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 nishpanna(words
that have local origin), thadbhava (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 1: Syntactic categories

<table>
<thead>
<tr>
<th>Nama</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kriya</td>
<td>Verbs</td>
</tr>
<tr>
<td>Nipatha (particles)</td>
<td>(Conjunctions, Prepositions, Interjections)</td>
</tr>
<tr>
<td>Kruantha</td>
<td>Gerunds</td>
</tr>
</tbody>
</table>

1 There is no exact category in English grammar that matches with Nipatha.
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>
<thead>
<tr>
<th>Word</th>
<th>Spelling variants(optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part of Speech</td>
</tr>
<tr>
<td></td>
<td>Pronunciation</td>
</tr>
<tr>
<td></td>
<td>Root Word</td>
</tr>
<tr>
<td></td>
<td>Etymology</td>
</tr>
<tr>
<td></td>
<td>Definition(s) (meaning(s))</td>
</tr>
<tr>
<td></td>
<td>Synonyms and cross references</td>
</tr>
<tr>
<td></td>
<td>Tamil and English translation equivalents</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>
</table>

Further description on this classification is available in Appendix D.
<table>
<thead>
<tr>
<th>Nama</th>
<th>Simple</th>
<th>Compound</th>
<th>Complex</th>
<th>Animateness</th>
<th>Gender</th>
<th>Number</th>
<th>Person</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kriya</th>
<th>Finite Verbs:</th>
<th>Number</th>
<th>Gender</th>
<th>Tense</th>
<th>Person</th>
<th>Voice</th>
<th>Transitivity</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constructed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex constructed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Infinite Verbs:     |         |        |        |       |       |       |               |       |
|---------------------|---------|--------|        |       |       |       |               |       |
| Purva Kriya (Past participle) |        |        |        |       |       |       |               |       |
| Mishra Kriya (Present participle) |        |        |        |       |       |       |               |       |
| Asambhavya Kriya (Conditional) |        |        |        |       |       |       |               |       |
| Prajojya Kriya (Causative) |        |        |        |       |       |       |               |       |
| Vidhi Kriya (Imperative) |        |        |        |       |       |       |               |       |
| Ashirvada Kriya (Blessing) |        |        |        |       |       |       |               |       |

| Nipatha |        |        |        |       |       |       |               |       |
|---------|--------|--------|        |       |       |       |               |       |

| Krudantha |        |        |        |       |       |       |               |       |
|-----------|--------|--------|        |       |       |       |               |       |
| Atheetha Krudantha (past) |        |        |        |       |       |       |               |       |
| Varthamana Krudantha (present) |        |        |        |       |       |       |               |       |

| Nama Vishesana |        |        |        |       |       |       |               |       |
|----------------|--------|--------|        |       |       |       |               |       |

| Kriya Vishesana |        |        |        |       |       |       |               |       |
|----------------|--------|--------|        |       |       |       |               |       |
| Akarartha (manner) |        |        |        |       |       |       |               |       |
| Kalartha (time) |        |        |        |       |       |       |               |       |
| Sthanartha (place) |        |        |        |       |       |       |               |       |
| Hethvarth (reason) |        |        |        |       |       |       |               |       |

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 researches, 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 Getapadaya
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
Dhahamgetamalaya

KANDY PERIOD
Akaradiya
Shabdha Mukthawaliya
Shuddha Sinhalaaya
Pali Sinhala Artha Kathana Granthaya
Pali Sinhala Katha Malawa
Perani Sidath Sangara Sannaya
Niganda Grantha
Vararuchi
Saraswathi
Vasudewa
Siddhuwadhara

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

APPENDIX B
Nama
This is the classification proposed by Cumaratunga Muniadasa in Vyakarana Vivaranaya for Nouns.

Vyakarana Vivaranaya –Cumaratunga Muniadasa

Bhinnaya Nama (Referring for specific things)
Jathi Nama (Referring for general things)
Ekaika Nama (Referring for one thing)

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

W. Bidgnell’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 Shabdakoshaya – 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 Shabdakoshaya
Rev. Veliwitiye Soratha – 1952

Nuthana Sinhala Paribhashika Shabdakoshaya
Harishchandra Wijetunge – 1978

Prayogika Sinhala Shabdakoshaya
Harishchandra Wijetunge – 1982

Abhinna Nama/ Sarva Nama (Pronoun)

Paurusha (Person)
Uttama Purusha (First Person)
Madyama Purusha (Second Person)
This is the classification proposed by Prof. W.S. Karunatilaka in Sinhala Bhasha Vyakaranaya for Nouns.

**Purusha (Third Person)**
- Prathama
  - Suchaka (Locative)
  - Prashna (Interrogative)
  - Aniyam (Indefinite pronoun)
- Bhedaka
  - Samudaya (Collective)
- Avadharana (Emphatic)

**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)**
- Prathama
- Karma
- Kartru
- Karana
- Sampradana
- Avadhi
- Sambandha
- Alapana

---

**Nama (Nouns)**
- Sarala Nama (Simple Nouns)
- Sankirana Nama (Complex Nouns)
- Samasa Nama (Compound Nouns)

**Nama Nama**
- Jathi Nama (Referring for general things)
- Kevala Nama (Referring for one thing)
- Samuharthavachi Nama (Referring more than one thing)

**Dravya Nama** (Referring for inanimate things)
- Guna Nama (Referring for modifiers)
- Bhava Nama (Referring for modifiers or verbal nouns)
- Sagna Nama (Proper Noun)

**Nama (Pronoun)**
- Sankya Nama (Countable Noun)
- Aganya Nama (Uncountable Noun)

**Pranavachi Nama** (Animate)
- Purusha Linga (Masculine)
- Stree Linga (Feminine)
- Manava (Human)
- Amanava (Non-Human)

**Napunsaka Linga** (Neuter)
- Podu Linga (Common Noun)

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

**Niyatha/Aniyatha** (Definite/Indefinite)

**Vibhakthi (Case)**
- Prathamad (Nominative)
- Karma (Accusative)
- Kartru (Instrumental)
- Karana (Auxiliary)
- Sampradana (Dative)
- Avadhi (Ablative)
- Sambandha (Genitive)
- Adhara (Locative)
- Alapana (Vocative)
This is the classification proposed by Prof. J.B.Disanayaka in *Nama Padaya* for Nouns.

**Nama Padaya – J.B.Disanayaka**

- **Sarala Nama** (Simple Nouns)
- **Sanyuktha Nama** (Compound Nouns)
  - Prana Bheda (Animateness)
  - Pranavachi Nama (Animate)
    - Sachethanika (Animate)
    - Pranabhasa (Inanimate things use as animate)
    - Pranaropitha (Inanimate things use as animate)
  - Apranavachi (Inanimate)
- **Vachana** (Number)
  - Eka Vachana (Singular)
  - Bahu Vachana (Plural)
- **Linga** (Gender)
  - Purusha (Person)
- **Niyatha** (Definiteness)
  - Niyatha (Definite)
  - Purusha Vachi (Masculine)
  - Stree Vachi (Feminine)
  - ApranaVachi (Inanimate)
- **Pranava** (Animateness)
- **Vachana** (Number)
- **Samadaya Nama** (Collective Noun)
  - Jathi Nama (Referring the names of trees)
  - Nipatha Nama (to convey the feelings of human beings)
  - Sagna Nama (Proper Noun)
- **Kriya Prakruti** (According to appending suffixes to the stem of the verb)
- **Bhava Nama** (Verbal Noun)
- **Krudantha Nama** (Gerund Noun)
Verbs

This is the classification proposed ancient “Sidath Sangara” in 13 century A.D.

**Kriya Padaya**

_Sidath Sangarawa_

“اهل كارما، عامل، دارا في الحالات الثلاث؛ كارما، نقص، وفظ.”

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akarmaka kriya (Intransitive)</td>
<td>अकार्मका क्रिया</td>
</tr>
<tr>
<td>Sakarmaka kriya (Transitive)</td>
<td>भार्मका क्रिया</td>
</tr>
<tr>
<td>Thraikalika kriya (Tense)</td>
<td>थ्राकिलिका क्रिया</td>
</tr>
<tr>
<td>Vidhi kriya (Imperative)</td>
<td>विद्धि क्रिया</td>
</tr>
<tr>
<td>Ashirvada kriya</td>
<td>अशिर्वदा क्रिया</td>
</tr>
<tr>
<td>Asambhavya kriya (Conditional Mood)</td>
<td>असमभाव्या क्रिया</td>
</tr>
<tr>
<td>Prayojya kriya (Causative)</td>
<td>प्रायोज्यो क्रिया</td>
</tr>
<tr>
<td>Karmakaraka kriya (Passive)</td>
<td>कर्मकरका क्रिया</td>
</tr>
<tr>
<td>Purva kriya (Past Participle)</td>
<td>पुर्व भागीरथिका क्रिया</td>
</tr>
<tr>
<td>Mishrakiya (Present Participle)</td>
<td>भूषर्किया प्रस्ततिकर्मका</td>
</tr>
</tbody>
</table>

This is the classification proposed by Cumaratunga Munidas in _Vyakarana Vivaranaya_ for Verbs.

**Kriya Vivaranaya – Cumaratunga Munidasa.**

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shuddha kriya (Simple verbs)</td>
<td>शुद्ध क्रिया</td>
</tr>
</tbody>
</table>

This is the classification proposed by Prof. J.B. Disanayaka in _Kriya Padaya for Verbs._

**Kriya Padaya – J.B. Disanayaka**

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>एका वचना – Eka Vachana (Singular)</td>
<td></td>
</tr>
<tr>
<td>बहु वचना (Plural) – Baha Vachana</td>
<td></td>
</tr>
<tr>
<td>परुष – Purusha (Person)</td>
<td></td>
</tr>
<tr>
<td>उत्तम परुष – Uttama Purusha (First Person)</td>
<td></td>
</tr>
<tr>
<td>मध्यम परुष – Madyama Purusha (Second Person)</td>
<td></td>
</tr>
<tr>
<td>तीर्थ प्राचार्य – Pratha Purusha (Third Person)</td>
<td></td>
</tr>
<tr>
<td>काल ब्देय – Kala Bhedaya (Tense)</td>
<td></td>
</tr>
<tr>
<td>अनाथिथा काला – Anathitha kala (Non-past tense)</td>
<td></td>
</tr>
<tr>
<td>अथ्थिथा काला – Athitha kala (Past tense)</td>
<td></td>
</tr>
<tr>
<td>पद ब्देय – Pada Bhedaya</td>
<td></td>
</tr>
<tr>
<td>पारस्म औप – Parasmai pada</td>
<td></td>
</tr>
<tr>
<td>अत्थिथा औप – Athmane pada</td>
<td></td>
</tr>
<tr>
<td>कारका ब्देया – Karaka Bhedaya (Voice)</td>
<td></td>
</tr>
<tr>
<td>कार्त्रुकरका – Kartrukarakaka (Active voice)</td>
<td></td>
</tr>
<tr>
<td>कार्मकरका – Karmakarakaka (Passive voice)</td>
<td></td>
</tr>
<tr>
<td>नाम क्रिया – Nama Kriya (Verbal Noun)</td>
<td></td>
</tr>
<tr>
<td>निपथा क्रिया – Nipatha Kriya</td>
<td></td>
</tr>
</tbody>
</table>

This is the classification proposed by Prof. J.B. Disanayaka in _Kriya Padaya for Verbs._

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarala kriya pada (Simple verbs)</td>
<td>सरला क्रिया पदादायी</td>
</tr>
<tr>
<td>Sarala Shuddha kriya pada (Simple verbs)</td>
<td>सरला शुद्धा क्रिया पदादायी</td>
</tr>
</tbody>
</table>
Appendix C

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

Table C.1: Syntactic categories used in the lexicon structure

<table>
<thead>
<tr>
<th>Sinhala</th>
<th>English</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>වශයුතු ප්‍රතිපාදය ප්‍රතිපාදය</td>
<td>Spelling variants</td>
<td>The alternative ways of writing the same word</td>
</tr>
<tr>
<td><strong>Table C.2: Syntactic categories used in the lexicon structure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sruthi / Nama</strong></td>
<td>Noun</td>
<td>A word that can be used to refer to a person or place or thing.</td>
</tr>
<tr>
<td><strong>Sruthi / Kriya</strong></td>
<td>Verb</td>
<td>Expresses some state or action affecting some person or thing.</td>
</tr>
<tr>
<td><strong>Sruthi / Nipatha</strong></td>
<td>Nipatha</td>
<td>The syntactic categories of prepositions, conjunctions and interjections.</td>
</tr>
<tr>
<td><strong>Sruthi / Krudantha</strong></td>
<td>Gerund</td>
<td>Words formed by adding suffixes to roots of verbs.</td>
</tr>
<tr>
<td><strong>Sruthi / Nama Visheshana</strong></td>
<td>Adjectives</td>
<td>The word class that denotes some quality or attribute of the noun.</td>
</tr>
<tr>
<td><strong>Sruthi / Kriya Visheshana</strong></td>
<td>Adverbs</td>
<td>The word class that qualifies verbs or clauses.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Number</td>
<td>Singularity or Plurality of nouns and verbs.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Singular</td>
<td>The form of a word that is used to denote a singleton.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Plural</td>
<td>The form of a word that is used to denote more than one.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Person</td>
<td>A grammatical category of pronouns and verb forms.</td>
</tr>
<tr>
<td><strong>Sruthi / Srutha</strong></td>
<td>First person</td>
<td>Pronouns and verbs used to refer to the speaker.</td>
</tr>
<tr>
<td><strong>Sruthi / Srutha</strong></td>
<td>Second person</td>
<td>Pronouns and verbs used to refer to the person addressed.</td>
</tr>
<tr>
<td><strong>Sruthi / Srutha</strong></td>
<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><strong>Sruthi</strong></td>
<td>Gender</td>
<td>A grammatical category based on sex or animateness.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Masculine</td>
<td>A grammatical category that refers to males or to objects classified as male.</td>
</tr>
<tr>
<td><strong>Sruthi</strong></td>
<td>Feminine</td>
<td>A grammatical category that refers to females or to objects classified as...</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>Tamil</td>
<td>English</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>மாலை</td>
<td>Neuter</td>
<td>A grammatical category that refers to inanimate objects (neither masculine nor feminine).</td>
</tr>
<tr>
<td>கலை</td>
<td>Common</td>
<td>A grammatical category that refers to both masculine and feminine.</td>
</tr>
<tr>
<td>அமர்க்கல்/ஏமார்க்கல்</td>
<td>Animateness</td>
<td>A grammatical category that denotes living beings or nonliving things.</td>
</tr>
<tr>
<td>அமர்க்கல் (அமர்க்கல்)</td>
<td>Animate</td>
<td>A grammatical category that denotes living beings.</td>
</tr>
<tr>
<td>அமர்க்கல் (அமர்க்கல்)</td>
<td>Inanimate</td>
<td>A grammatical category that denotes nonliving beings.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Case</td>
<td>A grammatical case (nouns or pronouns or adjectives related in some way to other words in a sentence).</td>
</tr>
<tr>
<td>நிறுவல் குருத்தர்</td>
<td>Nominative case</td>
<td>The category of nouns serving as the grammatical subject of a verb.</td>
</tr>
<tr>
<td>பொருண்டு குருத்தர்</td>
<td>Accusative case</td>
<td>The category of nouns serving as the direct object of a verb.</td>
</tr>
<tr>
<td>வைணவ குருத்தர்</td>
<td>Vocative case</td>
<td>The category of nouns used when a person or thing is addressed.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Tense</td>
<td>A grammatical category of verbs used to express distinctions of time.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Past tense</td>
<td>A verb tense that expresses actions or states in the past.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Non-past tense</td>
<td>A verb tense that expresses actions or states in the Present or Future.</td>
</tr>
<tr>
<td>வேளை</td>
<td>Voice</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>வேளை வேளை</td>
<td>Active voice</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>வேளை வேளை</td>
<td>Passive voice</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>குருத்தர்/குருத்தர் நிறுத்த குருத்தர்</td>
<td>Transitivity</td>
<td>the grammatical relation created by a transitive or intransitive verb.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Transitive</td>
<td>Denotes an action passing from the doer to an object.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Intransitive</td>
<td>Express a state or condition, or an action or feeling that does not pass over to an object.</td>
</tr>
<tr>
<td>குருத்தர்</td>
<td>Predicate</td>
<td>Final verb of a sentence.</td>
</tr>
<tr>
<td>போரு குருத்தர் / Akyatha Kriya</td>
<td>Present Participle</td>
<td>The predicates of the subject which precede its final predicate.</td>
</tr>
<tr>
<td>போரு குருத்தர் / Asambhavya Kriya</td>
<td>Conditional Mood</td>
<td>Expresses uncertainty or the time when an action is done.</td>
</tr>
<tr>
<td>போரு குருத்தர் / Mishra Kriya</td>
<td>Present Participle</td>
<td>Expressing present action or the actions doing in the same time.</td>
</tr>
<tr>
<td>போரு குருத்தர் / Atheetha Krudantha</td>
<td>Gerund – past</td>
<td>Words formed by adding suffixes of past tense to roots of verbs.</td>
</tr>
<tr>
<td>போரு குருத்தர் / Varthamana Krudantha</td>
<td>Gerund – present</td>
<td>Words formed by adding suffixes of present tense to roots of verbs.</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.

Animate nouns
- Masculine-
  - Singular – Definite/Indefinite
  - Plural
- Feminine-
  - Singular – Definite/Indefinite
  - Plural
- Common-
  - Singular- Definite/Indefinite
  - Plural

Inanimate nouns
- Neuter-
  - Singular – Definite
  - Indefinite
  - Plural

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 (ඉංග්‍රීසි)
2. Accusative (ඉංග්‍රීසි)
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 (ඉංග්‍රීසි)
The auxiliary case is marked with *visin* (ඉංග්‍රීසි) 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>Case</th>
<th>Case marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental (ඉංග්‍රීසි)</td>
<td>sin pl</td>
</tr>
<tr>
<td>-en-</td>
<td>-walin-</td>
</tr>
<tr>
<td>-in-</td>
<td>-hata-</td>
</tr>
<tr>
<td>Dative (ඉංග්‍රීසි)</td>
<td>-ta-</td>
</tr>
<tr>
<td>-walata-</td>
<td>-hata-</td>
</tr>
<tr>
<td>Ablative (ඉංග්‍රීසි)</td>
<td>-gen-</td>
</tr>
<tr>
<td>-ener-</td>
<td>-keren-</td>
</tr>
</tbody>
</table>

Table D.1: Case Markers
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 ජාති, guna nama ඇති, අධි අධි etc. And will also focus on the features like onomatopoeia and reduplication.

Verb
- Verb
  a. Finite verb
  b. Non finite verb
- Finite verbs: a. Simple constructed verbs
  b. Complex constructed verbs

<table>
<thead>
<tr>
<th>Possessive/genitive (මිතුමා)</th>
<th>-ගේ</th>
<th>-ගේ</th>
<th>-වාල-</th>
<th>Locative (ලීත)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocative (ප්‍රශ්න)</td>
<td>Zero</td>
<td>-ි</td>
<td>-ගේ</td>
<td>-වාල-</td>
</tr>
</tbody>
</table>

Table D.2: Tense

<table>
<thead>
<tr>
<th>Non past</th>
<th>Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>singular</td>
<td>1st person balami balannemi balimi baluvemi</td>
</tr>
<tr>
<td></td>
<td>2nd person balahi balannehi balihi baluvehi</td>
</tr>
<tr>
<td></td>
<td>3rd person balai Balanneya – mas/neuter Balanniya-fem</td>
</tr>
<tr>
<td>Plural</td>
<td>1st person balamu balannemu balimu baluwemu</td>
</tr>
<tr>
<td></td>
<td>2nd person balahu balannahu baluhu baluvahu</td>
</tr>
<tr>
<td></td>
<td>3rd person balati Balannahu/balannoya baluha baluvoya</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)
  - Active voice (kartru karaka)
  - Passive voice (karma karaka)

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

- Ohu veDa karai.  \(\text{Active}\)
- Ohugen veDa kerei.  \(\text{Passive}\)

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 involutive forms.

- Lamaya natai \(\text{volitive}\)
- Bera gahana vita lamaya netavei.  \(\text{involutive}\)

Causative verb
Verb stem+ -va-

**Imperative verb**

<table>
<thead>
<tr>
<th>2\text{nd} person</th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero, va</td>
<td>vu</td>
<td></td>
</tr>
</tbody>
</table>

**Blessing verb**  \(\text{(Ashirvada kriya)}\)
Present/ future tense+ va

<table>
<thead>
<tr>
<th>1\text{st} person</th>
<th>Sin</th>
<th>Pres/fut tense+ mi/-im</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\text{st} person</td>
<td>pl</td>
<td>Pres/fut tense+ nau/-mo-</td>
</tr>
<tr>
<td>3\text{rd} person</td>
<td>pl</td>
<td>Pres/fut tense+ t/-tu-</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

\(\text{Pata} \text{(noun)}+ \text{karanava}\)

**Pasu (adj)+ (karanava)**
**Kapal\(\text{(non-finite verb form)}+ \text{danava}\)**

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 \(\text{(krudanrha)}\)
- b. past participle \(\text{(purva kriya)}\)
- c. present participle \(\text{(misra kriya)}\)
- d. conditional verbs \(\text{(asambhavya kriya)}\)

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

- **The past participle** \(\text{(purva kriya)}\)
  Verb stem+ -aa/-x/-ii-
  Balana/ xda/ xdi
  
- **present participle** \(\text{(misra kriya)}\)
  verb stem+ -min/-mina/-mini/-mni
  balamin/balamina/balamna/balamini

- **conditional verbs** \(\text{(asambhavya kriya)}\)
  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**

<table>
<thead>
<tr>
<th>Case</th>
<th>Particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>auxiliary</td>
<td>-visin-. ෂ්ද්</td>
</tr>
<tr>
<td>Ablative</td>
<td>-gen-.keren  මිඩය</td>
</tr>
<tr>
<td>dative</td>
<td>-(ha)ita-. මුරු</td>
</tr>
<tr>
<td>possessive</td>
<td>-ge. විඩය</td>
</tr>
<tr>
<td>locative</td>
<td>-kereni-. මුරු, විඩය</td>
</tr>
<tr>
<td>tula, matha, ud</td>
<td></td>
</tr>
</tbody>
</table>
Comparative particles
- -du- ꂧ
- -th- ꂧ
- -saha- ꂧ
- -ha- ꂧ

Particles for ‘with’ meaning
- Saha ꂧ
- Samaga ꂧ
- Ha ꂧ
- katuva/kativa ꂧ/(cb)

Emotive particle
- Vismayartha- ha ,aha, ah, aa,oho ꂧ/cb/ cb/ cb/ cb
- Khedartha- ane,ando, apoi, ayyo ꂧ/cb/ cb/ cb/ cb
- Vedanartha- uai,uai,uu ꂧ/cb/ cb/ cb
- Kuthisithartha- ch, chih, ꂧ/cb

Particles expressing possibility
- ho/ho ꂧ/cb/ cb
- do ꂧ
- doho ꂧ/cb/ cb

Particles expressing similarity
- Men ꂧ
- Se ꂧ
- Vani ꂧ
- Bandu ꂧ
- Van ꂧ
- Ayuru ꂧ
- Lesa ꂧ

Comparative particles
- vaDi ꂧ
- vaDu ꂧ
- vaDatha ꂧ/cb/ cb

Temporal particles
- Nithi ꂧ
- Nithin ꂧ
- Nithora ꂧ/cb/ cb
- Ikbihi ꂧ
- Dan ꂧ
- Yali ꂧ
- Puna ꂧ
- Vali ꂧ
- Ithin ꂧ

Particles expressing duration
- Sita ꂧ
- Sitan ꂧ
- Patan ꂧ
- Thek ꂧ
- Dakva ꂧ

Quantitative particles
- Sa ꂧ
- Thak ꂧ
- Thuvak ꂧ

Particles expressing meanings of ‘too’/’at least
- Pava ꂧ
- Da ꂧ
- Ho ꂧ
- Vath ꂧ

Particles expressing exception
- Vina ꂧ
- Hara ꂧ
- Misa ꂧ
- Math ꂧ

Question particles
- Kima ꂧ
- Kim ꂧ
- Ay ꂧ
- Manda ꂧ
- Da ꂧ

Particles expressing alternative meaning
- Hevath ꂧ
- Nohoth ꂧ/cb/ cb
- Hofo ꂧ/cb/ cb

Time adverbial particle
- Patha ꂧ
- Pasa ꂧ
- Gaane ꂧ/cb/ cb

Adjective/ adverb emphasizing particle
- Thava ꂧ
- Itha ꂧ
- Boho ꂧ/cb/ cb

Emphasizing particles
- -ma- ꂧ

Quoting particle
- -yi- ꂧ
- -yayi- ꂧ

Sentence ending particles
- -ki- ꂧ
- -ya- ꂧ
- -yi- ꂧ
- -yae- ꂧ

Imperative particles
- Epa ꂧ

Negative particles
- No+ noun base ꂧ+ ꂧ
- No+ verb ꂧ+ ꂧ
- A ꂧ

Particles expressing manner of adjectives and adverbs
Particles stand for the meaning of 'according to'.

- **Vocative particles**
  - emba එම
  - embala එමළ
  - kola ගොල
  - bola අොල

- **Particles expressing a suggestion**
  - nam නාම
  - vanahi වණහි
  - kali මාලි
  - kalාම
  - vikal මාලී

- **Particles expressing straightforwardness**
  - indura දුර
  - kelinna මැටීනා

- **Particles focusing on attention**
  - anna අන්න
  - onna අණන
  - menna මේනන

- **Reported speech particles**
  - –la- බු
  - –lu- බු

- **Particles stand for ‘for’**
  - Udesa උදෙසා
  - Pinisa පිනිසා
  - Sandaha සංදහා

- **Reason particles**
  - Arabaya අරංබයා
  - Pinisa පිනිසා
  - Nisa නිසා
  - Bavin බවීනා
  - Heyin භේයිනා

- **Honoric particles**
  - Sek සෙක
  - Seka සීක

- **Reporting particles**
  - Gana ගාන

- **Particles stand for the meaning of ‘according to’**
  - Anuva අනුව
  - Paridi පරිදී

- **Conjunction particles**
  - Vita විට
  - Nisa නිසා
  - Bavin බවීනා
  - Thek පේක
  - Thuru මොක

- **Conditional particles**
  - –oth- බු
  - –both- බාතා
  - –nam- බාම

- **Particles expressing reason for an action**
  - Pinisa පිනිසා
  - Sandaha සංදහා
  - Vas සව
  - Men සොන
  - Lesa සොන
  - Paridi බු

- **Particles used to link clauses**
  - namuth නමුත්
  - eheika එහාම

(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- = balana

- **gerund –past**
  - There are six gerund suffixes can be identified in past gerund forms. They are:
    - -uu- bala+uu- bāľaľu
    - -அ- vasi+அ vänśa
    - uuna sinase+uන sīnasunu
    - -L- athura athula
    - -T- basi baTa
    - -n- upadi upan

- **Bhavartha krudantha /Emotive gerund**

The Emotive gerund is normally used to express emotions.

- –im- nata+ im -nātīm
Most of the verb forms inflect according to these five procedures.

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

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

- **Voice and Gerund**

  The voice of a sentence is determined by the gerund suffixes.
  - Kapa+ -unu- kapidunu
  - Kapa+ -ana- kapanana

- **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 – natya+karu
    - Feminine – kreedika
    - Honorific – puth+aNu 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 – metirial nouns**
- **Sagna nama – proper nouns**
- **Guna nama – qualitative**

**Nominal noun adjectives**

- **Jathi nama – minis+ ruva** uttha ruva
- **Dravya nama – lii vella** velle
- **Sagna nama – Mugalan +thera** maula theru
- **Guna nama – svanda +mala** savanda mala

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

- **Pronoun adjectives**
  Apa + raTa tha+ defaya

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

- **Thaddhitha adjectives**
  Gunavath minisa
  Silvath himi

- **Numeral adjectives**
  Thun pola
  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 vegaven dava**
- **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/ अवयस्)
- Inflected adverbs
- Non-inflected adverbs (avyas)

These adverbs normally do not inflect according to the context.

\textit{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

\textit{Adin}

- Adverbs compounded with adjectives

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

Further information

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

\textit{Men, vashayen, thek, thuvak, tharam, pamana, sa, va, kota, vara, vaara, vita.}

In some cases numerals reduplicate to make an adverb.

\textit{Amma mal thuna thuna keduva}

\textit{Denna denna enna}