**Chapter 5**

**Result and Analysis**

**5.1.1. Interrogative Corpus**

This corpus is the most important corpus for our experiments in rule-based approach. To analysis the sentences of the corpus we have calculated the average number of words, average length of words and average number of letters per sentences. The calculation is reported in TABLE II. The position of the Bangla question words serves a great importance in our rule-based approach. Therefore, we have made a Bangla question word list which contains 20 words. The list is depicted in Figure 2(Bangla Question Word List). We have calculated the position of these question words in the interrogative corpus. This calculation is reported in TABLE III. We have also seen that the word “*Naki*” is used as the question word in 24 sentences

Ki, Keno, Kivabe, Kothay, Koto, Kar, Kon, Kobe, Kisher, Kokhon, Ke, Kemon, Koy, Ke Ke, Kake, Kara, Kader, Koi, Koyta, Kotha.

1. Bangla Question Word List (figure 2)
2. Interrogative Corpus Analysis

|  |  |
| --- | --- |
| Average Number of Words | 4.7 |
| Average number of Letters per Word | 4 |
| Average Number of Letters | 19 |

a. All the averages denotes the average per sentence

1. Bangla Question word Position Analysis in interrogative corpus

| Position of Bangla Question Word in Sentences | Corpus Information | |
| --- | --- | --- |
| Number of Sentences |
| 1st Word | 112 | |
| Last word | 171 | |
| 2nd Word | 167 | |
| 3rd Word | 37 | |
| 4th Word | 4 | |
| 5th Word | 4 | |
| 6th Word | 2 | |
| 7th Word | 1 | |
| 8th Word | 1 | |
| Just Before the Last Word | 52 | |

## **Other Mega Corpus**

In this corpus all the sentences without question mark remain present. Like the interrogative corpus, we have calculated the average number of words, average length of words and average number of letters per sentences for this corpus. The calculation is reported in TABLE IV.

1. other mega Corpus Analysis

|  |  |
| --- | --- |
| Average Number of Words | 5 |
| Average number of Letters per Word | 4 |
| Average Number of Letters | 11 |

* 1. All the averages denotes the average per sentence
  2. **Making of Dictionary**

We have made the list of words that appeared in the corpora. The dictionary made from the words have depicted the variations found in the spelling of the transliterated words as there is no standard form or rules. This varies from people to people.

Example:

The word “আপনি” is found as follows:

-apni

-aapni

-apny

-apne

## **5.3 Evalauation of Rule Based Approaches**

From the set of rules prescribed in section 3.1, we have tested the interrogative corpus. The main basis of this approach is the position of the Bangla question words in the sentence. From TABLE III, we observe the positions of the Bangla question words. Excluding the last position for the other position we find the mean position of the Bangla question words which is 1.90.

At first, we have tested the corpus according to the rule 1, that is the presence of Bangla question word at the 1st or last position of the sentence then gradually we integrated the other rules and observe the combined effects of the other rules in the improvement of accuracy. The evaluation is given in TABLE V.

1. Rule based Approach Evaluation

| Method | Accuracy % |
| --- | --- |
| Rule 1 | 40.42 |
| Rule 1 + Rule 2 | 64.29 |
| Rule 1 + Rule 2 + Rule 3 | 71.71 |
| Rule 1 + Rule 2 + Rule 3 + Rule 4 | 75.14 |

From this level of accuracies from a known set we can say the rule based approach is not satisfactory. As the number of rules can’t be made concrete due to the diversity and varieties of sentences. From TABLE III, we have seen that the maximum likely position of Bangla question word is the last word of the sentence. In the sentence “*Tader jete bollam kothay ar tara gelo kothay*”, the question word “*Kothay*” is at the last position but the sentence is not an interrogative sentence whatsoever.

From this, we say that the rule based approach is not pragmatic and there should be a learning based approach for the identification.

* 1. **Making Corpora for Learning-Based Approach**

For learning based approach, we have used four classifiers, SVM, KNN, MLP and linear regression. We need