LIST OF FIGURES

Figure 1 Phonetic Sound of Gurmukhi vowels ............................................................. 4
Figure 2 Shahmukhi and Gurmukhi Syllable Structures .............................................. 5
Figure 3 Grapheme-based Transliteration Model ....................................................... 6
Figure 4 IBM Source-channel Model for Statistical MT .............................................. 7
Figure 5 Phoneme-based Transliteration Model ........................................................ 9
Figure 6 Hybrid Approach of Transliteration ............................................................ 12
Figure 7 Combination-based Approach of Transliteration ...................................... 14
Figure 8 Overview of Shahmukhi-to-Gurmukhi Transliteration System ................... 18
Figure 9 Historical Script Families and Derived Scripts ......................................... 68
Figure 10 Shahmukhi Alphabet ............................................................................. 72
Figure 11 Non-Joiners in Shahmukhi Script .......................................................... 73
Figure 12 Joiners in Shahmukhi Script .................................................................. 73
Figure 13 Character Rank with Frequency Distribution of Gurmukhi and Shahmukhi
Corpus ............................................................................................................... 85
Figure 14 Character Groups Distribution in Gurmukhi Corpus ............................... 86
Figure 15 Character Groups Distribution in Shahmukhi Corpus ............................. 86
Figure 16 Comparison of Entropy and Redundancy of Gurmukhi, Shahmukhi and
English ............................................................................................................. 89
Figure 17 Components of Pre-Processing Phase .................................................... 94
Figure 18 A Sample of InPage with Unicode Conversion ....................................... 95
Figure 19 Internal Text Representation of InPage .................................................. 96
Figure 20 Implementation of first Pass for InPage to Unicode Converter ............... 98
Figure 21 InPage file to Unicode Converter Architecture ...................................... 99
Figure 22 Canonical Equivalence ........................................................................ 102
Figure 23 Compatibility Equivalence .................................................................... 102
Figure 24 Canonical Composites .......................................................................... 103
Figure 25 Compatibility Composites ...................................................................... 104
Figure 26 Sample Entries of Shahmukhi-Gurmukhi Dictionary ............................ 112
Figure 27 Dictionary Database Design .................................................................. 113
Figure 28 Architecture of Transliteration Engine .................................................. 120
Figure 29 Word Joiner Phase of Transliteration ..................................................... 126
Figure 30 Implementation of Gurmukhi Soundex Coding Guide ......................... 139
Figure 31 Word Merger Phase of Transliteration ..................................................... 144
Figure 32 Number of Gurmukhi Interpretations of Ambiguous words in Top 10K Dataset .............................................................................................................. 155
Figure 33 Word level Ambiguity in Transliterated Text ........................................ 156
Figure 34 A Fragment of Second Order HMM ....................................................... 168
Figure 35 A Fragment of Second Order HMM for ਹੁਣ ਤੂੰ ਦੱਸ .................................. 169
Figure 36 Structure of ±5 Window Context ............................................................. 170
Figure 37 Disambiguation using Context Window of Size ±5 ............................... 171
Figure 38 A Sample of Maximum Matching .......................................................... 180
Figure 39 MMS for Identical Candidate and Reference Text ................................ 181
Figure 40 System Stages with overall Accuracy and Incremental Improvement .. 187
Figure 41 Segmentation Accuracy along with Occurrence of Merged words ......... 196
Figure 42 Word Error Rate (WER) and Word Accuracy of the System ................. 197
Figure 43 Error Analysis of Test Data Set-1 (Book) .............................................. 197
Figure 44 Error Analysis of Set-2 (Likhari.org) ...................................................... 198
Figure 45 Error Analysis of Set-3 (Wichaar.com) ................................................. 198
Figure 46 Error Analysis of Overall Test Data ....................................................... 199
Figure 47 Country-wise Usage of Shahmukhi-to-Gurmukhi Transliteration System ......................................................................................................................... 205