List of Figures and Tables

CHAPTER 1

Table 1: Statistics of the corpus 30

Table 2: IL-POS Tagset (Baskaran et al., 2008) 39

Table 3: Proposed POS Tagset 41

Figure 1: Approaches to POS Tagging 14

Figure 2: Granularity of the POS Tagset 44

CHAPTER 2

Table 1: Types of Bangla Nouns (Ghosh, 1995), (Thompson, 2010) 49

Table 2: List of Pronouns in Bangla 53

Table 3: Agreement markers 59

Table 4: Honorificity 60

Table 5: Aspect markers 61

Table 6: Word classes and their respective classifiers 70

Table 7: Classifiers of Bangla 77

Table 8: 1st Person pronouns 78

Table 9: 2nd Person Pronouns 79
Table 10: 3rd Person pronouns with [+human] feature  80

Table 11: 3rd Person pronouns with [-human] feature  81

Table 12: Relative pronouns  82

Table 13: Reflexive pronouns  82

Table 14: [+human] interrogative pronouns  83

Table 15: [-human] interrogative pronouns  83

CHAPTER 3

Table 1: Tagged output of the three approaches  102

Table 2: The final tagged output of the HPT.  103

Table 3: Statistics of the Raw, punctuation-split and tokenized corpus  107

Table 4: Tags used to annotate the training corpus  108

Table 5: Statistics of the training and test corpus  110

Table 6: Statistics of the tagged files  123

Table 7: Unsorted tagged corpus  126

Table 8: Sorted tagged data  127

Table 9: Statistics of the sort corpus of CRF and Hybrid models  128

Flowchart 1: CRF tagging  111

Flowchart 2: Look-up tagging  116
Flowchart 3: The Hybrid POS Tagger (HPT) 124

Figure 1: Output of CRF module 110

Figure 2(a): CRF output 114

Figure 2(b): List of conjunctions 114

Figure 2(c): The output of the look-up tagger 115

Figure 2(d): Input of the RBM 118

Figure 2(e): Output of the RBM 118

Figure 2(f): The final output 121

Figure 3: Inaccurate CRF tags 120

Screenshot 1: Accuracy rate of CRF module 129

Screenshot 2: Accuracy rate of Hybrid POS tagger 129

CHAPTER 4

Table 1: List of homonyms from the corpus 134

Table 2: List of inaccurate bare nouns, and adjectives 142

Table 3: Inaccurate tags of Hybrid tagger 145

Table 4: Statistics of the percentage of error output of HPT 149

Figure 1: Hudhud’s 135
Figure 2: Bonsai’s

Figure 3: Abbreviation of ‘BDO’

Figure 4: Abbreviation of ‘CIA’

Figure 5: Error sample of CRF tagging

Figure 6: Incorrect Numeral tags of CRF module

Figure 7: Incorrect Numeral tags of Hybrid tagger

Figure 8: Non-word forms

Figure 9: Pie chart of the inaccurate tags generated by the Hybrid tagger