|  |
| --- |
| **Chapter 4: Findings** |

* 1. **Sentence Identification**

Since we have created rules specifically for the sports domain our queries and examples are focused on the players, tournaments and all types of area where sport are the primary category.

From the wh-question we can identify the POS tags which describes in the previous section this paper. In below example, if we can identify the sentence by whom, when and what we can get the result. Our algorithm was not efficient enough to identify all the words but small types of sentences can be identified and get the result of POS tags and its expression.

A full parse tree with details breakdown shown below -

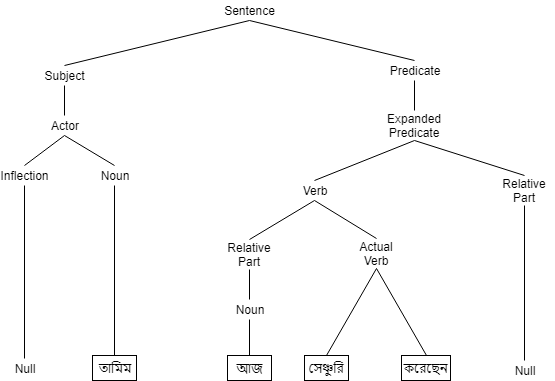


Figure 4-A: Sentence identification parse tree

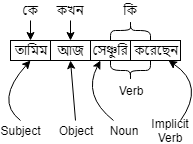


Figure 4-B: Sentence identification

* 1. **Position of Verb in Bangla Sentence**

Verb is the most difficult to detectable part-of-speech in Bangla sentence [5]. We have found that almost every Bangla sentences verb position hold the last position. It is quite easy to

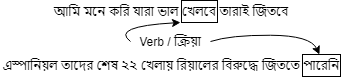


Figure 4-C: Position of verb in Bengali sentence

identify the verb but in Bengali verbs are formed with 2 or more words. So, it’s very difficult to understand the expression and also harder to make a link in machine learning. In English this verb position varied from sentence to sentence and way of we express itself.

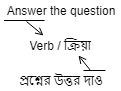


Figure 4-D: Verb position in Bangla & English

* 1. **Data Analysis**

For our thesis we have done quantitative analysis using various methods of statistics. We have done the total statistical analysis of our corpus, word frequency analysis and the most frequent used raw words in the corpus

|  |  |
| --- | --- |
| Number of sources | 3 |
| Number of source type | 1 |
| Total raw words in DB | 300138 |
| Total number of sentences | 28531 |
| Total number of articles | 1475 |
| Data collection range | 4 months |
| Total number of POS tags | 6 |
| Total number of respondents who were participated in 1st phase annotation | 30-40 |
| Total number of respondents who were participated in 2nd phase annotation | 3 |

***Table 4-A: Overall Statistics***

|  |  |
| --- | --- |
| Noun | 61.94 |
| Verb | 15.94 |
| Adjective | 7.61 |
| Preposition | 1.35 |
| Adverb | 4.71 |
| Pronoun | 6.39 |
| Undefined | 2.06 |

***Table 4-B: Most frequent POS tags in corpus***