# Guidelines for Annotators – V 1.1

## Basic PoS tags in Sinhala

1. Nouns

2. Verbs

3. Adjectives

4. Adverbs

5. Nipātha

## First Level Extended Tag Set for Sinhala

1. Common Noun (NNC) - සරල / සංයුක්ත නාම පද

Examples: මිනිසා, ලේකම්, පාසැල, ආයතනය, සභාව, දැනුම, පදනම, මිනිසෙක්, මිනිසෙකු, මිනිසාගේ, මිනිසාට, මිනිසාගෙන්, මිනිසුනේ, පාසලට, පාසලෙහි, පාසලෙන්, පාසලක්, එරට, මෙරට.

Example usage: ප්‍රාදේශීය **සභාවේලේකම්**, පාසැල් **ගොඩනැගිල්ල** පරීක්ෂා කරයි.

1. Pronoun (PRP) - සර්ව නාම

Examples: මම, අපි, ඔහු, ඇය, ඈ, තමා, ඌ, ඔවුන්, මට, මගේ, මගෙන්, එය, මෙය, මේවා, මෙයට, මෙයින්, මෙහි, දෙය, දෙයක්

My examples: එතෙර, මෙතෙර, අතන, මෙතන

Example Usage: **ඔහු** නව පාසල් ගොඩනැගිල්ල විවෘත කළේය. **එය** ඉතා අලංකාර ගොඩනැගිල්ලක් වන අතර **එයින්** පාසලට නව පෙනුමක් ලැබී ඇත.

1. Proper Noun (NNP) - සංඥා නාම

Examples: නිමල්, විමලා, ලංකාව, මොරටුව, ඇමරිකා එක්සත් ජනපදය, රුපියල, සිංහල, නිමල්ට, නිමල්ගේ, නිමල්ගෙන්, රුපියල්, රුපියලට, රුපියලෙන්

1. Determiner (DET)

Determines the kind of reference a noun or noun group has.

Examples: මේ, ඒ, මෙම, බොහෝ, එක්, සමහර, ඔය, යම්, අර, ඇතැම්, ටික, සියලුම, මුළු, කවර, හැම, අන්, කිසි,තව, වෙන, කිසිම, වෙනත්, අනෙක්, කිසිඳු, මුළු, සෑම, එම, කිසියම්, අනෙක්,

My examples: මෙවැනි, කීපයකින්, වම්, දකුණු,

Example Usage: **මෙම** පාසැල ගමේ **අනෙක්** පාසල් අතර කැපී පෙනෙයි.

**ඇතැම්** දරුවන් ක්‍රීඩා කරති

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

1. Deterministic Pronoun (NDT)

These are words of which meaning include a determiner and a pronoun, compounded as a single word

Examples: ඇතැමෙක් (ඇතැම්+අයෙක්), සමහරෙක් (සමහර+දෙනෙක්), කිසිවෙක් (කිසි+කෙනෙක්), සමහරක් (සමහර+දෙයක්), කිසිවක් (කිසි+දෙයක්)

1. Questioning Pronoun (QUE)

Examples: කුමක්ද, කෙසේද, කවදාද, කවරක්ද, කවරෙක්ද,

1. Question Base Pronoun(QBE)

කිම, කිමෙක්, කුමන, කුමක්, කුමකට, කුමකින්, කුමක,කවර, කවරක්, කවරකට, කවරකින්, කවරක, කවරෙක්, කවරෙකුට, කවරෙකුගෙන්, කවරෙකුගේ,

1. Verb Finite (VFM) - අවසාන ක්‍රියා

Verbs used at the end of a sentence.

Examples: බලමි, බලමු, බලති, බලන්නේය, බලන්නීය, බැලුවේය, බැලූහ, බැලිණ, බලන්න, බලනු.

1. Verb Participle (VP)

Examples: බැලූ, බලන, බැලුවේ, බැලිය, බැලෙන, බලන්නේ, **~~බැලීමේ~~**, බලනු

My examples: බැලුව, බැලුවද

Example Usage:

සතියක් තුල අත්හදා **බැලූ** දෙවැනි මිසයිලය මෙයයි

එබී **බලන** මිනිසා

සොයා **බැලිය** යුතු බව

1. Verb Non Finite (VNF)

Examples: බලා, බලමින්, බලද්දී, බලතොත්, බැලුවාම, බැලුවත්

My examples: පූර්ව ක්‍රියා, මිශ්‍ර ක්‍රියා, අසම්භාව්‍ය ක්‍රියා

Example Usage:

තම නිවෙස් **බලා** යාමට

පිට වීමට මාන **බලමින්** සිටි

1. Verbal Nouns (VNN)

Examples: බැලීමට, බැලීමටත්, බැලීමේ, බැලීමෙන්, බැලීම, බැලීමට, බැලීමක්, බැලුම්, බැලිලි

My examples: බලන්නා

Example Usage:

සිය මව **බැලීමට** බුරුමයට පැමිණි විට

පස් මහා **බැලුම්** බලා

1. Sentence Ending (NVB)

Conventional sentences ends with a verb. But there are words in sinhala which are NOT verbs, but are used to end a sentence. Such words are tagged as **Sentence Ending**

Examples: පොතකි, ගසකි, ශාලාවකි, ගෙයකි, විශේෂිතයි, පමණි, අවශ්‍යයි

Example Usage: ඔහු කියවා ඇත්තේ **පොතකි**.

එය ඉතා **විශේෂිතයි.**

1. Adjective (JJ) - නාම විශේෂණ

Examples:

* රළු, සුමුදු, වැඩි, හොඳ, විශේෂ, අලුත්, සුදු, අඩු, විශාල, කුඩා, ඉහල, දිග, පුංචි, ලොකු, අධික, දියුණු, සුළු, යට, සුදුසු, උඩ, මහා, උසස්, ප්‍රධාන, නියම, මුලික, තරුණ, පහළ, ප්‍රකට, මහත්, පැරණි, නිසි, මානව, විවෘත
* පළමු, දෙවන, තෙවන
* ප්‍රාදේශීය, නාගරික, ළමා, අධ්‍යාපන

1. Adjectival Noun (NNJ)

A noun, used to describe another noun. This same word can be used as a noun itself depending on the context.

Examples with usage: චිත්‍ර පොත, ගල් කණු, පාසල් වේලාව,

Example Usage: ප්‍රාදේශීය සභාවේ ලේකම්, **පාසැල්** ගොඩනැගිල්ල පරීක්ෂා කරයි.

TIP: Adjectival Nouns always take the base form of the noun. In Sinhala, base form of all countable nouns are plural.

Countable Noun Base Forms: චිත්‍ර, පාසල්, ගල්, ඇත්

Mas Noun Base Forms: වතුර, දැනුම, අධ්‍යාපන

TIP: **Adjective or Adjectival Noun??**

Adjectives (JJ) can only be used as adjectives regardless of context. Whereas Adjectival Nouns (NNJ) can be used as nouns (NNP, NNC) alone, without any change to the word.

Ex: **Adjectival Noun**

පාසැල් can be an Adjectival Noun or Common Noun.

මම **චිත්‍ර\_NNJ**පොතක **චිත්‍ර\_NNC**අන්දෙමි.

1. Adverb (RB): ක්‍රියා විශේෂණ

Examples: වේගයෙන්, සෙමින්, උද්යෝගයෙන්, එකට, එසේ, මෙසේ, කෙසේ, එබැවින්, දැන්, තවමත්, නැවත, නැවතත්, කලින්, එහෙයින්, එසේම, මෙතෙක්, එහෙම, විශේෂයෙන්, යලි, දැනට, වහා, දැනටමත්

My examples: පසුගිය, නියමිත, තුරු

**Note: Compound Verbs**

There are some **Verbs** in Sinhala which are **compounds of two words**. In such situations, second word when taken alone is always a Verb whereas first word can be a Noun, Adjective or a word which does not belong to any other form.

First word of such compounds are categorized for three tags as below.

1. Noun in Compound Verbs (NCV)

පාඩම් කරනවා

සෙල්ලම් කරනවා

වර්ග කරනවා

ප්‍රතිඥා දෙනවා

බලාපොරොත්තු වෙනවා

My examples: ඉදිරියෙන් සිටිනවා

1. Adjective in Compound Verbs (JCV)

අඩු කරනවා/ කල/ වූ

වැඩි කරනවා

සමාන කරනවා

ක්‍රියාකාරී වෙනවා

අඩපන වී

අසාර්ථක වීමේ

My examples: අවසන් කරනවා

1. Preposition in Compound Verbs (PCV)

ඉටු කරනවා/වෙනවා/ කල/ වූ

සිදු කරනවා

අත් වෙනවා

My examples: අවතීර්ණ වෙනවා, බඳවා ගන්නවා, පුද කරනවා

TIP: **Noun/Adjective/ Preposition in Compound Verb** is always followed by a Verb type (VFM, VNF, VP or VNN)

1. Nipātha Conjunction (CC)

Words used to combine two words or phrases to build more complex sentences

Examples: හා, හෝ, නමුත්, යැයි, යයි, දැයි, එහෙත්, ඒත්.

1. Nipātha Postposition (POST)

Used **after** a Noun, Noun phrase, Verb or a Verb phrase to show the relationship with other words or phrases in the sentence

Examples: සඳහා, අතර, ලෙස, සිට, වැනි, තුල, විසින්, අනුව, ගැන, මෙන්, මත,පිලිබදව, වෙත, සේ, මිස, උදෙසා, සම්බන්දයෙන්, මගින්, දක්වා, තුලින්, පිණිස, යටතේ, පිට, වෙනුවෙන්, අතරින්, ඔස්සේ, පටන්, හරහා, හට, හැර, මෙන්ම, වෙනුවෙන්, වෙනුවට, තෙක්, පරිදි, හෙයින්, හෙවත්, බැහැර, හැටියට, පුරා, කරා, සමග, එක්ක, ආදී,පමණක්, සහිත, රහිත , සහිතව, රහිතව, Number + වන/වැනි/වෙනි.

My examples: පහළ, සතු

1. Nipātha Particle (RP)

Nipaath which does not play the role of conjunction or postposition.

Examples: ය, ම, ද, දී, යි, ලු, ඉතා, වඩා, වඩාත්, ක

My example: ත්, මහත්

1. Nipātha Interjection

Examples: අහෝ, චිහ්, ෂික්, අහ්

1. Nipathana (NIP)

Examples: ඇති, නැති, ඕනෑ, එපා, බැරි, පුළුවන්, ඇතිව, නැතිව, ඇත්නම්.

1. Case Marker (CM)

There are set of suffixes in Sinhala, used to represent case (විභක්ති). According to Sinhala experts, it is wrong to write this suffixes as separate words. But in reality some Sinhala writers tend to do this

Following are two ways of writing, where correct way is underlined

**නිමල්ට**/ නිමල් ට , **නිමල්ගේ**/ නිමල් ගේ, **පොත්වලට**/ පොත් වලට

Since we need to handle both of these inputs, a POS tag called ‘Case Marker’ is introduces to tag the suffix, if written separately.

Example: ට, ගේ, ගෙන්, වලට, වලින්, වල, හිදී,

My examples: න්

1. Number Quantifier (NUM)

Example: එක, දෙක, 128, 1.28, එක්

1. Modal Auxiliary (AUX)

Example: හැකි, යුතු, නොහැකි

1. Abbreviation (ABB)

Example: අ පො ස

1. Unknown (UNK)

Use this tag if you think a particular word cannot be categorized under any of the above tags. Unknown words require further study.

1. Full Stop / Period (FS)
2. Punctuation (PUNC)

TIP: **Full Stop or Punctuation?**

Tag “Full stop / Period” is to mark the end of sentence. When “.” Is used for other purposes (Ex: 22.34, අ. පො. ස., ඒ. බි. සී. පෙරේරා), it should be tagged as punctuation

1. Foreign words (FRW)