

Translation of a Given Simple English Sentence into its Equivalent in Sinhala using a Speech Synthesizer

R. I. P. Wickramasinghe¹, K. H. Kumara² and N. G. J. Dias³

Machine Translation (MT) or Automatic Translation is generally concern about automating all or part of the process of translating one human language to another language. These human or natural languages bear similarities as well as differences due to the way these languages have organized. Therefore, translating from one natural language into another natural language depends on their vocabulary, grammar, and conceptual structure. The translation Text-To-Speech (TTS) can be considered as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter. To our knowledge there is no such a system in Sri Lanka that could translate simple English sentences into its equivalent in Sinhala with the relevant speech synthesis. In a country like Sri Lanka where the language barrier is a major issue, this type of systems will definitely help to reduce these language problems. With language translation coupled with TTS synthesis would be a good Computer Aided Learning Technique, that will provide a tool to learn English effectively.

This paper discusses an approach to translation with a speech synthesizer of a given simple English sentence into its Sinhala equivalent. The problem of translation is handled in two phases, namely the lexical selection, where appropriate target-language lexical items are chosen for each source-language lexical item and then the lexical reordering, where the chosen target-language lexical items are arranged to produce a meaningful target language string. Together with translator, here we use the concatenative synthesizer which is embedded in the speech units to be chained up. In our speech synthesizer, speech units that are typically smaller than words are used to synthesize speech from arbitrary input text. Speech units are algorithmically extracted from a phonetically transcribed speech data set. The unit selection process involves a combinatorial search over the entire speech corpus using the search algorithms.

Due to the practical difficulties and complexities this translation and TTS is tested only for the simple English sentences in which only SVO (Subject/Verb/Object) structure can be seen.

Key Words: Language Translation; TTS; English; Machine Translation; Sinhala

¹ Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka. E-mail: indika3@kln.ac.lk

² Department of Mathematical Sciences, Wayamba University of Sri Lanka, E-mail:harshakumaralk@yahoo.com

³ Department of Statistics & Computer Science, University of Kelaniya, Sri Lanka