

Machine translation of English text to Kashmiri text

Neeha Ashraf

Department of Computer Science, University Of Kashmir, Srinagar-190006, India.

E-mail: neehaashraf@gmail.com

Abstract: Language of human beings is diverse and multifarious. More than 5000 dialects exist on earth which mirror the phonetic differing qualities. In annuals of Anthropology, language is considered as a primary tool for studying the culture of a civilization. Culture is transmitted through language. Humans learn their culture through language. Translation is important for communication but the problem is that the demand for translation in the modern world far outstrips any possible supply. Another problem is that there are very few human translators and their productivity can be increased to a specific limit without automation. Therefore Information communication and technology (ICT) has facilitated development of efficient procedures of Machine Translation. Research endeavors have been on to learn the likelihood of programmed translation of one dialect (source content) to another dialect (target content). The translation process may be stated as: decoding the meaning of source text and re-encoding this meaning in the target language. Machine Translation (MT) has got importance in day to day activities. In real time communications where it would not be practical for humans to translate, machine translation is required there. In case of high value content that is changing every hour and every day like news, machine translation plays a vital role. In today's world the ability to obtain information in our own language is more important than money

The main purpose of our research work will be the translation of English text to Kashmiri text. In case of translation of Kashmiri language, not much has been achieved. In contrast to other Indian languages, very few computational resources are available for Kashmiri language. Dictionaries, Corpora, POS Taggers, Chunkers, Parser are yet to be developed for Kashmiri.

Our objective is to develop a solution for translation of English Language into Kashmiri language by implementation of Grammatical Framework and Machine Learning. Comparison of quality of translations deduced from above MT techniques will be our secondary objective. Through Grammatical Framework (GF), programming language framework, we will develop programs to parse and generate texts in several languages simultaneously while working from a language-independent representation of meaning. Programming through GF will enable us to have a reusable library for dealing with the morphology and syntax of Kashmiri language. For employing machine learning for Kashmiri language translation, factual translation models are assumed whose parameters stem from the examination of monolingual and bilingual corpora. Building statistical translation models depends on these corpus. A Kashmiri Language will be translated according to the probability distribution.

$p(e|f)$ is a string e in the target language i.e. **Kashmiri** is the translation of a string f in the source language i.e. **English**. Modelling the probability distribution $p(e|f)$ may be approached with **Bayes Theorem**.

$$p(e|f) \propto p(f|e)p(e).$$

5th International Conference on Nanotechnology for Better Living (NBL-2019), Jointly organized by National Institute of Technology (NIT) Srinagar and Indian Institute of Technology (IIT) Kharagpur; April 7 – 11, 2019 at SKUAST, Shalimar Srinagar, Jammu and Kashmir, India; Editor: Dr. M. A. Shah and Prof. P. Banerji, ISBN:978-81-939516-0-6; Published by Applied Science Innovations Pvt. Ltd., Pune, Maharashtra, India.

$p(f|e)$ is the probability that the source string is the translation of the target string and the language model $p(e)$ is the probability of seeing that target language string or the kind of sentences that are likely in the target language.

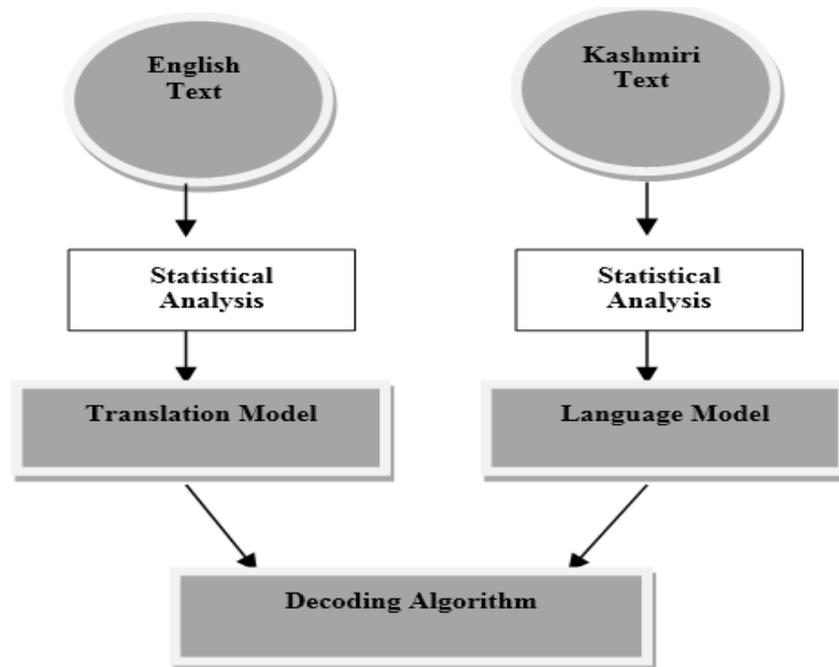


Figure 1: Depiction of components of SMT.

Unparalleled expansion into the field of machine translation is exploding in every corner of the globe. This technology has grown multitudinously and has reached to the mainstream with mobile devices, and hand-held electronics that translate the words as they are spoken. Machine Translation efficiently bridges the barriers of language. There is not a single language which can be treated as lingua franca in this ever shrinking global village. Customers are nurtured and appreciated from every region of the globe. Language hurdles should not be an obstruction to fortifying business relations with those who are attracted to your goods and services. Documents, communication, etc can all benefit with this translation solution. Progress in the area of machine translation has been seen in a variety of areas. The speed has increased with accuracy being higher than ever. The amount of languages presented is on the increase each year. Turnaround time is a fraction of what it was. The main obstacle left in the field is to break down the preconceptions that people have held onto concerning machine translation