

# A Hindi - Telugu Bi-directional Machine Translation System

Uma Maheshwar Rao<sup>1</sup>, G., K.V. Subbarao<sup>2</sup>, Christopher Mala<sup>3</sup> and Bindu Madhavi<sup>4</sup> B.

Center for Applied Linguistics and Translation Studies

University of Hyderabad

Hyderabad, India

[guraohyd@yahoo.com](mailto:guraohyd@yahoo.com), {<sup>2</sup>kvs2811,<sup>3</sup>efthachris,<sup>4</sup>madhavihcu}@gmail.com

## Abstract

We present the development of a Machine Translation (MT) System which translates texts from Telugu to Hindi and vice-versa (Bi-directional). This MT system was developed as part of IL-ILMT consortium project funded by Govt. of India at CALTS, University of Hyderabad. This Machine Translation system uses Transfer Based Approach. System's Architecture is divided into three stages i.e. Source language Analysis module (SL), Source language to Target language Transfer module (SL-TL) and Target language generation module (TL). The computational Modules that are used in the building of this system were developed mainly by CALTS-UoH, IIIT-H and AUKBC research teams[5]. We also use the statistical open source engine i.e. CRF++ for POS-Tagging, Chunking and Named Entity Recognizer (NER). Hence the architecture is a hybrid one.

The development of Machine Translation (MT) is one of the most challenging tasks of Natural Language Processing Applications[1]. In MT there are a number of methods that are being practiced all over the world, chiefly, they are *Direct Methods*, *Interlingual Methods*, *Transfer Based Approach* and a combination of these beside the statistical and corpus based methods. Hindi and Telugu are two Indian languages, which belong to the Indo-Aryan and the Dravidian language family respectively. These languages exhibit a considerable degree of diversity at every level viz. morphological, syntactic, semantic and lexical levels[2]. Keeping these in mind, building a Machine Translation System for this language pair using Transfer based Method can be non-trivial and challenging[1]. The present paper discusses the successful implementation of the Transfer Based Approach to the Machine Translation (MT) System for the Telugu-Hindi pair. The system is an assembly of various linguistic modules run on specific engines whose output is sequentially maneuvered and modified by a series of modules till the output is generated. The most crucial linguistic modules include, a Morphological Analyzer (MA)[1][4][6], Parts of Speech Tagger (POS-T), a Simple Parser (SP), the Transfer Grammar Component (TG), a Lexical Transfer module consisting of a Bilingual Dictionary and a Conceptual Dictionary, an Agreement module (AGR) and a Morphological Generator (Wordgen)[3][6]. The system is already built and is now being tested and evaluated. The presentation would involve the demonstration of a randomly selected text from the internet.

## References:

- 1.Akshar Bharathi, Vineet Chaitanya and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. New Delhi:Prentice Hall of India.
2. Parameshwari K, Uma Maheshwar Rao. G, Krupanandam. N, Christopher Mala. 2010. *New Perspectives in Linguistics (Silver Jubilee Commemoration Volume)*. pg.74-84 Department of Linguistics, Bharathiar University, Coimbatore.
- 3.Uma Maheswar Rao G. and Christopher M. 2010. *Word Synthesizer Engine*. In Morphological Analyzer and Generators. Mona Parakh (ed.) Page 73-81. Mysore; CIIL.
- 4.Uma Maheswar Rao G. and Parameshwari K. 2010. *On the Description of Morphological Data for Morphological Analysers and Generators: A case of Telugu, Hindi and Kannada*. In Morphological Analyzer and Generators. Mona Parakh (ed.) Page 114-123. Mysore; CIIL.
- 5.ILMT Consortium. 2007. *ILMT SRS and Functional Specifications (mimeo)*. Hyderabad.
- 6.Uma Maheswar Rao G, Amba P. Kulkarni and Christopher M. 2007. *Functional Specifications of*

*Morphology (mimeo)*. Hyderabad.