[Draft] Complete sinhala POS tag set

Contents

[Nouns (නාම පද) 2](#_Toc445912186)

[Common Noun Categorization 2](#_Toc445912187)

[Pronoun categorization 3](#_Toc445912188)

[Numeric 3](#_Toc445912189)

[Noun in Kriya Mula 3](#_Toc445912190)

[Proper Noun Categorization 3](#_Toc445912191)

[Verbs (ක්‍රියා පද) 7](#_Toc445912192)

[Finite Verbs (අවසාන ක්‍රියා) 7](#_Toc445912193)

[Non Finite Verbs (අනවසාන ක්‍රියා) 7](#_Toc445912194)

[Adjectives (නාම විශේෂණ) and Adverbs (ක්‍රියා විශේෂණ) 9](#_Toc445912195)

[Adjective in Kriya Mula 9](#_Toc445912196)

[Particles, Postposition, Conjunction and Interjection (නිපාත) 10](#_Toc445912197)

[Particle 10](#_Toc445912198)

[Postposition 10](#_Toc445912199)

[Conjunction 10](#_Toc445912200)

[Interjection 10](#_Toc445912201)

# Nouns (නාම පද)

There are three common noun types in languages and Sinhala is not different: Common nouns (සරල/සංයුක්ත නාම), Pro nouns (සර්ව නාම) and Proper nouns (සංඥා නාම). These three noun types in Sinhala are further categorized in different ways based on following factors.

1. ප්‍රාණ භේදය (Animacy)

Based on animacy, Sinhala nouns are categorized in to two areas. ප්‍රාණවාචී (Animate) and අප්‍රාණවාචී (Inanimate)

1. වචන භේදය (Number)

Based on number, Sinhala nouns are categorized in to two areas: ඒක වචන (Singular) and බහු වචන (Plural)

1. ලිංග භේදය (Gender)

Based on gender, Sinhala nouns are categorized in to two areas. පුරුෂ ලිංග (Masculine), ස්ත්‍රී ලිංග (Feminine)

1. නියත භේදය (Definiteness)

Based on definiteness, Sinhala nouns are categorized in to two areas. නියත (Definite) and අනියත (Indefinite)

1. කාර්ය භේදය (~Case : Nominative and Accusative)

Sinhala nouns can be categorized in to two areas based on the subject and object of a sentence: උක්ත (Nominative/Subjective) and අනුක්ත (Accusative / Objective)

1. විභක්ති (Case)

Sinhala nouns are inflated in four(4) ways based on case: සම්ප්‍රදාන (Dative : E.g. to man, to tree), සම්බන්ධ/සම්බන්ධාධාර(Genitive: E.g. man’s, on tree), අවධි/කරණාවධි(Instrumental: E.g. from man, from tree), ආලපන (Vocative – Words are not inflated for this case in English: E.g. John! Man! )

1. පුරුෂ භේදය (Person)

Sinhala nouns are categorized in to three areas based on Person. උත්තම පුරුෂ(First person), මධ්‍යම පුරුෂ (Second Person), ප්‍රථම පුරුෂ(Third Person)

## Common Noun Categorization (NN)

Common nouns in Sinhala are first categorized in to two broad areas based on animacy : Animate and Inanimate

Animate Common nouns in Sinhala are further categorized based on Number, Gender, Definiteness and Case (Nominative, Accusative, Dative, Genitive, Instrumental, Vocative).

Inflections of Animate Common Nouns

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Case | Singular - ඒක | | | | Plural - බහු | |  |
|  | Masculine - පුරුෂ | | Feminine - ස්ත්‍රී | | Masculine | Feminine | Common |
|  | Def  නියත | Indef  අනියත | Def | Indef | Indefinite | Indefinite |  |
| Nominative  කතෘ | මිනිසා  ළමයා | මිනිසෙක් | කාන්තාව  දැරිය | කාන්තාවක් | මිනිස්සු  ගොවියෝ  ළමයි | කාන්තාවෝ  අම්මාවරු  ස්ත්‍රීහු | දෙමවුපියෝ |
| Accusative  කර්ම | මිනිසා / මිනිසාව  ළමයා / ළමයාව | මිනිසෙකු | කාන්තාව  දැරිය | කාන්තාවක | මිනිසුන් | කාන්තාවන් | දෙමවුපියන් |
| Dative  සම්ප්‍රදාන | මිනිසාට | මිනිසෙකුට | කාන්තාවට | කාන්තාවකට | මිනිසුන්ට | කාන්තාවන්ට |  |
| Genitive  සම්බන්ධ | මිනිසාගේ | මිනිසෙකුගේ | කාන්තාවගේ | කාන්තාවකගේ | මිනිසුන්ගේ | කාන්තාවන්ගේ |  |
| Instrumental  අවධි | මිනිසාගෙන් | මිනිසෙකුගෙන් | කාන්තාවගෙන් | කාන්තාවකගෙන් | මිනිසුන්ගෙන් | කාන්තාවන්ගෙන් |  |
| Vocative  ආලපන | මිනිස! දරුව! | | දැරිය! | | මිනිසුනේ! | අම්මාවරුනේ! |  |

Animate common nouns sometimes belongs to a common gender which represent both masculine and feminine. (Eg: මිනිස්සු / දෙමවුපියෝ)

Inanimate Common Nouns in Sinhala are further categorized based on Number, Definiteness and Case (Direct, Dative, Genitive, and Instrumental).

Inflections of Inanimate Common Nouns

|  |  |  |  |
| --- | --- | --- | --- |
| Case | Singular | | Plural |
| Definite | Indefinite |
| Direct | පොත  ශාලාව | පොතක්  ශාලාවක් | පොත්  ශාලා |
| Dative | පොතට  ශාලාවට | පොතකට  ශාලාවකට | පොත්වලට  ශාලාවලට |
| Genitive | පොතේ / පොතෙහි  ශාලාවේ / ශාලාවෙහි | පොතක  ශාලාවක | පොත්වල  ශාලාවල / ශාලාවන්හි |
| Instrumental | පොතෙන්  ශාලාවෙන් | පොතකින්  ශාලාවකින් | පොත්වලින්  ශාලාවලින් |

## Pronoun categorization (PRP)

Pronouns are first categorized in to two broad areas based on animacy : Animate and Inanimate

Animate Pronouns in Sinhala are further categorized based on Animacy, Number, person, Gender and Case.

Inflection of Animate Pronouns

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Case | First Person | | Second Person |  | Third Person | | | |
| Singular | Plural | Singular | | |  | Plural |
| Masculine | | Feminine | Common |
| Nominative | මම | අපි | ඔබ | ඔහු | | ඈ / ඇය | තමා / ඌ | ඔවුහු |
| Accusative | මා | අප | ඔබ | ඔහු | | ඈ / ඇය | තමන් / උන් | ඔවුන් |
| Dative | මට | අපට | ඔබට | ඔහුට | | ඈට / ඇයට | තමන්ට | ඔවුන්ට |
| Genitive | මගේ / මාගේ | අපගේ / අපේ | ඔබගේ / ඔබේ | ඔහුගේ | | අගේ / ඈගේ / ඇයගේ | තමන්ගේ | ඔවුන්ගේ |
| Instrumental | මගෙන් / මාගෙන් | අපගෙන් / අපෙන් | ඔබගෙන් / ඔබෙන් | ඔහුගෙන් | | ඇගෙන් / ඈගෙන් / ඇයගෙන් | තමන්ගෙන් | ඔවුන්ගෙන් |

Inanimate Pronouns in Sinhala are further categorized based on Number.

Inflections of Inanimate pronouns

|  |  |  |
| --- | --- | --- |
| Case | Singular | Plural |
| Direct | මෙය  එය | මේවා  ඒවා |
| Dative | මෙයට / මීට | මේවාට |
| Genitive | මෙහි | මේවායේ / මේවායෙහි |
| Instrumental | මෙයින් / මින් | මේවායින් |

Apart from above mentioned pronouns there is an additional category for pronouns used for questioning.

Questioning Pronouns

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| කවුද | කවරක්ද | කුමක්ද | කොහේද | කවදාද |
| කොපමණද | කොතරම්ද | කී දෙනෙක්ද | කීයද |  |

Determiners are another set of pronouns which is used as a modifying word that determines the kind of reference a noun or noun group has.

Determiners

|  |  |  |  |
| --- | --- | --- | --- |
| මේ | ඒ | මෙම | බොහෝ |
| සමහර | කිසි | ඇතැම් | කවර |
| අන් | වෙන |  |  |

Examples

**ඇතැම්** දරුවන් ක්‍රීඩා කරති

අද උත්සවයක් පැවැත්විණි. **කිසිඳු** කෙනක් **ඒ** සඳහා පැමිණියේ නැත.

## Numeric

Numeric refers to a special category of nouns which are related to numbers. Table 4.2 provides some examples.

Numeric

|  |  |  |  |
| --- | --- | --- | --- |
| එක | එකට | එකම | එකක් |
| දුසිම | නවය හමාර | දෙදහස් දහසය | දහයෙන් එකක් |
|  |  |  |  |

## Noun in Kriya Mula (NVB)

There are some phrases in Sinhala which consists of a Noun and a Verb which are viewed semantically as a single unit and act as a verb. Noun in such a construct is known as Noun in Kriya Mula.

Noun in Kriya Mula

Combination of a noun and a verb which acts together as a verb

|  |  |  |  |
| --- | --- | --- | --- |
| පාඩම් කරනවා | සෙල්ලම් කරනවා |  |  |

Examples

ජන රැලි කීපයක්ම **වාර්තා** කිරීමට යොමු ව සිටි මාධ්‍යවේදීන්

මැතිවරණ නීති රීති **උල්ලංඝනය** කිරිම් වේ

රජය වෙත **දැනුම්** දුන් බවත්

පුවත්පත **වාර්තා** කරයි

## Proper Noun Categorization (NNP)

Proper nouns in Sinhala are further categorized based on Animacy and case. Table 5 summarizes the categorization

Inflection of Proper Nouns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Case |  | Direct | Dative | Genitive | Instrumental |
|  |  |  |  |  |  |
| Animate | Singular | නිමල් | නිමල්ට | නිමල්ගේ | නිමල්ගෙන් |
| Inanimate |  | රත්නපුරේ | රත්නපුරේට | රත්නපුරේ | රත්නපුරෙන් |
|  | සිංහල | සිංහලට | සිංහලේ / සින්හලෙහි | සිංහලෙන් |
|  | රුපියල / රුපියල් | රුපියලට | රුපියලේ / රුපියලෙහි | රුපියලෙන් |
|  | බෞද්ධ |  |  |  |
|  |  | ලංකාව |  |  |  |
|  |  |  |  |  |  |

Proper nouns are only categorized as animate and inanimate. Spoken Sinhala has a plural like form for animate nouns (Eg: නිමල්ලා) but this is not taken in to consideration.

# Verbs (ක්‍රියා පද)

Sinhala verbs can be mainly categorized in to two types based on the place of usage in a sentence: whether it appears as \_\_\_\_\_\_\_\_\_\_\_ (ආඛ්‍යාතය) or in any other place. Verbs used as \_\_\_\_\_\_\_\_\_\_\_\_ (ආඛ්‍යාතය) is known as Finite Verbs (අවසාන ක්‍රියා) where others are known as Non Finite Verbs (අනවසාන ක්‍රියා)

## Finite Verbs (අවසාන ක්‍රියා) (VFM)

Finite verbs are further categorized based on Person (පුරුෂ), Number (වචන), Tense (කාල) and Gender (ලිංග).

Inflection of Finite Verbs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | Present/Future | Past |
| First Person  ප්‍රථම | Singular | | බලමි  බලන්නෙමි | බැලීමි  බැලුවෙමි |
| Plural | | බලමු  බලන්නෙමු | බැලීමු  බැලුවෙමු |
| Second Person  මධ්‍යම | Singular | | බලහි  බලන්නෙහි | බැලීහි  බැලුවෙහි |
| Plural | | බලහු  බලන්නෙහු | බැලීහු  බලුවෙහු |
| Third Person  උත්තම | Singular | Common | බලයි | බැලීය  බැලිණ |
| Masculine | බලන්නේය | බැලුවේය |
| Feminine | බලන්නීය | බැලුවාය |
| Plural | | බලති  බලන්නෝය  බලන්නාහ | බැලුහ  බැලුවෝය  බැලුවාහ |

There is another category of finite verbs used for commanding. Table 6.1 provides examples

Finite Verbs Commanding

|  |  |  |  |
| --- | --- | --- | --- |
| දෙනු/ දෙන්න | බලනු/ බලන්න | කරනු / කරන්න |  |

# 

## Non Finite Verbs (අනවසාන ක්‍රියා)

All other types of verbs are non finite verbs which is used in different parts of sentences. Non finite verbs can further categorized in to three sections based on its usage.

**Verb participle** is a form of a verb that is used in a sentence to modify a noun, noun phrase, verb or a verb phrase.

Non Finite verb participle (VP)

Verb which modifies a noun, noun phrase, verb or a verb phrase. Acts as an adjective or adverb.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| වූ | කල | කරන | සිටි | දුන් |
| බලන | ගන්නා | ලැබෙන | පවතින | එන |
| කියන |  |  |  |  |

**Verbal Non Finite Noun** is a form of a verb that is used basically in the place of a noun and behave as a noun

Verbal Non Finite Noun (VNN)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| බැලීම් | බැලිලි | කිරීම | කිරීමට | ගැනීම |
| වීමෙන් | දීමට |  |  |  |

Other (**Verb Non Finite**) are all other non-finite nouns

Following is a list of examples for above two categories

Verb Non Finite (VNF)

|  |  |  |  |
| --- | --- | --- | --- |
| බලමින් | බලද්දී | බලනු | බලන්නට |
| බලතොත් | බලතත් | බලත්ම | බලත |
| බලන්නේ | බැලුවොත් | බැලුවත් | බලන්නාක් |
| බැලුවාම | බලනවාට/ බැලුවාට | බලනවා / බැලුවා | බලනවායින් |
| බලන්නාවූ | බලා | බැලූ |  |
|  |  |  |  |

# Adjectives (නාම විශේෂණ) and Adverbs (ක්‍රියා විශේෂණ)

Adjectives and adverbs in Sinhala is similar to any other language. Adjective is a word that used to describe a Noun and will come before the noun. Similarly adverb is a word that describe the verb and will come u the verb in a sentence.

Adjectives (JJ)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| රළු | සුමුදු | හොඳ | විශේෂ | කුඩා | විශාල |
| ප්‍රධාන | පැරණි |  |  |  |  |

## Adjective in Kriya Mula (JVB)

There are some phrases in Sinhala which consists of a Adjective and a Verb which are viewed semantically as a single unit and act as a verb. Adjective in such a construct is known as Adjective in Kriya Mula. Table 7.1 provides some examples.

Adjective in Kriya Mula

|  |  |  |  |
| --- | --- | --- | --- |
| කීකරු වෙනවා | එකඟ වෙනවා | අඩු කරනවා |  |
|  |  |  |  |

Examples

සුනඛයින් විසින් පන්ති කාමර **අපවිත්‍ර** කිරීම නිසා

වෙන්ව යෑමේ යෝජනාවට මම **එකඟ** වෙමි

5% දක්වා **පහත** වැටී ඇති අතර

විශ්ව විද්‍යාල සමයේ **කැපී** පෙනෙන ශිෂ්‍යයකු නොවූ

## Adjectival Noun (JJN)

When a noun is used to describe another noun, first noun acts as an adjective. In Sinhala these category of words can either act as a noun or an adjective based on the context.

Adjectival Noun

|  |  |  |  |
| --- | --- | --- | --- |
| ඇපල් | කෙස් | වීදුරු | කුරුළු |
|  |  |  |  |

Examples

සුනඛයින් විසින් **පන්ති** කාමර අපවිත්‍ර කිරීම නිසා

**මැතිවරණ** ප්‍රචණ්ඩ ක්‍රියා හා නීති රීති කඩකිරීම් සම්බන්ධයෙන් **පොලිස්** මූලස්ථානය වෙත සිකුරාදා උදෑසන වන විට ලැබී තිබූ මුළු පැමිණිලි සංඛ්‍යාව

ඉතාමත් සාර්ථක **චිත්‍ර** ප්‍රදර්ශන කිහිපයක් ම පැවැත්වූවෙකි

සරණාගතයන් රැගෙන එන **නීති** විරෝධී භාණ්ඩ

Adverbs (RB)

Adverbs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| වහා | සෙමින් | තවම | නැවත | විශේෂයෙන් | යලි |
| දැනට |  |  |  |  |  |

# Particles, Postposition, Conjunction and Interjection (නිපාත)

Sinhala has a word class called Nipatha which represent words that ***does not have a meaning by its own and does not get inflected by gender, number, tense, case or person***. Nipatha has different types of uses based on how and where it is used in a sentence. Even though there are no proper rules in Sinhala for further categorization of this word class, difference authors have their own categorizations.

We propose the following categorization based on available Sinhala POS tag set and expert opinion on applicability of this tags in Sinhala language.

## Particle (RP)

A word or a part of a word that has a grammatical purpose but often has little or no meaning

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ද | ය | නම් | ව | කොට | නො | වත් |
| පවා | ඊට |  |  |  |  |  |

## Postposition (POST)

A word used after a noun, pronoun or a verb in order to show the relation with other words or phrases of the sentence

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| සඳහා | අතර | ලෙස | සිට | වැනි | තුල |
| විසින් | ගැන | මගින් | දක්වා | තුලින් | හෙවත් |

Examples

දිවා ආහාරය **සඳහා** තරඟය නතර කර ඇත

දෙපිරිසක් **අතර** හට ගත ගැටුම නිරාකරණය කරන **ලෙස** දැනුම් දී තිබිණි.

වරද තමන් **පිට** වැටෙනවාට ඔහු කැමති නැත

ජයග්‍රහණය **කරා** ලඟා වීමට ඔහු වසරක් **පුරා** මහන්සි ගත්තේය

## Conjunction (CC)

A word used to connect clauses or sentences or to coordinate/subordinate words/phrases in the same clause

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| හා | හෝ | සමග | නමුත් | එහෙත් | තෙක් |
| දක්වා | පටන් |  |  |  |  |

## 

Examples

බැංකු **හා** වෙළඳ සල් එදින විවුර්ත කර තිබිණි

මාලා **සමඟ** කමලාත් පාසල් ගියාය.

ගුරුවරුන් **හෝ** ශිෂ්‍යයින් එදින පාසල් පැමිණියේ නැත

ඔහු ක්‍රීඩා කල **නමුත්** ජයග්‍රහණය ලැබීමට නොහැකි විය

ඔහු පාසල් ගියේ **යැයි** මම දැන සිටියේ නැත

## Interjection (UH)

Word or sound that used to express a particular emotion or sentiment

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| අහෝ! | අහ්! | ෂික් |  |  |  |

## Case Marker (CM)

A word that is used to mark the case of the previous word

|  |  |  |
| --- | --- | --- |
| Case | Case Marker | |
| Direct |  |  |
| Dative | ට | වලට |
| Genitive | ගේ | වල |
| Instrumental | ගෙන් | වලින් |