Polygot: An Approach Towards Reliable Translation By Name Identification And Memory Optimization Using Semantic Analysis

Md. Adnanul Islam*, A. B. M. Alim Al Islam[†] and Md. Saidul Hoque Anik[‡]
Department of CSE, BUET
Dhaka 1000, Bangladesh
*Email: *islamadnan2265@gmail.com, †alim_razi@cse.buet.ac.bd, ‡onix.hoque@gmail.com*

***Abstract*—We present a study of improving the efficiency, complexity and performance of language translation process by a translator. The goal of this research is to develop an efficient translation system for any language by optimizing the memory consumption and also identifying the names as nouns efficiently. Although a number of researches can be found on natural language processing in different areas, those were performed keeping English as the only target language mostly. However, a good number of languages remain nearly unexplored in the research fields yet. This study basically focuses on Bengali Language as an example of the unexplored languages. Some noticeable studies on Bengali language are on Bangla keyboard layout design, English to Bangla translator, etc. so far. However, very few researches have been done to translate Bengali text to English till now. To develop an efficient translation system is very complex and expensive as it requires huge amount of time and resources. In all the languages, there are many words having multiple meanings and multiple forms and also some sentences having multiple grammatical structures to express the same meaning. Besides, the names of people may not be easily identified due to the vast diversity of names and also the tags (prefix/ suffix) attached to emphasize names. Therefore, it remains a great challenge to recognize a sentence of a particular language with accurate semantic analysis. However it is very important to have a generalised translation system which can compute various possible outputs in reasonable time and space. In this paper we focus on the correct interpretation of the names as nouns in a sentence and also the optimization of space requirement using semantic analysis.**

***Keywords*—NLP, OpenNLP, Wordnet**

1. Introduction

Language learning requires motivation, time, and dedication. One has to read, write, listen, and speak regularly to learn a new language effectively. Learning a new language is exciting and beneficial at all ages as it offers practical, intellectual and many aspirational benefits. However, human beings generally get to know one language in early childhood almost subconsciously which is known as the mother tongue. Human beings can sense the meaning or expression of any sentence in his mother tongue easily by identifying the basic grammatical tags of each word in that sentence. For example, let us consider the simple sentence - “I eat rice”. The sentence can automatically be recognized by a child whose mother tongue is English without any formal knowledge on that language. This inherent subconscious human nature also teaches a child to detect the semantic error of the sentence - “I eat Football”. Primarily, perhaps visual perceptions of different objects and activities e.g., rice, Football, books, eating, running, etc., assist a child to sense the context and meaning of different words in a sentence. The ability of human to sense and learn one language appropriately enables him to learn another language by translation. This is where a machine lags behind human in the field of translation as a machine, which is dictated only by logic or proofs, cannot learn any language inherently or subconsciously.

While progress has been made in language-translation software and allied technologies, the primary language of the ubiquitous and all-influential World Wide Web is English. English is typically the language of latest-version applications and programs and new freeware, shareware, peer-to-peer, social media networks and websites. Manuals, installation guides and product fact sheets of various consumer electronics and entertainment devices usually are made available in English first before being made available in other languages.

Bengali, the native language of around 189 million people worldwide, mostly from Bangladesh, is considered as low-resource language for machine translation as it lacks different language resources like electronic texts and parallel corpus. Around 38% of Bengali speaking people are monolingual. Since significance of learning English is unavoidable at present, it is important to have a well developed Bengali to English translation learning system.

In this study, we take Bengali to English translation system as an example to propose such a generalised translation skeleton. The main focus of this work is-

- Proper identification of names as nouns by detection of emphasized tags at the end of the names
- Memory optimization by semantic analysis of verbs

978-1-5386-3288-8/17/$31.00 © 2017 IEEE

2. Motivation

Millions of immigrants travel the world from non-English-speaking countries every year. For obvious reasons, learning to communicate in English for the immigrants is very important to enter and also succeed in mainstream English speaking countries. Working knowledge of English language enhances many opportunities in international markets. However, a major group of people lacks proficiency in English. Also, there is no well developed translation system till now to translate many native languages to English. Therefore, importance of a generalised translation skeleton is noteworthy.

There are different existing systems for automatic translation process. Machine translation is the most popular among them. European Commission have been using the Machine Translator to convert text from one language to another language since 1976. This broad usage spreads its necessity widely with its developed translation technique for regular uses.

Now-a-days, Google translator is one of the pioneer applications supporting a number of languages to translate from one to another. Although it works successfully for many languages, it merely can translate Bengali to English. Bing translator, another popular translator, does not even recognize Bengali. The other translators e.g., Yahoo Babel Fish, Systran Language Translation, SDL Free Translation, etc., support multilingual translation like Danish, English, Chinese, Italian, Japanese, French, Greek, Korean, etc., however, not Bengali and many other widely used languages.

Natural languages like English, Spanish, and even Hindi are rapidly progressing in processing by computers. However, Bengali, being among the top ten languages in the world, is yet in quite delinquent stage in the area of computational linguistics and machine translation. Bengali lags behind in some crucial areas of research like parts of speech tagging, information retrieval from texts, text categorization, and most importantly in the area of syntax and semantic checking [1].

Therefore, our motive of this study is not only to efficiently translate one language to other by proper semantic analysis but also to teach the translated language by explaining the translation mechanism step by step.

3. Related Work

Bengali is one of the most widely spoken languages throughout the world with nearly 230 million total speakers. However, Bengali still lacks significant research in the area of natural language processing unfortunately. Bangla to English translation was first proposed by Sk. Borhan Uddin, Dr. Md. Fokhray Hossain and Kamanashis Biswas using opennlp tool. They proposed a simple technique for synthesizing Bengali words. However, they used opennlp tool after translating the Bengali word to corresponding English word which caused erroneous Parts Of Speech (POS) tagging for different words and generated wrong outputs for very simple sentences.

Dasgupta et al., [6] proposed to use syntactic transfer. They converted CNF trees to normal parse trees and using a bilingual dictionary, generated output translation. However, this research did not consider translating the unknown words which did not appear in the bilingual dictionary.

Chunk parsing was first proposed by Abney (1991). Although EBMT (Example based Machine Translation) using chunks as the translation unit is not new, it has not been widely explored for low-resource language like Bengali yet. Naskar et al., [13] reported a phrasal EBMT for translating English to Bengali without any evaluation of their EBMT. Besides, they did not clearly explain their translation mechanism, specially the word reordering process.

Saha et al., [14] reported an EBMT for the translation of differnet news headlines. The work showed that EBMT can be a positive approach for Bengali language. However, their approach relied mostly on news headlines. Moreover, Gangadharaiah et al., [3] proposed that templates can be useful for EBMT to obtain longer phrasal matches if coordinated with statistical decoders. His study showed that it is a time consuming task to cluster the words manually and would be less time consuming to use standard available resources such as, WordNet for clustering.

Kim et al., [4] used syntactic chunks as units of translation for improving insertion or deletion of words between two distant languages. However, an example base with aligned chunks in both source and target language is missing in this approach.

4. Proposed Mechanism

Our previous work was on going beyond database driven and syntax based translation [1]. The work basically focused on the translation of simple sentences. Simple sentence analysis and recognition is the preliminary step which leads to the advancement towards the translation of improved or more complex sentences. However, analyzing and recognizing a simple sentence of a language correctly, not only syntactically but also semantically, requires enormous exploration and exploitation on that particular language. For examples,

- "The complex houses married and single soldiers and their families."

  This is called a garden path sentence. Though grammatically correct, the reader's initial interpretation of the sentence may be nonsensical.
  Here, *complex* may be interpreted as an adjective and *houses* may be interpreted as a noun. Readers are immediately confused upon reading that the complex houses *married*, interpreting *married* as the verb. How can houses get married? In actuality, *complex* is the noun, *houses* is the verb, and *married* is the

adjective. The sentence is trying to express the following: Single soldiers, as well as married soldiers and their families, reside in the complex.

- “All the faith he had had had had no effect on the outcome of his life.”

  This sentence is an example of lexical ambiguity. Although this sentence might sound strange, it is actually grammatically correct. The sentence relies on a double use of the past perfect. The two instances of *had had* play different grammatical roles in the sentences the first is a modifier while the second is the main verb of the sentence.

From the above examples, we can see how a simple sentence can become so complex to deal with. Besides, there are many phrasal sentences which indirectly indicate different special semantic meanings.

One of the most challenging tasks is identification of names properly as nouns since the names cannot be included in vocabulary may contain some emphasized tags which need to be separated from the name accurately.

4.1. Methodology

Basically, translation of a sentence consists of six major steps as shown in figure 1.

- Input Bengali text.
- Analyze the input sentence by tokenizing.
- Tagging (parts of speech, number, person) of the tokens.
- Word by word translation of the tokens.
- Apply necessary suffixes and words to the verb.
- Rearrange the words applying grammatical rules to output the translated sentence.

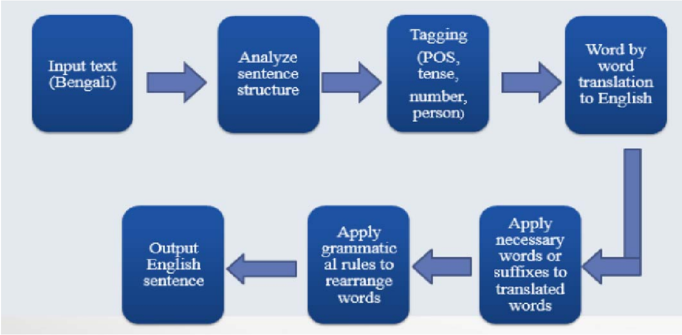

Input text (Bengali)
Analyze sentence structure
Tagging (POS, tense, number, person)
Word by word translation to English
Apply necessary words or suffixes to translated words
Apply grammatical rules to rearrange words
Output English sentence

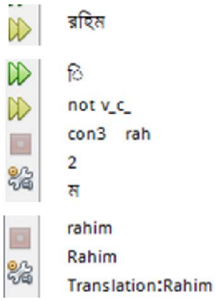
Figure 1. Proposed translation skeleton

We will mainly focus on tagging the tokens by identifying the names as nouns efficiently. Also, we will optimize the database required for tagging the verbs having different forms. Therefore, we will improve the subsequent steps of the translation methodology by efficient tagging mechanism proposed in this paper.

4.2. Name Identification

One major and unique improvement by our system is name identification and name translation. Names need to be identified as nouns correctly first to tokenize the input sentence and analyze the tokens properly. Names of people are generally considered as nouns in the sentences which dictate person, number and gender of the subject. Therefore, properly identifying the names as nouns is very crucial for accurate translation. Google translator completely misunderstands the names of the people in Bengali. Therefore, it fails to identify the number and the person of the subject which ultimately leads to failure for even translation of basic sentences. Names cannot be translated by using database containing the vocabulary. However, our system can recognize the names by applying its specific grammatical rule set to identify the subject.

After we have detected the names as subjects, we can assign the subject with the tag - third person, singular number. Then our system can modify the verbs accordingly. However, we do not have any translation for names in the vocabulary. Thus we have developed a Bengali to English phonetic mapping conversion algorithm in the proposed system which enables translation of the names from Bengali to English. We show a sample procedure of name processing for Bengali to English translation in figure 2.



```
রহিম

হি
not v_c_
con3   rah
2
ম

rahim
Rahim
Translation:Rahim
```

Figure 2. Name translation

4.3. Emphasized Name Identification

There are also names with different emphasizing tags used in many languages. The emphasizing tags, associated with the names, are actually not any part of the original name. They mean to emphasize the names as adjectives. They have separate meaning and use in the sentence. If these are not identified correctly and separated from the names, the resulting translation may become faulty as shown in the figure 3. In the figure, the system misinterprets the subject as a name (Amio) due to the omission of suffix checking. The translated sentence should have been - “I will also play Football”.

In Bengali names, we can always perform a check in the suffix of the subject and separate the emphasizing tags as

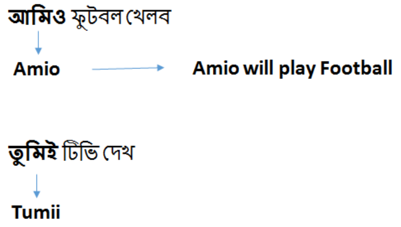


Figure 3. Emphasized name

suffixes from the actual name. After that, we can translate the name as earlier and take care of the emphasizing tags (suffixes) according to grammatical rule set as shown in figure 4.

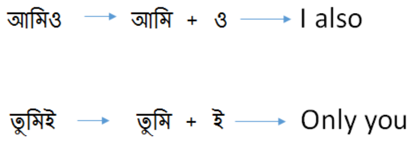


Figure 4. Emphasized name identification methodology

Although we can apply this mechanism for name identification appropriately, this may not generate the accurate result for all the cases. We will come back to this point later with relevant examples. However, considering probable faulty translations for some exceptional cases, we can apply the proposed mechanism for emphasized name identification effectively.

4.4. Memory Optimization

This is another important feature of this work. In Bengali, same verbs may occur in multiple forms depending on the number and person of the subject and the tense of the verb. If we need to store the word translation of each form of the same verb then the database will become very large due to the repeated verbs contributing to massive consumption of memory.

However, we can avoid the multiple insertions of the same verb having different forms in our proposed database optimization technique. We can only store the main format of a verb and apply semantic analysis to detect the main format from other formats of the verb depending on number, person and tense as shown in figure 5. The figure shows how one word (verb) can take different forms depending on tenses and suggests insertion of only that particular word (verb) instead of inserting all of its different forms. This will avoid multiple insertions in the database for the same verb with multiple forms.

More quantitatively, every ASCII character consumes one byte (8 bits) of memory and every Unicode character consumes more than one byte space. Every word we insert in the database for vocabulary contains multiple letters which are either ASCII or Unicode characters. Inserting ten different forms of the same word can be compared to inserting ten unnecessary words consuming more than ten times required space as different forms of one verb can be even bigger words containing arrays of Unicode or ASCII characters. Therefore, the memory (space) consumption should improve significantly by carefully avoiding these redundant insertions in database as proposed in our memory optimization technique.

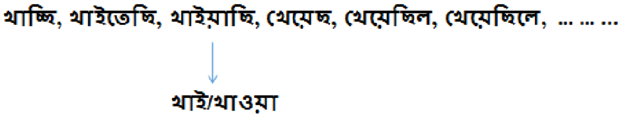

খাচ্ছি, খাইতেছি, খাইয়াছি, খেয়েছ, খেয়েছিল, খেয়েছিলে, … … …
খাই/খাওয়া

Figure 5. Verbs mapping

5. Experimental Evaluation

This section reflects the evaluation and outcomes of the experimental results according to our proposed methodology.

5.1. Tools and Settings

Experimental set ups cost significant amount of time and energy. This was the most critical period of our research. For experimentation, we used the following features in our implemented system:

- Language : JAVA
- Platform/ IDE : Netbeans
- Database : Sqlite
- Tool : Opennlp tools

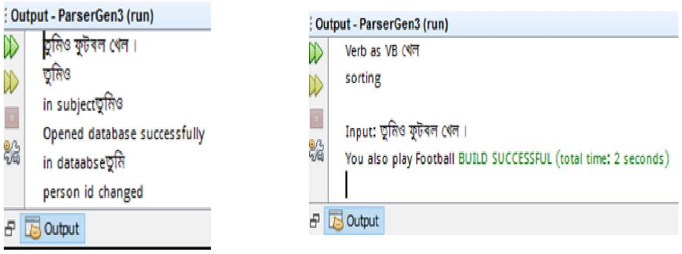
A major issue arose while taking input and parsing Bengali texts in java. We set up text encoding to UTF-8, however Bengali texts failed to appear in netbeans. After a lot of research, we had to change the font settings and some other settings to work with Bengali texts successfully in netbeans.

We used Sqlite with java for database in our system. Since we had to use a Bengali to English dictionary, we needed a database to retrieve the word translations. To do this we installed Sqlite and aso added a jar file for Sqlite in our project.

Regarding Bengali to English dictionary, we could not find any well defined dictionary format or API so that we could integrate it to our database directly in a time efficient way. Since inserting the words by brute force is a huge time consuming issue, we had to do a huge amount of exhaustive work to insert a reasonable amount of words in our database.

5.2. Results

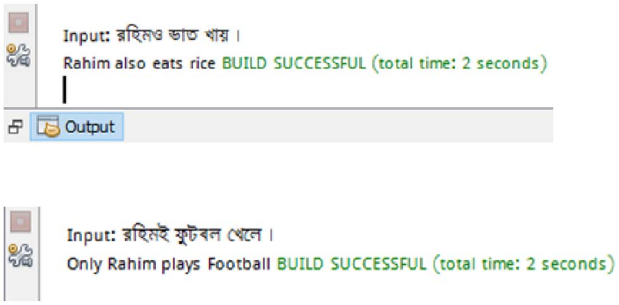
We dealt with various types of sentences. However, our system works perfectly with basic simple sentences and complex sentences. Particularly in this work, we tested our system with different emphasized names and verbs with different tenses. A scenario of our experimental results is shown as follows:



```
Output - ParserGen3 (run)
তুমিও ফুটবল খেল ।
তুমিও
in subjectতুমিও
Opened database successfully
in dataabseতুমি
person id changed
```

```
Output - ParserGen3 (run)
Verb as VB খেল
sorting

Input: তুমিও ফুটবল খেল ।
You also play Football BUILD SUCCESSFUL (total time: 2 seconds)
```

Figure 6. Example on emphasized name identification

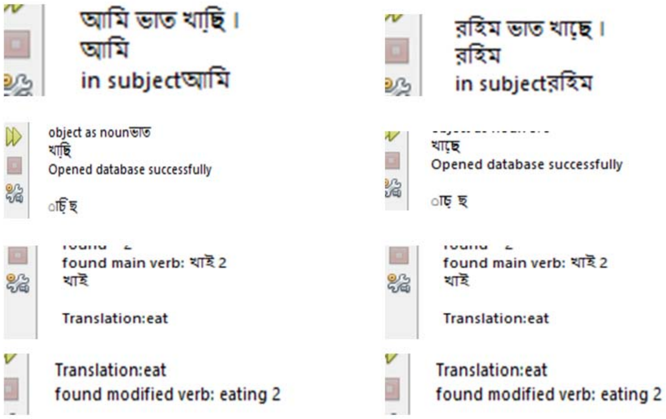


```
Input: রহিমও ভাত খায় ।
Rahim also eats rice BUILD SUCCESSFUL (total time: 2 seconds)
```

```
Input: রহিমই ফুটবল খেলে ।
Only Rahim plays Football BUILD SUCCESSFUL (total time: 2 seconds)
```

Figure 7. Emphasized name identification results

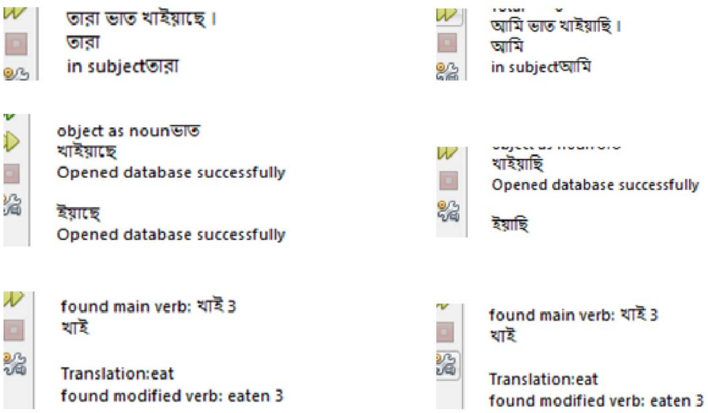
Figure 6 and 7 reflect the results of our emphasized name identification methodology. The figures show how the emphasizing tags, ’also’ and ’only’, have been identified and separated from the subjects, ’you’ and ’Rahim’, which ultimately leads to the construction of correct form of verbs by recognizing the number and the person of the subjects accurately.



```
আমি ভাত খাচ্ছি ।
আমি
in subjectআমি

object as nounভাত
খাচ্ছি
Opened database successfully

◌চ্ছি ছ

found main verb: খাই 2
খাই

Translation:eat

Translation:eat
found modified verb: eating 2
```

```
রহিম ভাত খাচ্ছে ।
রহিম
in subjectরহিম

খাচ্ছে
Opened database successfully

◌চ্ছে ছ

found main verb: খাই 2
খাই

Translation:eat

Translation:eat
found modified verb: eating 2
```

Figure 8. Database optimization technique (Present Continuous)



```
তারা ভাত খাইয়াছে ।
তারা
in subjectতারা

object as nounভাত
খাইয়াছে
Opened database successfully

ইয়াছে
Opened database successfully

found main verb: খাই 3
খাই

Translation:eat
found modified verb: eaten 3
```

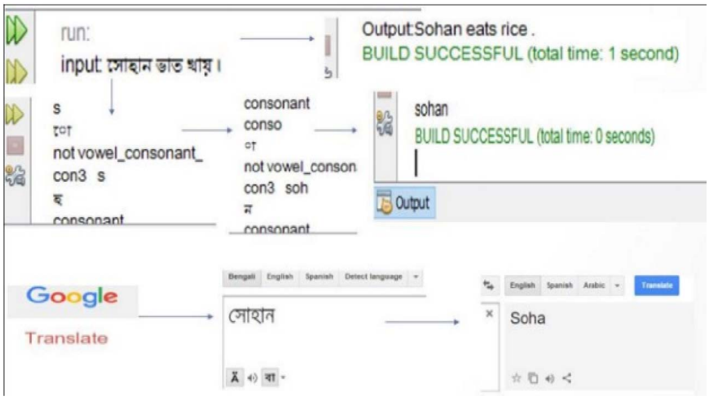
```
আমি ভাত খাইয়াছি ।
আমি
in subjectআমি

খাইয়াছি
Opened database successfully

ইয়াছি

found main verb: খাই 3
খাই

Translation:eat
found modified verb: eaten 3
```

Figure 9. Database optimization technique (Present Perfect)

In the figures 8 and 9, we show the database optimization technique using mapping of different forms of verbs to one main verb. The figures illustrate the mapping of different forms of the verb, ’eat’, in present continuous tense and present perfect tense. Also, different forms of ’eat’, occurring in Bengali, due to the subjects having different persons and numbers have been mapped to ’eat’ here by appropriate semantic analysis.

5.3. Findings

In this section we show a comparison of our proposed system with Google translator. Google translator completely fails to identify both the names (in figure 10)and the emphasized tags from the names (in figure 11).



```
run:
input: সোহান ভাত খায় ।
```

```
Output:Sohan eats rice .
BUILD SUCCESSFUL (total time: 1 second)
```

```
s
হো
not vowel_consonant_
con3   s
হ
consonant
```

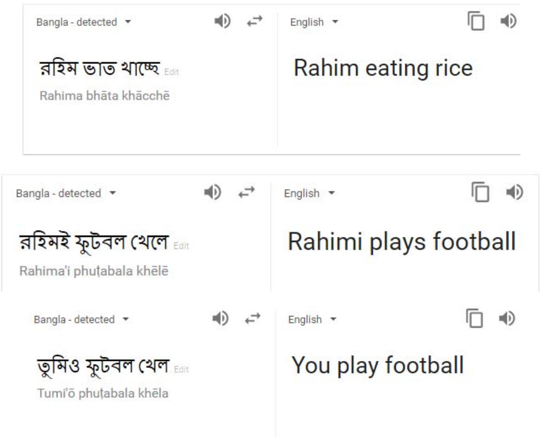
```
consonant
conso
হা
not vowel_conson
con3   soh
ন
consonant
```

```
sohan
BUILD SUCCESSFUL (total time: 0 seconds)
```

সোহান

Soha

Figure 10. Comparison of Google Translator with our proposed system for name identification



রহিম ভাত খাচ্ছে
Rahima bhāta khācchē

Rahim eating rice

রহিমই ফুটবল খেলে
Rahima'i phuṭabala khēlē

Rahimi plays football

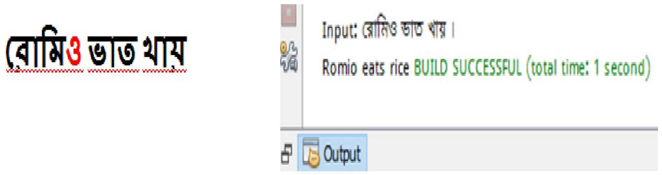
তুমিও ফুটবল খেল
Tumi'ō phuṭabala khēla

You play football

Figure 11. Faulty translation with Google Translator

Also, Google translator does not follow any optimization technique for identifying or predicting verbs from different forms of verbs.

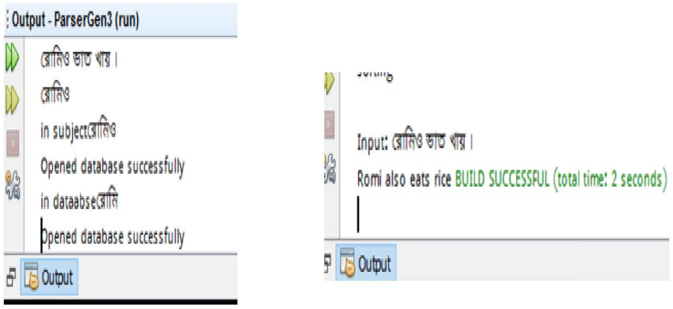
We have already shown the improvement achieved by our proposed system in figure 6 and 8. Figure 6 reflects the identification of emphasizing tags in subjects and figure 8 shows the database optimization technique for present continuous tense. However, one important observation regarding emphasized name identification is that the emphasized tag itself may also be the part of a valid name.



রোমিও ভাত খায়

```
Input: রোমিও ভাত খায় ।
Romio eats rice BUILD SUCCESSFUL (total time: 1 second)
```

Figure 12. Desired accurate translation



```
Output - ParserGen3 (run)
রোমিও ভাত খায় ।
রোমিও
in subjectরোমিও
Opened database successfully
in dataabseরোমি
Opened database successfully
```

```
sorting

Input: রোমিও ভাত খায় ।
Romi also eats rice BUILD SUCCESSFUL (total time: 2 seconds)
```

Figure 13. Ambiguous translation by our proposed system

Our system would separate that tag from the name which can lead to faulty name identification in such cases. However, the improvement achieved for most of the cases is much more significant than these very rare cases. Therefore, we need to tolerate the ambiguous name identification resulting to the faulty or unexpected translation here as a tradeoff as shown in figure 12 which shows the desired correct translation and figure 13 which shows the faulty translation by the proposed system in that particular case.

6. Future Works

- One of the main challenges in Bengali to English text conversion remains in implementing its vast grammatical rules. If we can track the core rules to acquire a generalized format for all rules and exceptions then the translation task will be simpler and compact.
- There is a great deal of research opportunities in language processing. Grammars keep changing as the language builds its grammar. Therefore, we need to find a translation process to update new sentence rules anytime. Machine Learning using Statistical Machine Translation can be one way to achieve it.
- Statistical language model can improve the translation quality. We plan to experiment on this model in future.
- Efficient AI techniques, indexing and searching mechanisms should improve the total system that might result in better output.
- Another idea is to extend the preposition handling component by adding more post positional words and inflectional suffixes.
- As wordnet does not have all sufficient information yet, preparing a Bengali wordnet can be a progressive approach.
- Developing Opennlp tools of Bengali sentences for parts of speech tagging of Bengali words efficiently is one of the most crucial tasks in Bengali to English translation. At present, Opennlp tools can recognize and process English sentences successfully.
- We also plan to make a machine translation system so that user can train it using AI techniques.
- Initially, our aim was to build a translation model for Bengali to Arabic conversion. However due to lack of proficiency in Arabic language, we had to start with conversion to English language. Therefore, we want to implement our proposed generalised skeleton of language processing for Arabic language soon so that we can help a large group of people to learn and understand Arabic language.

7. Conclusion

Natural language processing tasks are always complex and challenging due to a number of critical issues. Even the most sophisticated software cannot substitute the skill of a professional translator. There are so many reasons why machine translations are not as satisfactory as human translations. Ambiguity in translation mainly occurs due to one word having different meanings depending on the context. Also, there may be human emotions and expressions associated with a sentence causing ambiguity in expected meaning of that sentence. Disambiguation requires use of either shallow approach that uses statistical techniques to remove ambiguity or the more intelligent approach that involves comprehensive knowledge of a word. The former still leaves plenty scope for translation error while the latter is almost impractical to implement.

One of the reasons that translator cannot replace professional human translation is the same reason that plain old bilingual laypeople, for many tasks, cannot replace professional human translation. Most of the translation tasks require more than just knowledge of two languages. The idea that one can simply create one-to-one equivalencies across languages is wrong. Translators are not walking dictionaries. They recreate language. They craft beautiful phrases and sentences to make them have the same impact as the source. Often, they devise brand-new ways of saying things, and to do so, they draw upon a lifetime's worth of knowledge derived from living in two cultures. Machines or machine translators cannot exactly do that.

Almost in every language, the normal rules of grammar always consist of a number of exceptions. And keeping track of all those situations is a difficult task even for the intelligent beings. It massively demands wide and complex application of Artificial Intelligence to build up a near-accurate translator which on the contrary, may result in degradation of the overall performance of the system drastically. Hence, the efficiency in translating languages with complex grammatical rules is not too high. In our proposed system, we found that for simple sentences the system can easily respond with correct answer (or, may be just with an answer) immediately.

Our system currently focuses on only Bengali to English translation. However, it has limited knowledge base and vocabulary. By increasing the vocabulary and the knowledge base we can improve its efficiency by testing over wide range of different cases for general purpose use.

This is a very creative research topic. People generally thrive not only for translations between two languages but also for learning multiple languages equal effectively. There are many existing language learning websites at present. As they already have implemented some important concepts and features, exploiting and exploring them more, and even may be integrating them would be a useful task. Besides, many other creative features still remain to be thought of for making the learning easier, faster and importantly, interesting. Considering the limitations of a translator, our preference is always towards making the learning of a language easier by implementing all the basic translation procedures which is our proposed translation system all about.

References

[1] M. Islam and A. Islam, Polygot: Going Beyond Database Driven And Syntax-based Translation, ACM DEV '16: Proceedings of the 7th Annual Symposium on Computing for Development, November 2016.

[2] Z. Anwar, Developing a Bangla to English Machine Translation System Using Parts Of Speech Tagging: A Review, Vol. 1. No. 1, Journal of Modern Science and Technology, May 2013.

[3] R. Gangadharaiah, R. D. Brown, and J. G. Carbonell., Phrasal equivalence classes for generalized corpusbased machine translation. In Alexander Gelbukh, editor, Computational Linguistics and Intelligent Text Processing, volume 6609 of Lecture Notes inComputer Science, pages 1328. Springer Berlin / Heidelberg, 2011.

[4] S. Raphael, J. D. Kim, R. D. Brown, J. G. Carbonell, Chunk-Based EBMT. EAMT, 2010.

[5] M. Roy, A Semi-supervised Approach to Bengali-English Phrase-Based Statistical Machine Translation, Proceedings of the 22nd Canadian Conference on Artificial Intelligence, 2009.

[6] S. Dasgupta, A. Wasif, and S. Azam, An Optimal Way Towards Machine Translation from English to Bengali, Proceedings of the 7th International Conference on Computer and Information Technology (ICCIT), 2004.

[7] M. Anwar and M. Bhuiyan, Syntax Analysis and Machine Translation of Bangla Sentences, International Journal of Computer Science and Network Security, 09(08),317326; 2009.

[8] Optimal Way Towards Machine Translation from English to Bengali, In the Proceedings of the 7th International Conference on Computer and Information Technology (ICCIT), Bangladesh, 2004.

[9] Improving Example Based English to Bengali Machine Translation using WordNet; 2009.

[10] Bangla to English Text Conversion using opennlp Tools; Daffodil International University Journal Of Science & Technology, Vol. 8, Issue 1, JANUARY 2013 .

[11] G. Doddington, Automatic Evaluation of Machine Translation Quality Using N-gram CoOccurrence Statistics, Proceedings of the second international conference on Human Language Technology Research, 2002.

[12] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhou, A Study of Translation Edit Rate with Targeted Human Annotation, Proceedings of Association for Machine Translation in the Americas, 2006.

[13] S.p K. Naskar and S. Bandyopadhyay, A Phrasal EBMT System for Translating English to Bengali, Proceedings of the Workshop on Language, Artificial Intelligence, and Computer Science for Natural Language Processing Applications (LAICSNLP), 2006.

[14] D. Saha, S. K. Naskar, S. Bandyopadhyay, A Semantics-based English-Bengali EBMT System for translating News Headlines, MT Xummit, 2005.

[15] N. Karamat, Verb Transfer For English To Urdu Machine Translation, FAST-Lahore, 2006

[16] N. Chatterjee, S. Goyal, A. Naithani, Resolving Pattern Ambiguity for English to Hindi Machine Translation Using WordNet, Department of Mathematics, Indian Institute of Technology Delhi, Published in Workshop Modern Approaches in Translation Technologies, Borovets, Bulgaria, 2005.

[17] Example Based English to Bengali Machine Translation Thesis work of Khan Md. Anwarus Salam completed in August 2009.

[18] J. Tiedemann and L. Nygard, The OPUS corpus - parallel and free, Proceedings of LREC, 2004.

[19] OpenNLP, www.maxnet.sourceforge.net, accessed on July 13,2017 & Google Translator.

[20] D. Melamed, A Geometric Approach to Mapping Bitext Correspondence, Proceedings of the First Conference on Empirical Methods in Natural Language Processing (EMNLP), 1996.