**Chapter 4 Implementation**

The main objective of our thesis is to detect interrogative Bangla sentences from transliterated Bangla sentences. In general perspective the topic seems very facile but after taking the whole overview, we have discovered that this is a very sophisticated and convoluted problem.

**4.1 Making of Primary Corpora for Rule-Based Approach**

The dataset we have worked on has been formed by extracting 44,538 comments from cricket -based Facebook public groups. The comments we have extracted contain various types of sentences. We have excluded the comments written in English and in Bangla alphabet and made a dataset and named it as the “Primary Corpus”. From this corpus we have excluded the sentences with the question mark and made the “Interrogative mega Corpus”. From this corpus we have taken the distinct sentences and formed the “Distinct Interrogative Corpus” and the rest of the primary corpus is named as the “Other Mega Corpus”. We have used these corpora for the rule-based approach of analysis.   
 The information about these corpora are reported in TABLE I.

1. Corpora Analysis

| Corpus | Corpus Information | |
| --- | --- | --- |
| Number of Sentences | Number of Words |
| Primary Corpus | 29,883 | 145,009 |
| Interrogative Mega Corpus | 4,624 | 23,785 |
| Other Mega Corpus | 25,259 | 121,224 |
| Interrogative Corpus | 700 | 3,073 |