Aligarh Journal of Linguistics

Editor-in-Chief
Shabana Hameed

Editors
Masood A. Beg
Nazrin B. Laskar

CONTENTS

Theoretical

TAMIL PRONOUNCING DICTIONARY: A PROMISING TREND IN LANGUAGE TECHNOLOGY DEVELOPMENT

Thennarasu Sakkan*
thennarasus@gmail.com

Abstract:

This paper aims to present a machine readable Tamil pronouncing dictionary. This dictionary contains 1,700 unique basic words along with their category, transliteration, phonetic transcriptions and word with sentence example. Issues in the development of the pronouncing dictionary are described in this paper. The use of these resources is totally free for any research purposes in order to speed up Tamil language speech technology research.

1. Introduction:

Natural language is very important for human interaction and also it is necessary for communication with machines. In the past years there has been a significant growth in research including the area of speech recognition; but the speech recognition systems are mainly developed for English and Tamil language although the used techniques are mostly language independent. The lack of freely usable linguistic resources is the main problem in development of large vocabulary continuous speech recognition (LVCSR) systems for other languages (Domokos, J., et al. 2014).

Though large amount of research works have gone into the Tamil language it appears to be a language with very limited language resources such as transcribed and annotated speech corpora or pronunciation dictionaries that have no copyright. Thus only an insignificant number of speech recognition systems exist to deal with the Tamil language. Excepting the recently introduced Google Voice Search, there is no other fully functional free-to-use LVCSR system for Tamil although it is reported that there are several Tamil language automatic speech recognition (ASR) systems (Nayeenmulla et.al 2001,

Affiliation: *Dept of Linguistics, Central University, Kerala