

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	English	Sinhala	English	Sinhala	English	Sinhala	English
මිනිසා (minisā)	the man	මිනිසු (minissu)	men	ගස (gasa)	the tree	ගස් (gas)	trees
මිනිසෙක් (minisek)	a man	මිනිසුන් (minisun)	men	ගසක් (gasak)	a tree	ගස්වල (gasvala)	in/on trees
මිනිසාගේ (minisāgē)	mans'	මිනිසුන්ගේ (minisungē)	Men's	ගසෙහි (gasehi)	in/on the tree	ගස්වලින් (gasvalin)	from trees
මිනිසාගෙන් (minisāgen)	from the man	මිනිසුන්ගෙන් (minisungen)	From men	ගසෙන් (gasen)	from the tree	ගස්වලට (gasvalaṭa)	to trees
මිනිසාට (minisāṭa)	to the man	මිනිසුන්ට (minisunṭa)	To men	ගසට (gasaṭa)	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	English	Sinhala	English
ජෝන් (jōn)	John	ජපානය (japanaya)	Japan
ජෝන්ට (jōnṭa)	To John	ජපානයට (japanayata)	To Japan

ජෝන්ගේ (jōnge)	John's	ජපානයේ (japanaye)	In Japan
ජෝන්ගෙන් (jōngen)	From John	ජපානයෙන් (japanayen)	From Japan

2.1.3 Pronouns

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

Sinhala	English	Sinhala	English	Sinhala	English	Sinhala	English
මම (mama)	I	අපි (api)	we	ඔහු (ohu)	he	ඔවුන් (ovun)	them
මට (maṭa)	To me	අපිට (apiṭa)	For us	ඔහුට (ohuṭa)	To him	ඔවුන්ට (ovunṭa)	To them
මගේ (magē)	mine	අප/අපේ (apa / apē)	our	ඔහුගේ (ohugē)	his	ඔවුන්ගේ (ovungē)	theirs
මගෙන් (magen)	From me	අපෙන්/ අපගෙන් (apen / apagen)	From us	ඔහුගෙන් (ohugen)	From him	ඔවුන්ගෙන් (ovungen)	From them
එය (eya)	that	එහි (ehi)	In/on that	එවා (ēvā)	those	එවායේ (ēvāyē)	In those
එයින් (eyin)	From that			එවායින් (ēvāyin)	From those		
එයට (eyaṭa)	To that			එවාට (ēvāṭa)	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'

පාඩම කරනවා (pāḍam karanavā)	studying	සෙල්ලම කරනවා (sellam karanavā)	playing
ක්‍රියා කරනවා (kriyā karanavā)	Acting / doing		

Examples of Noun in compound verb

පාඩම (pāḍam)	Lesson	සෙල්ලම (sellam)	Play	ක්‍රියා (kriyā)	Actions
--------------	--------	-----------------	------	-----------------	---------

2.1.5 Questioning Pronouns

Questioning Pronouns are words used to ask a question.

Sinhala	English	Sinhala	English	Sinhala	English
කුමක්ද/කුමක්දැයි (kumakda / kumakdæyi)	What	කෙසේද/කෙසේදැයි (kesēda / kesēdæyi)	How	කවදාද/ කවදාදැයි (kavadāda / kavadādæyi)	When
කොහේද	Where	කොතනද	Where	කවරෙක්ද	Who

(kohedīda)		(kotanadīda)		(kavarekda)	
කවරක්ද (kavarakda)	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	English	Sinhala	English	Sinhala	English
සමහරෙක් / සමහරුන් (samaharek / samaharun)	Some of them	ඇතැමෙක් (ætæmek)	Some of them	සමහරක් (samaharak)	Some of those
සමහරුන්ට (samaharunṭa)	To some of them	සමහරුන්ගේ (samaharungē)	Some of their	සමහරුන්ගෙන් (samaharungēn)	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	English	Sinhala	English	Sinhala	English
කුමක් (kumak)	What	කවර (kavara)	Which	කෙසේ (kesē)	How
කොහේදී (kohēdī)	Where				

2.2 Categories of Verbs

2.2.1 Verb finite

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

Sinhala	English	Sinhala	English	Sinhala	English
කළේය (kaḷēya)	(he) did	කළාය (kaḷāya)	(she) did	කරන්නේය (karannēya)	Is doing
තිබේ (tibē)	Have/has	තිබුණි / තිබිණ (tibuṇi / tibiṇa)	had	කරමි (karami)	Am doing
කරති (karati)	Are doing	කළෝය (kaḷōya)	(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	English	Sinhala	English	Sinhala	English
බැලූ (baḷu)	looked	බැලිය (baḷiya)		බැලෙන (baḷēna)	

(bælu)		(bæliya)		(bælena)	
බලන (balana)	Is looking at	බලන්නේ (balannē)	Is looking	බැලීමේ (bælimē)	
බලනු (balanu)		බැලුවේ (bæluvē)	looked		

2.2.3 Verbal noun

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

Sinhala	English	Sinhala	English	Sinhala	English
බැලීම (bælimā)	The look	බැලිලි (bælili)	looks	බැලීමෙන් (bælimēn)	From the look
බැලීමක් (bælimak)	A look	බැලීමට (bælimāṭa)	To look	බැලීමටත් (bælimāṭat)	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 (hæki)	Can / could	යුතු (yutu)	Must /should	නොහැකි (nohæki)	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	English	Sinhala	English	Sinhala	English
බලා (balā)	Looked and	බලමින් (balamin)	While looking	බලද්දී (baladdī)	when looking at
බලනොත් (balatot)	If looked	බැලුවාම (bæluvāma)	When (someone) look at	බැලුවත් (bæluvat)	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.

Sinhal	English	Sinhala	English	Sinhala	English	Sinhala	English
රළු (raḷu)	rough	විශේෂ (viśēṣa)	Special	සුදු (sudu)	White	ඉහල (ihala)	Up
සුමුදු (sumudu)	Smooth	අලුත් (alut)	New	කුඩා (kuḍā)	Small	පහල (pahala)	Down

ප්‍රාදේශීය (prādēśī ya)	regional	නාගරික (nāgarika)	urban	පළමු (paḷamu)	first	දෙවන (devana)	second
-------------------------------	----------	----------------------	-------	------------------	-------	------------------	--------

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’

Sinhala	චිත්‍ර (chitra)	පාසල් (pāsal)	ගල් (gal)	කාර්යාල (kāryāla)	ගොඩනැගිලි (goḍanægili)	වැව (væv)	විද්‍යාල (vidyāla)
English	Drawings	Schools	Stones	Offices	Buildings	Tanks	Colleges

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

Sinhala	අඩු කරනවා (aḍu karanavā)	වැඩි කරනවා (væḍi karanavā)	අඩපන වෙනවා (aḍapana venavā)
English	Reducing / decreasing	increasing	weakening

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.

Sinhala	සඳහා (saṇḍahā)	අතර (atara)	ලෙස (lesa)	සිට (siṭa)	වැනි (væni)	තුළින් (tulini)	දක්වා (dakvā)
English	For	while	As	From	Such as/like	Through	To
Sinhala	මෙන් (men)	මත (mata)	පිළිබඳ (pilibaṇḍa)	මිස (misa)	උදෙසා (udesā)	පරිදි (paridi)	තෙක් (tek)
English	Like	On	Of/about	Except	For	As	Until

2.5.2 Conjunction

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

Sinhala	හා (hā)	හෝ (hō)	නමුත් (namut)	යයි (yayi)	දැයි (dæyi)	එහෙත් (ehet)	එත් (ēt)
English	And	Or	But	That	Weather	But	but

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	මික	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	English	Sinhala	English	Sinhala	English
මෙම	This	මෙ	This	එක්	One
සියලු	All	බොහෝ	Most	සමහර	Some
යම	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	අති (æti)	නැති (næti)	ඕනෑ (ōnæ)	එපා (epā)	බැරි (bæri)
English	Enough / having	Not having	Want/requi re	Do not want	unable
Sinhala	පුළුවන් (puḷuvan)	ඇතිව (ætiva)	නැතිව (nætiva)	ඇත්නම් (ætnam)	
English	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 ම - 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. Case marker does not have an English meaning on its own.

ට (ṭa)	ගේ (gē)	ගෙන් (gen)	වලට (valaṭa)	වලින් (valin)	වල (vala)	හිදී (hidī)
-----------	------------	---------------	-----------------	------------------	--------------	----------------

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. Prepositi

ඉටු (iṭu)	සිදු (sidu)	පත් (pat)	මිය (miya)	ගත (gata)
--------------	----------------	--------------	---------------	--------------

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. When such a word is used as a sentence ending, original word is inflected with prefixing a particle like යි (yi) or යා (ya).

Sinhala	සඳහායි (saṇḍahāyi)	ගොඩනැගිල්ලකි (goḍanægillaki)	විශේෂිතයි (viśēṣitayi)	පමණයි (pamaṇayi)
English	For	A building	Special	Only

2.6.2 Number

All numbers written in numerics or letters

2.6.3 Abbreviation

2.6.4 Full stop

2.6.5 Punctuation

2.6.6 Foreign word

A word that is written in any other language or a word that can not be categorized under any of the above POS tag class.

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.

Animate common nouns can be inflected based on number, gender, definiteness and case (Nominative, Accusative, Dative, Genitive, Instrumental).

Inflections of Animate Common Nouns

විභක්ති Case	ඵක Singular				බහු Plural		
	පුරුෂ- Masculine		ස්ත්‍රී - Feminine		පුරුෂ Masculine	ස්ත්‍රී Feminine	පොදු Common
	නිශ්චිත Definite	අනිශ්චිත Indefinite	නිශ්චිත Definite	අනිශ්චිත Indefinite	අනිශ්චිත Indefinite	අනිශ්චිත Indefinite	අනිශ්චිත Indefinite
කතෘ Nominative	මිනිසා (minisā)	මිනිසෙක් (minisek)	කාන්තාව (kāntāva)	කාන්තාවක් (kāntāvak)	මිනිසු (minissu)	කාන්තාවෝ (kāntāvō)	දෙමවුපියෝ (demavupiyō)
	The man	A man	The woman	A woman	Men	Women	Parents
කර්ම Accusative	මිනිසා/ මිනිසාව (minisā/minisā va)	මිනිසෙකු (miniseku)	කාන්තාව (kāntāva)	කාන්තාවක (kāntāvaka)	මිනිසුන් (minisun)	කාන්තාවන් (kāntāvan)	දෙමවුපියන් (demavupiyan)
සමප්‍රදාන Dative	මිනිසාට (minisāṭa)	මිනිසෙකුට (minisekuṭa)	කාන්තාවට (kāntāvaṭa)	කාන්තාවකට (kāntāvakaṭa)	මිනිසුන්ට (minisunta)	කාන්තාවන්ට (kāntāvanṭa)	දෙමවුපියන්ට (demavupiyanṭ a)
	To the man	To a man	To the woman	To a woman	To men	To women	To parents
සම්බන්ධ Genitive	මිනිසාගේ (minisāgē)	මිනිසෙකුගේ (minisekugē)	කාන්තාවගේ (kāntāvagē)	කාන්තාවකගේ (kāntāvakagē)	මිනිසුන්ගේ (minisungē)	කාන්තාවන්ගේ (kāntāvangē)	දෙමවුපියන්ගේ (demavupiyan gē)
	The mans	A mans	The womans	A womans	Mens	Womens	Parents
අවධි Instrumental	මිනිසාගෙන් (minisāgen)	මිනිසෙකුගෙන් (minisekugen)	කාන්තාවගෙන් (kāntāvagen)	කාන්තාවකගෙන් (kāntāvakagen)	මිනිසුන්ගෙන් (minisungen)	කාන්තාවන්ගෙන් (kāntāvangen)	දෙමවුපියන්ගෙන් (demavupiyan gen)
	From the man	From a man	From the woman	From a woman	From men	From women	From parents
ආලපන Vocative	මිනිස! (minisa!)		කාන්තාව! (kāntāva!)		මිනිසුනේ! (minisunē!)	කාන්තාවනි! (kāntāvani!)	දෙමවුපියනි! (demavupiyani !)
	Man!		Womn		Men! People!	Women!	Parents!

Animate common nouns sometimes belongs to a common gender which represent both masculine and feminine. (Eg: මිනිස්සු - people / දෙමව්පියයෝ - 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		Plural	
				Masculine	Feminine		Common
Nominative	මම	අපි	ඔබ	ඔහු	ඇ / ඇය	තමා / ඌ	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 / ඇගේ / ඇයගේ	තමන්ගේ	ඔවුන්ගේ
Instrumental	මගෙන් / මාගෙන්	අපගෙන් / අපෙන්	ඔබගෙන් / ඔබෙන්	ඔහුගෙන්	ඇගෙන් / 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 / සිත්හලෙහි	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.

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 උන්තම	Singular	Common	බලයි	බැලිය බැලිණ
		Masculine	බලන්නේය	බැලුවේය
		Feminine	බලන්නීය	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

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