

**East West University**

A project proposal on

**English to Bangla Machine Translation System**

**Course Title:** Machine Learning

**Course Code:** CSE475

**Section:** 2

**Submitted To:**

Md. Mohsin Uddin

Lecturer,   
Dept. of Computer Science & Engineering

East West University

**Submitted By:**

Shaykh Siddique

ID: 2016-1-60-053

Tahmid Ahmed

ID: 2016-1-68-007

Rifayet Azam Talukder

ID: 2016-1-60-030

**Designing English to Bangla Machine Translation system with Neural Network.**

**Introduction:**

Translation is the communication of meaning from one language to another language. A translation could be word by word translation and another one is translation by sentence. In sentence translation we get more information rather than word by word translation. In this paper our main focus is to translate from English to Bengali by using a machine. Machine translation refers to translation of text or speech. Our task is to translate from English language to Bangla language. With approximately 228 million native speakers and another 37 million as second language speakers. Bengali is the fifth most-spoken native language and the seventh most spoken language by total number of speakers in the world [1].

**Motivation:**

Almost all translator translates word by word and the accuracy is not that much good? Almost all the model fails to translate English to Bengali accurately. Our proposed system will be able to solve this problem. It will help millions of Bengali people to make their translation work automatically with less headache and more accurately.

**Aim:**

The aim is to design an architecture of English to the Bangla Machine Translation system with a demo application.

**Objectives:**

1. To synthesize the parsing factors and grammatical structures of the sentence.
2. To specify the structure of the complex sentence and coreference relation.
3. To design intelligent system architecture.
4. To develop a demo application based on the architecture.

**Related Works:**

Natural language processing is used to make the machine intelligent. The way of language processing is enriching day by day. Many studies defined the architecture for natural language processing [2] but few deal with the improvement of English to Bangla language translation. A team works in tense based structure of English to Bangla translation[3]. Another study is on simple sentence structure and comparison of different machine translation systems [4]. Still, now there is a lack of studies with complex sentence structure and recurrent meaning of a sentence.

To make machine intelligent natural language processing is used. Language translation is getting improved but there are a few improvements is getting on English to Bangla translations. There are many ways that translations are solved. Here for Hindi translations, two encoder-decoder neural machine translation architectures are used, which convolutional sequence to sequence model (ConvS2S) and recurrent sequence to sequence model (RNNS2S) [5]. One is for English to Hindi and another is to do the opposite. Here in training data, 1492827 sentences used where 20666365 words for English and 22164816 words for Hindi. The RNNS2S model was trained using Nematus framework and for ConvS2S model was trained using Fairseq5, an open-source library developed by Facebook for neural machine translation using CNN or RNN networks. Their result showed that ConvS2S performed better on English to Hindi translation which would help to solve our problem. In the Corpus-Based Method using one subject file and one verb file, the translations are solved [4]. Here for each subject, there is a flag corresponded to its verb and the most suitable and meaningful sentences are selected for final translations. The result showed better compared to Google Translator. For another English to Hindi translation feed-forward back-propagation artificial neural network used[6]. For the implementation java used for main programming language to implement the rules and all the modules apart from the neural network model which have been implemented in Matlab. Here, training data is encoded into numeric form by the Encoder which is also implemented in Java. They have used BLEU [7] to calculate the score of the system.

**Data Collection Method:**

We need data to prepare our intelligent system. The main dataset of our research will be English and Bangla parallel sentences. For each English sentence, we need some co-responding Bangla Sentence to train and test the intelligent system. We collect data from some article which are manually written in English and Bangla by human.

*Data Type:*

After collecting raw data English and Bangla parallel sentences, we have to clean that and divide into small instances and store it in the CSV file. The data type is nominal. Working with Bangla Language we have to deal with the Unicode writing system.

**Raw Dataset Link:** [github.com/shaykhsiddique/Machine-Translation-with-RNN-Python/tree/master/Dataset](https://github.com/shaykhsiddique/Machine-Translation-with-RNN-Python/tree/master/Dataset)

**References:**

[1] S. K. Chatterji, “Bengali phonetics,” *Bull. Sch. Orient. African Stud.*, vol. 2, no. 1, pp. 1–25, 1921.

[2] R. Collobert and J. Weston, “A unified architecture for natural language processing,” 2008, pp. 160–167.

[3] K. Muntarina, G. Moazzam, and A. Bhuiyan, “Tense Based English to Bangla Translation Using MT System,” *Int. J. Eng. Sci. Invent.*, vol. 2, no. 10, pp. 30–38, 2013.

[4] S. Nahar, M. N. Huda, M. Nur-E-Arefin, and M. M. Rahman, “Evaluation of machine translation approaches to translate English to Bengali,” *20th Int. Conf. Comput. Inf. Technol. ICCIT 2017*, vol. 2018-Janua, pp. 1–5, 2018.

[5] S. Singh, R. Panjwani, A. Kunchukuttan, and P. Bhattacharyya, “Comparing Recurrent and Convolutional Architectures for {E}nglish-{H}indi Neural Machine Translation,” *Proc. 4th Work. {A}sian Transl.*, no. 2014, pp. 167–170, 2017.

[6] S. Shahnawaz and R. B. Mishra, “A Neural Network based Approach for English to Hindi Machine Translation,” *Int. J. Comput. Appl.*, vol. 53, no. 18, pp. 50–56, 2012.

[7] K. Papineni, S. Roukos, T. Ward, W. Zhu, and Y. Heights, “IBM Research Report Bleu : a Method for Automatic Evaluation of Machine Translation,” *Science (80-. ).*, vol. 22176, pp. 1–10, 2001.