METHODOLOGY

A direct machine translation system will be developed on windows platform, in which words from source language are chosen, their equivalents in target language are found out from the lexicon and are replaced to get target language. The source text will be passed through pre-processing and translation phase and output should go through a post processing phase for translation, as shown in figure 1.

![Proposed architecture for Hindi to Marathi Translation System](image)

Fig. 1 Proposed architecture for Hindi to Marathi Translation System
The architecture for Hindi to Marathi Machine Translation System will have 3 basic modules, such as

1. Pre-processing phase
2. Translation phase and
3. Post Processing phase.

The pre-processing phase is a collection of operations that are applied on input data to make it processable by the translation engine. Pre-processing phase will comprise input text normalization, replacing collocations and proper nouns. If proper nouns (names of person, bank, river, city, country, ocean, days of week, months of year, university, cooperative society etc.) are present in the input text, it will be replaced. After tokenization, translation phase will identify titles, surnames, etc and transliterates. Lexicon lookup will be followed after this to replace word to word directly. For ambiguous and unknown words, either translation or transliteration will be performed. During the third phase, generated output text will be corrected grammatically achieving target text. Testing and evaluation will be performed on target text to find out accuracy of translation.

Lexical Resources planned to be used are Root word Lexicon, Inflectional form lexicon, Ambiguous word lexicon, Bigram & Trigram Tables for Hindi to Marathi languages.

Utility:

Being Hindi and Marathi closely related languages, Hindi-Marathi translation system will be very beneficial for removing language barrier within the country’s sociological structure. The system can be used by the researchers working in Hindi and Marathi natural language processing tasks.
Central governments Hindi documents, Hindi Literature and other documents such as web pages in Hindi of ones interest can be translated into Marathi for the readers use and ease of off field communities.

The large repository of historical, literary and religious work done by generation will be available for easy transformations and critique for all.

The target groups will be media organizations, literary and literacy promotional organizations, writers & NGOs working in rural areas, virtual Hindi-Marathi speaking communities, schools and colleges, publishers, translators and news media.

Digital dictionary for Hindi-Marathi words would be developed in the process of completion of this work can be used by Hindi and Marathi speaking communities.

The development will bring Marathi on the map of Machine Translation system.

The system can be integrated to other existing translation systems e.g. English to Hindi provided by Google, so that system will produce the output in Marathi from English text.