International Journal of Computer Vision and Signal Processing, 6(1), 1-9(2016) ORIGINAL ARTICLE

PVBMT: A Principal Verb based Approach for English to Bangla Machine Translation

Masud Rabbani

Department of Computer Science and Engineering
Daffodil International University, Dhaka-1207, Bangladesh

Kazi Md. Rokibul Alam, Muzahidul Islam

Department of Computer Science and Engineering
Khulna University of Engineering and Technology, Khulna-9203, Bangladesh

Yasuhiko Morimoto

Hiroshima University, Higashi-Hiroshima 739-8521, Japan

Abstract

This paper proposes principal verb based machine translation (PVBMT), a new approach of machine translation (MT) from English to Bangla (EtoB) that runs in both web-based and mobile applications. The key mechanism is to detect the principal verb from any form of English sentence and then to transform it into the simplest form of English sentence *i.e*, subject plus verb plus object; identical to rule based MT (RBMT). Also while a 'prepositional phrase (PP)' or an 'idiom and phrase (I&P)' exists in a sentence, PVBMT uses its own corpus to tag and bind it properly; identical to statistical MT (SMT). While only RBMT is employed, often it generates feeble output because it requires the matching of various forms of English sentences with established grammatical rules stored in the knowledge-base. Therefore PVBMT employs hybrid machine translation (HMT) paradigm, a hybrid of RBMT and SMT. Finally the performance of PVBMT has been compared with a number of existing on-line EtoB translators employing a syntactic and a semantic analyzer. The experimental result shows that PVBMT can translate any form of English sentence *i.e*, Interrogative, Imperative, Exclamatory, Active, Passive, Complex or Compound along with an 'I&P' or a 'PP' with better accuracy than others.

Keywords: Machine Translation, English to Bangla, Natural Language Processing, Human Language Technology, Semantic Analysis, Syntactic Analysis.

ISSN: 2186-1390 (Online)
http://www.ijcvsp.com

Article History:
Received: 12 October 2015
Revised: 26 December 2015
Accepted: 11 March 2016
Published Online: 18 June 2016

1. INTRODUCTION

Machine translation (MT) is a process that enables the automatic translation of one natural language to another one employing computing device. Linguistic rules are used in MT for translating a source language into a targeted one. Thus MT helps to establish convenient communication among the inhabitants of different native languages. Also to preserve and expose the tangible and intangible heritage of any nation, language plays an important role. Bangla is one of the most popular Indo-Aryan languages. Now it is worlds sixth ranked language and has about 220 million native and 250 million total speakers [1] all over the world. Besides, international mother language day [2] is now observed only for Bangla language; nevertheless it lags behind in research areas like parts-of-speech (POS) tagging, text summarization, and most importantly in MT from English to Bangla (EtoB). Nowadays natural languages like English, Hindi, Japanese etc have been rapidly progressing in these aspects. Bangla has a great

Email addresses: masud.cse@diu.edu.bd (Masud Rabbani), rokibcse@yahoo.com (Kazi Md. Rokibul Alam), ashraf6892@gmail.com (Muzahidul Islam), morimoto@mis.hiroshima-u.ac.jp (Yasuhiko Morimoto)

opportunity to work in EtoB MT, because it has demand in numerous applications.

The main approaches of MT are: Rule-based MT (RBMT), Statistical MT (SMT), Example-based MT (EBMT), and Hybrid MT (HMT) [3]. RBMT (also known as “Knowledge-Based MT“) comprises of a group of semantic, morphological and syntactic rules that translates the structure of a source sentence into the structure of a target sentence. Although RBMT can generate new rules, usually to change an existing rule or to generate a new one is costly and may generate poor accuracy. SMT is used in statistical methods based on bilingual text corpora for translating similar text. However sometimes, corpus is highly expensive and rare for many language pair. EBMT is a mode of MT which use bilingual corpus with its knowledge-base for translation analogy. HMT is not a unique approach, is a combination of RBMT and SMT which leverages the strength of MT.

In any natural language there are variations in case of structures of sentences and it may change for various entities like people, place, time, etc. Therefore in many cases, it is difficult to execute MT with predefined rules. However principal verb based MT (PVBMT) [4, 5, 6] approach proposed in this paper, contains only a simple structural rule. It always pre-processes the words of an English sentence and then generates some nominal and verbal groups. If there is more than one nominal group, then one nominal group is used as a subject and the other one is used as an object, and the verbal group is used as the principal verb to translate from EtoB; identical to RBMT. Besides, while a ‘prepositional phrase (PP)’ or an ‘idiom and phrase (I&P)’ exists in a sentence, PVBMT seeks it within its database and binds it with appropriate meaning; identical to SMT. Thus PVBMT is a hybrid of RBMT and SMT, belongs to HMT paradigm, and can translate any form of English sentence even a complex or compound one along with a ‘PP’ or an ‘(I&P)’, without any complexity.

The rest of the paper is organized as follows. Section 2 summarizes some related works. Section 3 describes the methodology of PVBMT for EtoB MT. Section 4 illustrates the experimental studies. Finally Section 5 concludes the paper.

2. RELATED WORKS

Recently extensive research on EtoB MT has been conducted. The approach proposed in [7] uses RBMT for EtoB and follows methodology especially ‘fuzzy method’. At first it splits words from a sentence, and then lexically adds attributes to words. These lexemes are essential to determine the grammatical and the sentence structure of the source sentence. After that for the given English sentence, a fuzzy rule (r) is found out which may be matched fully or partially. Here a dictionary is used to find the corresponding Bangla words for the lexemes. Finally, Bangla sentence is reconstructed according to the rule (r) which has predefined correspondence with (r). However the approach cannot translate different types of English sentences properly because all types of sentences do not match fully with (r).

A phrase based EtoB MT approach has been proposed in [8]. It is based on SMT which needs millions of parallel bilingual text corpora. For better translation, it emphasizes to generate rules for preposition binding. The preposition handle module of this approach is divided into two parts: (1) pre-process sub-module and (2) post-process sub-module. To handle out-of-vocabulary (OOV) words, a module named ‘Transliteration‘ is added. However the existence of parallel corpora for EtoB is very few, therefore the quality of this MT is not so high, only sufficient for short sentences.

Another statistical phrase-based MT approach proposed in [9] employs some novel active learning (AL) strategies of statistical translation for better performance. Here, a small amount of parallel text and a large amount of monolingual source language text have been used as novel AL strategies. At first it creates a large noisy parallel text, and then improves it by using small injections of human translation. Thus before experiments, it avoids the use of any knowledge in AL. However for better accuracy, the approach needs to increase the coverage of bilingual training data which is also tough.

In [10], an EBMT based approach has been proposed that operates in five steps. These are: (1) Tagging, (2) Parsing, (3) Preparing chunks of the sentence using sub-sentential EBMT, (4) Matching the sentence with adapting scheme rule, and (5) Translating the chunk to generate the output with morphological analysis. But it can translate only simple sentences, cannot translate sentences that do not match to the knowledge-base. Also it cannot determine words which are not stored in the dictionary and cannot choose the appropriate meaning for multi-meaning words. However, it has defined a way to translate complex sentences using sub-sentential EBMT.

The approach proposed in [11] improves the approach of [10] employing WordNet and International Phonetic Alphabet (IPA) based transliteration. For an unknown word, first it tries to find semantically related English words form WordNet. If the unknown word is not found in the English IPA dictionary, finally it uses Akkhor transliteration mechanism, and thereby improves the quality of translation. However to generate chunk-string templates (CSTs) it uses a small parallel corpus that decreases its performance. For better accuracy, still it needs a more balanced parallel corpus.

Another approach proposed in [12] is based on EBMT where “Translation Memory (TM)” technique is used for reusing the example from the existing translations. So at first they develop a parallel corpus in a particular field (*i.e.* patient-receptionist dialogue). The approach consists of three steps. At first in the matching step, it finds the closest sentence (S_c) from the source language example for the given input sentence (S). Then in the adapting step, the mismatch portions of S are extracted from S_c and its target equivalent (Sct). At last in the recombination step, necessary segments are added or substituted from S with Sct for getting Bangla translation. Though the approach has accuracy about 57.56% according to BLEU, it is tough to develop a parallel corpus for EtoB and to build a high quality TM is very expensive. Moreover, errors can easily propagate for wrong detection of S_c for any S.

A syntactic transfer based EtoB MT approach has been proposed in [13] which uses Cockey-Younger-Kasami (CYK) algorithm for parsing. It consists of five steps, which are: (1) Tagging, (2) Parsing, (3) Change CNF parse tree to normal parse tree, (4) Transfer of English parse tree to Bangla parse tree, and (5) Generation with morphological analysis. It is suitable for simple English sentences, and requires grammar to be in Chomsky Normal Form (CNF). But its problem is that, directly transferring from English parse tree to Bangla parse tree is not so easy. Therefore it needs to change the English parse tree generated via CYK parsing algorithm into another form of English parse tree.

In [14], an approach of handling English prepositions (Ps) for EtoB MT has been proposed where English Ps are handled in Bangla using inflections and / or post-positional words. It translates through anyone of the following three separate steps. These are: (1) Translating English Ps using inflections in Bengali, (2) Translating English Ps using inflections and Ps in Bangla, and (3) Translation of English idiomatic Ps.

The approach proposed in [15] adopts EBMT to translate from English to Hindi and consists of three steps. These are: (1) Building a parallel corpus, (2) Matching and retrieval, and (3) Adapting and recombination. It has a trained corpus which is a parallel database and consists of 677 sentences. Hence for translation, it relies on the words stored in its corpus. As a result, its overall performance mainly depends on this parallel corpus. It also has challenges in case of matching and adaptation. It performs well for sub-language phenomena like - phrasal verbs, not so perfect for English sentence structure.

Analyzing the above approaches it has been observed that EBMT or SMT performs well while an appropriate corpus exists. Similarly, RBMT also needs to match any form of English sentence with an established grammatical rule of knowledge-base. Therefore the proposed PVBMT is in HMT paradigm, a hybrid of RBMT and SMT. Herein, to bind any form of English sentence it uses rules to convert it as subject plus verb plus object, for proper Bangla meaning. It also needs a corpus to translate a 'PP' and / or an 'I&P' within a sentence.

3. PRINCIPAL VERB BASED MACHINE TRANSLATION (PVBMT)

PVBMT consists of several steps and proceeds as follows. Fig. 1 is the flow chart of it.

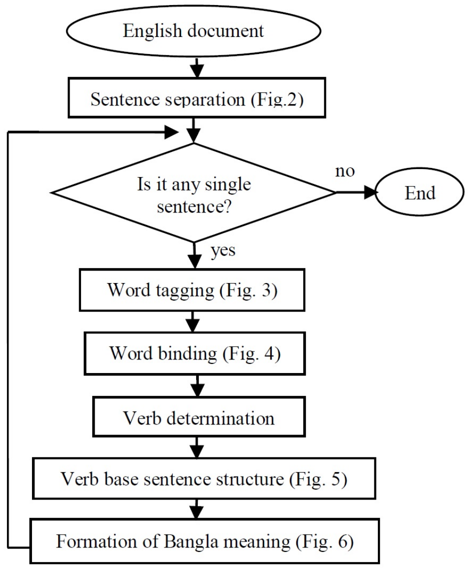


Figure 1: Flow chart of PVBMT for EtoB MT.

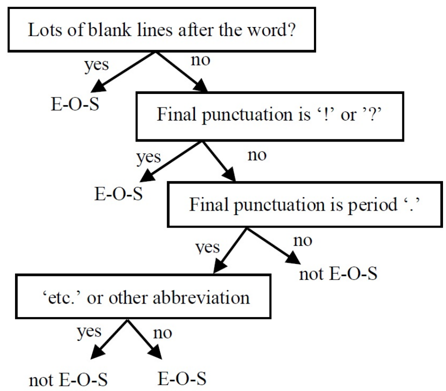


Figure 2: Decision tree to detect the end of a sentence (E-O-S).

3.1. Sentence separation

To translate any single English sentence, the first step of PVBMT is to separate individual English sentence if there exists a paragraph. Initially, a single sentence has been split from the given paragraph according to the rule proposed in [16] which is shown in Fig. 2.

If there is lots of blank space after any word, then it is the end of a sentence (E-O-S). If there is any punctuation symbol such as '?' or '!' after any word, then it is also an E-O-S. Otherwise it needs to check if there exists any '.' such as 'Mr.' or 'Mrs.' or 'etc.'. If so, it is not an E-O-S; otherwise it is an E-O-S.

3.2. Word tagging

The second step of the proposed PVBMT is lexical analysis. In this step the words of a sentence are separated as tokens. Normally words are split according to white

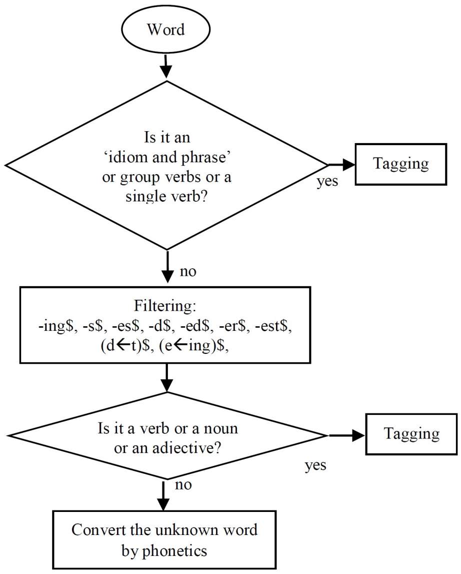


Figure 3: Flow chart for word tagging.

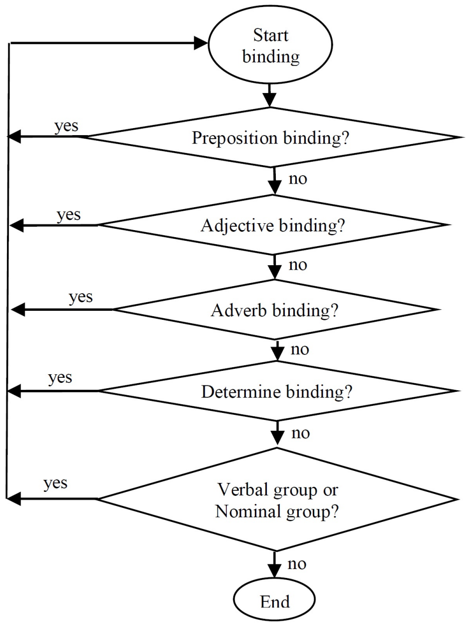


Figure 4: Flow chart for word binding.

space. Then the words are tagged with their corresponding meanings as well as POS. The procedure is shown in Fig. 3.

At first PVBMT searches a word either it is a 'PP' or an 'I&P' or group verbs or a single verb. If the matching exists it is tagged with the meaning and POS. To match any 'PP' or any 'I&P', PVBMT maintains a 'json' [17] database as a repository. At first PVBMT tries to match every token one after another consecutively (words within a sentence) with the repository word list. If there is any exact matching with an 'I&P', this group of tokens is tagged as an 'I&P' as well as with its' meaning. Here it is mention worthy that PVBMT maintains a rich repository for 'I&P'.

If not an 'I&P', PVBMT filters and matches the word either as a verb or a noun or an adjective. For this purpose some rules are employed as described in [18] to find the original words. A few examples of them are given below:

- Elimination of 'ed', 'd', 's', 'es', 'ing', 'r', 'er', 'st' from the last portion of the word to search it in the database. Some examples are: 'added → (add - ed) → add', 'agreed → (agree - d) → agree', 'boxes → (box - es) → box', 'harder → (hard - er) → hard'.
- Elimination of 'ing' and addition of 'e' at the last portion of the word to search it. An example is: 'coming → (com - ing + e) → come'.
- Elimination of 'ing' from the last portion of a word to determine, whether the last letter is 'y' or not. If so, add 'ie' at the last portion to search it. For example, 'lying → (ly - ing) → (l - y + ie) → lie'.
- Elimination of 'ied' / 't' and addition of 'y' / 'd' respectively at the last portion of the word to search it. Examples are: 'fried → (fr - ied + y) → fry', and 'sent → (sen - t + d) → send'.
- Elimination of 'ed' or 'ing' or 'er' or 'est' from a word and then to determine whether the last two letter are double consonant or not. If so, then it also needs to eliminate the last letter to search the word. An examples is: 'dropped → (dropp - ed) → (dropp - p) → drop'.

Normally nominal groups are failed to be tagged, therefore PVBMT simply converts these words with the phonetics according to the above procedure. For example, any proper noun like 'Masud Rabbani' is just converted to its phonetics 'মাসউদ রাববানি' (: 'masud rabbani').

3.3. Word binding

The third step of PVBMT is word binding. Here the words that are tagged in the previous step are binded according to the mechanism presented in Fig. 4. This is an iterative process, and it iterates until the tagged words are formed as subject or verb or object.

This is an important step of PVBMT where at first the preposition is bound. In our database, words are stored with the properties of the meaning and POS. Here POS are described with various properties, namely a 'noun' may be categorized into five types. These are: proper noun, common noun, material noun, abstract noun and collective noun. So all POS are tagged with various properties and these properties are used for preposition binding as well as to determine the proper meaning. For example, the preposition 'at' is different for different properties of words and some of which are presented in Table I.

Table I: PREPOSITION BINDING

Property of word	**Example**	**Meaning**
Place/Position	At the capital At the home	রাজধানীতে (rajdhanite) বাসায় (basay)
Time	At 10 a.m. At noon	সকাল (shokal) ১০ (dosh) টায় (tai) দুপুরে (dupure)
Direction/Movement	Go to school	স্কুলে (schoola) যাও (jao)

Table II: ADJECTIVE BINDING

Adjective	**Example**	**Meaning**
Attributive Adjective (Adj+Noun)	The Intelligent girl stood first in the examination.	বুদ্ধিমতি (buddi-moti) মেয়েটি (meyeti) পরীক্ষায় (porik-khay) প্রথমস্থান (prothom-sthan) অধিকার (odhikar) করেছিলো (kora-silo)
Predicative Adjective (Verb+Adj)	The girl is very intelligent	মেয়েটি (meyeti) বেশ (besh) বুদ্ধিমতি (buddi-moti)

After preposition binding, PVBMT binds adjective with other words. There are two kinds of adjective according to their binding which are shown in Table II.

Then PVBMT binds adverb with other words. It is known that an adverb is a word used to modify any POS (except a noun or pronoun) or a whole sentence. So it is very important to ensure the correct binding of an adverb. According to [19] adverbs are divided into three categories. These three categories are also divided into many sub-categories and some of them are shown in Table III.

After adverb binding, the next step of binding is to bind a determine. There are two types of determines:

- Definite determine——The
- Indefinite determine—-A, An

And their meanings are as follows: A, An = একটি (ekti) /একজন (ekjon); The = টি (te) / টা (ta).

'A/An' is always placed before a noun and makes the noun indefinite. If the category of the noun is material, the meaning of 'A/An' is একটি ('ekti'). If the word is related to a person type, the meaning is একজন ('ekjon'). Definite determine 'The' is usually placed before a common noun. When it is a countable common noun, the meaning of 'The' is টি/টা ('te/ta'). On the other hand, if the noun is uncountable, the meaning of 'The' is avoided.

Table III: CATEGORY OF ADVERB

Adverb	**Classification**	**Example**
Simple	Time	Now, Then, Before, Since etc.
	Place	Here, Far, Near Locally etc.
Relative / Conjunction		Where, When, Why, How.
Interrogative		Where, When, Why, How.

Table IV: EXAMPLE OF A GROUP VERB

Group Verb	**Meaning**
1. Lila laughs.	লিলা (Lila) হাসে (hasha)
2. Lila laughed at the poor	লিলা (Lila) দরিদ্রকে (doridro-ka) বিদ্রুপ (bidrup) করলো (korlo)

The next step of word binding is to bind any group verb. If a preposition is added with a verb or a collection of verbs (known as compound verb) and it gives idiomatic meaning, then it is called a group verb or prepositional verb or a phrasal verb. It is very important to bind group verb because different group verb has different meaning. Table IV presents an example of a group verb.

3.4. Verb determination

The fourth step of PVBMT is to determine the verbs in a sentence. There may be three types of verbs within a sentence. These are: auxiliary verb (AV), finite verb (FV) and non-finite verb (NV). If a sentence has more than one verb, then PVBMT creates different sets for different types of verbs with their meanings and positions. So for determining a verb, a sentence is scanned from the left to the right and then all verbs *i.e.* AV, FV and NV are detected. Then each verb is determined whether it is an AV or FV or NV according to their position and structure in the sentence. The following examples show that a sentence can possess one or more verbs.

- I saw (FV) him running (NV)
- I am eating (FV) and writing (FV)
- I have (FV) a cat.

In the above, the first sentence has two verbs (*i.e.* saw and running). However according to word tagging as described above in Step 2, 'running' (run+ing) is detected as a verb which has no AV. Here according to English grammar without an AV like: 'verb + ing', a verb is called a NV. Therefore here, 'saw' is FV and 'running' is NV.

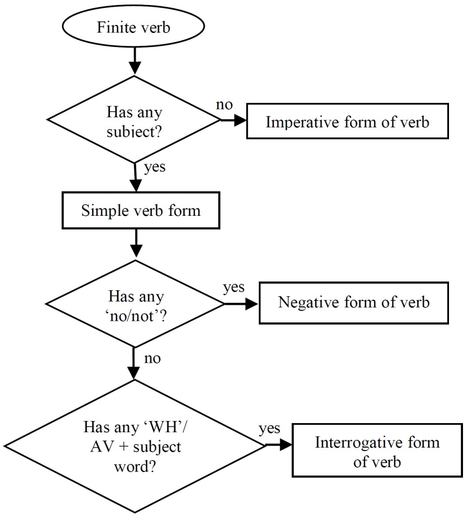


Figure 5: Flow chart to decide the form of a sentence based on verb.

Therefore while tagging any NV; it needs to add 'এ' ('a') at the end of Bangla meaning. In the above, the second sentence has two verbs. Here both of them are connected with a conjunction 'and' and as a form of 'verb + ing'. Since there is an AV 'am' before these verbs, therefore both of them are FV. In the above third sentence, there is only one verb 'have' and it is FV.

3.5. Verb base sentence structure

In this step PVBMT defines the Bangla sentence structure corresponding to the English sentence. If there exists more than one FV in a sentence, then at first the sentence is scanned from the right side to the finite verb and then subject, object, AV or NV (if exists) are collected to construct the sentence structure for that particular sentence or clause. An example of this is given below. Fig. 5 is the flow chart of this process.

- I saw him running.

  sub + FV + obj + NV

  ←

- sub + obj +NV + FV

  I him running saw.

If any FV exists without any subject, the sentence is formed as an imperative structure. Otherwise it is constructed in simple form (*i.e.* either negative or interrogative). If there exists any negation word (*i.e.* no, not etc), then the sentence is formed as a negative sentence structure. And if there is any 'WH' question word or 'AV + subject' form, then an interrogative sentence structure is constructed.

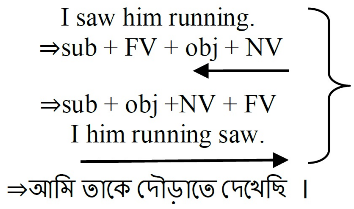


Figure 6: Formation of a Bangla sentence from an English sentence.

3.6. Formation of Bangla meaning

The last step of PVBMT is to generate the final Bangla sentence according to the modified English sentence structure obtained from step 5. Here all English words are translated into Bangla from the left to the right. Fig. 6 presents an example of this process. PVBMT doesn't translate any AV within a sentence because usually an AV (except have) has no meaning in case of EtoB MT.

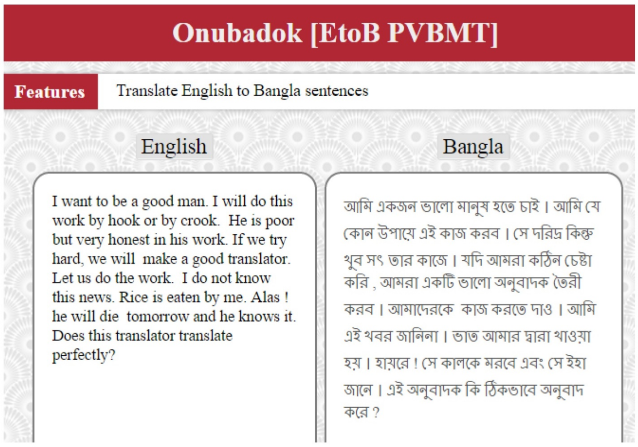
4. EXPERIMENTAL STUDIES

4.1. Experimental Setup

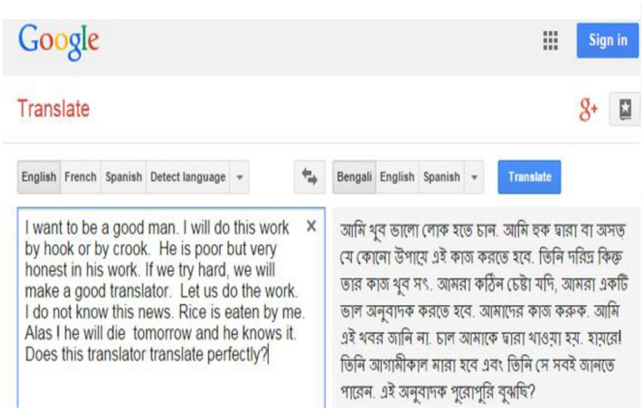
PVBMT has been developed under the environment on Intel Core i5-2.30 GHz processor with 4.0 GBytes of RAM running on Windows 8 operating system. Here the desktop application of PVBMT has been developed in JavaScript employing Microsoft Visual Studio Express 2013 [20] where 'Json' [17] & LocalStroage [21] have been used to store data. Also, the web application has been developed by using PHP [22] and JavaScript with Aptana Studio 3 [23] where MySql [24] has been used for maintaining the database, and available in the web as "Onubadok [EtoB PVBMT]" [4]. Besides, the android based application has been developed in minimum API level-15 and target API level-17. Also according to our knowledge, there is no benchmark dataset to test the output of EtoB translators. Therefore a test dataset containing samples of various forms of English sentences has been created and available at [25]. Among them several samples have been used to show the output of translators, and all have been used to compare their accuracy semantically and syntactically.

4.2. Experimental Results and Comparisons

For translating from EtoB, the experimental results employing PVBMT has been presented in Fig. 7 (a). Herein different forms of English sentences have been chosen for EtoB MT namely, the 1st one is a 'PP', the 2nd one is an 'I&P', the 3rd one is a compound, the 4th one is a complex, the 5th one is an imperative, the 6th one is an negative, the 7th one is a passive, the 8th one is an exclamatory, and the last one is an interrogative sentence. Also the same input has been applied for another two translators available in the web. One of them is Google



(a)



(b)

(c)

Figure 7: EtoB MT for several forms of sentences employing: (a) PVBMT; (b) Google translator; and (c) Anubadok online.

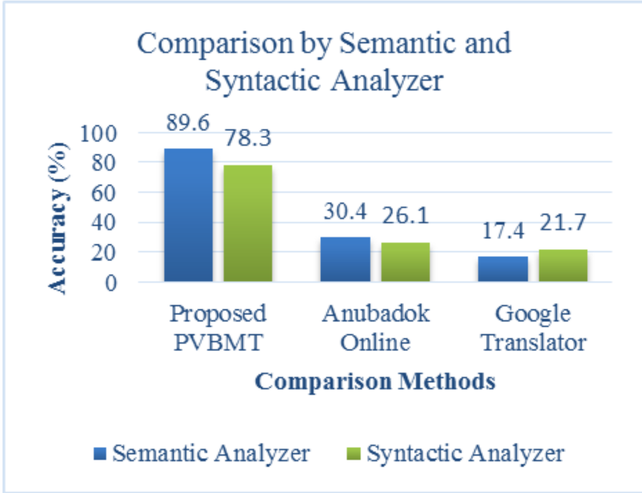


Figure 8: Employing syntactic [28] and semantic analyzer [29], the accuracy of PVBMT, Anubadok online, and Google translator.

translator [26] and the other one is Anubadok Online [27]. Fig. 7 (b) and 7 (c) shows their MT output respectively.

Moreover according to our knowledge up to now, although there is no standard platform to test the accuracy of EtoB translators' semantically or syntactically, a syntactic analyzer proposed in [28] and a semantic analyzer proposed in [29] have been used to compare the accuracy of output of these translators through the same input which has been shown in Fig. 8.

4.3. Discussions

The experimental results presented in the above section shows that PVBMT is capable enough to translate any form of English sentence including 'I&P'. Moreover, the effect of MT of an 'I&P' within a sentence is trained also. Herein, PVBMT tries to bind a group of words as an 'I&P' and then tags it with proper POS and meaning in assistant with the database. For example, in the 2nd sentence of Fig. 7 (a), there exists a phrase 'by hook or by crook'. Here, PVBMT binds and tags the group of words 'by hook or by crook' as a phrase with its' Bangla meaning 'যে কোন উপায়ে (ja kono upaya)'. To do this, at first PVBMT finds 'I' and tags it as "noun", which means 'আমি (ami)' and selects it as the 'subject' of the sentence. Then PVBMT finds 'will' which is an AV, and tries to bind it with the next word *i.e.* with 'do'. Now it binds these two words *i.e.* 'will do' and tags it as "verb" which means 'করব (korbo)'. Next PVBMT finds 'this' and tries to bind it with the successive words *i.e.* with 'work', and binds the words 'this work' and tags it as "noun", which means 'এই কাজ (ei kaj)' and selects it as the 'object' of the sentence. Now PVBMT finds 'by' which is a preposition, so it takes the next words and tries to bind them as either a 'PP' or an 'I&P'. Then it finds 'hook', and searches it in the database with 'by hook' and finds that it is a portion of 'by hook or by crook'. So it tries to find the next words as 'or by crook', and finds the words as 'or by crook'. Now it can bind the words as an 'I&P' and tags it as "adjective", with meaning 'যে কোন উপায়ে (ja kono upaya)'. Thus in this sentence the sequence of words that PVBMT finds is: noun (subject) + verb + noun (object) + adjective (phrase). Now for MT, the sequence that PVBMT follows is: noun (subject) + adjective (phrase) + noun (object) + verb. However in case of other two approaches [26, 27], they cannot translate this sentence properly for the lack of proper binding effect of the 'I&P'.

To evaluate PVBMT and to compare the output of other translators presented above, the syntactic [29] and the semantic [28] analyzers that have been used, are not known as the standard evaluation technique. Except these, even no human judge has been employed for this purpose, which seems to be the limitation of this work. A future plan is to employ BLEU [30] and other standard MT evaluation metrics for this purpose.

5. CONCLUSION

The proposed PVBMT can translate utmost forms of English sentences from EtoB, along with a 'PP' or an 'I&P' with proper meaning while existing in a sentence. The experimental result shows that, the overall performance of PVBMT is better than other considered EtoB translators. Eventually PVBMT is simple and efficient enough. The reason is that it handles any form of English sentence in a simple and intuitive way; just it detects the principal verb of the sentence and then transforms it into the simplest English sentence structure i.e., subject plus verb plus object. Also to translate a 'PP' or an 'I&P' within a sentence, it uses its own corpus. Thus PVBMT is a HMT paradigm. Traditional paradigms like RBMT, EBMT, SMT, etc are usually efficient for simple sentences. Using RBMT, it is difficult to define every rule of Bangla language. Also to develop a parallel corpus for EtoB is difficult and expensive. Thus RBMT, EBMT, SMT, etc are not so perfect while deployed alone. Although PVBMT is good, still it has limitations in case of correct word tagging because of inappropriate matching of words from the database. It is expected that from the database the manual tuning of words with expert, would make the database more accurate and thereby the limitation would be resolved. Although time consuming, it is expected that very soon it would be implemented. Another future plan is to adopt the Transliteration to handle unknown words.

References

[1] W. T. F. Encyclopedia, Bengali language, Tech. rep., Wikimedia Foundation Inc (2015).

[2] A. Rahim, Canadian immigration and south asian immigrants, booktitle = Xlibris Corporation, optpages = 102, optyear = 2014.

[3] W. T. F. Encyclopedia, Machine translation, Tech. rep., Wikimedia Foundation Inc (2015).

[4] M. Rabbani, A. K. M. R. Alam, M. Islam, Onubadok [pvbmt etob], Tech. rep., Khulna University of Engineering and Technology (2015).

[5] M. Rabbani, A. K. M. R. Alam, M. Islam, A new verb based approach for english to bangla machine translation, in: Proceedings of International Conference on Informatics, Electronics and Vision (ICIEV14).

[6] M. Rabbani, A. K. M. R. Alam, M. Islam, Onubadok [pvbmt etob], Tech. rep., Khulna University of Engineering and Technology (2015).

[7] J. Francisca, M. M. Mia, S. M. M. Rahman, Adapting rule based machine translation from english to bangla, IJCSE.

[8] M. Z. Islam, J. Tiedemann, A. Eisele, English to bangla phrase-based machine translation, in: 14th Annual Conference of the European Association for Machine Translation.

[9] H. Gholamreza, M. Roy, A. Sarkar, Active learning for statistical phrase-based machine translation, in: Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

[10] K. M. A. Salam, M. Khan, T. Nishino, Example based english-bengali machine translation using wordnet, in: Triangle Symposium on Advanced.

[11] T. N. K. M. A. Salam, S. Yamada, How to translate unknown words for english to bangla machine translation using transliteration, in: Journal of Computers.

[12] S. Dandapat, S. Morriessy, S. K. Naskar, H. Somers, Statistically motivated example-based machine translation using translation memory, in: 8th International Conference on Natural Language Processing.

[13] S. Dasgupta, A. Wasif, , S. Azam, An optimal way of machine translation from english to bengali, in: 7 th International Conference on Computer and Information Technology (ICCIT).

[14] S. K. Naskar, , S. Bandyopadhyay, Handling of prepositions in english to bengali machine translation, booktitle = Third ACL-SIGSEM Workshop on Prepositions, optpages = 89-94, optyear = 2006.

[15] M. B. C. R. A. Sinhal, A pure example based approach for english to hindi sentence mt systems, in: Journal of Computational Linguistics and Natural Language Processing.

[16] D. Jurafsky, Sentence segmentation and decision tree, Tech. rep., Stanford University (2012).

[17] Json, Tech. rep. (2015).

[18] Applied English Grammar and Composition, Kala Chand Dey, 2015-2016.

[19] A Passage to the English Language, by Rohel Publication, Dhaka, 1999.

[20] Visual studio, Tech. rep. (2015).

[21] Localstroage, Tech. rep. (2015).

[22] Php, Tech. rep. (2015).

[23] Aptana studio 3, Tech. rep. (2015).

[24] Mysql, Tech. rep. (2015).

[25] Test dataset for etob translators, Tech. rep. (2015).

[26] Google translate, Tech. rep. (2015).

[27] Anubadok online, Tech. rep. (2015).

[28] H. K. M. Azharul, M. Hozaifa, S. Dutta, Detection of semantic errors from simple bangla sentence, booktitle = 17th International Conference on Computer and Information Technology (ICCIT-17), optyear = 2014.

[29] M. I. R. Shuvo, R. Z. Rabbi, Shift reduce parsing for bangla grammar, Master's thesis, Department of Computer Science and Engineering, Khulna University of Engineering & Technology (KUET) (2014).

[30] T. W. W. Z. K. Papineni, S. Roukos, Bleu: A method for automatic evaluation of machine translation, in: Proc. of 40th Annual Meeting of the Association of Computational Linguistics.