Malayalam VerbFrames

Jisha P Jayan
Centre for Development of
Imaging Technology
Thiruvanathapuram
jishapjayan@gmail.com

Asha S Nair
Centre for Development of
Imaging Technology
Thiruvanathapuram
ashanaircdit@
gmail.com

Govindaru V
Centre for Development of
Imaging Technology
Thiruvanathapuram
neithalloor@gmail.com

Abstract

Verbs acts as a major role in describing a sentence meaning. Capturing of the syntactic distributions of occurrence of a verb in a sentence is the VerbFrame. This paper tests the applicability of verbframe approach that has been developed for Hindi language in Malayalam. Around 255 verbs were selected for this study, showing the basic argument structure of words with these verbs.

Keywords- verbframe; karaka relations; semantic; syntactic;

1 Introduction

Verbs are the most important grammatical category in any language. With the help of an action, activity and state are denoted. The arguments of the verb indicate various participants required by the verb. Verbs play a noteworthy part in interpreting meaning of a sentence, therefore, the study of the argument structure of a verb and their syntactic behavior will provide the needed knowledge base for intelligent NLP applications. Verbframe is the gathering of the syntactic distribution of the verb occurrence in any sentence. Paninian Grammatical Framework (PGF) is followed in creating a Verbframe as verb plays the important role in the sentence analysis.

The relation of verb with the alternate units of a sentence, in a language may be encoded in various ways. Among them, the word order and the presence of case markers on the arguments are very often used by computational linguists. There are, however, languages in which the marking can be present of the verb itself rather than its arguments (Butt, 2010). Such types of relations frequently reflect semantics of a verb, that-means the syntactic behavior of the verb provides a good support to understand its semantics. Researchers also

encode other information such as tense, aspect, modality, gender, number, person etc., with verb, that allow language specific variations.

This paper is intended to develop verbframe for Malayalam language which has got grammatical roots from Dravidian and Aryan languages. This paper presents the work in different stages, beginning in Section 2 with the major works related. Section 3 introduces the Verb Frame and its description . Section 4 describes the Verb frame for Malayalam. Finally, Section 5 concludes the paper.

2 State of Art

Some of the famous linguistic sources related to verb argument structure, are discussed briefly in this section. Levin's work on verb classes (Beth, 1993) indicates the relationship between semantic and syntactic behavior of the English verbs. The verb behavior can be used to get an insight into linguistically applicable aspects of the verb meaning (Beth, 1995). VerbNet (VN) (Kipper, 2000) (Kipper, 2005) is a domain-independent; hierarchical, wide-coverage of online verb dictionary which extends Levin's verb classes (Beth, 1993) and providing syntactic and semantic information for English verbs. It is mapped to various language resources such as Wordnett (Fellbaum, 1998), FrameNet, and PropBank. Each class of verbs in VN is described by thematic roles, selectional restrictions on the arguments, and syntactic frames (Beth, 1993).

PropBank (PB) (Palmer, 2003) (Palmer, 2005) is a corpus, annotated with verbal propositions and their arguments. This has been extensively used for semantic role labeling task in recent times (CoNLL shared task 2004-05 and 2008-2009). PB gives a layer of semantic annotation upon the syntactic structures. PB represents the verb argument depending on the valency of the verb relations by Arg0, Arg1, Arg2, etc., (Palmer,

236

2002). Each set of argument labels and their definitions is called a frameset. As an example, consider the frameset for the verb *dance*. This verb takes the dancer:Arg0, dance:Arg1 , partner:Arg2 and audience:Arg3 as essential roles. It also has non-essential roles such as location:Argm-loc and time:Argm-tmp. This is for capturing spatio-temporal aspects of verbs.

FrameNet (FN) (Baker et al., 1998) is an on-line lexical resource for English, based totally on frame semantics and supported by means of corpus evidence. FN groups words in accordance to the conceptual structures, i.e., frames that underlie them (Arun, 2008) . The paper describes three major components such as: (1) Lexicon; (2) Frame Database; (3) Annotated Example Sentences. The Frame database deals with the descriptions of each frame's basic conceptual structure, and provides the names and descriptions of the elements participating in such structure (Begum, 2017). Annotated Sentences are marked to illustrate the semantic and morpho-syntactic properties of the lexical items. Each frame contains numerous elements, i.e., core (core arguments) and non-core (adjuncts or peripheral roles) elements which are considered as semantic roles. For example, core elements of the frame Getting-up are person/animal getting up from sleep and place of sleeping; non-core elements are time, purpose, etc.

All these resources looks into the argument structure of English verbs. They gives the syntactic and semantic information, and correlation between them. These resources are also mapped to each other making individual resources much richer. In the work of creating verb frames for Hindi, the argument structure of verb is captured using Karaka relations which capture both syntactic and semantic information about the verbs. Between Karaka relations, thematic roles and Propbank annotation, a mapping is done . Begum et al. (Begum, 2008) mentioned their experience with the creation of Hindi verb frames. These frames are further classified based on a Paninian grammar framework using 6 Karaka relations. This method considered the morphology, syntactic variations and semantics of the verb to divide it into various classes.

Based on similar approach, Ghosh (Ghosh, 2014) created a resource for verb frames for compound verbs in Bengali language. The main aim of the paper is to investigate if the vector verb

from the compound verb is able to retain its case marking properties and argument structure or not. Additionally the knowledge and syntax associated with verb frames can be utilized for categorizing and analyzing the verb words for various NLP applications.

Soni et al. (Ghosh, 2013) explores the application of verb frames and the conjuncts in sentence simplification for Hindi language. The method proposed by the authors includes usage of conjuncts as a first level of sentence simplification. This is followed by using verb frames enhanced with tense, aspect and modality features. It is a rule based system and its output is evaluated manually and automatically using the BLEU score for the ease of readability and simplification.

A semi-automatic annotator tool for verb frames was developed by Hanumant et al (Redkar, 2016). The tool is used for extracting and generating the verb frames automatically from the example sentences of Marathi wordnet. The paper explains the concept and working of the verb - frame tool with its advantages and disadvantages. Other related work by Schulte (Walde, 2009) has also explored verb frames for the English language.

3   Verb Frames

In all languages, verb plays the major part-of-speech category. Verbs are used to define actions, activities and states. Ability of the verbs to choose their arguments and/or adjuncts is termed as 'verb sub-categorization'or 'verb valency'. Combination of functional units that are elicited by a verb is refered to as verb frames. In linguistics, verb-framing and satellite-framing are typological descriptions of how verb phrases in different languages describe the path of motion or the manner of motion, respectively (Redkar, 2016).

Verb frame generally constitutes verbal propositions and arguments of words surrounding a verb in a given sentence. Each of the prepositional words in a verb frame has arguments such as an arc-label, otherwise called a semantic role label, its necessity in a frame, case markers or the suffixes, lexical type, relation of the word with head verb, position with respect to head verb, etc. These verb frames are developed to generate dependency tree structures in a given language. Verb frames on the basis of their argument demands categorization of any verb. The verb frames show mandatory

Karaka[1] relation for a verb. They are:

1. *Karaka : dependency arc labels.*

2. *The necessity of the argument whether it is mandatory (m) or desirable (d).*

3. *Case Markers / Vibhakti: post-position or the case associated with the nominal.*

4. *Lexical category of the arguments.*

5. *The Position of the demanded nominal with respect to verb whether it is left(l) or right(r).*

Verb frames are built for the base form of a verb. The demands undergo a subsequent change based on the tense, aspect and modality (TAM) of the verb used in the sentence. Knowledge about the transformations induced on the base form of a verb by TAM is stored in the form of transformation charts for each distinct TAM.

In the present work we develop verbframe for Malayalam based on Karaka theory developed by IIIT-Hyderabad for Hindi.

4 Malayalam Verb Frame

Amid the semantic analysis, verb is taken as the central, element of the sentence. According to Paninian viewpoint, there are four levels in understanding any sentence (Bharati, 1995) namely the surface level (uttered sentence), the vibhakthi level, the Karaka level and the semantic level. The Karaka level has related to semantics on one side and on the other side with the syntax. Karaka relation can be identified from markers/suffixes or case endings after the noun. The Karaka relations in Malayalam are analyzed from the point of vibhakthi and the postpositions that associate with it. The types of verb and the vibhakthi markers in Malayalam are illustrated in Figure 1 and Table 1 respectively.

The roles and the dependency relation based on IIIT_H approach, are shown in Table 2.

The genitive noun does not have any direct grammatical or semantic relation with the verb but only the noun modified by the genitive is related to the verb. The Genetive case "സംബന്ധികാവിഭ ക്തി" *saMbhndhikaavibhakti* otherwise Possessive takes the markers "ന്റെ" *nRe*, "ഉടെ" *uTe*.
Eg (a): രാമന്റെ അനിയൻ വന്നു.

No	Case	Case markers
1	nirddeeSika നിർദ്ദേശിക Nominative	ϕ
2	prathigraahika പ്രതിഗ്രാഹിക Accusative	എ -e
3	samyojika സംയോജിക Sociative	ഓട് -ooTu
4	uddeeshika mal Dative	ക്ക്, ന് kku ,nu
5	sambandhika സംബന്ധികാ Gentive	ന്റെ , ഉടെ -nRe,-uTe
6	aadhaarika ആരാധിക Locative	ഇൽ, കൽ -il, -kal
7	prayoojika പ്രയോജിക Instrumental	ആൽ aal
8	sambhoodana സംബോധിക Vocative	ീ , ാ, േ long forms
9	മിശ്രവിഭക്തി Ablative	ഇൽ നിന്ന് il ninn

Table 1: Case and Case Markers

ramanRe aniyan vannu.
Raman's brother came.
Eg (b): അവളുടെ അമ്മ പറഞ്ഞു .
avaLuTe amma paranjnju.
Her mother said.
Because of this, the genitive noun can be removed from the sentence without affecting the grammaticality of the sentence

Dependency annotated data are used for developing Malayalam verb frames. The dependency annotation is a collective process of Tokeniser, Morphological Analyser, POS tagger, Chunker and Dependency annotation. A raw text will be given as the input and the text is converted into tokens, identifies grammatical features of the individual words, assigns parts of speech (POS) tags to each word , groups them to phrases and the dependency tree diagrams are drawn. Malayalam has tendency to join a wide variety of suffixes with a single word forming compound words, which makes the process more complicated. Therefore complicated words are spllited and then analysed in the present analysis. As an example, consider the following sentence.

മറ്റ് ഭക്ഷണ സാധനങ്ങളെ അപേക്ഷിച്ച് പഴങ്ക ഞ്ഞിയിൽ ബി6 , ബി12 വൈറ്റമിനുകൾ ധാരാളം അടങ്ങിയിട്ടുണ്ട് .
maRRu bhakshaNa saadhanangngaLe apeek-shiccu pazhangkanjnjiyil b6 , b12 vaiRRaminukaL

---

[1]karakas are the typed dependency labels in Computational Paninian Framework (Bharati, 1993)

Verbs (ക്രിയകൾ) in Malayalam

Verb (കൃതി/ക്രിയ)

അർത്ഥം പ്രമാണിച്ച് (Meaning) — അകർമ്മകക്രിയ (Transitive Verb); അകർമ്മകക്രിയ (Transitive Verb)

പ്രകൃതി പ്രമാണിച്ച് (Steam) — കേവല ക്രിയ (Ordinary Verb); പ്രയോജകം ക്രിയ (Causative Ver)

രൂപം പ്രമാണിച്ച് (Morph) — കാരിത ക്രിയ (Strong Verb); അകാരിത ക്രിയ (Weak Verb)

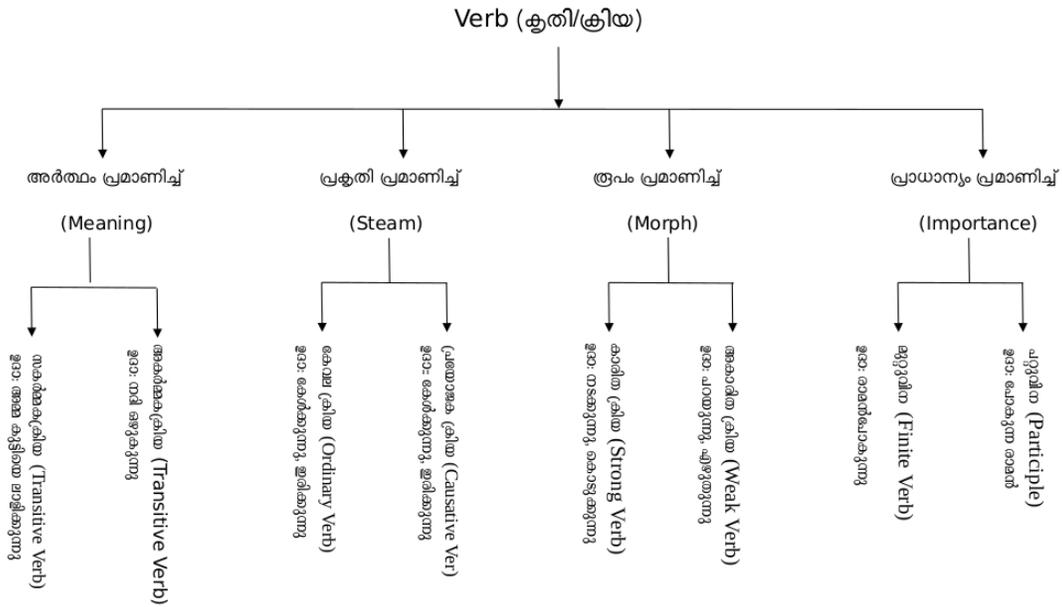പ്രാധാന്യം പ്രമാണിച്ച് (Importance) — മുറ്റക്രിയ (Finite Verb); പറ്റുക്രിയ (Participle)

Figure 1: Verb types in Malayalam

dhaaraaLaM aTangngiyiTTuNTu.

In comparison to other food items, rice gruel is rich in vitamins B-6 and B-12.

The sentenence is annotated as follows:
<Sentence id="1">
1 ((NP <fs af='സാധന,n,ne,pl,3,d,എ,NGaLe' head='സാധനങ്ങളെ' name='NP' drel='k2:VGF'>
1.1 മറ്റ് QT_QTF <fs af='മറ്റ്,qtf,,,,,,' name='മറ്റ്'>
1.2 ഭക്ഷണ JJ <fs af='ഭക്ഷണ,adj,,,,,,' name='ഭക്ഷണ'>
1.3 സാധനങ്ങളെ N_NN <fs af='സാധന,n,ne,pl,3,d,എ,NGaLe' name='സാധനങ്ങളെ'>
1.4 അപേക്ഷിച്ച് PSP <fs af='അപോക്ഷിച്ച്,psp,,,,,,' name='അപേക്ഷിച്ച്'>
))
2 ((NP <fs af='പഴങ്കഞ്ഞി,n,ne,pl,3,d,ഇൽ,il' head='പഴങ്കഞ്ഞിയിൽ' name='NP2' drel='k7:VGF'>
2.1 പഴങ്കഞ്ഞിയിൽ N_NN <fs af='പഴങ്കഞ്ഞി,n,ne,pl,3,d,ഇൽ,il' name='പഴങ്കഞ്ഞിയിൽ'>
))
3 ((NP <fs af='ബി6,n,ne,sg,3,d,0,0' name='NP3' drel='ccof:NULL_CCP'>
3.1 ബി6 N_NN <fs af='ബി6,n,ne,sg,3,d,0,0'

name='ബി6'>
3.2 , RD_PUNC <fs af='&comma,punc,,,,,,' name=','>
))
4 ((NULL_CCP <fs af=',,,,,,,' dmrel='k1:VGF' name='NULL_CCP'>
4.1 NULL CC <fs af=',,,,,,,' name='NULL'>
))
5 ((NP <fs af='വൈറ്റമിന്,n,ne,pl,3,d,കൾ,kaLu' head='വൈറ്റമിനുകൾ' name='NP4' drel='ccof:NULL_CCP'>
5.1 ബി12 N_NN <fs af='ബി12,n,ne,sg,3,d,0,0' name='ബി12'>
5.2 വൈറ്റമിനുകൾ N_NN <fs af='വൈറ്റമിന്,n,ne,pl,3,d,കൾ,kaLu' name='വൈറ്റമിനുകൾ'>
))
6 ((JJP <fs af=',qtf,,,,,,' head='4' name='JJP' drel='pof:VGF'>
6.1 ധാരാളം QT_QTF <fs af='ധാരാളം,qtf,,,,,,' name='ധാരാളം'>
))
7 ((VGF <fs af='അടങ്ങ്,v,,,,,ഇഉണ്ട്,iTTuNTu' head='അടങ്ങിയിട്ടുണ്ട്' name='VGF' Participles_m='ഇട്ട്' Participles='yes'>
7.1 അടങ്ങിയിട്ടുണ്ട് V_VM_VF <fs af='അടങ്ങ്,v,,,,,ഇഉണ്ട്,iTTuNTu' name='അടങ്ങിയിട്ടുണ്ട്'

239

Karakas	Case	Case Marker		Role
Karthru Karakam	Nominative	ϕ	k1	Agent/Subject/Doer
Karma Karakam	Accusative	e എ	k2	Object /Patient/Causer
Karna Karakam	Instrumental	aal ആൽ	k3	Instrument
Sampradana Karakam	Dative Sociative	kku ,nu (ക്ക്, ന്) ooTu(ഓട്)	k4	Recepient/Experiencer
Apadana Karakam	Ablative	il ninn ഇൽ നിന്ന്	k5	Source
Vishayadhikarana Deesaadhikarana Kaladhikarana	Locative	ഇൽ il	k7 k7p k7t	Locative (in general) Location in space Location in time

Table 2: Karakas and Role (IIIT-H)

Participles_m='ഇട്' Participles='yes'>
    ))
8 ((BLK <fs af='.,punc,,,,,,' head='.36'
name='BLK' drel='rsym:VGF'>
8.1   . RD_PUNC <fs af='.,punc,,,,,,' name='.'>
    ))
</Sentence>

The dependency annotation is tree is given in Figure 2.

4.1 Diagnosis

Diagnosis of Malayalam verbfame is illustrated with an example of a verb entry with the description and verb frame. Gloss explains meaning of the particular verbal root. Arc label is to show the dependency relation between any words and the verb that exist in a sentence. Necessity is valency. Valency is the the number of grammatical aspects of verbs which combines other words in that sentence. On the other hand, it is the capacity of verbs that how many arguments, it can combine with itself at time. The distinction among the modifiers and complements is mostly defined using valency, which is a central notion in the theoretical tradition of dependency analysis (Theoritical tradition of dependency analysis has limitation in Computational Linguistics that has discussed widely in recent years. So it is not detailed here). Although the exact characterization of this notion differs from one theoretical framework to the other, valency is usually related to the semantic predicate-argument structure associated with certain classes of lexemes, in particular verbs but sometimes also

nouns and adjectives (Nivre, 2005). The idea is that the verb imposes requirements on its syntactic dependents that reflect its interpretation as a semantic predicate. Dependents that correspond to the arguments of the predicate can be mandatory or optional (Nivre, 2005). The valency frame of the verb is generally taken to incorporate argument dependents, however some theories also allow desirable non-arguments to be included. Position basically refers to on which side of the verb, the particular word takes place in the sentence. That is, the word can occur either on the left side of the verb or right side of the verb. So, 'l' stands for the word left and 'r' stands for the word right.

Verb::അടങ്ങ് *aTangngu*
SID:: അടങ്ങ് *%VT%S1*
Verb Sense::
Eng_Gloss::to contain
Verb Class::
Verb_in_Same_Class::
TAM for the verb root::ഇഇട്ട്ഉണ്ട് *i iTT uNTu*

Frames::
Example::മറ്റ് ഭക്ഷണ സാധനങ്ങളെ അപേക്ഷിച്ച് പഴങ്കഞ്ഞിയിൽ ബി6 , ബി12 വൈറ്റമിനുകൾ ധാരാളം അടങ്ങിയിട്ടുണ്ട് .
maRRu bhakshaNa saadhanangngaLe apeekshiccu pazhangkanjnjiyil b6 , b-12 vaiRRaminukaL dhaaraaLaM aTangngiyiTTuNTu.
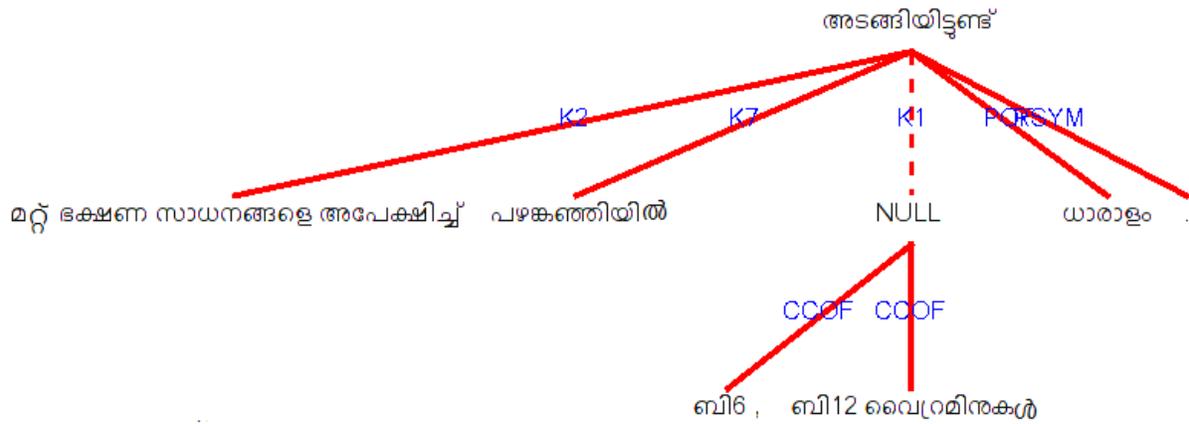In comparison to other food items, rice gruel is rich in vitamins B-6 and B12.
FRAME_ID::1

Figure 2: Dependency tree for given example

arc_label	nec-essity	Vibhakti	Lex Type	posn	reln
k2	m	കൾഎ (kaLe)	n	l	c
k7	m	ഇൽ (il)	n	l	c
k1	m	കൾ (kal)	n	l	c
pof	m	0	n	l	c

In the verbframe file above as example, the first feild gives the name of the verb. SID is the unique sense identification number. It is represented as *verb_root%verb_type%sense number*. The verb_types in Malayalam are distinguished into transitive, intransitive and causative. Here in the example the type of the verb is transtitive and is represent by *VT*. Verbs_in_Same_Class field gives the list of all the verbs that have same meaning as the given verb. Since Malayalam is a verb final language, by default, all the words are kept normally on the left side of the verb. Rarely, it happens that particular word occurs on the right side of the verb. As an example consider the following sentence:

എങ്കിലും ഭരതനാട്യത്തിന് തന്നെ ശ്രദ്ധ ചെലുത്താൻ ആണ് നീലിമ ഇപ്പോൾ ശ്രമിക്കുന്നത് .

enkiluM bharatanaaTyattinu tanne shRaddha celuttaan aaNu niilima ippoL shRamikkunnatu.

Nilima is now trying to focus on Bharatanatyam itself.

In the above example, "ആണ്" *aaN* is the finite verb and it occurs in between the given sentence. These type of constructions are found mostly. In such cases, the the words that follow the final

verbs are positioned on the right side. The verb frame for above example is as follows :

arc_label	nec-essity	Vibhakti	Lex Type	posn	reln
ccof	m	0	avy	l	c
k4	m	ഉ് (u)	n	l	c
vmod	m	ആൻ (aan)	v	l	c
vmod-emph	m	ഉന്നത് (unnatu)	v	l	c

Verbframe of a similar verb in different sentences varies according to argument relations (Karaka relations) change. For example, for the verb "കാണാം" *kaaNaaM*, different frames are shown below.

Verb::കാണ് *kaaN*

SID::കാണ%VT%S1

Verb Sense::

Eng_Gloss::to see

Verb Class::

Verb_in_Same_Class::

TAM for the verb root::ആം *aaM*

Frames::

Example::1

വള്ളുവനാടൻ ഗ്രാമങ്ങളും നെൽവയലുകളും ഗത കാല പ്രൗഡിയോടെ നിൽക്കുന്ന മനകളും എങ്ങും കാണാം .

vaLLuvanaaTan graamangngaLuM nelvay-alukaLuM gatakaala prauDiyooTe nilkkunna manakaLuM engnguM kaaNaaM.

Valluvanadan villages, paddy fields and the abode houses of Malayali Brahmins with its historical pride can be seen.
FRAME_ID::1

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
k1	m	ഉം (uM)	n	l	c
adv	d	0	adv	l	c

Example::2
ഇവിടെ ഉള്ള വ്യൂപോയിന്റിൽ നിന്ന് ഷില്ലോങ്ങ് നഗരത്തിന്റെ , പ്രത്യേകിച്ച് രാത്രിയിൽ ദീപാലംകൃതം ആകുന്ന ഉജ്ജ്വലമായ കാഴ്ച കാണാം .
iviTe uLLa vyuupooyinRil ninnu Silloongng nagarattinRE , pratyeekiccu raatriyil diipaalaMkRItaM aakuna ujjvalamaaya kaazhcha kaaNaM.
From the view point here, the city of Shillong, especially the magnificent view of the city decorated with lights at night can be seen.
FRAME_ID::2

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
k5	m	ഇൽനിന്ന് (ilninn)	n	l	c
k1	d	0	n	l	c

Example::3
വിദേശത്ത് ചെന്നാൽ ശ്രീലങ്കയിലും ബാലിയിലും ഒക്കെ രാമയൻ സർക്യൂട്ടുകൾ കാണാം .
videeshattu cennaal sRIIlangkayiluM baaliyiluM okke raamayan sarkyuuTTukaL kaaNaaM.
If you go abroad, the Ramayan circuits can be seen especially in Sri Lanka and Bali.
FRAME_ID::3

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
vmod	m	0	v	l	c
k7p	m	ഇൽ (il)	n	l	c
k1	d	കൾ (kal)	n	l	c

Example::4
ബക്സറിൽ നിന്ന് അഞ്ച്കിലോമീറ്റർ വടക്കുകിഴക്കു മാറി അഹല്യാദേവിയുടെ അമ്പലം കാണാം .
baksaRil ninnu anjcu kiloomiiRRar vaTakkukizhakku maaRi ahalyaadeeviyuTe ampalaM kaaNaaM.

The temple of Ahalya Devi is located five kilometers north-east from Buxar.
FRAME_ID::4

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
k5	m	ഇൽനിന്ന് (ilninn)	n	l	c
vmod	m	ഇ (i)	v	l	c
k1	d	കൾ (kal)	n	l	c

Example::5
ഭാരത്തിൽ മൂന്നു പ്രധാന വിഭാഗങ്ങളിൽ പെട്ടവരെ കാണാം : ഇന്തോ - ആര്യൻ വംശജർ , ദ്രാവിഡ വംശ ജർ , മംഗോൾ - ആര്യൻ വംശജർ .
bhaaratattil muunu pradhaana vibhaagangngaLil peTTavare kaaNaaM : intoo-aaryan , draaviDa vaMshajar , maMgooL - aaryan vaMshajar.
There are three main groups in India: Indo-Aryan tribes, Dravidian and Mangol-Aryan tribes.
FRAME_ID::5

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
k1	m	അർഎ (are)	v	l	c
k1	d	0	n	r	c

Example::6
ആംഗികം , വാചികം , സാത്വികം , ആഹാര്യം എന്നീ അഭിനയരീതികൾ ഏറ്റക്കുറച്ചിലുകളോടെ മുടിയേറ്റിൽ കാണാം
aaMgikaM, vaacikaM, saatvikaM, aahaaryaM enni abhinjayariitikal eeRRakkuRaccilikaLoTe kaaNaaM.
Different styles of actings like Agikam, Vachikam, Satvikam, Aharyam are found in a ritualistic art form Mudiyettu.
FRAME_ID::6

arc_ label	nec- essity	Vibhakti	Lex Type	posn	reln
k1	d	കൾ *(kal)*	n	l	c
k4	d	കൾഓട്എ (kaLooTe)	n	l	c
k7	m	ഇൽ (il)	n	l	c

It is clear from the above example that verbframe of similar verb is different from the

other verb frames as the argument relations namely the Karaka relations are changing. In the present study, we have taken 3000 dependency annotated sentences for generating the verb frames. Verb frames for 255 verbs [2] were generated from these sentences.

There are some sentences which have 2 finite verbs. Such sentences are not considered in the present study. Examples for such a sentence:

അതിൻെറ അർഥം മനസ്സിലാക്കാൻ ആർക്കും നിയമ സഹായം തേടേണ്ടി വരും എന്നും തോന്നുന്നില്ല .

atinRe arthaM manassilaakkaan aarkkum niyama sahaayaM teeTeeNTi varuM ennuM toonnunnilla.

It does not appear to have any legal assistance to understand the meaning.

In the above sentence, the "എന്നും"(ennuM) is the connector. To this connector, the two finite verbs "വരും"" *varuM* and "തോന്നുന്നില്ല "*toonnunnilla* is joined in the dependency tree. The sentenence is annotated as follows and the dependency tree diagram is depiced in Figure 3.

5 Conclusion and Future Directions

5.1 Conclusion

Application of approach for generating verbframe developed by IIIT-H seems to be adopted for Malayalam languages. However we have to explore how does this can be made applicable for the sentences which have two or more finite verbs.

5.2 Future Directions

This work can be further extended to classify the verb frames according to the semantic nature of the verb. Also attempts can be made to extract the verb frames from dependency annotated corpora though some machine learning approaches.

Acknowledgement

References

Beth Levin. 1993. *English Verb lasses and Alternations. A Preliminary Investigation*, University of Chicago press.
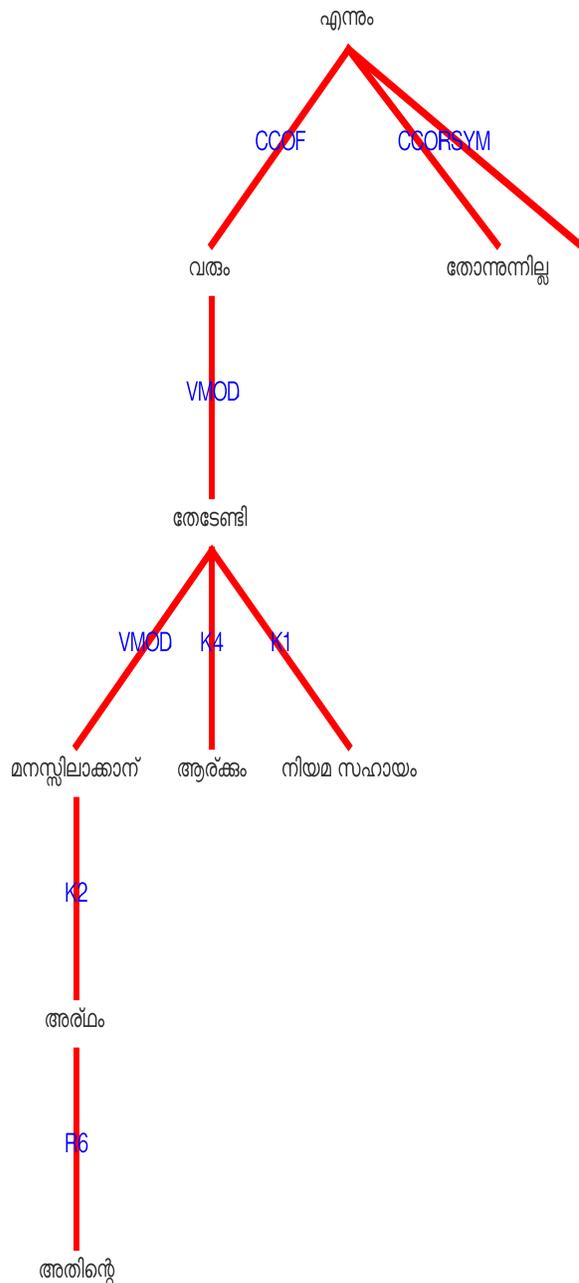
Figure 3: Dependency tree for given example

Beth Levin and Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface.*, MIT press. 26.

Karin Kipper Schuler. 2005. *VerbNet: A Broadcoverage, Comprehensive Verb Lexicon*, PhD dissertation, Computer and Information Science Department*, University of Pennsylvania.

---

[2]Only finite verbs are considered.

Karin Kipper, Hoa Trang Dang and Martha Palmer . 2000. *VerbNet: Class based construction of a verb lexicon*, In *AAAI/IAAI*, 691 696.

Christiane Fellbaum. 1998. *WordNet An Electronic Lexical Database*, MIT press.

Paul Kingsbury and Martha Palmer. 2003. *PropBank: the next level of treebank*, In *Proceedings of Treebanks and lexical Theories*, Vol. 3.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. *The proposition bank: An annotated corpus of semantic roles.*, In *Computational linguistics*, Vol. 31(1) , 71-106.

Paul Kingsbury and Martha Palmer. 2002. *PropBank: the next level of treebank*, In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pp. 1989-1993.

Rafiya Begum, Samar Husain, Lakshmi Bai, and Dipti Misra Sharma. 2008. *Developing Verb Frames for Hindi.*, In *LREC*.

Collin F Baker, Charles J Fillmore, and John B Lowe . 1998. *The Berkeley FrameNet Project*, In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* , Association for Computational Linguistics. 6:86-90.

Rafiya Begum and Dipti Misra Sharma. 2017. *Development and Analysis of Verb Frame Lexicon for Hindi.*, In *Linguistics and Literature Studies*. 5(1):1-22.

Rafiya Begum, Samar Husain, Arun Dhwaj, Lakshmi Bai, cc and Rajeev Sangal. 2008. *Dependency Annotation Scheme for Indian Languages.*, In *IJCNLP*. 721-726.

Sanjukta Ghosh. 2014. *Making Verb Frames for Bangla Vector Verbs.*, In *Proceedings of 11th Intl. Conference on Natural Language Processing.* 305314.

Ankush Soni, Sambhav Jain, and Dipti Misra Sharma. 2013. *Exploring Verb Frames for Sentence Simplification in Hindi.*, In *IJCNLP*. 1082-1086.

Hanumant Redkar, Sandhya Singh, Nandini Ghag, Jai Paranjape, and Nilesh Joshi. 2016. *Verbframator: Semi-Automatic Verb Frame Annotator Tool with Special Reference to Marathi.*, In *Proceedings of 13th International Conference on Natural Language Processing.* pp-299.

Schulte im Walde and Sabine. 2009. *The induction of verb frames and verb classes from corpora.*, *Corpus Linguistics. An International Handbook.* Mouton de Gruyter, Berlin.

Akshar Bharati, Vineet Chaitanya, Rajeev Sangal, and K V Ramakrishnamacharyulu. 1995. *Natural language processing: a Paninian perspective.* Prentice-Hall of India New Delhi.

Joakim Nivre. 2005. *Dependency grammar and dependency parsing.*, In *MSI report*. 5133(1959):1-32.

Akshar Bharati and Rajeev Sangal. 1993. *Parsing free word order languages in the Paninian framework*, In *Proceedings of the 31st annual meeting on Association for Computational Linguistics* Association for Computational Linguistics. 105-111.

Miriam Butt. 2010. *The light verb jungle: still hacking away.*, In *Complex predicates in cross-linguistic perspective* Cambridge University Press. 48-78.