

Comprehensive Multi Level POS Tag Set for Sinhala

This document presents a comprehensive, multi-level POS tag set for Sinhala. Examples are given for each tag with their (most suitable) English translation whenever possible.

1. Level 1

Sinhala words are broadly categorized in to five part of speech at level 1

1. Nouns
2. Verbs
3. Adjectives
4. Adverbs
5. Nipāta

2. Level 2

At level 2, each tag at level 1 is further categorized based on context definitions.

2.1 Categories of Nouns

2.1.1 Common Nouns

Common Noun denote a class of objects or a concept. At second level of tag set, both animate and inanimate common nouns and their inflections belongs to the same category.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| මිනිසා | the man | මිනිස්සු | men | ගස | The tree | ගස් | trees |
| මිනිසෙක් | A man | මිනිසුන් | men | ගසක් | A tree | ගස්වල | In/on trees |
| මිනිසාගේ | Mans | මිනිසුන්ගේ | Men's | ගසෙහි | In/on the tree | ගස්වලින් | From trees |
| මිනිසාගෙන් | From the man | මිනිසුන්ගෙන් | From men | ගසෙන් | From the tree | ගස්වලට | To trees |
| මිනිසාට | To the man | මිනිසුන්ට | To men | ගසට | To the tree | | |

2.1.2 Proper Nouns

Proper noun identifies an exact entity (person, place or thing). Since it refers to an exact known entity, proper nouns cannot have an indefinite form.

| Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|
| ජෝන් | John | ජපානය | Japan |

| | | | |
|----------|-----------|----------|------------|
| ජෝන්ට | To John | ජපානයට | To Japan |
| ජෝන්ගේ | John's | ජපානයේ | In Japan |
| ජෝන්ගෙන් | From John | ජපානයෙන් | From Japan |

2.1.3 Pronouns

Pronouns are words that can be substituted for a noun or a noun phrase.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|----------------|---------------------|--------------|---------------------|--------------|---------------------|
| මම | I | අපි | we | ඔහු | he | ඔවුන් | them |
| මට | To me | අපිට | For us | ඔහුට | To him | ඔවුන්ට | To them |
| මගේ | mine | අප/අපේ | our | ඔහුගේ | his | ඔවුන්ගේ | theirs |
| මගෙන් | From me | අපෙන් / අපගෙන් | From us | ඔහුගෙන් | From him | ඔවුන්ගෙන් | From them |
| එය | that | එහි | In/on that | එවා | those | එවායේ | In those |
| එයින් | From that | | | එවායින් | From those | | |
| එයට | To that | | | එවාට | To those | | |

2.1.4 Nouns in Compound Verb

If a compound verb is a combination of Noun + Verb, such nouns are identified as Noun in Compound verbs. Noun in compound verb is always the base, plural form of the particular noun. There is no corresponding translation in English, since all compound verbs in Sinhala is a normal verb in English

Examples of compound verbs. First word of following compound verbs are identified as 'Noun in Compound Verb'

| | | | |
|---------------|----------------|---------------|---------|
| පාඩම් කරනවා | studying | සෙල්ලම් කරනවා | playing |
| ක්‍රියා කරනවා | Acting / doing | | |

Examples of Noun in compound verb

| | | | |
|-------|---------|---------|--|
| පාඩම් | සෙල්ලම් | ක්‍රියා | |
|-------|---------|---------|--|

2.1.5 Questioning Pronouns

Questioning Pronouns are words used to ask a question.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|------------------|---------------------|----------------|---------------------|--------------------|---------------------|
| කුමක්ද/කුමක්දැයි | What | කෙසේද/කෙසේදැයි | How | කවදාද/ කවදාදැයි | When |
| කොහෙදීද | Where | කොතනදීද | Where | කවරෙක්ද | Who |
| කවරක්ද | What | | | | |
| | | | | | |

2.1.6 Deterministic Pronouns

Deterministic pronouns are words built up from a combination of a determiner (discussed below) and a pronoun. For example සමහරෙක් (some of them) is a word in Sinhala derived from සමහර (some), which is a determiner and දෙනෙක් (them), which is a pronoun.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|-------------------|---------------------|--------------|---------------------|--------------|---------------------|
| සමහරෙක් / සමහරුන් | Some of them | ඇතැමෙක් | Some of them | සමහරක් | Some of those |
| සමහරුන්ට | To some of them | සමහරුන්ගේ | Some of their | සමහරුන්ගෙන් | From some of them |

2.1.7 Question Based Pronouns

Question base pronouns are used to show the uncertainty of a noun/noun phrase of interest.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| කුමක් | What | කවර | Which | කෙසේ | How |
| කොහේදී | Where | | | | |

2.2 Categories of Verbs

2.2.1 Verb finite

Finite Verbs are verbs used at the end of a sentence.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|----------------|---------------------|--------------|---------------------|
| කළේය | (he) did | කළාය | (she) did | කරන්නේය | Is doing |
| තිබේ | Have/has | තිබුණි / තිබිණ | had | කරමි | Am doing |
| කරති | Are doing | කළෝය | (they) did | | |

2.2.2 Verb participle

Inflected form of a verb that is used in a sentence to modify a noun, noun phrase, verb or a verb phrase. Thus it plays the same role as adjective or adverb.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| බැලූ | looked | බැලිය | | බැලෙන | |
| බලන | Is looking at | බලන්නේ | | බැලීමේ | |
| බලනු | | බැලුවේ | | | |

2.2.3 Verbal noun

Inflected form of a base verb that acts as a noun.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| බැලීම | The look | බැලිලි | looks | බැලීමෙන් | From the look |
| බැලීමක් | A look | බැලීමට | To look | බැලීමටත් | Even to look |

2.2.4 Modal auxiliary

Verbs which expresses necessity or possibility.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| හැකි | Can / could | යුතු | Must /should | නොහැකි | Could not / can not |

2.2.5 Verb Non Finite

Verbs that does not belong to any of the above categories are considered as verb non finite

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|------------------------|--------------|------------------------|
| බලා | Looked and | බලමින් | While looking | බලද්දී | when looking at |
| බලනොත් | If looked | බලුවාම | When (someone) look at | බලුවත් | Even (someone) look at |

2.3 Categories of Adjectives

2.3.1 Adjective

Adjectives are words used to describe nouns or noun phrases. Whole purpose of an adjective is to describe a noun and could not be used as any other word form.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| රළු | rough | විශේෂ | Special | සුදු | White | ඉහල | Up |
| සුමුදු | Smooth | අලුත් | New | කුඩා | Small | පහල | Down |
| ප්‍රාදේශීය | regional | නාගරික | urban | පළමු | first | දෙවන | secons |

2.3.2 Adjectival Noun

Adjectival noun is a common noun that acts as an adjective to describe another noun. When a common noun is used as an adjectival noun, it always take the base, plural form of the common noun.

So in a noun phrase like ‘පාසල් වත්ත’, ‘පාසල්’ is an adjectival noun which describes the main common noun ‘වත්ත’

| | | | | | | | |
|--------|-------|-----|-----|---------|-----------|------|----------|
| විත්‍ර | පාසල් | ගල් | ඇත් | කාර්යාල | ගොඩනැගිලි | වැව් | විද්‍යාල |
|--------|-------|-----|-----|---------|-----------|------|----------|

| | | | | | | | |
|--|--|--|--|--|--|--|--|
| | | | | | | | |
|--|--|--|--|--|--|--|--|

2.3.3 Adjective in Compound Verbs

If a compound verb is a combination of Adjective + Verb, such adjectives are identified as Adjective in Compound verbs.

Compound verb examples. First word in such compound verbs will be tagged as Adjective in compound verbs

| | | |
|-----------------------|------------|------------|
| අඩු කරනවා | වැඩි කරනවා | අඩපන වෙනවා |
| Reducing / decreasing | increasing | |

2.4 Categories of Adverbs

Adverbs are not further sub categorized in second level

2.5 Categories of Nipatha

2.5.1 Postposition

Postpositions in Sinhala are words used after nouns, verbs and sometimes even after adjectives and adverbs to show their relationship to other words in order to build up a meaningful sentence.

| | | | | | | |
|------|-----|--------|-----|-------|--------|-------|
| සඳහා | අතර | ලෙස | සිට | වැනි | තුලින් | දක්වා |
| මෙන් | මත | පිලිබඳ | මිස | උදෙසා | පරිදි | තෙක් |

2.5.2 Conjunction

Conjunctions are words used to connect words, phrases or sentences.

| | | | | | | |
|----|----|-------|-----|------|-------|-----|
| හා | හෝ | නමුත් | යයි | දැයි | එහෙත් | ඒත් |
|----|----|-------|-----|------|-------|-----|

2.5.3 Particle

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

| | | | | | | | |
|-------|---|---|----|----|---|-----|-----|
| ය | ම | ද | දී | යි | ල | ඉතා | වඩා |
| වඩාත් | ක | | | | | | |

2.5.4 Interjection

Interjections are words used to show the emotion or feeling.

| | | | |
|-----|------|-------|--|
| අහෝ | මික් | අහ්හ් | |
|-----|------|-------|--|

2.5.5 Determiner

Determiners are words that are used before a noun to show which particular example of the noun is referred to.

| Sinhala Word | English Translation | Sinhala Word | English Translation | Sinhala Word | English Translation |
|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| මෙම | This | මේ | This | එක් | One |

| | | | | | |
|-------|------|-------|------|--------|------|
| සියලු | All | බොහෝ | Most | සමහර | Some |
| යම් | some | ඇතැම් | some | සියලුම | all |
| අර | that | ඊක | | මුළු | all |

2.5.6 Nipathana

Nipathana can be used alone in some contexts, and can be used as a *postposition* as well if needed.

| Sinhala Word | ඇති | නැති | ඕනෑ | එපා | බැරි | පුළුවන් | ඇතිව | නැතිව | ඇත්නම් |
|---------------------|-----------------|------------|------|-------------|--------|---------|------|---------|------------------------|
| English Translation | Enough / having | Not having | want | Do not want | unable | able | with | without | If [someone] is having |

2.5.7 Case marker

Sinhala nouns are morphologically inflected based on the case. A suffix is added to the noun to show the case. For animate nouns and inanimate singular nouns, suffix ට - ta is added for dative case, suffix ගේ - “ge” is added for genitive case and suffix ගෙන් - “gen” is added for instrumental case. For inanimate plural nouns, suffix වලට - “valata” is added for dative case, suffix වල - “vala” is added for genitive case and suffix වලින් - “walin” is added for instrumental case. According to Sinhala language rules, it is wrong to separate these case marking suffixes from the main noun. However, some Sinhala writers tend to separate this case marking suffix from the main noun. If separated, such suffixes are categorized under case markers

| | | | | | | |
|---|----|------|-----|-------|----|------|
| ට | ගේ | ගෙන් | වලට | වලින් | වල | හිදී |
|---|----|------|-----|-------|----|------|

2.5.8 Preposition in compound verb

Preposition in compound verb is a words that do not have a meaning by themselves but, when combined with another verb, make up a compound verb.

| | | | | | | | |
|-----|------|-----|-----|----|--|--|--|
| ඉටු | සිදු | පත් | මිය | ගත | | | |
|-----|------|-----|-----|----|--|--|--|

2.6 Extra tags at Level 2

2.6.1 Sentence Ending

A word (Noun, Postposition, Adjective, etc) that is used to end the sentence, but is not a finite verb

| | | | |
|--------|--------------|-----------|-------|
| සඳහායි | ගොඩනැගිල්ලකි | විශේෂිතයි | පමණයි |
|--------|--------------|-----------|-------|

- 2.6.2 Number
- 2.6.3 Abbreviation
- 2.6.4 Full stop
- 2.6.5 Punctuation
- 2.6.6 Foreign word

3. Level 3

Inflection based grammatical variations of the language is captured using tags at third level. Sinhala words (Nouns and verbs) can be inflected based on Animacy, Number, Gender, Person, Case, Tense and Definiteness.

3.1 Common Noun Inflections

Common nouns can be either Animate or Inanimate. Those two categories are further sub categorized based on other inflecting factors. Each cell of the following tables is a separate tag

Inflections of Animate Common Nouns

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

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

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 | | Common | Plural |
| | | | | Masculine | Feminine | | |
| Nominative | මම | අපි | ඔබ | ඔහු | ඇ / ඇය | තමා / උ | ඔවුහු |
| Accusative | මා | අප | ඔබ | ඔහු | ඇ / ඇය | තමන් / උන් | ඔවුන් |
| Dative | මට | අපට | ඔබට | ඔහුට | ඇට / ඇයට | තමන්ට | ඔවුන්ට |
| Genitive | මගේ / මාගේ | අපගේ / අපේ | ඔබගේ / ඔබේ | ඔහුගේ | ඇගේ / ඇයගේ | තමන්ගේ | ඔවුන්ගේ |
| Instrumental | මගෙන් / මාගෙන් | අපගෙන් / අපෙන් | ඔබගෙන් / ඔබෙන් | ඔහුගෙන් | ඇගෙන් / 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 |

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 / සින්හලෙනි | සිංහලෙන් |
| | | රුපියල / රුපියල් | රුපියලට | රුපියලේ / රුපියලෙනි | 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.

Finite Verbs (අවසාන ක්‍රියා)

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 | බැලිමි බැලුවෙමි |
| | Plural | | බලමු බලන්නෙමු | බැලිමු බැලුවෙමු |
| Second Person මධ්‍යම | Singular | | බලහි බලන්නෙහි | බැලිහි බැලුවෙහි |
| | Plural | | බලහු බලන්නෙහු | බැලිහු බැලුවෙහු |
| Third Person උත්තම | Singular | Common | බලයි | බැලිය බැලිණ |
| | | Masculine | බලන්නේය | බැලුවේය |
| | | Feminine | බලන්නිය | බැලුවාය |
| | Plural | | බලති බලන්නෝය බලන්නාහ | බැලූහ බැලුවෝය බැලුවාහ |

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

Finite Verbs Commanding

| | | | |
|---------------|---------------|---------------|--|
| දෙන්න / දෙන්න | බලන්න / බලන්න | කරන්න / කරන්න | |
|---------------|---------------|---------------|--|