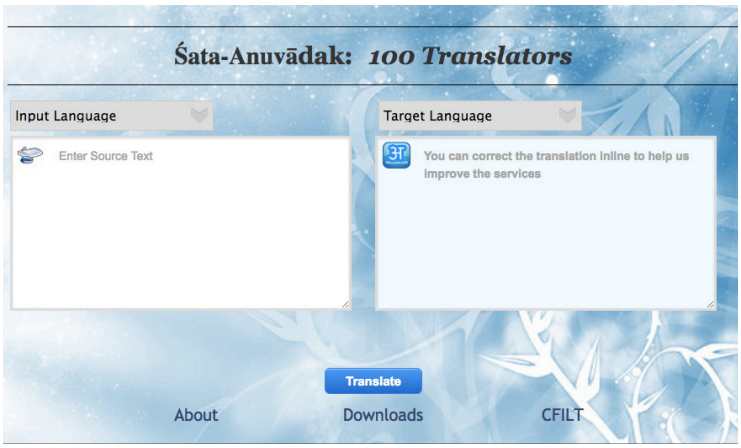
Machine translation

Machine translation (MT) deals with automatic translation of text from one natural language to another. It is one of the most challenging problems in natural language processing (NLP), requiring knowledge from all sub-areas of NLP. In an increasingly connected world, human interaction requires crossing language barriers in the government, business, social and cultural spheres. MT is a key technology to overcome these language barriers.

For nearly 20 years, we at the Centre for Indian Language Technology (CFILT), has been working on these problems using various MT paradigms: (i) inter-lingua based (ii) transfer-based (iii) statistical (iv) neural based. Currently, the MT research revolves around the state-of-the-art statistical methods and promising neural methods based on deep learning. Indian languages are a major focus, where several key use-cases have been explored: (i) Inter-Indian language translation, (ii) English-to-Indian language translation, and (iii) Indian language-to-English translation. The research in this area also formed the foundation for Prof. Bhattacharyya's book 'Machine Translation' published by CRC Press.

Via projects sponsored by Ministry of IT, Ministry of MHRD, Xerox Research and Accenture Global R&D, CFILT's contribution to machine translation research includes:

- Studies of translation divergences and translation evaluation measures in the context of Indian languages
- One of the first statistical MT systems in Indian languages
- Addressing syntactic differences in statistical MT via source reordering
- Incorporating semantic information for generation of fluent translations
- Addressing morphological richness of Indian languages
- Leveraging the similarity between Indian languages for reducing resource usage
- Parallel corpora for multiple Indian languages in collaboration with multiple institutes
- Shata-Anuvaadak, an web-based statistical machine translation system for 110 Indian language pairs (www.cfilt.iitb.ac.in/indic-translator)
- Transliteration system and corpora - transliteration is a sub-problem to be addressed for MT (www.cfilt.iitb.ac.in/brahminet)



Prof. Pushpak Bhattacharyya, Department of Computer Science and Engineering, pb@cse.iitb.ac.in