



Language Technology and Natural Language Processing: A Developmental Study of Roman(WX) Notation of Ladakhi

Irfan ul Salam

Humaira Khan

Afreen Nazir

Department of Linguistics

University of Kashmir

ABSTRACT

In the present world of Digitalization, Language Technology has become a vital medium to preserve, populate and modernize one's language and culture simultaneously. This is known that computers can understand the Roman Script, in processing and editing, it has become a very important task to develop a standard Roman Notation for non-roman languages to be able to use for various NLP (Natural Language Processing) applications. This motivates to develop such a notation called as WX-Notation schema for Ladakhi language. Ladakhi is a Tibetic language, classified under Tibeto-Burman language family. It is spoken in Leh district of Jammu & Kashmir state. Ladakhi uses Tibetan script called 'yige' or Western Tibetan Archaic. WX-Notation is a transliteration scheme for representing Indian languages in ASCII (American Standard Code for Information Interchange,) for computational processing of Indian languages, and is widely used among the natural language processing (NLP) community in India. this transliteration scheme is productive as every consonant and every vowel has a single mapping into Roman. Hence it is a prefix code, advantageous from computation point of view.

It is high time for Ladakhi language, to come into the world of language computation in order to sustain the linguistic and cultural aspects. In this process there are other roman notations available, but how scientifically they are, raises some important questions.

1. Can these notations be case sensitive?
2. Are they anywhere near to the writing style of Ladakhi script?
3. How complex are these to enter in computers?
4. How easy they are for a new learner?
5. Do we have technology to convert from Roman to the Target Script?

In this research paper, an attempt is made to answer these questions and a usable WX notation scheme is discussed. A demonstration of the converter will also be discussed.

Keywords: *Computational Linguistics, Ladakhi script, Language technology, NLP, WX- Notation,*