J

J Contains -a, -u, -l, -ī, -n ending stems and ŋnu is taken as the past tense marker. Among this group the stems ending with, -ṛ, -l, -n are akāraithas and -a, -u, -l are kaarithas. -l forms are not geminating unlike the akāraitha forms. J1 contains 96 and J2 has 30 verb stems.

3.11 Class K

K class stems are ending in consonants -k, -c, -l, -p, -ŋ, -n, -m, -v, -q, -ɣ, -l, -i, -u, -z. Along with these forms, ending in geminated forms kk, cc, pp, tt, ll, nn and Consonants Clusters like mpa, nca, ṇka 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

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem end</td>
<td>peja</td>
<td>ūŋu</td>
</tr>
<tr>
<td>Past</td>
<td>peju</td>
<td>upu</td>
</tr>
<tr>
<td>Present</td>
<td>pejjuŋu</td>
<td>uŋunu</td>
</tr>
<tr>
<td>Future</td>
<td>Pejum</td>
<td>uŋum</td>
</tr>
</tbody>
</table>

3.1.2. Table 2: Aspect Forms

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressives 1, Present tens + ungu</td>
<td>pejjuŋungu</td>
<td>uŋungu</td>
</tr>
<tr>
<td>2. Infinitive + āŋa</td>
<td>pejukuŋungu</td>
<td>uŋukuŋungu</td>
</tr>
<tr>
<td>Iterative 1, kongo + irikk + Tense markers</td>
<td>pejukonširikkumun</td>
<td>uŋukonširikkumun</td>
</tr>
<tr>
<td>2. kongo + ē + irika + Tense markers</td>
<td>pejukonširikkumun</td>
<td>uŋukonširikkumun</td>
</tr>
<tr>
<td>Perfect 1, Simple- Past + irunnunu</td>
<td>pejirunnunu</td>
<td>uŋirunnunu</td>
</tr>
<tr>
<td>2. Contemporary - Past + irikkunnu</td>
<td>pejirannunnu</td>
<td>uŋirannunnu</td>
</tr>
<tr>
<td>3. Remote- Past + āŋa + ungu</td>
<td>pejirannunnu</td>
<td>uŋirannunnu</td>
</tr>
</tbody>
</table>

3.1.3. Table 3: Modality Forms

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperative</td>
<td>Pejǔ, pejanam</td>
<td>uŋu, uŋanam</td>
</tr>
<tr>
<td>Compulsive</td>
<td>Peitē prṟu</td>
<td>uŋitē prṟu</td>
</tr>
<tr>
<td>Negative Compulsive</td>
<td>Pejite pririlla</td>
<td>uŋite pririlla</td>
</tr>
<tr>
<td>Promissive</td>
<td>Pejām</td>
<td>uŋām</td>
</tr>
<tr>
<td>Permissive</td>
<td>pejukolla</td>
<td>uŋukolla</td>
</tr>
<tr>
<td>Optative</td>
<td>pejiŋate</td>
<td>uŋate</td>
</tr>
<tr>
<td>Precautive</td>
<td>pejjanē</td>
<td>uŋanē</td>
</tr>
<tr>
<td>Negative Precautive</td>
<td>Pejaranē</td>
<td>uŋaranē</td>
</tr>
<tr>
<td>Desiderative</td>
<td>pejanamāŋijnunu</td>
<td>uŋanamāŋijnunu</td>
</tr>
<tr>
<td>Abilitative</td>
<td>pejāvunntējul</td>
<td>uŋāvunntējul</td>
</tr>
<tr>
<td>Irrealis</td>
<td>Pejēnē</td>
<td>uŋēnē</td>
</tr>
<tr>
<td>Purposive</td>
<td>Pejān</td>
<td>uŋān</td>
</tr>
<tr>
<td>Conditional</td>
<td>Pejāllo</td>
<td>uŋāllo</td>
</tr>
<tr>
<td>Satisfactive</td>
<td>Pejāllo</td>
<td>uŋāllo</td>
</tr>
<tr>
<td>Monitorry</td>
<td>Pejuŋme</td>
<td>uŋume</td>
</tr>
<tr>
<td>Epistemic Modality</td>
<td>Pejām</td>
<td>uŋām</td>
</tr>
<tr>
<td>Evidential Modality</td>
<td>Pejām</td>
<td>uŋām</td>
</tr>
<tr>
<td>Judgmental Modality</td>
<td>Pejām</td>
<td>uŋām</td>
</tr>
<tr>
<td>Alethic</td>
<td>pejapeŋunu</td>
<td>uŋapeŋunu</td>
</tr>
<tr>
<td>Quotative</td>
<td>pejatate</td>
<td>uŋatate</td>
</tr>
<tr>
<td>Reportive</td>
<td>pejutuŋūnum</td>
<td>uŋutuŋūnum</td>
</tr>
</tbody>
</table>

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

\[
\text{Verb Stem + Vowel} \\
\begin{align*}
(a) &= (ji\overline{a}) \\
(\overline{a}) &= (ji\overline{a}) \\
(u) &= (jiu) \\
(\overline{u}) &= (ji\overline{u}) \\
\end{align*}
\]

4.1 Picture 2: V2 Morphophonemic Alteration

\[
\text{Verb Stem + Vowel} \\
\begin{align*}
(a) &= (\overline{a}) \\
(\overline{a}) &= (a) \\
(u) &= (u) \\
(\overline{u}) &= (\overline{u}) \\
\end{align*}
\]

4.2 Class B

Picture 3: Morphophonemic Alteration of Class B

\[
\text{Verb Stem + Vowel} \\
\begin{align*}
(a) &= (\overline{a}) \\
(\overline{a}) &= (a) \\
(u) &= (u) \\
(\overline{u}) &= (\overline{u}) \\
\end{align*}
\]
4.3 Class C

**Picture 4: Morphophonemic Alteration of Class C**

Verb Stem + Vowel

- (a) = (ja)
- (i) = (jä)
- (u) = (ju)
- (ü) = (jü)

J Assimilation in vowel context

Verb Stem + Consonant

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

4.4 Class D

**Picture 5: Morphophonemic Alteration of Class D**

Verb Stem + Vowel

- (a) = (va)
- (i) = (vi)
- (u) = (vu)
- (ü) = (vü)
- (e) = (ve)

J Assimilation

Verb Stem + Consonant

(Labial Vowel) + (tu) = (ntu) # Assimilation

4.5 Class E

**Picture 6: Morphophonemic Alteration of Class E**

Verb Stem + Vowel

- (a) = (qä)
- (i) = (qü)
- (u) = (qu)
- (ü) = (qü)

Verb Stem + Consonant

(qä) + (tu) = (qii) # Rertoflexation

4.6 Class F

**Picture 7: Morphophonemic Alteration of Class F**

**Verb Stem + Vowel**

(a) = (ja)

(i) = (ji)

(u) = (ju)

(ü) = (jü)

4.7 Class G

4.7.1 Subclass G1

**Picture 8: Morphophonemic Alteration of Class G1**

Verb Stem + Vowel

- (a) = (a)
- (i) = (i)
- (u) = (u)
- (ü) = (ü)

Verb Stem + Vowel

(i) + (tu) = (ttu) # Consonant Gemination

4.7.2 Subclass G2

**Picture 9: Morphophonemic Alteration of Class G2**

Verb Stem + Vowel

(a) = (lka)

(i) = (likä)

(u) = (lku)

(ü) = (likü)

(ə) = (lku)

4.8. Class H

4.8.1 Subclass H1

**Picture 10: Morphophonemic Alteration of Class H1**

Verb Stem + Vowel

(a) = (iia)

(i) = (iī)

(u) = (iű)

(ü) = (iü)

Verb Stem + Consonant

(iia) + (tu) = (ittu) # Stem End Gemination
4.8.2 Subclass H₂

*Picture 11: Morphophonemic Alteration of Class H₂*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad \text{(a)} = (kkä) \\
\text{ Verb Stem + Consonant} & \quad (l) + (u) = (ku) \\
\text{kk infixation} & \quad (u) = (kkũ) \\
\end{align*}
\]

(l) + (tu) = (ttu) \# Palatalization and Gemination

4.9 Class I

4.9.1 Subclass I₁

*Picture 12: Morphophonemic Alteration of Class I₁*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad (a) = (ra) \\
\text{Verb Stem + Consonant} & \quad (a) = (κκa) \\
\text{Verb Stem + Consonant} & \quad (u) = (κku) \\
\text{Verb Stem + Consonant} & \quad (u) = (κκu) \\
\end{align*}
\]

(rₐ) + (tu) = (rru) \# Stem end Gemination

4.9.2 Subclass I₂

*Picture 13: Morphophonemic Alteration of Class I₂*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad (a) = (kka) \\
\text{Verb Stem + Consonant} & \quad (a) = (kκa) \\
\text{Verb Stem + Consonant} & \quad (u) = (kku) \\
\text{Verb Stem + Consonant} & \quad (u) = (kκu) \\
\end{align*}
\]

(l) + (tu) = (rru) \# i Gemination

4.10 Class J

4.10.1 Subclass J₁

*Picture 14: Morphophonemic Alteration of Class J₁*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad (a) = (κκa) \\
\text{Verb Stem + Consonant} & \quad (a) = (κκa) \\
\text{Verb Stem + Consonant} & \quad (u) = (κku) \\
\text{Verb Stem + Consonant} & \quad (u) = (κκu) \\
\end{align*}
\]

(rₙ) + (tu) = (nu) \# Nasalisation

4.10.2 Subclass J₂

*Picture 15: Morphophonemic Alteration of Class J₂*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad (a) = (κκa) \\
\text{Verb Stem + Consonant} & \quad (a) = (κκa) \\
\text{Verb Stem + Consonant} & \quad (u) = (κku) \\
\text{Verb Stem + Consonant} & \quad (u) = (κκu) \\
\end{align*}
\]

(rₙ) + (tu) = (nu) \# Nasalisation

4.11 Class K

*Picture 16: Morphophonemic Alteration of Class K*

\[
\begin{align*}
\text{Verb Stem + Vowel} & \quad (a) = (a) \\
\text{Verb Stem + Vowel} & \quad (a) = (κκa) \\
\text{Verb Stem + Vowel} & \quad (u) = (κku) \\
\text{Verb Stem + Vowel} & \quad (u) = (κκu) \\
\end{align*}
\]

(Plosive) (Nasal) (Approximants) (Trill) (Lateral) (Fricatives) (Consonants doubling) (Consonants clusters)

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


