



510/E1

On computing Sinhala grammar

B Hettige¹ and A S Karunananda²

¹*Department of Statistics and Computer Science,
University of Sri Jayawardenepura, Nugegoda.*

²*Faculty of Information Technology, University of Moratuwa, Moratuwa.*

The computational model of grammar for the Sinhala language has been developed by considering the morphology and the syntax of Sinhala language. Finite State Transducers (FST) and Context-free grammar (CFG) have been used to describe the computational grammar for Sinhala.

The Sinhala language is a morphologically rich language when compared to the English language. Prakurthi, Pratyā, Thaddhita and Upasarga are the morphological components of the Sinhala language. In addition to the above components, Sandhi rules are used to join two or more words in Sinhala. Further, Nama Gana and Kriya Gana show how each of the nouns and verbs are derived from its base form. These "gana" also gives the theoretical basics for the concept of Varanegeema (Conjugation) in the Sinhala language. To implement the Sinhala grammar, we have used 85 grammar rules for Sinhala Nouns considering the Nama gana in the Sinhala language. To implement the Kriya Gana, 18 rules have been considered. By using these computational grammars, the Sinhala morphological generator has been developed. This morphological generator can generate all the word forms for the given Sinhala base word.

Considering the syntax of the Sinhala language, a Sinhala sentence can be divided into eight components namely Attributive adjunct of Subject, Subject, Attributive adjunct of Object, Object, Attributive adjunct of Predicate, Attributive adjunct of the complement of Predicate, Complement of Predicate and the Predicate. These components are building blocks of the Sinhala sentence. By using the context-free grammar, Sinhala parser has been developed. This Sinhala parser can analyze the given simple sentences with the simple and complex subject and the object.

The Sinhala morphological generator and the Sinhala parser have been used to implement the English to Sinhala machine translation system. By using 400 sample sentences, the English to Sinhala machine translation system (BEES) has been evaluated. The Sinhala morphological generator and the Sinhala parser work with reasonable accuracy.

buddhitha@yahoo.com

Tel: 0723915391