Abstract:

In recent years, corpus based machine translation systems produce significant results for a number of language pairs. However, for low-resource languages like Urdu the purely statistical or purely example based methods are not performing well. On the other hand, the rule-based approaches require a huge amount of time and resources for the development of rules, which makes it difficult in most scenarios. Hybrid machine translation systems might be one of the solutions to overcome these problems, where we can combine the best of different approaches to achieve quality translation.

The goal of the thesis is to explore different combinations of approaches and to evaluate their performance over the standard corpus based methods currently in use. This includes:
2. Automatic extraction of lexical and syntactic rules using statistical methods to facilitate the Transfer-Based Machine Translation.

The novel element in the proposed work is to develop an algorithm to learn automatic reordering rules for English-to-Urdu statistical machine translation. Moreover, this approach can be extended to learn lexical and syntactic rules to build a rule-based machine translation system.