Abstract

Lexicons are one of the useful language resources both in localization and computational linguistics point of views. The extensive use of computers in language processing tasks demands the necessity for electronic lexicons. Sinhala language has a very long tradition in lexicon building in various domains. Existence of various linguistic traditions has given rise to several inconsistencies among the parts of speech and other linguistic data provided in them. Though efforts have recently been made to develop electronic lexicons at commercial level it appears that they were hindered by limitations of technology and insufficiency in linguistic research on lexicon structures for Sinhala. Traditional Sinhala grammar faces with various difficulties when it tries to analyze the modern behavior of language since structures of the language have undergone changes during the post independence era. This paper presents the importance revisiting traditional Sinhala grammar in order to keep abreast of current changes. The paper also proposes a reasonable classification for Sinhala words. It is expected that the proposed lexicon structure based on this classification will cater to the requirements of both ordinary users and computational linguists.

1. Introduction

1.1. Historical background

Sinhala language has a very long history in preparation of lexicon since 5th century A.D. Those were glossaries of technical terms, for example Buddhism, poetics, aesthetic theories etc and also as resources for language learning and teaching at different levels those were used by scholars, writers, teachers and scholars for their respective work. Words in these lexicons have been classified according to various linguistic traditions that were dominating at the time of preparation or to the compiler(s) belonged to. The quantity and the quality of these lexicons vary according to the scopes which the compiler(s) had in mind. Due to these reasons several inconsistencies exist among them with respect to orthography, spelling, parts of speech, word separation, collation sequence etc. Dictionaries/lexicons were studied under this research are given in Appendix A.

1.2. Electronic dictionaries (lexicons) in Sinhala

Even though a considerable number of dictionaries/lexicons have been available for Sinhala and computers have extensively been used for Sinhala word processing for more than two decades a few electronic dictionaries are available in the market. These are based on most of the commonly used English-Sinhala dictionaries and the glossaries of technical terms published by the government during the last four decades for the purpose of teaching technical subjects in local languages. These lexicons give higher level syntactic information such as whether a word is a Nama (noun), Kriya (verb), Nipatha (prepositions, conjunctions, interjections ), Krudanta (Gerund) etc, the meaning or synonyms and Tamil & English translation equivalents. It is clear that these applications have been developed in the ordinary person’s point of view and they have less importance in computational linguist’s point of view which requires a deep syntactic information such as in case of nouns gender, number, person, (in) definiteness, active/passive and case (vibhakthi) and in case of verbs number, person, gender, active/passive and tense.

1.3. Words in Sinhala

Words in the Sinhala language can basically be classified into three groups namely nisippana(words that have local origin), thadhava (words derived from other languages), thathsama (words borrowed from
other languages as they are). Pali and Sanskrit are said to be the mother language of Sinhala. Therefore most of the words used in Sinhala are borrowed from those two languages. Due to geographical proximity the influence of South Indian languages have contributed a considerable amount of words to Sinhala esp. Tamil. Portuguese, Dutch and English words entered the language during the period of colonialism. [1]

1.4. Problems of traditional Sinhala grammar

Traditional Sinhala grammar is based on the notions expounded in famous “Sidath Sangara” which was written in 13 century A.D. It was based on poetry language existed in that period. It was Munidasa Cumaratunga who revolutionized Sinhala grammar by critically analyzing the notions of “Sidath Sangara” and standardized Sinhala grammar as much possible at the beginning of the 20th century. His findings and recommendations are described in his two monumental works: “Vyakarana Vivaranaya” (Analysis of Grammar) and “Kriya Vivaranaya” (Analysis of Verbs). [2],[3] After the independence in 1948 Sinhala language was used extensively in various domains such as newspapers, literature, government publications, and text books. This gave rise to an exponential growth of Sinhala language both in grammatical structures as well as in words. Due to changes occurred in the post independence era neither “Sidath Sangara” nor Cumaratunga’s grammar is capable of analyzing modern Sinhala effectively. It is necessary to revisit the traditional grammar by linguists in order to describe the modern Sinhala.

2. Methodology

2.1. Survey on word classification

A literature survey was performed on word classification used in standard dictionaries/lexicons and the classification described in standard grammar books published since 13th century. It was also decided to consult a group of eminent scholars drawn from both linguistics and Sinhala regarding word classification of Sinhala since different schools hold different views on this issue. Results of this survey are given in Section 3.

2.2. Criterion for selection of words

It is very important to decide what kind of information is expected to be included in and what are the target groups of the present lexicon. At initial stage of this study it was decide that following characteristics should be taken into consideration when a set of words are selected for the lexicon:

a) Words should be drawn from almost all the domains where language is being actively used.

b) Considerable amount of words should represent the modern usage of language rather than old fashioned Sinhala.

c) Technical terms and loan words should also be included according to the proportion of their usage.

It is also expected to include meanings or definitions, synonyms, English and Tamil translation equivalents, etymology and morpho-syntactic information of each lexical entry which can be used for language teaching/learning, linguistic research purposes and computational linguistic activities such as parsing, machine translation, grammar/spell checking. This broad range of information will cater the requirements of students, teachers, linguists and computational linguists as well.

3. Results

3.1. Modern Sinhala & syntactic categories

Having made a substantial amount of effort to eliminate unnecessary formalisms in traditional grammar and preserve the positive aspects of it, the following classification has been developed after considering the classifications presented in traditional grammar books and views of the modern linguist. The requirements of the ordinary users of the lexicon and the computational linguistics processes were also kept in mind when it was designed. The observations made in dictionaries and standard grammar books during this study are given in Appendix B. [4]

<table>
<thead>
<tr>
<th>Table 1: Syntactic categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nama</td>
</tr>
<tr>
<td>Kriya</td>
</tr>
<tr>
<td>Nipatha(^1) (particles)</td>
</tr>
<tr>
<td>Kruantha</td>
</tr>
<tr>
<td>Nama Visheshana</td>
</tr>
<tr>
<td>Kriya Visheshana</td>
</tr>
</tbody>
</table>

\(^1\) There is no exact category in English grammar that matches with Nipatha.
Table 2: Attribute-value pairs of syntactic categories

<table>
<thead>
<tr>
<th>Category/Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nama:</td>
<td></td>
</tr>
<tr>
<td>Definiteness</td>
<td>definite, indefinite</td>
</tr>
<tr>
<td>Number</td>
<td>Singular, plural, both</td>
</tr>
<tr>
<td>Person</td>
<td>1st, 2nd 3rd person</td>
</tr>
<tr>
<td>Gender</td>
<td>masculine, feminine, common, neuter</td>
</tr>
<tr>
<td>Animateness</td>
<td>animate, inanimate</td>
</tr>
<tr>
<td>Case</td>
<td>nominative, accusative, auxiliary, instrumental dative, ablative, possessive, locative, vocative</td>
</tr>
</tbody>
</table>

Kriya:

<table>
<thead>
<tr>
<th>Category/Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definiteness</td>
<td>Singular, plural, both</td>
</tr>
<tr>
<td>Number</td>
<td>1st, 2nd 3rd person</td>
</tr>
<tr>
<td>Person</td>
<td>masculine, feminine, common, neuter</td>
</tr>
<tr>
<td>Tense</td>
<td>past, non-past</td>
</tr>
<tr>
<td>Voice</td>
<td>active, passive</td>
</tr>
<tr>
<td>Transitivity</td>
<td>transitive, intransitive</td>
</tr>
</tbody>
</table>

3.2. List of words

A list of 40,000 words has been prepared the National Institute of Education (NIE), Sri Lanka. NIE is the official government body of Sri Lanka that handles the issues related to curriculum development and educational publications for government schools. Having considered the quality and the quantity of this set of words and its availability in electronic form it was decided to adopt this list of words as the base of the present lexicon.

This list of words is based on the “Sinhala Sabdhakoshaya” which was published by the Ministry of Cultural Affairs in 26 volumes. It is the largest and comprehensive dictionary published in Sinhala. A group of scholars appointed by the NIE including the most senior linguist in Sri Lanka, Prof. Wimal G. Balagalle, who was also former Chief Editor of “Sinhala Sabdhakoshaya”, has carefully selected the present list of words in such a way that it includes practical, commonly used and important word of Sinhala.

4. Lexicon structure

4.1. Lexical entries

Lexical entries of the lexicon will be displayed according to the collation order of Sinhala. Lexical entries generated by the same lexical entry will be displayed under one lexical entry which may be the root or the most common form of them. This is due to the fact that Sinhala is an inflectionally rich language. Lexical entries which have alternative spellings will be given along with most common form as the head word. Each of these lexical entries will be associated with a set of useful information relevant to each lexical entry. Sections 4.2 and 4.3 describe the information structures which will be used in the present lexicon.

4.2. Information structures of a lexical entry

Table 3 shows the information associated with each lexical entry in the lexicon. Meaning of each of these fields is described in the Appendix C.

Table 3: Lexical entry information

<table>
<thead>
<tr>
<th>Word</th>
<th>Spelling variants(optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part of Speech</td>
<td></td>
</tr>
<tr>
<td>Pronunciation</td>
<td></td>
</tr>
<tr>
<td>Root Word</td>
<td></td>
</tr>
<tr>
<td>Etymology</td>
<td></td>
</tr>
<tr>
<td>Definition(s) (meaning(s))</td>
<td></td>
</tr>
<tr>
<td>Synonyms and cross references</td>
<td></td>
</tr>
<tr>
<td>Tamil and English translation equivalents</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: POS subcategories and attributes

<table>
<thead>
<tr>
<th>Part of Speech</th>
<th>Subcategory</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nama</td>
<td>Simple</td>
<td>Animateness</td>
</tr>
<tr>
<td></td>
<td>Compound</td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case</td>
</tr>
</tbody>
</table>

Further description on this classification is available in Appendix D.
5. Discussion

The main objective of this study is to study the required structure and the content of the Sinhala lexicon that will be developed under the PAN Localization Project. As an important language resource tool lexicon will be used by people with different objectives to ranging from students, language teachers and researchers, translators and computational linguists. It is hoped that the proposed information structure (in section 4) will meet the requirements of them adequately. The set of words going to be used will also cover a considerable number of words drawn from various areas since that has carefully been selected from the most comprehensive and the largest dictionary in Sinhala by a group of scholars.

It was identified the existing classification of words proposed by traditional grammar is insufficient to handle the modern language styles. The classes and the attributes that characterize each part of speech have been proposed after a rigorous study of Sinhala grammar and the classifications used in standard dictionaries.

The results of the present research can effectively be used in the lexicon development stage. The authors and the Language Technology Research Laboratory would like to acknowledge the support given by the scholars to conduct this survey by contributing their knowledge and valuable time.

5. References


Appendix A: Glossaries published in the Ancient Period

**Anuradhapura Period**

*Mahawamsayata Pali Atuwakaranaya – 5th Century A.D.*

*Abhidanppadipika*

*Helamuwa*

*Dampiya Atuwa Getapaday*

*Visuddhi Sannaya*

*Kudasika Sannaya*

*Abhidharmartha Varnanaya*

**Dambadeni Period**

*Piyummala – 1270-1293 A.D.*

**Kotte Period**

*Sinhala Namawaliya*

*Ruwanmala – 1420 A.D.*

*Nawa Namawaliya*

*Heladiv Abhidanawatha*

*Dhamamgetalaya*
Kandy Period

Akaradiya
Shadbha Mukthawaliya
Shuddha Sinhalaya
Pali Sinhala Artha Kathana Granthaya
Pali Sinhala Katha Malawa
Perani Sidath Sangara Sannaya
Nigandu Grantha
    Vararuchi
    Saraswathi
    Vasudewa
    Siddhaushadhha

Modern Period

Clough’s Dictionary
Benjamin Clough
    English-Sinhala Dictionary - 1821
    Sinhala-English Dictionary – 1830

John Callaway’s Dictionary
John Callaway
    English-Sinhala Dictionary - 1821
    Sinhala-English Dictionary - 1821

Lambrick Lexicon
Though this is not a dictionary, but the words of English, Sinhala and Portuguese are included.

Appendix B

Nama

This is the classification proposed by Cumaratunga Munidasa in Vyakarana Vivaranaya for Nouns.

Vyakarana Vivaranaya –Cumaratunga Munidasa.

Abhinna Nama/Sarva Nama (Pronoun)
    සංජු ලේ පෙරවවදා – Paurusha (Person)
    චංජු ලේ – Uttama Purusha
    කළඳ ලේ – Madyama Purusha
    මංජු ලේ – Prathama
    චංජු ලේ – Suchaka (Locative)
    චංජු – Prashna (Interrogative)
    චංජු – Aniyam (Indefinite pronoun)

W.Bidgell’s Dictionary
    English-Sinhala Dictionary – 1848

Nicholson’s Dictionary
    English-Sinhala Dictionary – 1864

Charles Carter’s Dictionary
Charles Carter
    English-Sinhala Dictionary – 1881
    Sinhala-English Dictionary – 1892

Sinhala Shabdahakoshaya – 1937
    Published by the Department of Cultural Affairs.

An Etymological Glossary of the Sinhala Language
    Prof.Wilhelm Geiger – 1941

Malalasekara English-Sinhalese Dictionary
    Prof.G.P.Malalasekara – 1948

Sri Sumangala Shabdahakoshaya
    Rev.Veliwitiye Soratha – 1952

Nuthana Sinhala Paribhashhika Shabdakoshaya
    Harishchandra Wijetunge – 1978

Prayogika Sinhala Shabdakoshaya
    Harishchandra Wijetunge – 1982

Draft Edition
    දුංජු ලේ – Dravya Nama (Referring for inanimate things)
    වාංජු ලේ – Guna Nama (Referring for modifiers)
    අංජු ලේ – Bhava Nama (Referring for modifiers or verbal nouns)
    සංජු ලේ – Sagna Nama (Proper Noun)

උප්පාංජු ලේ/ජුංජු ලේ

අංජු – Paurusha (Person)
    කලංජු ලේ – Uttama Purusha
    කලංජු ලේ – Madyama Purusha
    සංජු – Prathama
    චංජෑ – Suchaka (Locative)
    චංජු – Prashna (Interrogative)
    චංජිං – Aniyam (Indefinite pronoun)
This is the classification proposed by Prof. W.S. Karunatilaka in Sinhala Bhasha Vyakaranaya for Nouns.

- Niyaththa Nama (Proper Noun)
- Sarva Nama (Pronoun)
- Sankya Nama (Numerals)
- Krudantha Nama (Verbal Noun)
- Thadditha Nama (According to appending suffixes to the stem of the noun)
- Samasa Nama (Compound Noun)
- Ganya Nama (Countable Noun)
- Aganya Nama (Uncountable Noun)

- Linga (Gender)
  - Purusha Linga (Masculine)
  - Stree Linga (Feminine)
  - Napunsaka Linga (Neuter)
- Sadharana Linga (Common)

- Antha (Endings)
- Swarantha (Vowel endings)
- Halantha (Consonant endings)

- Vachana (Number)
  - Eka Vachana (Singular)
  - Bahu Vachana (Plural)

- Vibhakthi (Case)

- Pranavachi Nama (Animate)
- Purusha Linga (Masculine)
- Stree Linga (Feminine)
- Manava (Human)
- Amanava (Non-Human)
- Apranavachi Nama (Inanimate)
- Napunsaka Linga (Neuter)

- Podu Linga (Common Noun)

- Vachana (Number)
  - Eka Vachana (Singular)
  - Bahu Vachana (Plural)

- Niyaththa/Aniyaththa (Definite/Indefinite)
This is the classification proposed by Prof. J.B.Disanayaka in Nama Padaya for Nouns.

Nama Padaya – J.B.Disanayaka

This is the classification proposed ancient “Sidath Sangara” in 13 century A.D.
This is the classification proposed by Prof. J.B. Disanayaka in Kriya Padaya for Verbs.

This is the classification proposed by Cumaratunga Munidasa in Vyakarana Vivaranaya for Verbs.
Appendix C

Table 1 describes the meaning of syntactic categories used in the lexicon structure.

<table>
<thead>
<tr>
<th>Sinhala</th>
<th>English</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>වචන   ෙභ+ඳඔ0Dඒ70DCAදය</td>
<td>Vachana</td>
<td>Number</td>
</tr>
<tr>
<td>මැංක වචන – Eka Vachana</td>
<td>(Singular)</td>
<td></td>
</tr>
<tr>
<td>උපකාරක භ+ඳඔ0Dඒ80DD2භ+ඳඔ0DC1හrඹkඹrයා පද – Upakarakawkriyapada</td>
<td>(Reflexive verb)</td>
<td></td>
</tr>
<tr>
<td>බාවභ+ඳඔ0Dඒඒ0DD6පයා පද – Bhavarupaw kriyapada</td>
<td>(Paralel Verbs)</td>
<td></td>
</tr>
</tbody>
</table>

Table C.1: Syntactic categories used in the lexicon structure
<table>
<thead>
<tr>
<th>Particulars</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part of speech</td>
<td>One of the traditional categories of words intended to reflect their functions in a grammatical context.</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>The manner of someone utters a word.</td>
</tr>
<tr>
<td>Root Word</td>
<td>Base, stem - the form of a word after all affixes are removed.</td>
</tr>
<tr>
<td>Etymology</td>
<td>Structure of words.</td>
</tr>
<tr>
<td>Definition(s)(meaning(s))</td>
<td>Explanation of the meaning of a word.</td>
</tr>
<tr>
<td>Synonyms and cross references</td>
<td>Equivalent word.</td>
</tr>
<tr>
<td>Tamil and English translation equivalents</td>
<td>Tamil and English version of a word.</td>
</tr>
</tbody>
</table>

### Table C.2: Syntactic categories used in the lexicon structure

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>A word that can be used to refer to a person or place or thing.</td>
</tr>
<tr>
<td>Verb</td>
<td>Expresses some state or action affecting some person or thing.</td>
</tr>
<tr>
<td>Nipatha</td>
<td>The syntactic categories of prepositions, conjunctions and interjections.</td>
</tr>
<tr>
<td>Gerund</td>
<td>Words formed by adding suffixes to roots of verbs.</td>
</tr>
<tr>
<td>Adjectives</td>
<td>The word class that denotes some quality or attribute of the noun.</td>
</tr>
<tr>
<td>Adverbs</td>
<td>The word class that qualifies verbs or clauses.</td>
</tr>
<tr>
<td>Singular</td>
<td>Singularity or Plurality of nouns and verbs.</td>
</tr>
<tr>
<td>Plural</td>
<td>The form of a word that is used to denote a singleton.</td>
</tr>
<tr>
<td>Person</td>
<td>A grammatical category of pronouns and verb forms.</td>
</tr>
<tr>
<td>First person</td>
<td>Pronouns and verbs used to refer to the speaker.</td>
</tr>
<tr>
<td>Second person</td>
<td>Pronouns and verbs used to refer to the person addressed.</td>
</tr>
<tr>
<td>Third person</td>
<td>Pronouns and verbs that are used to refer to something other than the speaker or addressee.</td>
</tr>
<tr>
<td>Gender</td>
<td>A grammatical category based on sex or animateness.</td>
</tr>
<tr>
<td>Masculine</td>
<td>A grammatical category that refers to males or to objects classified as male.</td>
</tr>
<tr>
<td>Feminine</td>
<td>A grammatical category that refers to females or to objects classified as male.</td>
</tr>
<tr>
<td>Sinhala</td>
<td>English</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>මූල්ලම් මියිත්</td>
<td>Neuter</td>
</tr>
<tr>
<td>මූල්ලම් කාන්තා</td>
<td>A grammatical category that refers to inanimate objects (neither masculine nor feminine).</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Common</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical category that refers to both masculine and feminine.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Animateness</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical category that denotes living beings or nonliving things.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Animate</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical category that denotes living beings</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Inanimate</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical category that denotes nonliving beings.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Case</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical case (nouns or pronouns or adjectives related in some way to other words in a sentence).</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Nominative case</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The category of nouns serving as the grammatical subject of a verb.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Accusative case</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The category of nouns serving as the direct object of a verb.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Vocative case</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The category of nouns used when a person or thing is addressed.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Tense</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A grammatical category of verbs used to express distinctions of time.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Past tense</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A verb tense that expresses actions or states in the past.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Non-past tense</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>A verb tense that expresses actions or states in the Present or Future.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Voice</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The grammatical relation (active or passive) of the grammatical subject of a verb to the action that the verb denotes.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Active voice</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The voice used to indicate that the grammatical subject of the verb is performing the action or causing the happening denoted by the verb.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Passive voice</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The voice used to indicate that the grammatical subject of the verb is the recipient (not the source) of the action denoted by the verb.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Transiiteness</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>the grammatical relation created by a transitive or intransitive verb.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Transitive</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Denotes an action passing from the doer to an object.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Intransitive</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Express a state or condition, or an action or feeling that does not pass over to an object.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Predicate</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Final verb of a sentence</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Past Participle</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>The predicates of the subject which precede its final predicate.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Conditional Mood</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Expresses uncertainty or the time when an action is done.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Present Participle</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Expressing present action or the actions doing in the same time.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Gerund – past</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Words formed by adding suffixes of past tense to roots of verbs.</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Gerund – present</td>
</tr>
<tr>
<td>මූල්ලම්/කාන්තාමාණ</td>
<td>Words formed by adding suffixes of</td>
</tr>
</tbody>
</table>
Appendix D

Nouns

1. Simple nouns.
Sinhala noun roots and the nominative forms of the nouns are categorized under the section of simple nouns.

2. Complex nouns.
Complex nouns are produced by adding a word formation affix to the noun stem. Mainly, krudantha (gerund) and thaddhitha (derivative nouns) are included in this section. Apart from them adjectives and adverbs are also being made in this procedure.

3. Compound nouns
Compound nouns are made by adding number of other noun roots to the primary noun base.

<table>
<thead>
<tr>
<th>Nouns</th>
<th>Animate nouns</th>
<th>Inanimate nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculine</td>
<td>Singular – Definite</td>
<td>Indefinite</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td></td>
</tr>
<tr>
<td>Feminine</td>
<td>Singular – Definite</td>
<td>Indefinite</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td></td>
</tr>
<tr>
<td>Common</td>
<td>Singular – Definite</td>
<td>Indefinite</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td></td>
</tr>
</tbody>
</table>

Animate nouns

Inanimate nouns

- Neuter- Singular – Definite
- Indefinite
- Plural

A few nouns inflect in a different manner rather than the above mentioned criteria. Although the nouns like bonikka (doll) pambaya (scarecrow) are inanimate by their nature, they inflect like animate nouns. In contrary, the noun of ibba (containing two meanings – tortoise/padlock) bear animateness and inanimateness according to the context.

Ibbó dennek innawa. There are two tortoises.

Ibbó dennek dala thiyenawa. Two padlocks have been fixed.

Sinhala singular nouns are always marked with definiteness or indefiniteness.
The animate nouns can be conjugated within the framework of all three persons. The inanimate nouns only bear the third person.

- **Animate nouns** - First person
  - Second person
  - Third person

- **Inanimate nouns** - Third person

Case

Sinhala nouns inflect according to the case they bear. Nouns which make their primary manifestation within discourse commonly appear in the direct case.

1. Nominative/ Direct case (කභ+ඳඔ0Dඒජහrඹkඹrථමා)
2. Accusative (කභ+ඳඔ0Dඒඒ0DCAම)
The singular forms of the accusative case are similar to the singular forms of the nominative case. The plural forms have different inflection.

3. Auxiliary case (කභ+ඳඔ0Dඒඒ0DCAතෘ)
The auxiliary case is marked with visin (භ+ඳඔ0DC00DD2භ+ඳඔ0DC30DD2භ+ඳඔ0Dඒ10DCA) functional morpheme and it is added to the accusative form of the noun. Auxiliary case is not marked with any suffixation.

These three cases are considered as the preliminary forms within the hierarchy of case study. The other case markers, which have given below would be added to either Nominative case marked singular noun or accusative case marked plural noun.

<table>
<thead>
<tr>
<th>Table D.1: Case Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case marker</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Instrumental</strong></td>
</tr>
<tr>
<td>sin</td>
</tr>
<tr>
<td>pl</td>
</tr>
<tr>
<td>Animate</td>
</tr>
<tr>
<td>Inanimate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Dative</strong></td>
</tr>
<tr>
<td>-tā-</td>
</tr>
<tr>
<td>-tā-</td>
</tr>
<tr>
<td>-tā-</td>
</tr>
<tr>
<td>-walata-</td>
</tr>
<tr>
<td>-hata-</td>
</tr>
</tbody>
</table>
Further information on nouns. There will be a further classification on nouns regarding their semantic aspects. In that stage they will be categorized according to their semantic values. Ex. Jathi nama ජාභි නාම, guna nama ගුණ නාම etc. And will also focus on the features like onomatopoeia and reduplication.

Verb
- Verb  
  a. Finite verb  
  b. Non finite verb

Table D.2: Tense

<table>
<thead>
<tr>
<th>Non past</th>
<th>Past</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short past</td>
</tr>
<tr>
<td>singular</td>
<td></td>
</tr>
<tr>
<td>present</td>
<td>constructed</td>
</tr>
<tr>
<td></td>
<td>present</td>
</tr>
<tr>
<td>1st person</td>
<td>balami</td>
</tr>
<tr>
<td></td>
<td>balim</td>
</tr>
<tr>
<td>2nd person</td>
<td>balahi</td>
</tr>
<tr>
<td></td>
<td>balhi</td>
</tr>
<tr>
<td>3rd person</td>
<td>balai</td>
</tr>
<tr>
<td></td>
<td>Balaninya-fem</td>
</tr>
<tr>
<td>Plural</td>
<td></td>
</tr>
<tr>
<td>1st person</td>
<td>balamu</td>
</tr>
<tr>
<td></td>
<td>balim</td>
</tr>
<tr>
<td>2nd person</td>
<td>balahu</td>
</tr>
<tr>
<td></td>
<td>balahu</td>
</tr>
<tr>
<td>3rd person</td>
<td>balati</td>
</tr>
<tr>
<td></td>
<td>baluha</td>
</tr>
</tbody>
</table>

Sinhala verb is divided into three categories according to its function in the syntactic level. Although there are three categories, the same two forms of the verb occur in each category.
- Transitive and Intransitive verbs  
  *(Sakarmaka/ Akarmaka)*
In Sinhala some verbs are transitive by their nature. They always expect an object in the string. Some verbs like –handa-, -duva- and –siti- are originally intransitive.

- Voice (karaka bhedaya)
  o Active voice (kartru karaka)
  o Passive voice (karma karaka)

The verb’s relationship with either agent or object is depicted in the voice.

- Ohu veDa karai. පොඳ ලබ ලඟ Active
- Ohugen veDa kerei. පොඳ ලබ ලඟ Passive

- Volitive and Involitive verbs

If an action is done without any mediation of the agent that verb is an involitive verb. The Sinhala verb is presented in volitive and involitive forms.

- Lamaya natali කාලයේ පොඳ volitive
- Bera gahana vita lamaya netavei. එර අගන ලබ ලඟ ලඟ involutive

- Causative verb

Verb stem+ -va-
- Imperative verb

<table>
<thead>
<tr>
<th>2nd person</th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero, va</td>
<td>vu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>පොඳ/බල/බලව/බලභ+ඳඔ0Dඒජ0DD2යභ+ඳඔ0DC00DCA/BALAMIN/BALAMINA/BALAMINA/BALAMINI</td>
<td></td>
</tr>
</tbody>
</table>

- Blessing verb (Ashirvada kriya)

Present/ future tense+ va

<table>
<thead>
<tr>
<th>1st person</th>
<th>Sin</th>
<th>Pres/fut tense+ -mi/-m</th>
</tr>
</thead>
<tbody>
<tr>
<td>pl</td>
<td>Pres/fut tense+ -mu/-mo-</td>
<td></td>
</tr>
<tr>
<td>3rd person</td>
<td>pl</td>
<td>Pres/fut tense+ -t/-ti-</td>
</tr>
</tbody>
</table>

Complex constructed verbs

In respect of the morpho-syntactic constructions of the verbs, there are two main patterns in Sinhala Language. The first one may call as the simple constructed verbs and the other the complex constructed verbs. Simple constructed verbs behave like one word and complex constructed verbs connect two different grammatical categories.

- Simple constructed verbs
  Duvanava

- Complex constructed verbs:

An initial stem (verb, noun, adjective or etc) + main verb

_Pata_ (noun)+ karanava

_Pasu_ (adj)+ (karanava

_Kapala_(non-finite verb form)+ danava

According to the structural pattern of the complex verbs, the most common verbs used as the second constituents are following:
1. karanava
2. venava
3. gannava
4. denava
5. danava
6. thienava
7. gahanava

Most of the borrowed nouns are used with the above mentioned verbs to make complex verbs.

Non finite verbs

a. gerund (krudanrha)
b. past participle (purva kriya)
c. present participle (misra kriya)
d. conditional verbs (asambhavya kriya)

The analysis of the gerund structure is due to be presented in a separate paper.

- The past participle (purva kriya)

Verb stem+ -aa/-x/-ii-
_Balaa/ xdx /vxtii_ බලා/ ඇද/ වැභ+ඳඔ0DA70DD2

- present participle (misra kriya)

verb stem+ -min/-mina/-mini
_balamin/balamina/balamina/balamin_ බලමින/බලමින/බලමින/බලමින

- conditional verbs (asambhavya kriya)

a. verb stem+ -t/-ta+ (prep;di) (balat/balata+di)
b. verb stem+ -t/-da/-du- (baluvath, baluvada, balatudu)

Particles

Particles are used to link words within the structure of a sentence. They also could be utilized as conjunctions to link phrases in complex sentences.

- Case marking particles
• Collective particles
  o –da- ɕ
  o –th- ɕ
  o –saha- ɕ
  o –ha- ɕ

• Particles for ‘with’ meaning
  o Saha ɕ
  o Samaga ɕ
  o Ha ɕ
  o katuva/kativa 㠸/㠸

• Emotive particle
  o Vismayartha- ha, aha, ah, aa, oho 㠸/ 㠸/ 㠸/ 㠸/ 㠸
  o Khedartha- ando, apoi, ayyo 㠸/ 㠸/ 㠸/ 㠸/ 㠸/ 㠸/ 㠸/ 㠸
  o Vedanartha- aai, uui, uu 㠸/ 㠸/ 㠸
  o Kuthisitartha- chi, chih, 㠸/ 㠸

• Particles expressing possibility
  o ho/ho ㄉ/ㄉ
  o do ㄉ
  o doho ㄉ

• Particles expressing similarity
  o Men ㄉ
  o Se ㄉ
  o Vani ㄉ
  o Bandu ㄉ
  o Van ㄉ
  o Ayuru ㄉ
  o Lesa ㄉ

• Comparative particles
  o vaDi ㄉ
  o vaDa ㄉ
  o vaDathma ㄉ

• Temporal particles
  o Nithi ㄉ
  o Nithin ㄉ
  o Nithora ㄉ
  o Ikbithi ㄉ
  o Dan ㄉ
  o Yali ㄉ
  o Puna ㄉ
  o Vali ㄉ
  o Ithin ㄉ

• Particles expressing duration
  o Sita ㄉ
  o Sitam ㄉ
  o Patan ㄉ
  o Thek ㄉ
  o Dakva ㄉ

• Quantitative particles
  o Sa ㄉ
  o Thak ㄉ
  o Thuvak ㄉ

• Particles expressing meanings of ‘too’/’at least
  o Pava ㄉ
  o Da ɕ
  o Ho ㄉ
  o Vath ㄉ

• Particles expressing exception
  o Vina ㄉ
  o Hara ㄉ
  o Misa ㄉ
  o Muhㄉ

• Question particles
  o Kima ㄉ
  o Kim ㄉ
  o Ay ㄉ
  o Manda ㄉ
  o Da ɕ

• Particles expressing alternative meaning
  o Hevath ㄉ
  o Nohoth ㄉ
  o Ho/ho ㄉ

• Time adverbial particle
  o Patha ㄉ
  o Pasa ㄉ
  o Gaane ㄉ

• Adjective/ adverb emphasizing particle
  o Thava ㄉ
  o Itha ㄉ
  o Boho ㄉ

• Emphasizing particles
  o –ma- ㄉ

• Quoting particle
  o –yi- ㄉ
  o –yayi- ㄉ

• Sentence ending particles
  o –ki- ㄉ
  o –ya- ㄉ
  o –yi- ㄉ
  o –yae- ㄉ

• Imperative particles
  o Epa ㄉ

• Negative particles
  o No+ noun base ㄉ
  o No+ verb ㄉ
  o A ㄉ

• Particles expressing manner of adjectives and adverbs
Vocative particles
- emba එබ описание
- embala එබල описание
- kola කොල описание
- bola බොල описание

Particles expressing a suggestion
- nam නම описание
- vanahi වනාහි описание
- kali කාල описание
- kal කල описание
- vukali සූකාල описание

Particles expressing straightforwardness
- indura ආදුර описание
- kelinma කොලිනාම описание

Particles focusing on attention
- anna අනන описание
- onna අනන описание
- menna මනන описание

Reported speech particles
- –lu- එලු описание
- –luu- එලුු описание

Particles stand for ‘for’
- Udesa උඩසා описание
- Pinisa පිනිසා описание
- Sandaha සඳහා описание

Reason particles
- Arabaya අරබයා описание
- Pinisa පිනිසා описание
- Nisa නිසා описание
- Bavin බාවන описание
- Heyin හේයින описание

Honorific particles
- Sek සැක описание
- Seka සැකා описание

Reporting particles
- Gana ගැන описание

Particles stand for the meaning of ‘according to’
- Anuva එනුව описание
- Paridi පාරිඩි описание

Conjunction particles
- Vita බිටි описание
- Nisa නිසා описание
- Bavin බාවන описание
- Thek එත්කෙ описание
- Thuru පොතු описание

Conditional particles
- –oth- එත් описание
- –hoth- එහි описание
- –nam- නම описание

Particles expressing reason for an action
- Pinisa පිනිසා описание
- Sandaha සඳහා описание
- Vas වස описание
- Men මෙන описание
- Lesa බල описание
- Paridi පාරිඩි описание

Particles used to link clauses
- namuth නමුත описание
- eheth එහොත описание

(The terms given here are subject to change.)

Krudantha (gerund) and Thaddhitha (suffixes which are used in noun inflection)

Krudantha
Sinhala gerund is a word which bears a verb base and an inflectional suffix. This suffix will convert the verb base into a gerund.

- Gerund
  - Present
  - Past

- Gerund –Present
  This gerund form is made by adding –na- suffix into the verb base.
  Verb base+ -na- = balanə

- Gerund –past
  There are six gerund suffixes can be identified in past gerund forms. They are:
  - uu- bal+uu- baluū
  - -a- vasi+aa vassaa
  - unu sinas+unu sinasunu
  - -L- athura athuLa
  - -T- basi baTa
  - -n- upadi upan
    Sindi sun

- Bhavartha krudantha /Emotive gerund

The Emotive gerund is normally used to express emotions.
  - –iim- nata+ iim -nətəiim
Most of the verb forms inflect according to these five procedures.

- *Shakyarth* krudantna / gerunds with the meaning of competence

Verb stem + *inya* - *Bala* + *iya* - *baliya*
*Daki* + *iya* - *dakiy*

- *Voice and Gerund*

The voice of a sentence is determined by the gerund suffixes.

- *Kapa* + *unu* - *kunu*
- *Kapa* + *ana* - *kana*

- *Thaddhitha*

*Thaddhitha* is the counterpart of *Krudantha* (Gerund), while both come under the category of affixes. Gerund suffixes function on the verb base and *thaddhitha* suffixes on the noun.

*Thaddhitha*
- *Animate*
  - Masculine - *natyaparu*
  - Feminine - *kreeda* + *ika* - *kreedaka*
  - Honorific - *puth* + *ana* - *puthanu*

- Inanimate

The range of the *thaddhitha* allomorphs;

- *Noun affixes*
  These affixes will occur with a noun base.
  Eg. *vasana* + *vantha* - *Vasanavantha*

- *Verb affixes*
  These affixes will occur with a verb base.
  Eg. *Kara* + *va* + *i* - *karavai*.

### Adjectives

Adjectives (*Nama Visheshana*) is a wording which describes the nature of the noun. Nominal nouns, pronouns, gerund nouns, thaddhitha nouns and numeral nouns occur as adjectives in Sinhala language. The nominal nouns can be again classified in two sections.

- *Jathi nama* – categorical nouns
- *Dravya nama* – metrial nouns
- *Sagna nama* – proper nouns
- *Guna nama* – qualitative

Nominal noun adjectives

- *Jathi nama* – *minis* + *ruva* - *minis ruva*
- *Dravya nama* – *lii* *veda* - *ivi veda*
- *Sagna nama* – *Mugalan + thera* - *Mugalan thera*
- *Guna nama* – *suvanda + mala* - *suvanda mala*

These examples present the probability of occurrence all the noun roots as adjectives in the nominal noun category.

- *Pronoun adjectives*
  *Apa* + *raTa* - *apa rat*
  *tham* + *defaya* - *tham defaya*

- *Gerund adjectives*
  *Hamana* + *sulanga* - *hamana sulanga*
  *Hindina* + *asuna* - *hindina asuna*

- *Thaddhitha adjectives*
  *Gunavath minisa* - *gunavath minisa*
  *Silvath himi* - *silvath himi*

- *Numeral adjectives*
  *Thun pola* - *thun pola*
  *Pasvana mahala* - *pasvana mahala*

### Adverbs

Adverbs can be divided into 4 sections according to their meaning.

- *Akarartha* (how the action is done)
- *Kalartha* (when the action is done)
- *Sthanartha* (where the action is done)
- *Hethvartha* (cause for the action)

Adverbs can be also divided into three sections according to the quality of their occurrence.

- *Adverb + verb* – *ohu vegayan duvai*
- *Adverb + adjective + noun* – *Lalith itha honda lamayek*
- *Adverb + adverb + verb* – *ohu boho semin riya padavai*. 
The Sinhala adverb set could be mainly categorized into two parts.

- Non-inflected adverbs (avyas/අවයස)
- Inflected adverbs
- Non-inflected adverbs (avyas)

These adverbs normally do not inflect according to the context. *Vaha, ithin, semin, pasulu, indura, yali, pili, puna punaa* are examples for the above mentioned adverbs.

- Inflected adverbs

The inflected adverbs are exclusively made out of instrumental, ablative, locative and dative case markers.

- Inflected particles as adverbs
  - *Vetha*
  - *Vethata*
  - *vethin*

- Temporal nouns as inflected adverbs
  - *Ada*
  - *Adata*
  - *Adin*

- Adverbs compounded with adjectives

<table>
<thead>
<tr>
<th>Adjective</th>
<th>noun</th>
<th>adverb</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sudu</em></td>
<td>suda</td>
<td><em>Sudata, sudin, sude</em></td>
</tr>
<tr>
<td><em>Rathu</em></td>
<td>ratha</td>
<td><em>Rathata, rathin, rathe</em></td>
</tr>
<tr>
<td><em>Kota</em></td>
<td>kota</td>
<td><em>Kotata, kotin, kote</em></td>
</tr>
</tbody>
</table>

**Further Information**

Greater number of particles has been used as adverbs. They are as follows.

*Men, vashayen, thek, thuvak, tharam, pamana, sa, va*, *kota, vara, vaara, vita.*

In some cases numerals reduplicate to make an adverb.

* Amma mal thuna thuna keduva
  * Denna denna enna*