



# Characteristics of Sinhala Pronunciation

Ruvini Ramanayake  
University of Moratuwa, Sri Lanka



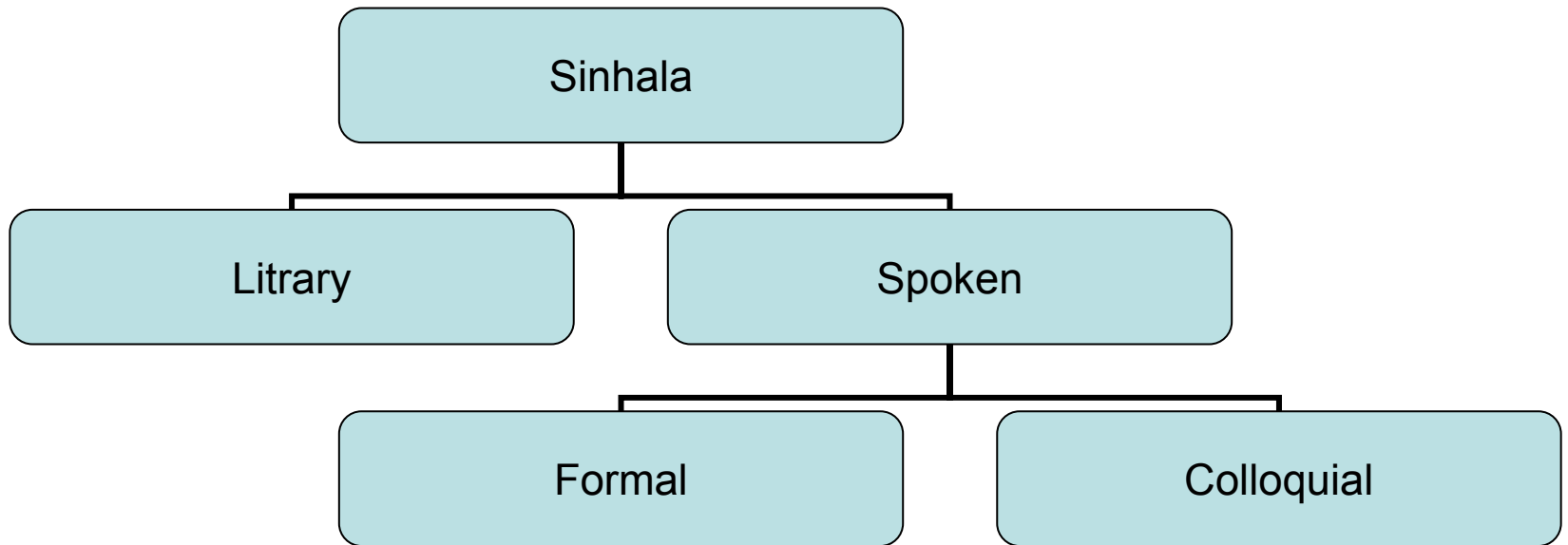
# Agenda

- Introduction to Sinhala sound system
- Sinhala Localization Initiatives
- Sinhala Unicode encoding Issues
- Impact of loan words
- Pronunciation exceptions in Sinhala
- Conclusion



# Introduction

Sinhala is a phonetic language. Letter to sound conversion for Sinhala usually has simple one to one mapping between orthography and phonemic transcription for most Sinhala letters.



# Sound System



- Sinhalese has 14 vowels sounds, seven of them are short and the other seven are long. Two of these vowels are unique to the language. They are represented as “æ” and “æ:” and not found in Indo-Aryan or Dravidian.
- There are 26 consonants of which four are prenasalized stops. The prenasalized sounds indicated as
  - G Sanyaka gayanna
  - D Sanyaka dayanna
  - Ḍ Sanyaka ḍayanna
  - B Amba bayanna

# Sinhala Localization Initiatives



- Encoding of Sinhala characters
- Development of Sinhala fonts
- Standardisation of a Sinhala keyboard and
- Standards-based applications and utilities (such as spelling checkers).

# Sinhala Unicode encoding Issues



## Deficiencies in the Unicode encoding

- It was observed that the Unicode encoding suffered from a number of shortcomings. These are:
- lack of encodings for conjunct letters such as
- lack of encodings for the *yansaya*, *rakaransaya* and *rephaya*,
- lack of guidance on the use of multiple vowel modifiers
- lack of guidance on the encoding of non-standard letters, such as and .

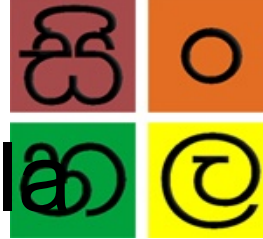


# Impact of loan words

The impact of English loan words on the structure of modern Sinhala is more on Spoken Sinhala (SS) than on Literary Sinhala (LS)

In the field of phonetics of SS the number of allophones was increased by the addition of the open back vowel [ɔ] as in <orange>, <toffee>, <office> and the addition of the long central vowel /: / as in <shirt>, <nurse>, <skirt>.

- <nurse> /nēs nōna /
- <purse> /pōs ekə /
- <sir> /sæ:r /



# Pronunciation exceptions in Sinhala

- Pronunciations of some words differ based on the context and their parts of speech.

Eg: “කර” /karə/ or /kara/.

“වන” /vanə/, /vənə

“කළ” /kalə/, /kələ/

- Some words are written differently from pronunciations.

Eg:





# Conclusion

- Though the sinhala script is phonetic there are exceptions that should be handled differently
- There's few documentations available in Sinhala speech domain
- Less research in Sinhala speech synthesis
- Pronunciation disambiguaties



# References

- Gihan Dias and Aruni Goonetilleke, “Development of Standards for Sinhala Computing”, in *Proc. 1st Regional Conference on ICT and E-Paradigms*, Colombo, Sri Lanka, 2004. Available: [http://www.fonts.lk/doc/sinhala\\_standards.pdf](http://www.fonts.lk/doc/sinhala_standards.pdf)
- Karunatilake, W.S., An introduction to spoken Sinhala, 3rd edn., M.D. Gunasena & Co. Ltd., 217, Olcott Mawatha, Colombo 11 (2004)
- Ruvan Weerasinghe, Asanka Wasala and Kumudu Gamage. 2005. A Rule Based Syllabification Algorithm for Sinhala, *Proceedings of 2nd International Joint Conference on Natural Language Processing (IJCNLP-05)*, p. 438-449, Jeju Island, Korea



Thank you