

Guidelines for POS Tags in Malayalam

POS Tagging is the process of assigning a Part-of-Speech to each word in a sentence. It is the act of assigning each word in a sentence a tag that describes how that word is used in the sentences. It expresses the grammatical category of each word in a sentence.

POS Tagging is done for the disambiguation of the morphologically analyzed result of the input text.

The following are the POS Tags in Malayalam:

1. Noun - N
Noun can be Common Noun or Proper Noun.

1.1 Common Noun - N_NN

For Common Nouns, NN is the tag given. വീട്, വെള്ളം, മരം etc: are Common Nouns in Malayalam.

1.2 Proper Nouns - N_NNP

Place names and Personal names are included in Proper Nouns and the tag given is NNP.

മോഹനൻ, സീത, കൊല്ലം, ഡൽഹി etc are examples.

1.3 N loc - N_NST

Nouns denoting special and temporal expressions are NST in Malayalam.

These words often occur as temporal and special arguments of a verb in a given sentence, taking the appropriate vibhaktis (case markers).

Eg:

വീടിന്റെ house- of	മുന്നിൽ	'in front of'
“	പിന്നിൽ	'on the back'
“	ഉള്ളിൽ	'in side'
“	മുകളിൽ	'on the top'
“	താഴെ	'below'
“	പുറകിൽ	'on the back'
“	അകത്ത്	'in side'
“	പുറത്ത്	'out side'

മുന്നിൽ, പിന്നിൽ, ഉള്ളിൽ, മുകളിൽ, താഴെ, പുറകിൽ etc:
are NST s in Malayalam NST comes after a noun with a
Genitive marker.

2. Pronoun - PRP

All pronouns are marked as PRP. All 1st person, 2nd person
and 3rd person pronouns are included in this list.

2.1 Personal Pronouns - PR_PRP

In Malayalam ,the 1st person pronoun is ഞാൻ ' I ', the 2nd
person pronoun is നീ ' You' and 3rd person Pronouns are ഇവൻ '
He (Prox)', ഇവൾ ' She (prox)', ഇത് ' It (prox)' , അവൻ 'He
(Rem)', അവൾ 'She (Rem)' and അത് 'It (Rem)'.

2.2 Reflexive Pronoun - PR_PRF

The Reflexive pronoun in malayalam is താൻ.

2.3 Relative Pronoun - PR_PRL

The Relative pronouns in Malayalam are ആരോ 'somebody',

ആരും 'anybody', എന്തും 'anything' , എപ്പോഴും 'at any time' etc.

2.4 Reciprocal Pronouns - PR_PRC

The Reciprocal Pronouns in Malayalam are തമ്മിൽ 'together', തമ്മിൽ തമ്മിൽ 'to words each other' and പരസ്പരം 'each other'.

അവർ തമ്മിൽ വഴക്കിടുന്നു
'they' 'together' 'quarell'

' They quarell with each other '

അവർ തമ്മിൽ തമ്മിൽ അടിക്കുന്നു.
'they', 'each other' 'quarell'

' They quarell with each other '

അവൻ പരസ്പരം വഴക്കിടുന്നു
'they' 'together' 'quarell'

' They quarell with each other '

2.5 Wh – word - PR_PRQ

Question words are included in PR-PRQ. ആര് 'Who', എന്ത് 'what', ഏത് 'what', എപ്പോൾ 'when', എങ്ങനെ 'How', എത്ര 'How much' etc: are question words in Malayalam.

3. Demonstratives - DM

Demonstrative in Malayalam is 'CuTTezhuttu (cuuNTunna ezhuttu)' which means a category that is used to demonstrate something or somebody.

There is a clear distinction between a Pronoun and a Demonstrative in Malayalam.

See these examples:

ആ പെൺകുട്ടി രാജന്റെമകൾ ആണ്
'that' 'girl' 'Rajan-of daughter' 'is'

'That girl is Rajan's daughter'

അവൾ എന്റെ മകൾ ആണ്
'She' 'my' 'daughter' 'is'

'She is my daughter'

From the above examples, it is evident that ആ 'that' in the first sentence is a Demonstrative and അവൾ in the second sentence is a Pronoun in Malayalam.

4. Verb

4.1 Verb Main — V-VM

The Verbal constructions in Malayalam are composed of a Verb root and one or more Auxiliary suffixes. In our language, the Auxiliaries are not free forms, but bound forms.

See the examples below:

ഓടിക്കൊണ്ടിരിക്കുന്നു 'running'

This verbal form consists of the verb root ഓട് - 'to run' and the following suffixes.

ഓട് -	Verbal root
- ഇ -	Past tense
-കൊണ്ട് -	Progressive Aspect
-ഇരി -	Progressive Aspect
-ക്ക് -	Link Morph
-ഉന്ന -	Present tense

4.1.1 Finite Verb - V_VM_VF

In Malayalam, പറഞ്ഞു 'said', ഓടുന്നു 'running', പാടും 'will sing' etc: are Finite Verbs.

See the examples below:

രാമൻ ഓടുന്നു
'Ram' 'run-pres.'
'Ram runs'

അവൻ എന്നോട് അത് പറഞ്ഞു
'He' 'me -to' 'that' 'say -past'
'He said that to me'

സീത ഇന്ന് പാട്ട് പാടും
'Sita' 'today' 'song' 'sing-will'
'Sita will sing the song today'

The action of the verb will be complete in the Finite Verb form.

4.1.2 Non-Finite Verb - V_VM_VNF

In Non-Finite Verb forms, the action will not be complete. The Non-finite Verb groups in Malayalam can be of the following types:

a. Relative Participle forms.

The Relative Participle form modifies a noun. See the following examples:

അവൻ പറഞ്ഞ കാര്യം ശരി ആണ്
'he' 'said – which' 'matter' 'true' 'be – pres.'
'The matter he said is true.'

The 'അ' in പറഞ്ഞ is the RP marker. Here, പറഞ്ഞ 'which he said' is Non-Finite (VNF).

b. Nominalized Verb forms.

Sentences with Nominalized verb forms is commonly found in Malayalam.

See the examples below:

അവൾ പറഞ്ഞത് സത്യം ആണ്
'She' 'Said-that' 'true' 'be-pres.'

'The matter she said is true'

In the above example, പറഞ്ഞത് 'the matter said' is the Nominalized form. Here, the Neuter Gender marker -ത് of ആത് 'that' is suffixed to the verb to form a Nominalized entity which functions like a Noun. But they are capable of taking arguments.

Nominalized forms are used to give emphasis also. See the examples below:

ഞാൻ സീതക്ക് ആണ് 100 രൂപ കൊടുത്തത്
I Sita-to is 100 rupees gave-NOML

'It is to Sita that I gave 100 rupees'

ഞാൻ ആണ് 100 രൂപ സീതക്ക് കൊടുത്തത്.

I is 100 rupees Sita-to gave-NOML

'It is me who gave 100 rupees to Sita'

100 രൂപ ആണ് ഞാൻ സീതക്ക് കൊടുത്തത്

100 rupees is Sita-to gave-NOML

'It is 100 rupees that I gave to Sita'

Nominalized verb forms have the Masculine, Feminine and Neuter Gender markers as suffixes. If we split പറഞ്ഞത് into പറഞ്ഞ 'that-which is said' and അത് 'that', then the whole sentence would be unacceptable to the native speaker of Malayalam.

See another example,

അത് ചെയ്തത് രാമൻ ആണ്
that do-past-NOML Ram is

'It is Rama who did that'

The above one is an emphatic type of sentence, giving emphasis to രാമൻ 'Rama'. അത് ചെയ്തത് 'who did that' is a clause in this sentence and the whole clause is Karma also (K2). Here also, we cannot split അത് ചെയ്തത് into അത് ചെയ്ത അത്.

If we split like that, the sentence would be as seen below:

അത് ചെയ്ത അത് രാമൻ ആണ്
that who did that Rama is

The above shown is not an acceptable grammatical sentence in Malayalam.

We can add suffixes to these Nominalized forms as ചെയ്തതിനെ, ചെയ്തതിൽ etc: We can add postpositions also as ചെയ്തതിനെ പറ്റി.

Another peculiarity is that these Nominalized verb forms are capable of taking Arguments. So, considering their verbal nature, these Nominalized verbs would be considered as Non- finite forms - VNF.

Like ചെയ്തത്, we have in Malayalam,

ചെയ്തവൻ 'he who did' and ചെയ്തവൾ 'she who did' also with Masculine and Feminine Gender suffixes. These are also included in Non finite Verb forms- VNF.

c. Verbal Participle forms

VP forms are Non-finite verbs, before a finite verb.

രാമൻ ബസ്സിൽ ചാടി കയറി

Rama bus-in having jumped get in

'Having jumped,Rama got into the bus'

In the above example, ചാടി 'having jumped' is a Verbal Participle form which is a Non- finite verb form (VNF).

d. Adverbial clauses

Some Adverbial clauses can have RP + Time marker as structure.

രാമൻ വന്നപ്പോൾ രാജൻ പോയി

Rama came-past-when Rajan went

'When Rama came, Rajan went'

Here, വന്നപ്പോൾ 'when came' is Adverbial clause and it is marked as VNF.

e. Conditional clause

Conditional verb forms have - ആൽ 'if' as suffix.

നീ പഠിച്ചാൽ പാസ്സാകും

you study-if pass-will

'If you study, you will pass'

In the above sentence, പഠിച്ചാൽ 'if study', is conditional and labelled as VNF.

f. Adverbial participle of ആക്-

The adverbial participle form of ആക്- is ആയി and it modifies a Noun.

ഞാൻ അവൾക്കായി ഒരു സാരി വാങ്ങി
I She-to-Advp one saree bought

'I bought a saree for her'

In the above sentence, ആയി is an Adverbialized form of ആക്-, and it modifies the Argument ഒരു സാരി and so ആയി is VNF.

g. Relativized form of ഉണ്ട് and ആണ്.

ഉള്ള is the Relativized form of ഉണ്ട് and ആയ is the Relativized form of ആണ് 'be -pres.'

എനിക്കുള്ള വീട്
me-Dat-Adj1 house

'The house which I have'

The R.P form ഉള്ള is used to Adjectivize sentences with ഉണ്ട്.

എനിക്ക് വീട് ഉണ്ട് > എനിക്കുള്ള വീട്
I -to house be-pres. I- to which house

'I have a house' 'The house which I have '

The relativized Adjectives are Non-finite forms and would be marked as VNF.

4.1.3 Infinitives - VINF

- a. The Purposive Infinitive forms would be marked as VINF.

അമ്മ കുളിക്കാൻ പോയി
mother take bath_INFIN went

'Mother went to take bath'

Here, കുളിക്കാൻ is Purposive Infinitive form and it would be marked as VINF.

- b. Pure Infinitives.

Infinitive forms like ഓടുക 'running', പറയുക 'saying' are marked as VINF.

രാവിലെ ഓടുക പ്രയാസം ആണ്
morning running difficult be-pres.

'Running in early morning is difficult.'

4.2 Verbal Noun - V_VN

Verbal Nouns are nouns derived from verbs functionally these are nominals.

eg: വിൽക്കൽ, കൊയ്ത്ത്.
'selling' 'harvest'

എന്റെ കാർ വിൽക്കൽ നടന്നില്ല
my car sell-NOM happened-not

'My car selling didn't happen'

Here, വിൽക്കൽ is V_VN. It is derived from the verb വിൽക്കുക 'to sell'

ഇന്ന് വയലിൽ കൊയ്ത്ത് ആണ്
today field-in harvest be-pres.

'There is harvest in the field today'

In the above sentence, കൊയ്ത്ത് is V-VN. It is derived from the verb കൊയ്യുക 'to harvest'.

5. Adjective - JJ

Adjectives are modifiers of nouns.

വലിയ 'big', ചെറിയ 'small', പുതിയ 'new', നീല 'blue' etc. Are Adjectives in Malayalam.

സീത വലിയ വീട് വാങ്ങി
Sita big house bought

'Sita bought big house'

Modifiers like ഭംഗിയുള്ള 'beautiful', സുന്ദരമായ 'beautiful' are Adjectives.

Comparative and Superlative Adjectives are also included.

6. Adverb - RB

Adverbs modify verbs. The tag RB will be used ONLY for manner Adverbs. പെട്ടെന്ന് 'suddenly', പതുക്കെ 'slowly', ഉടനെ 'immediately' etc. are Adverbs in Malayalam.

അവൻ പെട്ടെന്ന് വന്നു
he suddenly came

'He came suddenly'

The tag RB will not be used for the Time and Space expressions in Malayalam, such as അവിടെ 'there', ഇവിടെ 'here', അപ്പോൾ 'that time', ഇപ്പോൾ 'this time', അന്ന് 'that dat' and ഇന്ന് 'today'. These will be marked as PR.

7. Postposition – PSP

Postpositions in Malayalam occur as free forms after the case suffixes of Nouns and Pronouns. There are a good number of Postpositions in Malayalam. പറ്റി 'about', കുറിച്ച് 'about', കൊണ്ട് 'with', നിന്ന് 'from', വരെ 'to' etc. are some of them.

8. Conjunctions - CC

The tag CC will be used for both Co-ordinating and Subordinating conjuncts.

8.1 Co-ordinator - CC_CCD

പക്ഷേ 'but', അഥവാ 'but' are co-ordinators in Malayalam.

ഞാൻ വിളിച്ചു പക്ഷേ അവൻ വന്നില്ല.

I called , but he didn't come

'I called, but he didn't come'

8.2 Subordinator - CC_CCS

എന്നാൽ 'but', അതിനാൽ 'therefore', എങ്കിൽ 'if', എന്നിട്ട് 'after that ' etc. are subordinate conjuncts in Malayalam.

നാളെ ഹർത്താൽ ആണ്, അതിനാൽ ക്ലാസ് ഇല്ല.

Tomorrow harthal is therefore class be-Neg.

'Tomorrow is a harthal day, therefore there will not be class'

8.2.1. Quotative - CC_CCS_UT

എന്ന് 'that', and എന്ന 'that' are Quotatives in Malayalam.

അച്ഛൻ ഓഫീസിൽ പോയി എന്ന് അമ്മ പറഞ്ഞു

Father Office-in went that mother said

'Mother said that father went to office'

The above sentence quotes a matter and so എന്ന് 'that' is quotative form.

ഗോപി മരിച്ചു എന്ന വാർത്ത ശരി ആണ്.
Gopi died that news true is
'The news that Gopi died is true'

In the above example also, എന്ന് 'that' quotes a sentence and so എന്ന് 'that' is a quotative form.

9. Particles - RP

മാത്രം 'only', കൂടി 'also', etc. are particles in Malayalam.

9.1 Default - RP_RPD

Expressions such as കൂടി 'also', മാത്രം 'only', പോലും 'even' etc., in Malayalam would be marked as RP_RPD.

ഞാൻ കൂടി അമ്പലത്തിൽ വരുന്നു
I also Temple-in coming
'I am also coming to the Temple'

സീത കടയിൽനിന്ന് സാരി മാത്രം വാങ്ങി
Sita Shop-from saree only bought
'Sita bought only a saree from the shop'

9.2 Classifier - RP_CL

A Classifier is a word or morpheme, used to classify a noun according to its meaning.

In Malayalam, തുടങ്ങിയ, എന്നീ, മുതലായ are classifiers.

രമ, സുമ, സീത എന്നീ കുട്ടികൾ വന്നില്ല

'The students such as Rema,Suma,Sita didn't come'

9.3 Interjection - RP_INJ

The Interjections in Malayalam are ഹാവു ! ഹൊ!,അയ്യോ etc.

9.4 Intensifier - RP_INTF

Intensifier intensifies Adjectives or Adverbs. It modifies another modifier. വളരെ 'very', ഏറെ 'more' etc: are intensifiers in Malayalam .

വളരെ നല്ല സാരി
very good saree

' Very good saree'

ഏറെ ഭംഗിയുള്ള ചുരിദാർ
more beautiful churidar

'More beautiful churidar'

10. Quantifier - QT_QTF

Quantifiers denote quantity. കുറച്ച് 'a littile', ധാരാളം 'many' etc., are quantifiers in Malayalam.

10.1 Cardinals - QT_QTC

A word denoting a Cardinal number will be tagged as QC.

സമരത്തിന് 100 അംഗങ്ങൾ വന്നിരുന്നു.
strike-for 100 members came

' 100 members came for the strike.'

ഞാൻ അവൾക്ക് 1000 രൂപ കൊടുത്തു.
I she-to 1000 rupees gave

'I gave 1000 rupees to her.'

10.2 Ordinals - QT_QTO

Words denoting ordinals will be marked as QO.

അവൻ മൂന്നാമത്തെ നിലയിൽ ആണ്.
He third f loor-in is

'He is in the third floor.'

വിനു ഒന്നാം ക്ലാസ്സിൽ പഠിക്കുന്നു.
Vinu first standard-in studying

'Vinu is studying in the First standard.'

11. Residuals - RD

11.1 Foreign Word - RD_RDF

If the word is a Foreign word, RD_RDF is the tag.

For example, white – house, American President etc:

11.2 Symbol - RD_SYM

All those words which cannot be classified in any of the other tags will be tagged as SYM. Special symbols like \$, %, & etc: are treated as SYM.

11.3 Punctuation - RD_PUNC

. , , , “ , ” etc: are punctuations.

11.4 Unknown - RD_UNK

Unknown words will be included in this tag.

11.5 Echowords - RD_Ech

ചടപട, ചറപറ etc: are ech-words in Malayalam.

