Bangla Text and Spoken Language Technology

Professor Dr. Mohammad Nurul Huda

MSCSE Director, United International University, Dhaka-1212, Bangladesh

NLP team Consultant-eGeneration System Limited

Email: mnh@cse.uiu.ac.bd, Phone: +88-01673476433

1. History of Bangla Language

- Largely Spoken Languages in the world
- More than 220 million people as native language
- Ranked 7th as number of speakers of Bangla
- Phoneme based language

2. Problems of Bangla Research

- Lack of Bangla Corpus
- Dialect Effect
- Unstructured text data
- Kolkata and West Bengal research
- Pronunciation different
- Phonetic variation

3. Objectives

- To build a large scale Bangla corpus
- To start research on Bangla Language Technology

4. How to achieve this target?

- People Republic of Bangladesh allocates 160 crore BDT
- Launches 16 Bangla Projects
- Corpus construction for text and spoken languages
- Projects: Corpus, text to speech and speech to text
- Projects: IPA converter, Sentiment Analysis, Machine Translation
- Projects: Sign language for disable, Grammar and Spell checker, etc

5. My Research Area

- Small corpus construction
- Machine Translation
- Speech Recognition
- Speech Synthesis
- IPA conversion
- Document classification
- Sentiment Analysis
- Machine Learning
- Pattern Classification
- Algorithms
- Bangla Shadhu basha
- Bangla Cholito basha
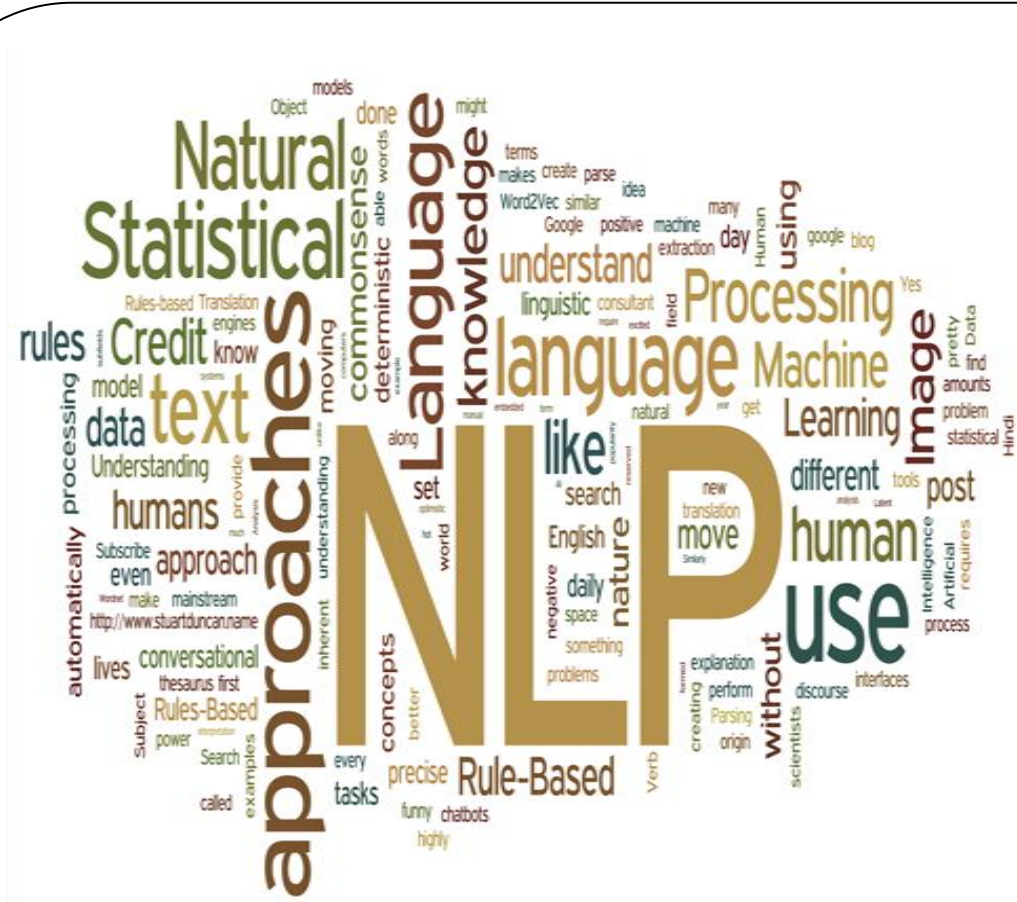- Bangla keyboard
- Bangla Font converter

6. My Research Papers

- Speech Recognition (60 papers)
- Speech Synthesis (2 papers)
- Machine Translation (3 papers)
- Document Classification (2 papers)
- Sentiment Analysis (2 papers)
- Question Answering (2 papers)
- Similarity Measure (1 paper)
- POS Tagger (2 papers)
- Morphology (1 paper)
- UNL (8 papers)
- Algorithms (7 papers)
- M. Learning (17 papers)
- Sign Language (1 paper)
- Traffic sign (1 paper)

7. NLP LOGO

8. NLP TOOLS DEVELOPMENT

- বাংলা Speech Recognizer
- বাংলা Speech Synthesizer
- বাংলা Answering Grading System
- Parallel Corpus for বাংলা – English
- Spell Checker for Bengali (বাংলা)
- Question-Answering System for বাংলা
- Text-to-Speech (TTS) for বাংলা
- IPA Converter for বাংলা
- Machine Translator for বাংলা to English
- Sentiment Analyzer for বাংলা
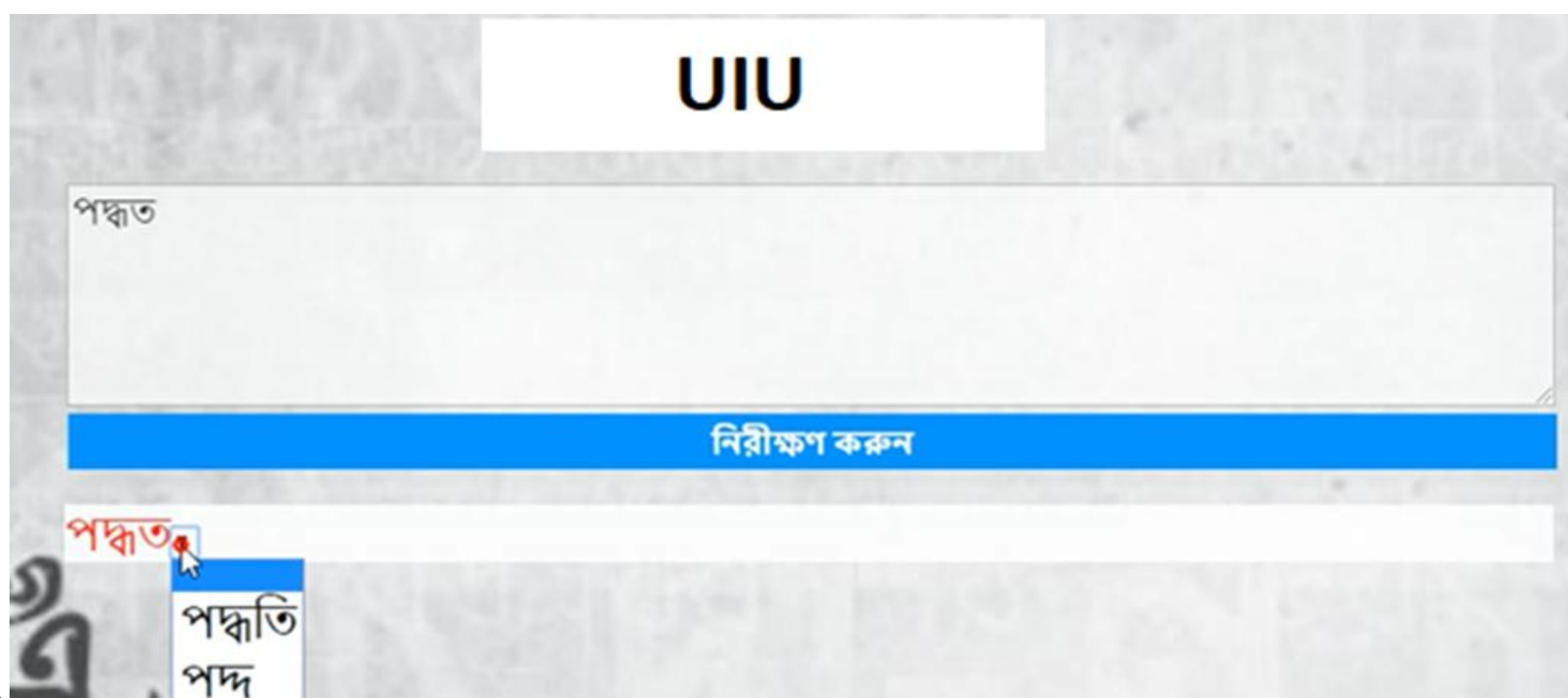- বাংলা Speaker Identification

9. MACHINE LEARNING & others

- Naive Bayes Classifier
- Recurrent Neural Network
- Time Delay Neural Network
- Support Vector Machine
- Hidden Markov Model
- Multi Layer Neural Network
- LSTM
- Cosine Similarity
- TF-IDF
- Soundex, Metaphone, Double Metaphone
- Edit Distance
- Trigram Language Model
- Python-NLTK

10. Some Developed NLP Tools

Spell Checker

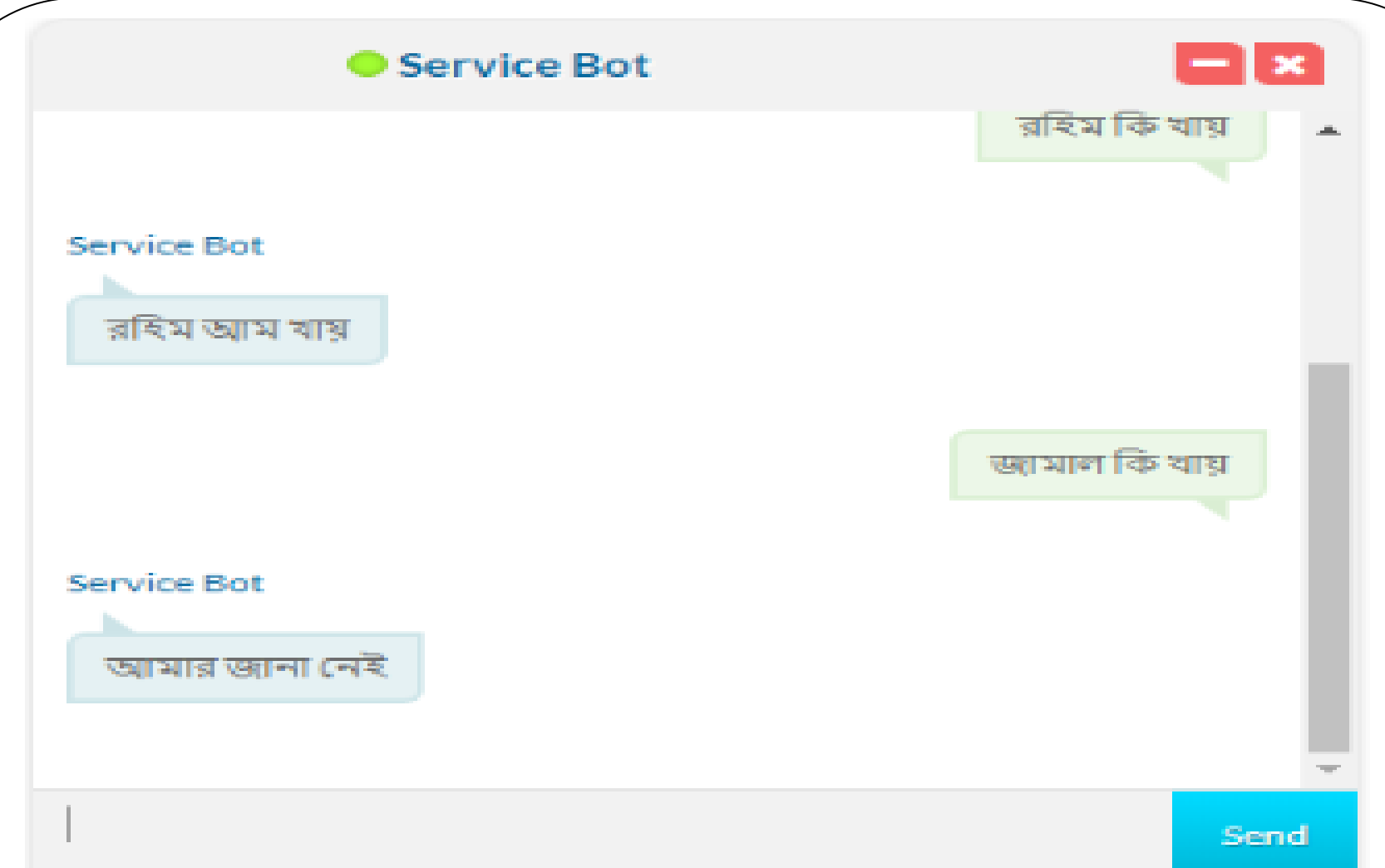- Check spelling for Bangla words
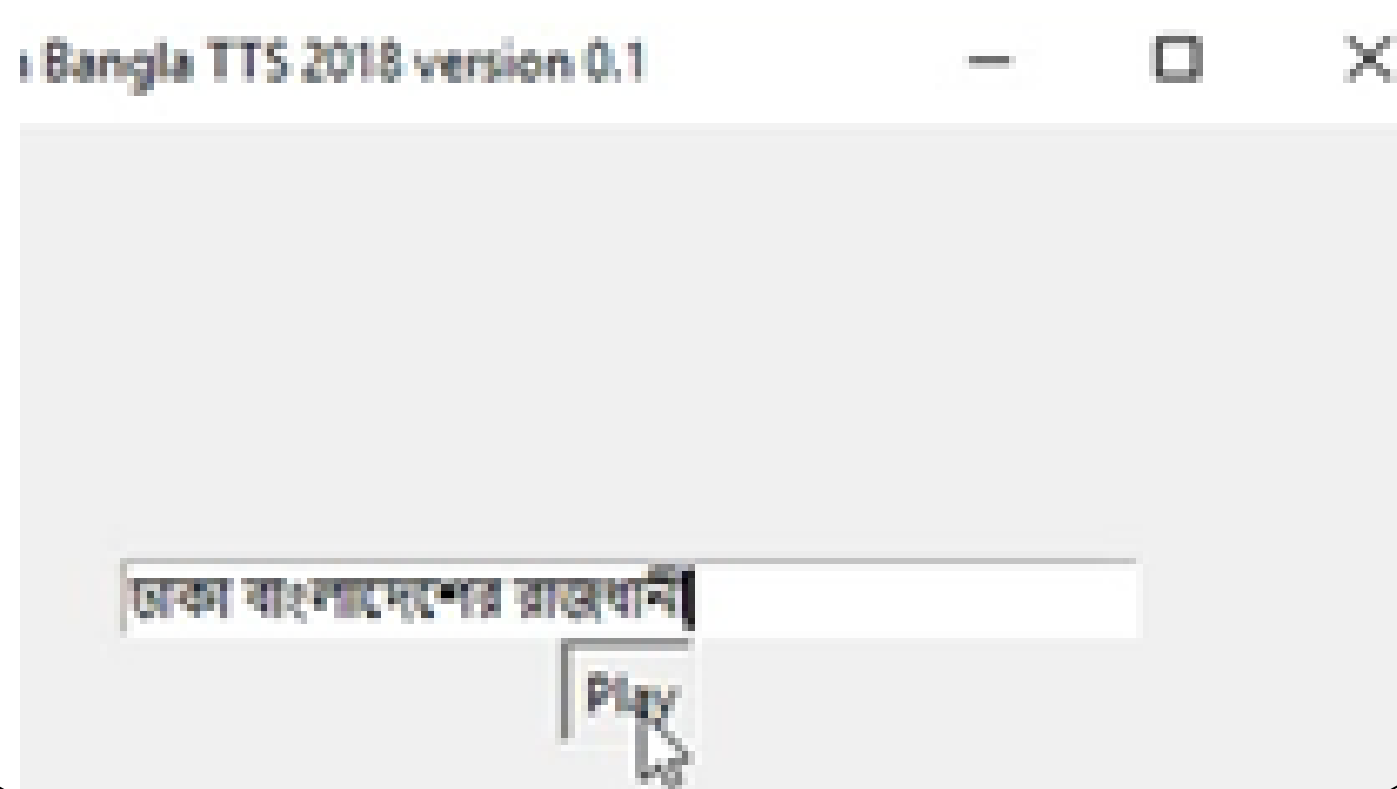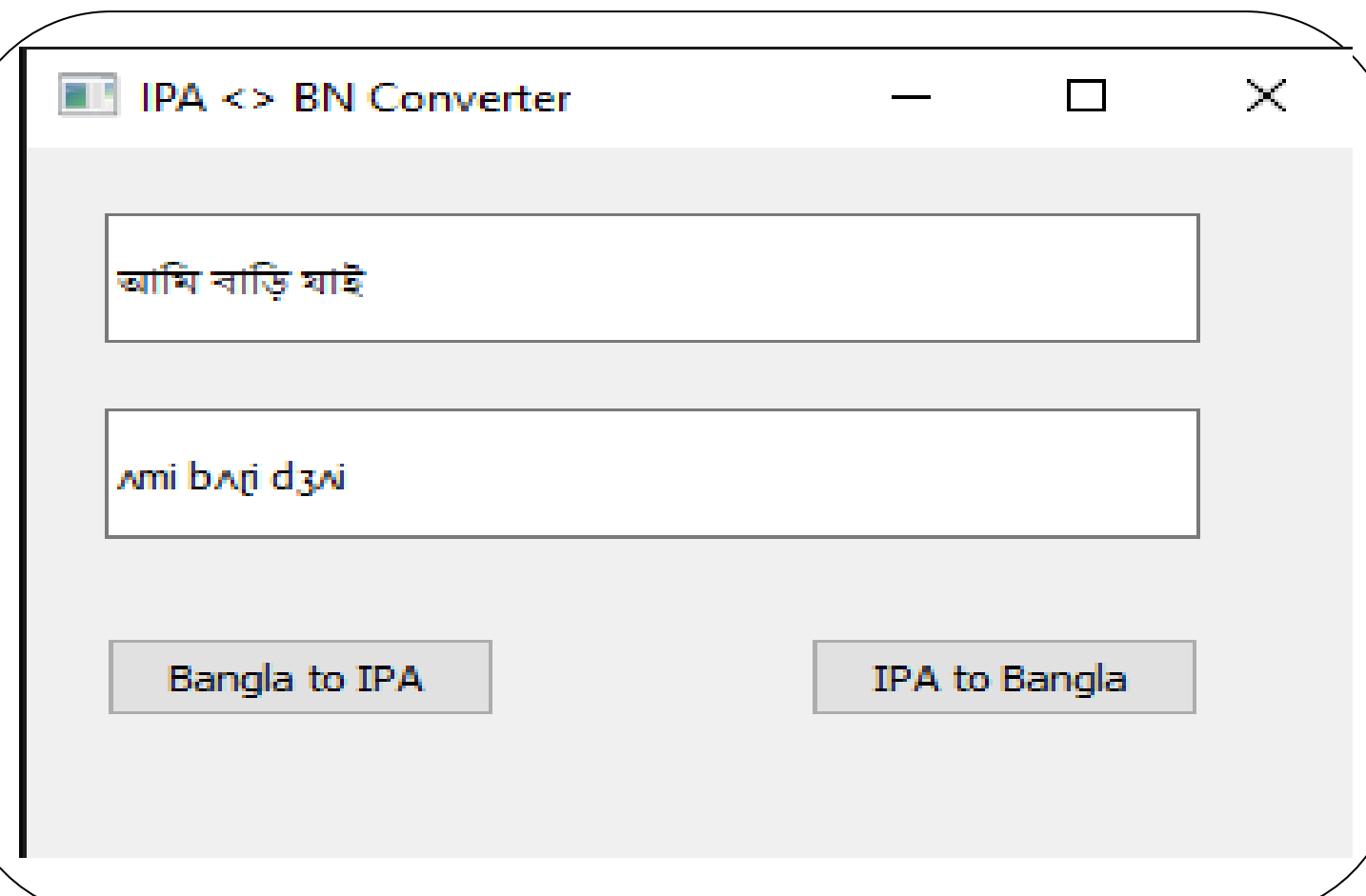- Suggest correction for incorrect words

Question Answering System

Bangla TTS

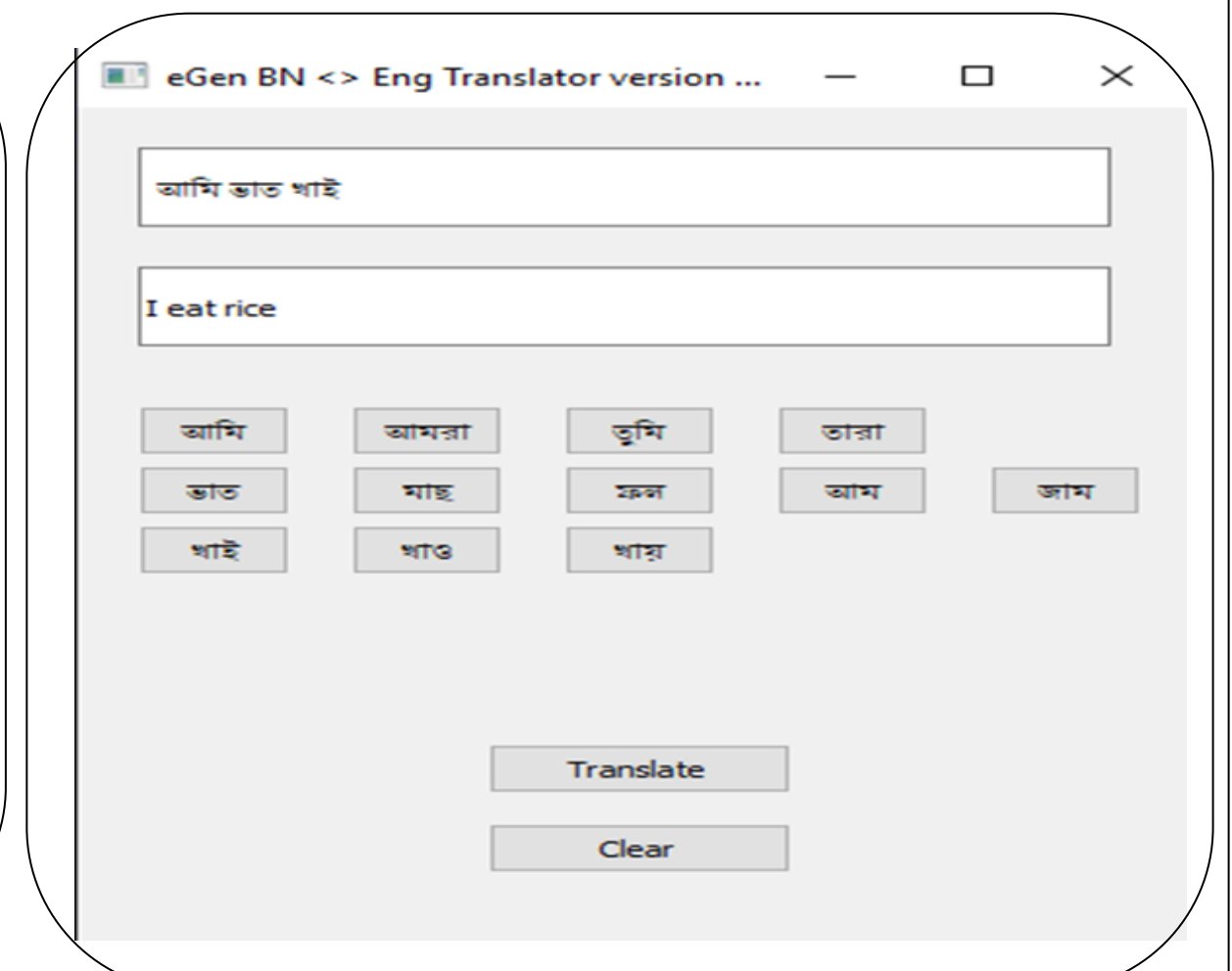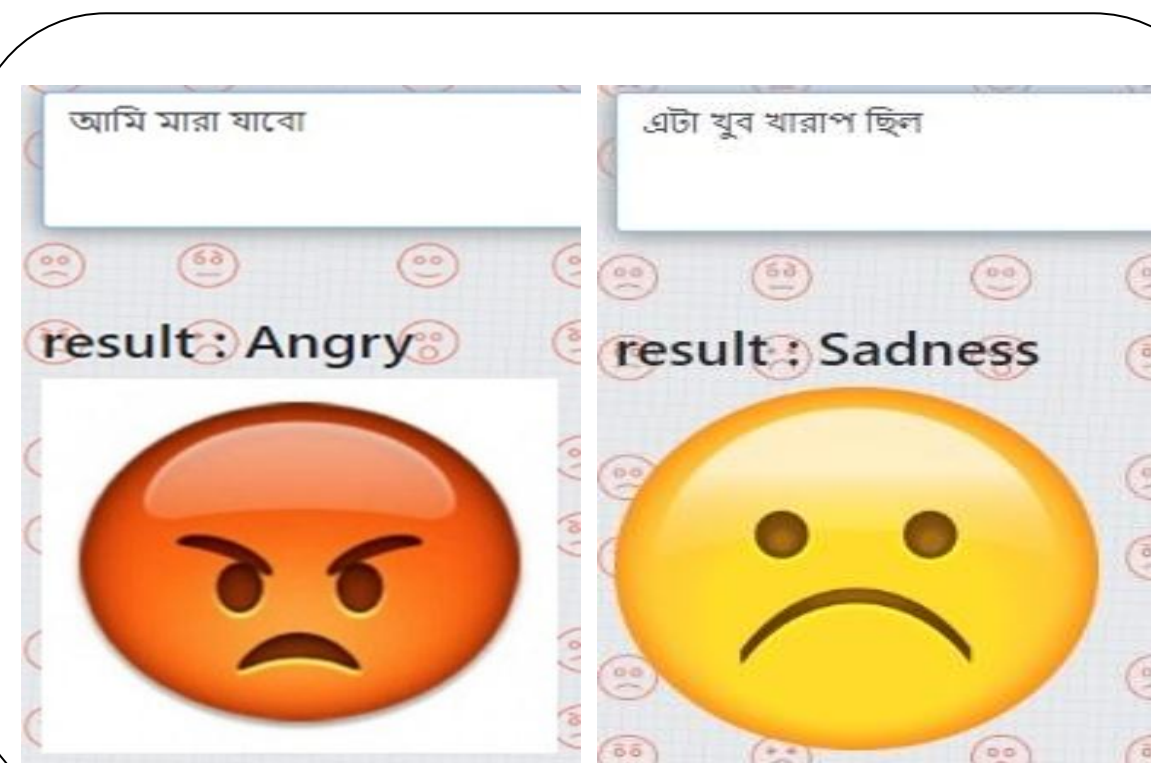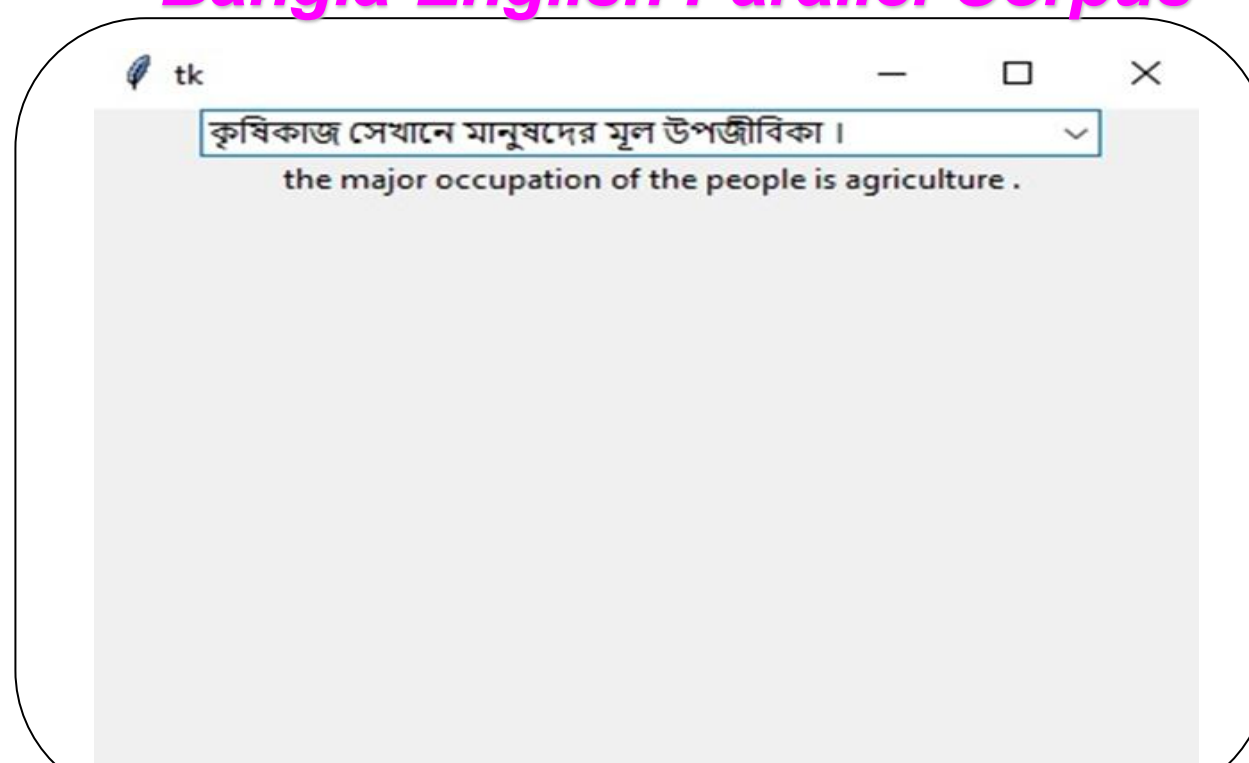### Bangla to IPA Converter

Machine Translator

Bangla Sentiment Analyzer

Bangla-English Parallel Corpus