

English-Telugu Machine Translation

Uma Maheshwar Rao¹, G., Rajyarama², K., Bharathi³, V.D., Santosh⁴, J.

CALTS, University of Hyderabad

¹guraohyd@yahoo.com, {²krajyarama, ³vijaya.anthony, ⁴santosh.jena@gmail.com}

Abstract

Greater amounts of information are being created and available only in a few dominant languages, which are unknown to vast majority of masses. Learning the dominant language to access information pushes the other languages to the marginal status and ultimately towards the loss of that language and the associated culture. MT has transcended from the early years when it was harshly criticized for its inefficiency and infeasibility. Modern approaches with linguistic knowledge at its core have greatly changed the perception [1]. There have been dozens of sustained efforts in making MT not only a possibility (constraints exists) but also a necessary activity. The approach that is considered is the transfer based approach but involves deeper amounts of analysis at different levels. The crucial and significant part of the system is focus on the disambiguation at every level of the analysis [2]. In this regard the current system geared up to take care of the divergence between English and Telugu [2]. The transfer grammar component with a considerable number of rules bridge the gap of the structural divergence between the two languages. The system's architecture involves Morphological Analyzer, POS tagger, Chunker, and parser which are from the open source platforms and adapted and incorporated into the system. In addition to these the system uses various modules such as, transfer grammar module, the bilingual English-Telugu lexical substitution module, module for computing tam and agreement and the Telugu generator. The lexical substitution module is supported by about 60k strong English – Telugu bilingual lexicon. Some important aspects of the system include the artificially created lexical equivalents to negative pronouns of English, viz. nobody, never, nowhere, nothing etc.

Ex. Eng. Nobody knows where the oranges were grown first.

Tel. ~ఎవడూ ఎరగడు నారింజలు మొదటగా ఎక్కడ పెంచబడ్డాయో

Currently, the complex syntactic structures like multiple embeddings and coordinate structures are usually considered hard to translate by the system. The present paper is an effort demonstrate and evaluate the current performance of the system in the context of the divergent linguistic structures of English and Telugu.

Reference :

1. Krishnamurti. Bh and J.P.L. Gwyun. 1985. *A grammar of modern Telugu*. Delhi; Oxford University Press.
2. Uma Maheshwara Rao, G. 1999. *A Morphological Analyser for Telugu* (electronic form) , Hyderabad: University of Hyderabad.
3. Bharati Akshar, Chaitanya, Vineet., and Sangal, Rajeev. 1995. *Natural Language Processing. A Paninian Perspective*. New Delhi : Prentice Hall of India.