Independent Study Report

Improving Example Based English to Bengali Machine Translation using WordNet

Supervisor Name: Dr. Mumit Khan

Student Name: Khan Md. Anwarus Salam

Student ID: 07141002

Email: kmanwar@gmail.com

January 2009

DECLARATION

I hereby declare that this Independent Study is based on the results found by myself. Materials of work found by other researcher are mentioned by reference. This Independent Study, neither in whole nor in part, has been previously submitted for any degree.

Signature of
Supervisor

Signature of
Author

ACKNOWLEDGMENTS

I have to thank specially Dr. Mumit Khan for over viewing my Independent Study work to completion. I also thank all team members of Center for Research on Bangla Language Processing for assisting me in every step.

1. ABSTRACT

The goal of this research topic is to develop efficient machine translation system for English to Bengali language by improving my Thesis work “Example Based English to Bengali Machine Translation”. Due to relevance of this work I kept the most of the texts of my thesis. To develop an efficient machine translation system is very important but it is really expensive as it requires a huge amount of time and resources. In all languages there are many words that may have multiple meanings and also some sentence may have multiple grammar structure to express the same meaning, it is a great challenge to do the right semantic analysis. But it is very important to have a machine translation system which can compute all possible outputs in reasonable time and able to choose the best option. We can dramatically improve the performance of English to Bengali Example Based Machine Translation using WordNet. For example the ‘have’ verb has more than ten different meaningful uses during English to Bengali translation. Using the word senses given by WordNet we can dramatically improve the performance of Example Based Machine Translation (EBMT) Depending on various characteristics of words. The proposed EBMT system has five steps: 1) Tagging 2) Parsing 3) Prepare the chunks of the sentence using sub-sentential EBMT 4) Using an efficient adapting scheme match the sentence rule 5) Translate from Source Language (English) to Target Language (Bengali) in the chunk and generate with morphological analysis with the help of WordNet.

Keywords: Machine Translation, WordNet, English to Bengali, EBMT, Example Based, Adaptation

TABLE OF CONTENTS

Page

Chapter 2: Introduction

The title of my research is "Improving Example Based English to Bengali Machine Translation using WordNet". The goal of this research topic is to develop efficient machine translation system for English to Bengali language by improving my Thesis work “Example Based English to Bengali Machine Translation”. Due to relevance of this work I kept the most of the texts of my thesis. To develop an efficient machine translation system is very important but it is really expensive as it requires a huge amount of time and resources. In all languages there are many words that may have multiple meanings and also some sentence may have multiple grammar structure to express the same meaning, it is a great challenge to do the right semantic analysis. But it is very important to have a machine translation system which can compute all possible outputs in reasonable time and able to choose the best option. We can dramatically improve the performance of English to Bengali Example Based Machine Translation using WordNet. For example the ‘have’ verb has more than ten different meaningful uses during English to Bengali translation. Using the word senses given by WordNet we can dramatically improve the performance of Example Based Machine Translation (EBMT) Depending on various characteristics of words. The proposed EBMT system has five steps: 1) Tagging 2) Parsing 3) Prepare the chunks of the sentence using sub-sentential EBMT 4) Using an efficient adapting scheme match the sentence rule 5) Translate from Source Language (English) to Target Language (Bengali) in the chunk and generate with morphological analysis with the help of WordNet.

In present there are many ways of machine translation system. Many researchers came up with different approaches. But still it is not possible to get the finest possible result. I want to use the example based machine translation system, to get all possible outputs. For achieving this I have to plan to prepare a dictionary with morphological analysis and a Parallel Corpus. Then from semantic analysis it may possible to choose the best desired output.

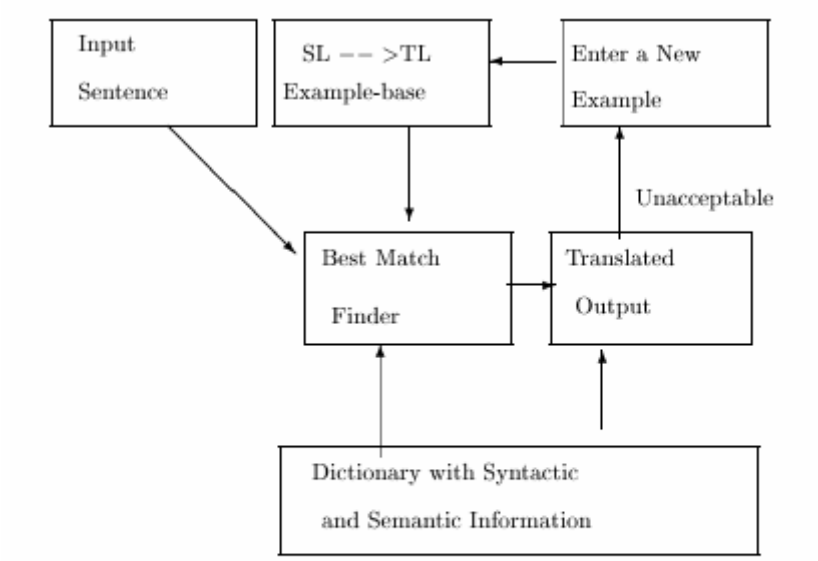


Figure 1: Role of Examples in Translation

Chapter 3: Background

3.1 What is Machine Translation?

Machine Translation is the process of translating text units of source language into a target language by using computers. The term Machine Translation can be defined as “translation from one natural language (source language (SL)) to another language (target language (TL)) using computerized systems, with or without human assistance” (Hutchins and Somers 1992, pg. 3).

3.2 Generations and Types of Machine Translation

Machine translation systems can be divided in two generations direct systems and indirect systems. First generation systems are known as direct systems. In such systems, translation is done word by word or phrase by phrase. In such systems very minimal linguistic analysis of input text is conducted (Hutchins and Somers 1992). This architecture is still being extensively used in commercial MT systems. The main idea behind direct systems is to analyze the input text to the extent that some transformational rules can be

applied. This analysis could be parts of speech of words or some phrasal level information. Then using a bilingual dictionary, source language words are replaced with target language words and some rearrangement rules are used to modify the word order according to the target language (Arnold et al. 1993).

This architecture is very robust because it does not fail on any erroneous or ungrammatical input. Since the analysis level is very shallow and the system contains very limited grammatical information, it hardly considers anything ungrammatical. In the worst case if the rule does not apply to the input, the input is passed on without any alteration as output. This kind of system is hard to extend because all the rules are written in one direction and are language specific. To make another language pair work, all the rules have to be re-written. Since the system does not perform very deep analysis, its time complexity is low. These systems work very well for closely related languages but are not suitable for modeling languages with diverse syntactic nature. Since the system does not explicitly contain the grammatical rules of the target language, there is a chance that the output will not be grammatical but it will be similar to the target language (Arnold et al. 1993)

Owing to the fact that linguistic information helps an MT system to produce better quality target language translation, with the advance of computing technology, MT researchers started to develop methods to capture and process the linguistics of sentences. This was when the era of second generation MT systems started. Second generation machine translation systems are called indirect systems. In such systems the source language structure is analyzed and text is transformed into a logical form. The target language translation is then generated from the logical form of the text (Hutchins and Somers 1992). The transition from direct systems to indirect systems is illustrated in Figure 2.1, taken from (Hutchins and Somers 1992, pg. 107).

SYSTRAN is one of the most well-known direct systems. It is described in Hutchins and Somers (1992) and Wilks (1992). Indirect systems can be further divided into Interlingua and Transfer based systems.

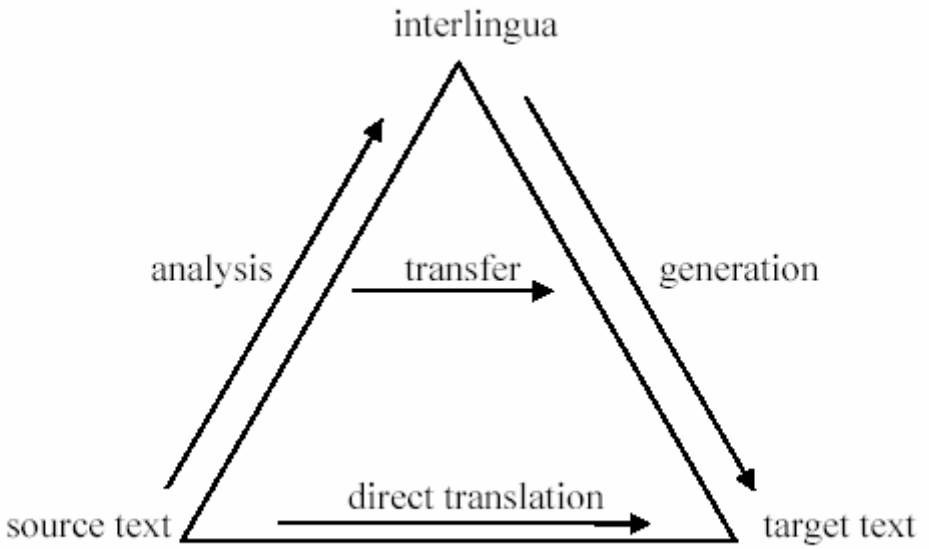


Figure 2.1: Transfer and interlingua 'pyramid' diagram

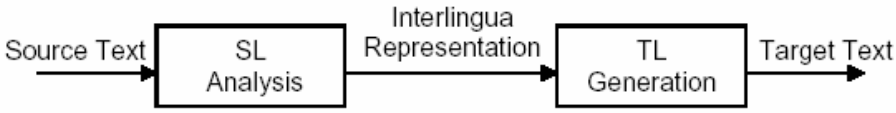


Figure 2.2a: Interlingua Based System

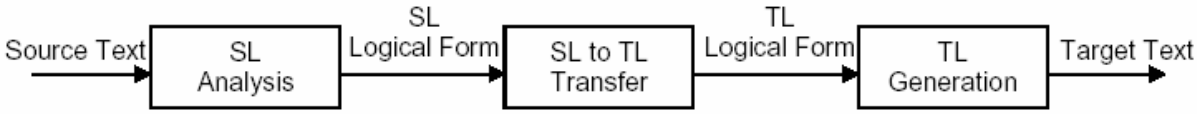


Figure 2.2b: Transfer Based System

In the transfer method, the source language is analyzed to an abstract level. Then, through a transfer module, this abstract form is converted to the corresponding abstract form in the target language through which the target translation text is generated. The module '*SL Analysis*' captures the required linguistic information about the source language sentences to aid the translation. '*SL to TL Transfer*' module transfers the representation generated by '*SL Analysis*' to a target language representation. The module '*TL Generation*' generates the translation text using this logical representation. Such a system requires independent grammars for the source and target languages. Moreover it requires a comparative grammar or transfer roles to relate source structures to target structures. Since the system

assumes full grammatical knowledge it does not allow ungrammatical sentences to be parsed, thus reducing the output of the system. This kind of system is easy to extend because to add a new language, grammar and transfer rules for the new language need to be written but the grammar of the other language is reusable. Such systems are theoretically reversible. The same grammars can be used in the reversed system. Practically there are problems in reversing the system because some transfer rules which are correct in one direction may not be correct in the other direction. The system has the explicit grammar of the target language, which ensures grammatical output (Arnold et al. 1993). Examples of transfer systems include ARIANE (Vauquois and Boitet 1985), SUSY (Maas 1987), MU (the Japanese National Project) (Nagao et al. 1986), METAL (Slocum et al. 1987; Bennett and Slocum 1988), TAUM-AVIATION (Isabelle 1987), ETAP-2 (Apresian et al. 1992), LMT (McCord 1989), EUROTRA (Arnold 1986; Arnold and des Tombe 1987; Copeland et al. 1991a,b), CAT-2 (Sharp 1988), MIMO (Arnold and Sadler 1990), MIMO-2 (van Noord et al. 1990) and ELU (Estival et al. 1990).

The Interlingua approach involves the use of an intermediate language (i.e. an Interlingua) for the transfer, with the source language text translated to the Interlingua and the Interlingua translated to the target language text. As suggested by Hutchins and Somers (1992), an Interlingua is an intermediate ‘meaning’ representation and this representation:

“*includes all information necessary for the generation of the target text without ‘looking back’ to the original text. The representation is thus a projection from the source text and at the same time acts as the basis for the generation of the target text; it is an abstract representation of the target text as well as a representation of the source text.*” (Hutchins and Somers 1992, p. 73)

Interlingua appears to be an attractive approach for machine translation due to several reasons. Firstly, from a theoretical point of view it is very interesting to establish a representation which is independent of language. Secondly, Interlingua systems are more easily extendable because only analysis and generation modules are required to add a new language and no language specific transfer information is needed. But it is difficult to

define such a language independent representation even for closely related languages (Arnold et al. 1993).

An attempt to define an Interlingua to represent the language in the form of a semantic relation is The Universal Networking Language (UNL) project. This project was initiated by the University of United Nations based in Tokyo in 1996. An utterance is represented as a hyper-graph in UNL. Normal nodes in the graph bear Universal Words (UWs) with semantic attributes and arcs bear semantic relations (deep cases, such as agt, obj, goal, etc.). UNL representation is being built in many languages including Arabic, Chinese, French, German, Hindi, Indonesian, Italian, Japanese, Mongolian, Portuguese, Russian, and Spanish. Some other Interlingua systems are Rosetta (Landsbergen 1987b,a), KBMT (Goodman 1989; Goodman and Nirenburg 1991). (Arnold et al. 1993).

There are new emerging approaches to MT known as the empirical approaches. They apply statistical or pattern matching techniques for MT. These techniques are called empirical since the knowledge for translation is derived empirically by examining text instead of linguistic rules. There are two such approaches, the ‘example’ or ‘analogy’ based approach, and the ‘statistical’ approach (Arnold et al. 1993).

In the ‘example-based’ approach, translation is done by matching the given text with stored example translations. The basic idea is to collect a bilingual corpus of translation pairs and then use a best match algorithm to find the closest example to the source phrase to be translated. This gives a translation template, which can then be filled in by a word for word translation. A limitation of this technique is that it requires a large bilingual aligned corpus. But these examples can also be built incrementally, increasing the quality of translation. Such systems are efficient because they need not to go through complex grammars to analyze the text, but if many examples match the input text then finding the best match can be a complex task. A pure example based system will include no linguistic knowledge but addition of some linguistic knowledge can improve the system by increasing its capability of dealing with more patterns concisely as one can specify categories instead of raw words (Arnold et al. 1993).

The second approach, the 'statistical approach', uses probabilistic analysis in MT as the name suggests. This term sometimes refers to the use of probability based techniques in parts of the MT task like word sense disambiguation or structural disambiguation. The other use of this term refers to a pure statistical machine translation system which uses probabilistic models to determine the correct translation of input text. In this approach, two statistical models, namely a 'language model' and a 'translation model' are built. This technique has been successfully used in speech recognition. A language model provides probabilities of occurrence of the sentence in the language, P(S) and a translation model provides probability of a target sentence given source sentence, P(T/S). An N-gram model is used to build the language model. Language models for both source and target languages are built. The translation model is computed using a word-level aligned bilingual corpus. For details of the modeling process, refer to Brown et al. (1990). Using language model probabilities and conditional probabilities of the translation model, P(S/T) is computed using the following formula:

$$P(S/T) = \frac{P(S)P(T/S)}{P(T)}$$

This approach does not require explicit encoding of linguistic information. On the other hand, it is heavily dependent on the availability of good quality bilingual data in very large proportions, which is currently not available for most languages (Arnold et al. 1993).

3.3 Why Example-based Machine Translation?

Example-based Machine Translation makes use of past translation examples to generate the translation of a given input. An EBMT system stores in its example base of translation examples between two languages, the source language and the target language. These examples are subsequently used as guidance for future translation tasks. In order to translate a new input sentence in SL, similar SL sentence is retrieved from the example base, along with its translation in TL. This example is then adapted suitably to generate a translation of the given input. It has been found that EBMT has several advantages in comparison with other MT paradigms (Sumita and Iida, 1991):

1. It can be upgraded easily by adding more examples to the example base;
2. It utilizes translators' expertise, and adds a reliability factor to the translation;
3. It can be accelerated easily by indexing and parallel computing;
4. It is robust because of best-match reasoning.

Even other researchers (e.g. (Somers, 1999), (Kit et. al., 2002)) have considered EBMT to be one major and effective approach among different MT paradigms, primarily because it exploits the linguistic knowledge stored in an aligned text in a more efficient way. We apprehend from the above observation that for development of MT systems from English to Bengali, EBMT should be one of the preferred approaches. This is because a significant volume of parallel corpus is available between English and Bengali in the form of Newsletters, Bi-lingual websites, government notices, translation books, advertisement material etc. Although all data is generally not available in electronic form yet, converting them into machine readable form is much easier than formulating explicit translation rules as required by an EBMT system.

3.4 Difficulties of Example-based Machine Translation

- Can not use in general translation
- But improvable by increasing Knowledge Base
- Match sentence rule is very difficult
- No tools available

3.5 Initial Requirement for EBMT

- Prepare Language Model
- Generate Sentence Rule
- Morphological Analysis
- English to Bengali Dictionary with Analysis

Morphological Analysis
For English normally we can have 4 forms of a word. Eg. Do, Did, Done, Does
But in Bengali we may have nearly 20 forms of the same word meaning “koro”. Below we give the figure which explains the word forms.

Table1: Sample knowledge base of the English to Bengali EBMT System

English	**English Chunk**	**Transfer to Bengali Chunk**	**Bengali**
He reads a book	[NP He/PRP] [VP reads/VBZ] [NP a/DT book/NN]	[NP সে/PRP] [VP পড়ে /VBZ] [NP একটি/DTবই/NN]	সে একটি বই পড়ে
The sun Rises in the east	[NP The/DT sun/NN Rises/NNS] [PP in/IN] [NP the/DT east/JJ]	[NP /DT সূর্য /NN উদিত /NNS] [PP হয়/IN] [NP /DT পূর্বে /JJ]	সূর্য পূর্বে উদিত হয়
He is reading a book	[NP He/PRP] [VP is/VBZ reading/VBG] [NP a/DT book/NN]	[NP সে /PRP] [VP /VBZ পড়ছে /VBG] [NP এক টি /DT বই /NN]	সে এক টি বই পড়ছে
I am reading a book	[NP I/PRP] [VP am/VBP reading/VBG] [NP a/DT book/NN]	[NP আমি /PRP] [VP /VBP পড়ছি /VBG] [NP এক টি /DT বঈ /NN]	আমি এক টি বঈ পড়ছি
They are reading a book	[NP They/PRP] [VP are/VBP reading/VBG] [NP a/DT book/NN]	[NP তারা /PRP] [VP /VBP পড়ছে /VBG] [NP এক টি /DT বই /NN]	তারা এক টি বই পড়ছে
I have done the work	[NP I/PRP] [VP have/VBP done/VBN] [NP the/DT work/NN]	[NP আমি /PRP] [VP /VBP কাজটি /VBN] [NP টি/DT কাজ/NN]	আমি কাজটি করেছি
He has gone to Dhaka	[NP He/PRP] [VP has/VBZ gone/VBN] [PP to/TO] [NP Dhaka/NNP]	[NP সে /PRP] [VP /VBZ গিয়াছে /VBN] [PP /TO] [NP ঢাকা /NNP]	সে ঢাকা গিয়াছে
They have livedat this house five years.	[NP They/PRP] [VP have/VBP] [PP livedat/IN] [NP this/DT house/NN] [NP five/CD years/NNS] ./.	[NP তারা /PRP] [VP করছে /VBP] [PP বাস /IN] [NP এই /DT বাড়িতে /NN] [NP পাচ /CD বছর /NNS] ./.	তারা এই বাড়িতে পাচ বছর যাবত বাস করছে
He has been reading the book for two hours	[NP He/PRP] [VP has/VBZ been/VBN reading/VBG] [NP the/DT book/NN] [PP for/IN] [NP two/CD hours/NNS] ./.	[NP সে /PRP] [VP /VBZ যাবত /VBN পড়ছে /VBG] [NP টি /DT বই /NN] [PP for/IN] [NP দুই /CD ঘন্টা /NNS] ./.	সে দুই ঘন্টা যাবত বই টি পড়ছে
I did the work	[NP I/PRP] [VP did/VBD] [NP the/DT work/NN] ./.	[NP আমি /PRP] [VP করছিলাম /VBD] [NP টি/DT কাজ/NN] ./.	আমি কাজটি করছিলাম
He went home yesterday	[NP He/PRP] [VP went/VBD] [NP home/NN] [NP yesterday/NN] ./.	[NP সে /PRP] [VP গেল /VBD] [NP বাড়ি /NN] [NP গতকাল /NN] ./.	সে গতকাল বাড়ি গেল
He wrote the letter	[NP He/PRP] [VP wrote/VBD] [NP the/DT letter/NN] ./.	[NP সে /PRP] [VP লিখলো /VBD] [NP এক /DT চিঠি /NN] ./.	সে এক টি চিঠি লিখলো
The boys were playing	[NP The/DT boys/NNS] [VP were/VBD playing/VBG] ./.	[NP /DT বালকগুলো /NNS] [VP /VBD খেলতে ছিল /VBG] ./.	বালকগুলো খেলতে ছিল
You will do the sum	You/PRP] [VP will/MD do/VB] [NP the/DT sum/NN]	[NP tumi/PRP] [VP /MD korobe/VB] [NP Ti/DT Angk/NN]	তুমি অংক টি করবে
He will be doing the work	[NP He/PRP] [VP will/MD be/VB doing/VBG] [NP the/DT work/NN]	[NP se/PRP] [VP thak/MD be/VB korIte/VBG] [NP Ti/DT kaj/NN]	সে কাজ টি করতে থাকবে
You will be reding The book	[NP You/PRP] [VP will/MD be/VB reding/VBG] [NP The/DT book/NN]	[NP tumi/PRP] [VP thak/MD be/VB pRite/VBG] [NP Ti/DT bo`i/NN]	তুমি বই টি পড়িতে থাকবে

3.6 *Ontology*

Ontology needed for word sense disambiguation. It makes explicit the semantic relations.

It also tries to link correctly the exact place of a particular sense in the structure of a language. Conceptual categories of nouns, verbs, adjectives and adverbs are placed in a directed acyclic graph structure.

3.7 *WordNet defines an ontology*

From the following example we can see the Hyponymy structure of the word quake:

earthquake, quake, temblor, seism --

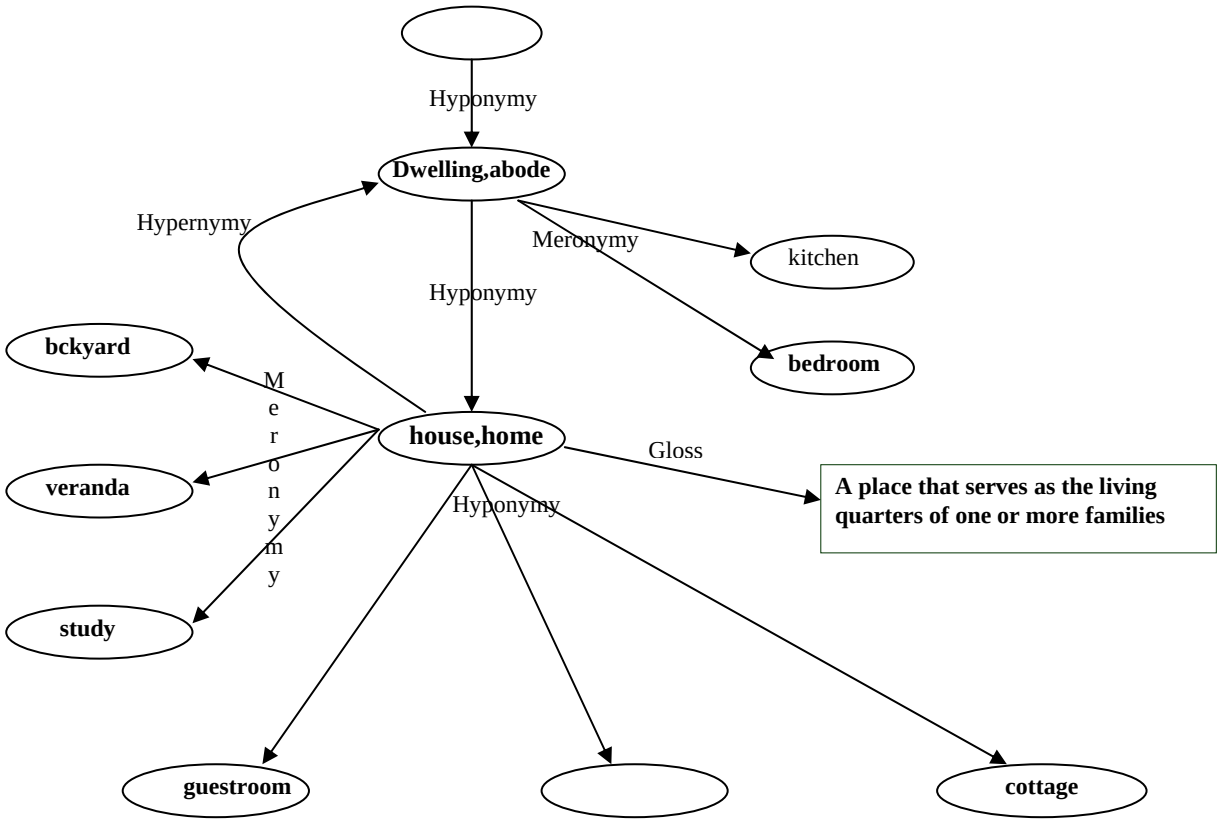
=> geological phenomenon --

=> natural phenomenon --

=> phenomenon --

Ontology is shallow for non-noun POS as we do not need any deep structure for non-noun POS. Property inheritance possible. Its important for sense disambiguation

Fig: WordNet Sub-Graph



3.8 Goals of the Independent Study

The primary goal of this Independent Study is to study various aspects of EBMT system for translation from English to Bengali. Detect Ambiguity, and translate from the rules with the help of WordNet. I will try to find suitable solutions for the following aspects:

a) Detect Ambiguity, and translate from the rules with the help of WordNet Information in tabular format. As wordnet does not have all sufficient information yet we can prepare bangla wordnet according to our need. So my interest lies to find the needs as well.

b) Study of Translation Patterns for English to Bengali translation, and how translation patterns can be effectively handled within an EBMT framework.

c) Prepared 500 Word Rules for Numbers (Bochon). Transliterating Nouns to improve the quality. Sample comparisons (Excel file given in CD)

উন্নত ও প্রাণীবাচক শব্দে

শিক্ষক	0	1	1	1	0
বালিকা	1	0	0	0	0
গরিব	1	0	0	0	0
ধনী	1	0	0	0	0
দেব	1	1	0	1	1
নর	1	1	0	0	0
জন	0	1	0	0	0
সুধী	1	1	1	1	1
ভক্ত	1	1	1	1	1
শিক্ষক	1	1	1	1	1
সম্পাদক	1	1	1	1	1
পণ্ডিত	1	1	0	0	1
মন্ত্রী	1	1	0	0	1

	কুল	নিচয়	সকল	সব	সমূহ	গুল্যে	গুলি
কবি	1	0	0	0	0	0	0
পক্ষি	1	0	0	1	1	1	0
মাতৃ	1	0	0	1	0	0	0
পরুত	0	1	1	0	1	1	1
কুসুম	0	1	0	0	0	0	0
মনুষ	1	0	1	1	1	1	1
ভাই	0	0	1	1	0	0	0
নথি	0	0	1	1	1	1	1
বৃক্ষ	1	0	1	1	1	1	1
মানুষ	1	0	1	1	1	1	1
পাখি	1	0	1	1	1	1	1
অত	0	0	0	1	0	1	1
আম	0	0	0	1	1	1	1
বই	0	0	0	1	1	1	1
খাতা	0	0	0	1	1	1	1

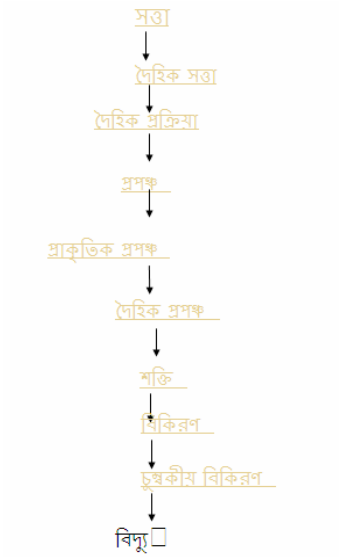
প্রাণীবাচক ও অপ্রাণীবাচক শব্দে

	আবালি	কুল	গণ	গ্রাম	চয়	জন	দাম	নিকর	নিচয়	মণ্ডল	মণ্ডলী	মালা	রাজি	লোক	বর্গ	বৃন্দ	সকল	সভা	সব	সমুচ্চয়	সমুহ	মহল
চরিত্ত	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
পদ	1																					
মনুষ	1	1																				
দেবতা	1	1																				
বিঘ্ন	1					1																
পণ্ডিত	1					1																
লতা	0						1															
বিদু	0						1															
মেঘ	1							1		1	1											
বৃক্ষ	1												1		1	1						
লেতৃ	1															1						
সুধী	0														1							
যুবতী	0																					
ভাই	0																					
বন্দু	0																					
রাজনৈতি	0																					

	অপ্রাণীবাচক	প্রাণীবাচক	সাধারণ	উদাহরণ	কুল	0	1	0	
আবালি	1	0	0	চরিতাবলি, পদাবলি		0	1	0	মনুষগণ, দেবতাগণ

গাম	1	1	0		লোক	0	1	0	
চয়	1	0	0		বর্গ	0	1	0	নেতৃবর্গ
জন	0	1	0	বিদ্বজন, পণ্ডিতজন	বৃন্দ	0	1	0	সুধীবৃন্দ
দাম	1	0	0	লতাদাম, বিদুদ্দাম	সকল	0	0	1	
নিকর	1	0	0		সভা	0	1	0	পণ্ডিতসভা যুবতীসভা
নিচয়	1	0	0		সব	0	0	1	ভাইসব
মণ্ডল	1	0	0	মেঘমণ্ডল	সমূচয়	0	0	1	
মণ্ডলী	0	1	0	পণ্ডিতমণ্ডলী	সমুই	0	0	1	
মালা	1	0	0	মেঘমালা	মহল	1	0	0	রাজনৈতিকমহল, বন্দুমহ
রাজি	1	0	0	বৃক্ষরাজি					

d) Wordnet Structure: Figure is showing a Hyponymy tree structure of Bangla WordNet which is direct translation of English WordNet. Full partial translation is given in the CD.



| সত্তা |

| দৈহিক সত্তা | সত্তা |

| দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| প্রাকৃতিক প্রক্রিয়া | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| শক্তি | দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| বিকিরণ | শক্তি | দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| বিদ্যু -চুম্বকীয় বিকিরণ | বিকিরণ | শক্তি | দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| আলোক-রাসায়নিক বিকিরণ | বিদ্যু -চুম্বকীয় বিকিরণ | বিকিরণ | শক্তি | দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

| আলো | আলোক-রাসায়নিক বিকিরণ | বিদ্যু -চুম্বকীয় বিকিরণ | বিকিরণ | শক্তি | দৈহিক প্রপঞ্চ | প্রাকৃতিক প্রপঞ্চ | প্রপঞ্চ | দৈহিক প্রক্রিয়া | দৈহিক সত্তা | সত্তা |

Chapter 4: Proposed Architecture of Improving Example Based English to Bengali Machine Translation using WordNet

The proposed EBMT system has five steps

1. Tagging the English sentence
2. Parsing the English sentence
3. Using sub-sentential EBMT prepare the chunks of the sentence
4. Using an efficient adapting scheme match the sentence rule.
5. Translate from Source Language (English) to Target Language (Bengali) in the chunk and generate with morphological analysis with the help of WordNet.

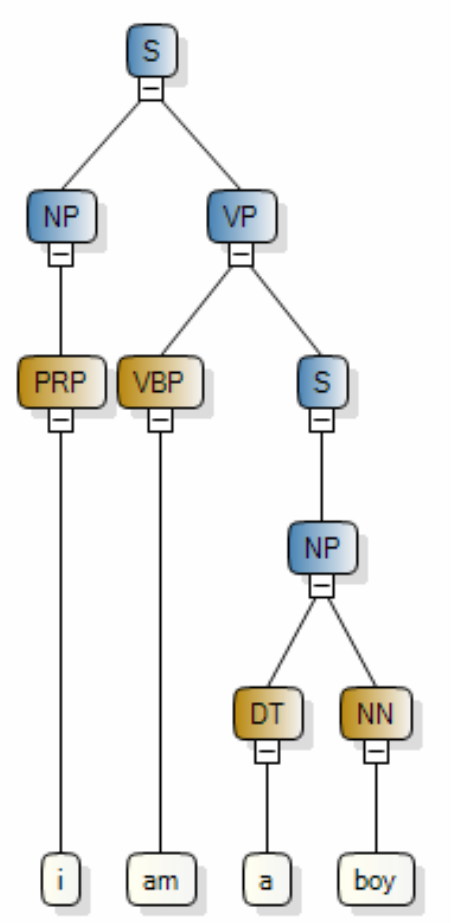
4.1 Tagging and Parsing

Tagging, is the process of marking up the words in a text as corresponding to a particular part of speech, based on both its definition, as well as its context—i.e., relationship with adjacent and related words in a phrase, sentence, or paragraph. Eg. I do-> I/PRP do/VBP

Parsing, is the process of analyzing a sequence of tokens to determine grammatical structure with respect to a given formal grammar. We used the tag set of Table2 for tagging the English sentence. Eg. Eg. I am a boy-> (S (NP (PRP I)) (VP (VBP am) (NP (DT a) (NN boy))))

Table2: Tag set used for English to Bengali EBMT

Level 1	Level 2	Tag
Noun	Common	NN
	Proper	NNP
	Compound Common Noun	NNC
	Compound Proper Noun	NNPC
	Verb Root	NNV
	Temporal	NNT
	Question Temporal	QNT
	Locative	NNL
	Question Locative	QNL
Pronoun	Personal Pronoun	PRP
	Question Pronoun	QPR
Adjective	Simple	JJ
	Verb Root	JJV
	Question Adjective	QJJ
Vocatives	Vocatives	VOC
Verb	Main Finite Verb	VB
	Nonfinite Nominal	VBM
	Nonfinite Conditional	VBC
	Nonfinite Perfective	VBT
	Nonfinite	VBF
	Past tense	VBD
	Gerund/present participle	VBG



	Past participle	VBN		Study	
	Non-3rd ps. sing. Present	VBP		Right ParenIndependent Study	)
	3rd ps. sing. Present	VBZ			
	Existential	VBE		Opening Left Quote	LQ
Adverb	Adverb	RB		Closing Right Quote	RQ
	Question Adverb	QRB		Preposition/subordinate conjunction	IN
Conjunction	Co-ordinating	CC			
	Compound Co-ordinating	CCC		Adjective, superlative	JJS
				Adjective, comparative	JJR
	Suspicion	CN		Modal	MD
	Eternal Joining	CET		Proper noun, plural	NNPS
	Subordinating	CS		Noun, plural	NNS
	Compound Subordinating	CSC		Predeterminer	PDT
				Possessive ending	POS
Numbers	Cardinal Numbers	CD		Possessive pronoun	PRP$
Adposition	Adposition	ON		Adverb, comparative	RBR
Interjection	Interjection	UH		Adverb, superlative	RBS
Particle	Particle	RP		to	TO
	Question Particle	QRP		wh-determiner	WDT
Determiner	Common	DT		wh-pronoun	WP
	Singular	DTS		Possessive wh-pronoun	WP$
	Question Determiner	QDT			
Quantifier	Quantifier	QF		wh-adverb	WRB
Foreign Word	Foreign Word	FW		Left open double quote	``
				Comma	,
Symbol	Symbol	SYM		Right close double quote	'
List Item Marker	List Item Marker	LS		Sentence-final punctuation	.
Suffixes	Adpositional	SFON			
	Accusative	SFAC		Colon, semi-colon	:
	Possessive	SF$		Dollar sign	$
Punctuation Marks & Others	Sentence Final Punctuation	.		Pound sign	#
				Left parenIndependent Study (	-LRB-
	Comma	,		Right parenIndependent Study	-RRB-
	Colon, Semi-colon	:			
	Dash, Double-Dash	-			
	Left ParenIndependent	(			

4.2 Handle Complex Sentence Using Sub-Sentential EBMT:

Handling complex sentence in general considered to be difficult to deal with in an MT system. Since exact sentence matches only occur in special domains, we want to extend this to sub-sentence matches. For this we need to:

- Find the most similar example (involves segmenting by preparing chunks)
- Alter source side to match current input.

Similarity requires a “distance metric” in the source language (English).
This can be closeness:

- of the lexical items in a hierarchy of terms/ concepts from ontology
- of the sequence of syntactic categories and function words,
- of the two syntactic structures,
- or combinations of these.

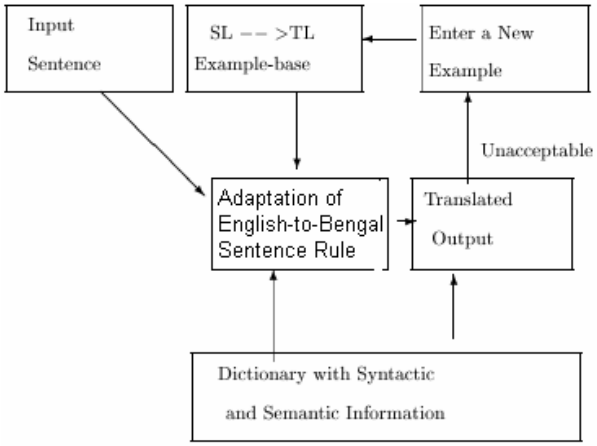


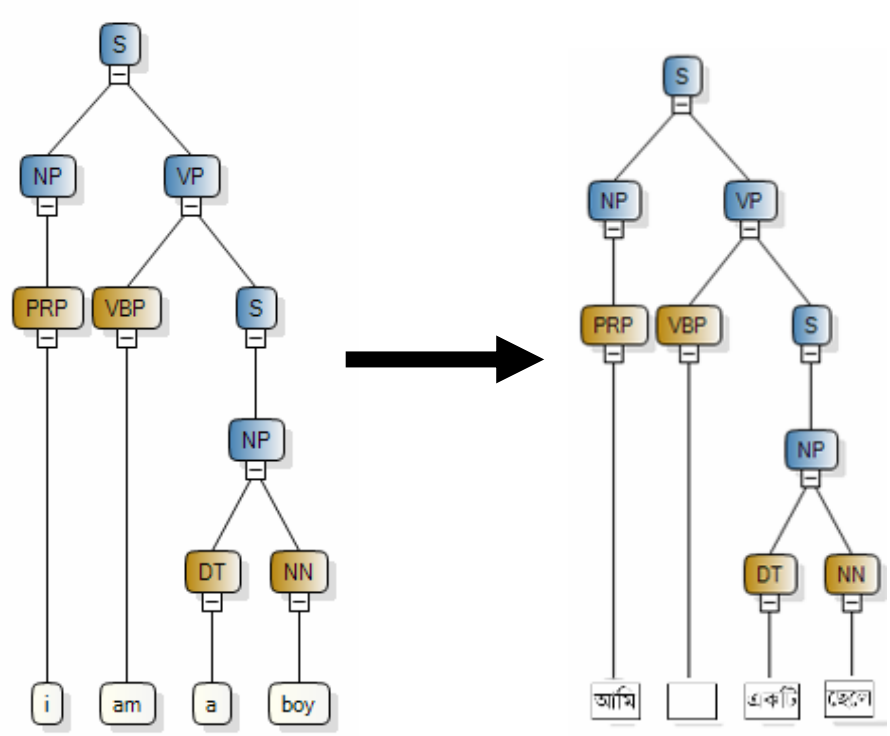
Figure 1: Role of Examples in Translation

4.3 Adapting Scheme to Match Sentence Rule

Efficient adaptation of past examples is a major aspect of an EBMT system. There are many adaptation schemes available for an EBMT system. Even an efficient similarity measurement scheme and a quite large example base cannot guarantee an exact match for a given input sentence. As a consequence, there is a need for an efficient and systematic adaptation scheme for modifying a retrieved example, and thereby generating the required translation. In section 5 we discuss details about our proposed adaptation scheme. In Table1 we gave a sample knowledge base of the English to Bengali EBMT System. During translation our adapting scheme chooses the best rule for the source sentence.

4.4 Match the sentence rule from the Knowledge Base

Figure: Tree Conversion for English to Bengali



4.5 Translate from Source Language (English) to Target Language (Bengali) in the chunk and generate with morphological analysis with the help of WordNet.

Study of divergence for English to Bengali translation is also required. Translation divergence can be effectively handled within an EBMT framework. As in earlier step we have the sample rule and the parsed sentence. Now we can easily translate the sentence by matching the rule.

I am a dog -> আমি একটি কুকুর
I am a man -> ~~আমি একটি মানুষ~~ আমি একজন মানুষ
i have a cat -> আমার একটি বিড়াল আছে
I have to love -> ~~আমার ভালোবাসতে আছে~~ আমাকে ভালোবাসতে হবে

Figure2: Improved Example of English to Bengali Translation

In first two example of figure2 “a” has different meaning in Bengali “jon” and “ti”. Here we can see that it has same sentence rule but different translation. Depending on the quality of the word “dog” / ”man” we are choosing the actual meaning. Using WordNet we are determining that word sense. This technique will dramatically improve the quality of EBMT.

Translation Patterns

- **Translation Pattern P1:** Here, genitive case ending *(tir, gulor)* is used to convey the sense of the "have" verb. Which of the genitive case endings (i.e. *tir, gulor)* will be used in a given case depends upon the number and gender of the object and *“gulor"* for masculine plural. For example:

The school has good name -> bidyaalay (school) tir(of) valo (good) naam (name) acche (has)

- **Translation Pattern P2:** In this pattern the object and its pre-modifying adjective in the English sentence are realized as the subject and subjective complement (SC). respectively, in the Bangla translation. The subject of English sentence is realized as possessive case of the subject of the Bangla translation. For example.

Gita has beautiful hair2 ~ *Gita* (Gita) -r (of) *chul*(hair) *sundar* (beautiful) -(are).

- **Translation Pattern P3:** Here a locative case ending "ke *paas"* is used instead of genitive postposition. For illustration, consider the following.

Mohan has a book ~ *Mohan* (Mohan) *er kache* (near to) *ekti* (a) *boi*(book) achhe (has)

- **Translation Pattern P4:** In this pattern a postposition "ko" is used in the Bangla translation of the given sentence. For example,

My uncle has asthma ~ *amar* (my) *chacha* (uncle) -r(to) *asthamaa* (asthma) *achhe*(has).

- **Translation Pattern P5:** Here the postposition "acche" is used for conveying the sense of the verb "have". For example.

This city has a museum ~ ei (This) shahar (city) -e(in) ek (a) jadughar(museum) acche(is).

- **Translation Pattern P6:** This translation pattern is similar to the pattern P5. For example consider the following:

The tiger has stripes ~ baagher(tiger) gaye (on) dorakata(stripes) acche(are).

- **Translation Pattern P7:** Here, upon translation in Bengali, the object of the English sentence is realized as an SC which is an adjective. The following translations illustrate this pattern. It adds “an”

She has grace ~ tar (She) aakarshan (graceful) achhe(has).

- **Translation Pattern P8:** This pattern occurs if the main verb of the Bengali translation is obtained from the object of the English sentence.For illustration, consider the following example:

Gita has regards for old men ~ Gitar (Gita) buzurgonder (old men) proti(of) sroddha(respect) acche(does).

- **Translation Pattern P9:** This pattern is similar to the translation pattern P8, For example,

I had tea ~ ami(I) cha (tea) pan korechhi(drank).

- **Translation Pattern P10**: In all the above cases the structure of the English sentences considered has been <SVO>. But. if the sentence has an additional component in the form of adjunct, then a variation in the translation may be noticed. For illustration, consider the two sentences: (a) Ram has two rupees (b) Ram has two rupees in his pocket. While the translation of the first one is "Ram er kachhe dui taka acche", the translation of the second one is "Ram er pocket e dui taka achhe".
- **Translation Pattern P11**: This pattern is observed if, along with the subject, verb and object, the sentence has an infinitive verb phrase. For example,

My children had me buy the car ~ amar (my) bachchra(children) amake(me) gaari (car) kine diyechhe(buy).

We can find nearly 13 use of the sentence pattern have/has. For illustration we can see:

Rule	English	Bengali
1.	I had tea/ I had rice	Ami cha pan korechi/ ami vat kheyechhi
2.	If I had known	Ami jodi jantam
3.	I haven’t many friends here	Ekhane amar khub beshi bondhu nei
4.	Do you often have colds?	Tomar ki prayi thanda lage?
5.	I must have my shirt ironed	Amar shirt obosshoi stri kora thakte hobe
6.	The school has good name	Schooltir valo nam kora
7.	Mohan has a book	Mohaner ekti boi **achhe**
8.	Gita has beautiful hair	Gita’r sundor chul **achhe**
9.	This city has a museum	Sohortite jadoghor acche
10.	The tiger has stripes	Bagher dorakata dag acche
11.	Gita has regards for old men	Gita boyosko lokder somman kore
12.	Ram has two takas	Ramer kacche dui taka acche
13.	My father had me buy the car	Amar baba amake ekti gari kine diyechhe

Another illustrated example is the use of word “run”:

Table 1: Different translations of “run”

Sentences	**Bengali**	**Translation of Verb**
He is running for election	*Se ebar nirbachone daracche*	*Darano*
They run an N.G.O.	*Tara ekti NGO chalay*	*Chalay*
The army runs from one end to another.	*Armi ek pranto theke arek prante zay*	*Zay*
The river ran into the sea.	*Nodi somudre giye mishe*	*Mishe*
He runs for his company.	*Se tar kompanir jonno kaj kore*	*Kaj kore*

He runs in sun.	*Se surjer majhe douray*	*Douray*
We ran the ad three times.	*Amra biggaponti tin bar prokash korechi*	*Prakaash*

Sense Number	Definition (As given by WordNet 2.0)	Translation Pattern	Example sentence
1	have or possess, either in a concrete or abstract sense	P1	Rita has *two* daughters.
		P3	She has a degree from IIT.
2	have as a feature	P1	This dog has three legs.
		P2	She has beautiful eyes
		P5	This car has an airbag.
		P6	The tree has flowers.
		Mixed P1 and P8	Ravi has a good grasp of subject.
3	of mental or physical states or experiences	P1	Ram has many dreams.
		P2	Mita has an idea.
		P3	Ram has sympathy for *the poor*.
		P8	She has regards for her father.
		P9	She had a difficult time.
4	have ownership or possession of	P1	Hemu has three houses.
		P3	Mohan has a car.
6	serve oneself to, or consume regularly	P9 ("*khaanaa*" or "*peenaa*")	I had an apple.
7	have a personal or business relationship with someone	P1	He has an assistant.
		P3	This professor has a research scholar.
8	organize or be responsible for	P1	John has a meeting.
9	have left	P3	Meera has two years left.
12	suffer from; be ill with	P4	Paul has fever.
14	receive willingly something given or offered	P9 ("*lenaa*" or "*sweekaar karnaa*")	Please have *this* gift.
15	get something; come into possession of	P9 ("*milnaa*" or "*prapt honaa*")	I have a letter from a friend.
16	undergo (as of injuries and illnesses)	Mixed P1 and P8	Rama had a fracture
		Mixed P4 and P8	His father had a heart attack.
17	achieve a point or goal	P9 ("*banaanaa*")	Sachin had a century.
18	give birth (to a newborn)	P9 ("*janam denaa*")	My wife had a baby boy yesterday.

Table 4: Rules for translation patterns for different senses of "

Bengali

রিতার দুই বোন আছে
তার আই.আই.টি থেকে ডিগ্রি আছে
কুকুরটির তিনটি পা আছে
তার সুন্দর চোখ আছে
গাড়িটির এয়ারব্যাগ আছে
গাছটিতে ফুল আছে
রবির এ বিষয়ে ভাল দক্ষতা আছে
রামের অনেক স্বপ্ন আছে
মিতার একটি ধারণা আছে
রামের গরিবের প্রতি সমবেদনা আছে
তার বাবার প্রতি তার সম্মান আছে
তার তখন কঠিন সময় ছিল
হিমুর তিনটি বাড়ি আছে
মোহনের গাড়ি আছে
আমার একটি আপেল আছে
তার একজন সহকারী আছে
প্রফেসরের একজন অনুদানপ্রাপ্ত গবেষক
জনের একটি মিটিং আছে
মিরার দুই বছর বাকি আছে
পালের অসুখ হয়েছে
অনুগ্রহ করে উপহারটি রাখুন
আমার বন্ধুর কাছ থেকে একটি চিঠি এসে
রামার হার ভেঙেছে
তার বাবার হার্ট এটাক হয়েছে

Chapter 5. Adaptation in English to Bengali Translation

A successful EBMT system requires a good adaptation scheme. The need for an efficient and systematic adaptation scheme arises for modifying a retrieved example, and thereby generating the required translation. Researcher came with various approaches to deal with adaptation aspect of an EBMT system. Overall the adaptation procedures employed in different EBMT systems primarily consist of four operations:

- **Copy**, where the same chunk of the retrieved translation example is used in the generated translation;
- **Add**, where a new chunk is added in the retrieved translation example;
- **Delete**, when some chunk of the retrieved example is deleted; and
- **Replace**, where some chunk of the retrieved example is replaced with a new one to meet the requirements of the current input.

5.1 Adaptability/Mappability for a chunk has 4 discrete values:
Depending on the Level 3: Source Language (SL): Target Language (TL) mapping is one-to-one for all words
Level 2: Syntactic Functions map, but not some POS tags
Level 1: Functions differ, but lexical correspondence still holds
Level 0: Cannot establish correspondence

5.2 Description of the Adaptation Operations
1. Constituent Word Replacement (WR): One may get the translation of the input sentence by replacing some words in the retrieved translation example. Suppose the input sentence is: ¥The bird was eating apples.", and the most similar example retrieved by the system (along with its Bengali translation) is: ¥The elephant was eating fruits."¥haathii phol khacchilo". The desired translation may be generated by replacing ¥haathii" with the Bengali of ¥birds", i.e. ¥pakhi" and replacing ¥phal" with the Bengali of ¥apples", i.e. ¥aapel". These are examples of the operation of constituent word replacement.

2. Constituent Word Deletion (WD): In some cases one may have to delete some words from the translation example to generate the required translation. For example, suppose the input sentence is: ¥Animals were dying of thirst". If the retrieved translation example is : ¥Birds and Animals were dying of thirst." ¥pakhi ebong pashu trishnay mara jacche", then the desired translation can be obtained by deleting ¥pakhi ebong" (i.e the Bengali of ¥birds and") from the retrieved translation. Thus the adaptation here requires two constituent word deletions.

3. Constituent Word Addition (WA): This operation is the opposite of constituent word deletion. Here addition of some additional words in the retrieved translation example is required for generating the translation. For illustration, one may consider the example given above with the roles of input and retrieved sentences being reversed.

4. Morpho-word Replacement (MR): In this case one morpho-word is replaced by another morpho-word in the retrieved translation example. For illustration, if the input sentence is ¥He eats rice.", and the retrieved example is: ¥He is reading a book." _ ¥se akte boi porChe", then to obtain the desired translation[4] first the morpho-word ¥Che" is to be replaced by “¥e”

5. Morpho-word Deletion (MD): Here some morpho-word(s) are deleted from the retrieved translation example.

6. Morpho-word Addition (MA): This is the opposite case of morpho-word deletion. Here some morpho-words need to be added in the retrieved example in order to generate the required translation.

7. Suffix Replacement (SR): Here the suffix attached to some constituent word of the retrieved sentence is replaced with a different suffix to meet the current translation requirements. This may happen with respect to noun, adjective verb, or case ending.

8. Suffix Deletion (SD): By this operation the suffix attached to some constituent word may be removed, and thereby the root word may be obtained.

9. Suffix Addition (SA): Here a suffix is added to some constituent word in the retrieved example.

10. Copy (CP): When some word (with or without suffix) of the retrieved example is retained in to in the required translation then it is called a copy operation.

5.3 Study of Adaptation Procedure for Morphological Variation of Active Verbs

Verb morphology variations are divided into four groups. These are:

1. Same tense same verb form
2. Different tenses same verb form
3. Same tense different verb forms
4. Different tenses different verb forms

Chapter6. Implementation of the proposed EBMT system

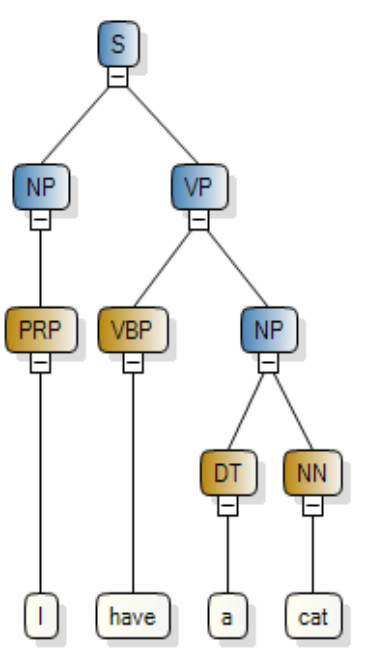
Currently our system is working for English to Bengali Translation. Our current system can translate simple sentences which are given in the knowledge base. We also defined a way to translate a complex sentence using sub-sentential EBMT. So while separating the chunk in the machine translation process we can use our knowledge base for translating the separated small chunks of the sentence. As this system can add more rules in the knowledge base, eventually it can be used for general purpose English to Bengali machine translation.

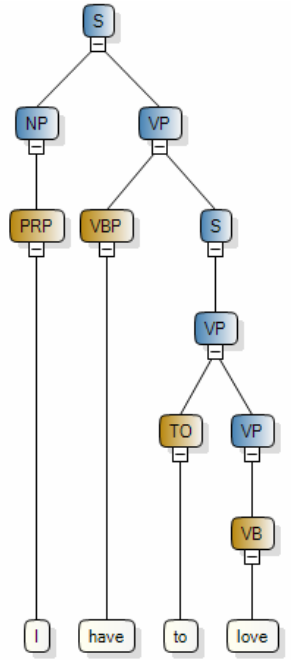
Sample Outputs

Our system can now do following types of English to Bengali translation: From this example we can say that EBMT is better as we can improve our system by adding more rules.

	Previous	Current
I am a man ->	আমি একটি মানুষ	আমি একজন মানুষ
i have a cat ->	আমার একটি বিড়াল আছে	
I have to love ->	আমার ভালোবাসতে আছে	আমাকে ভালোবাসতে হবে

Now we can see that,
I have a cat = PRP + VBP + NP
I have to love = PRP + VBP + S
If we add a new rule for “PRP + VBP + S” in our knowledge base then we can also translate “I have to love” properly. In this way we can update the system.





The screen shots of our implementation is given below:

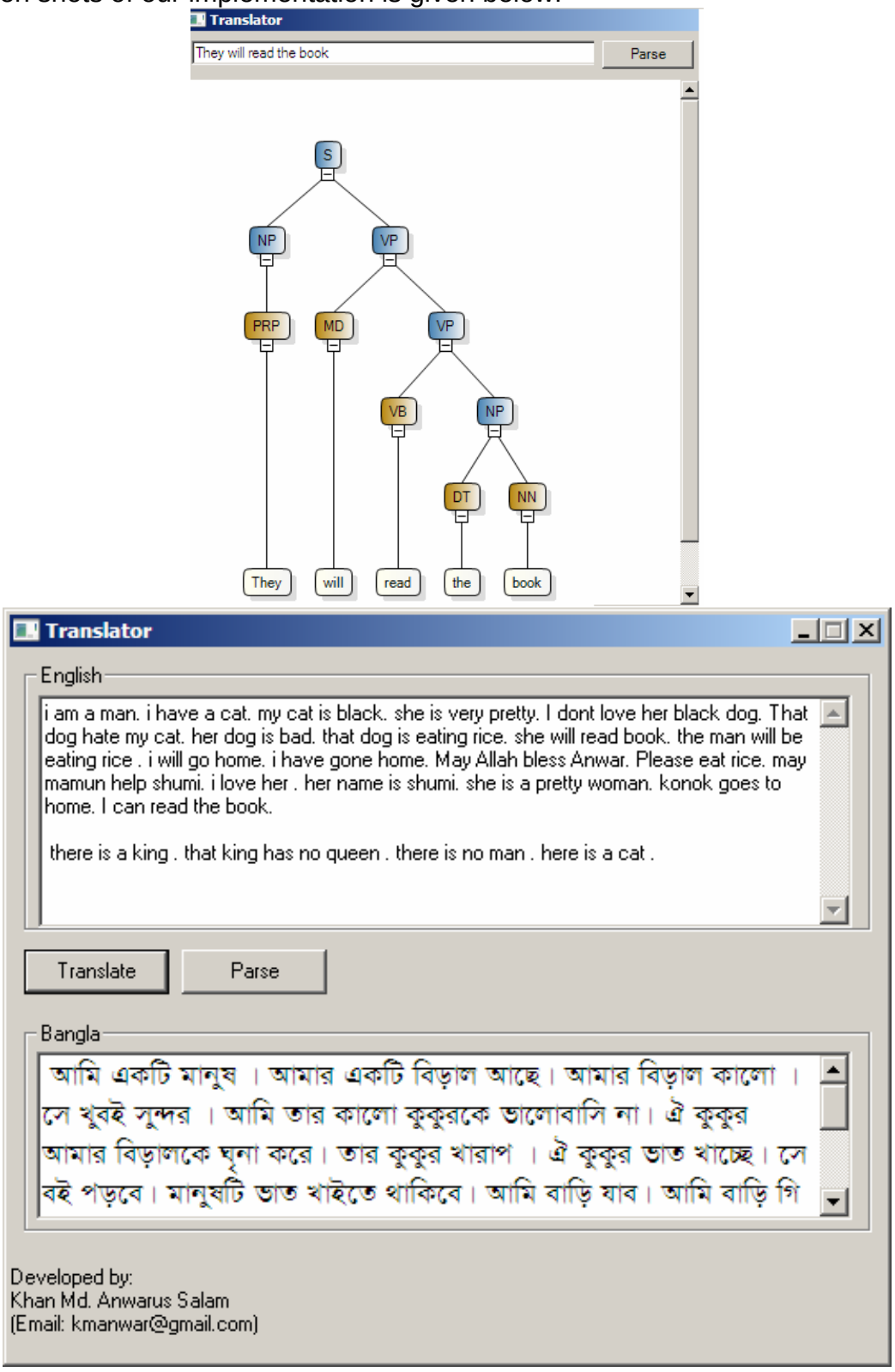


Figure2: Screen shot of the implementation

Chapter 7. Limitations of our System, Future Work and *Challenges*

Limitations:

- It can not differentiate different meaning of same word
    - Eg. Let the light light light (3 meaning)
- It can not identify the names unless it is given in the dictionary.
- It can not translate if the given sentence do not match in the knowledge base
- It can not determine different meaning of 'a' for pen/elephant. Different translation of animate/non animate objects etc (Need Bangla Wordnet)
- Can not do voice, narration or thematic translation

Random Translation of an article from The Daily Star (related to Bangla Software technology):

From the translation we can see it is 30% correct, but not understandable. Real world text is very complex as we can not define all possible correct sentence structure. EBMT may work well for a limited domain but as a tool its not reliable, however it is useful for educational purpose.

English Text:

"**WITH the development of software and fonts in Bangla to make the process of communication faster and easier among the local community, it is expected that people will become user-friendly using local language.**

Our neighbour India has already taken measures to make their language popular and beneficial to their nation. In a poor country that lacks educational skills, implementation of local language in new technologies could be a weapon to eliminate poverty and other deficiencies.

A recent research has shown some well-organised Bangla softwares helpful in daily use of computers. Akkhor Bangla, Avro Keyboard, Bangla Express, Interactive Dictionary and Online Bangla Obhidhan are examples of such products that are worth taking a look.

Akkhor Bangla: Developed by an eighteen-year-old tech genius Khan Md Anwarus Salam, Akkhor Bangla won the first prize in BCS Software Competition in 2003. Promoted and sponsored by Bangladesh Online Limited, at present the software is available in markets and also at www.akkhorbangla.com. Akkhor Bangla allows users to type words in Bangla with English letters without defining any Bangla font at the user interface. Predefined Bangla fonts are installed with the installation of the software. The software allows users to send emails in Bangla that the receipient receives as images. The software supports Unicode and includes a word processor. Unicode provides a unique number for every character regardless of the platform, programme or language. It also contains a Bangla keyboard manager, type tutor, calendar, media player and a converter. A dictionary is expected to be included in the software and is still under development. A converter in the software edits documents written in other Bangla fonts and software.

Talking to The Daily Star, Khan said he put in three years behind the development of the software and was finally released to the market on April 14, 2004.

Translated Bengali Text using my system:

the উন্নয়ন software এবং fonts ‡Z Bangla ‡Z সৃড়ি করা the পদ্ধতিয়া পদ্ধতি যোগাযোগ faster এবং easier অনেকের the ¯’vqীয় কোন ¯’vqে সীমাবদ্ধ community, ইহা expected ঐ দেশবাসী হয়ে ওঠা user-friendly using ¯’vqীয় কোন ¯’vqে সীমাবদ্ধ।

Our পদ্ধতিবেশী ভরতবক্ষ আছে ইতমধ্যে taken measures ‡Z সৃড়ি করা তা‡`I জনপ্রিয় এবং লাভজনক তে তাদেও একই রাড়েধও পদ্ধজাপুঅ। তে এক দরিদধ দেশ ঐ lacks educational skills, implementation পকপল ¯’vqীয় কোন ¯’vqে সীমাবদ্ধ তে নতুন technologies পারিত তথা ¯’vq একটি বমপ ইত্যাদি তে পরিহার করা অভাব এবং অন্য deficiencies।

A সভাপদ্ধতিক গ‡eIYv আ‡Q shown কিছু সংখ্যক well-organised Bangla softwares সহায্যকারী ‡Z ‰`wbK পদ্ধ‡qvM Kiv পকপল computers। Akkhor Bangla, Avro Keyboard, Bangla Express, পর¯’vের উপর গ্রিাশীল অভিযান এবং Online Bangla Obhidhan examples পকপল অনুরূc products ঐ , i‡Z taking একটি দেখে।

Akkhor Bangla: বিসকাশ পণ্ডাং কাছে কোন একটি eighteen-year-old tech পদ্ধতিভ Khan Md Anwarus Salam, Akkhor Bangla won the পথম পুর

গ্রার পারিতোষিক তে BCS Software পঙ্ক্তিযোগিতা তে ২০০৩। Promoted এবং sponsored কাiQ Bangladesh Online Limited, তে উপ ¯—স্থাপিত করা the software পাওয়া যায় iZ markets এবং মধ্যেও iZ www.akkhorbangla.com। Akkhor Bangla allows users iZ (মানুষ ব ¯— words iZ Bangla সাi? ইথমক্ষ letters (সীমার) বাইরে defining কোন কিছু Bangla খাঁড়ধমেপ দীক্ষিত করিবার জন্য যে পাওে জাল থাকে তে the ব্যবহারকারী ব্যক্তি interface। Predefined Bangla fonts installed সাi? the installation পকপল the software। The software allows users iZ পাঠানো emails তে Bangla ঐ the receipient receives ihinZz images। The software supports Unicode এবং includes এক শব্দ processor। Unicode provides এক অনুপম সংখ্যা কারণ পঙ্ক্তেক বৈশিড়্য regardless পকপল the platform, কমপস্টি বা ভাষা। ইহা সেইরুপ contains একটি Bangla keyboard manager, ধাতুনির্মিত ছাপার হরফ tutor, calendar, মাধ্যম সমূহ জুয়াড়ী এবং একটি converter। এক অভিযান expected iZ তাইinvK included iZ the software এবং ি—ন্ন তলে উনয়ন। এক converter iZ the software edits documents written iZ অপর Bangla fonts এবং software।

Talking তে The দৈনিক Star, Khan said সে ¯—স্থাপন করা তে তিনটি years অতিম করে the উনয়ন পকপল the software এবং পরিণাg released iZ the market উঠাiZ এপিল 14, 2004।"

Future Work:

- Lots of research opportunity in EBMT; In this changing world we need to find a way to update new sentence rules. Machine Learning using Statistical MT can be one way.
- Finding applications of WordNet in different areas of NLP.
- Make a machine learning system so that user can train it (HAMT, Suggest a translation). We can then improve its efficiency for general purpose use
- Connect with wordnet. (But wordnet does not have all sufficient information yet. But we can prepare the bangla wordnet)
- Make a machine learning system so that user can train it (HAMT, Suggest a translation). We can then improve its efficiency for general purpose use

Challenges:

The quality and effectiveness of EBMT depends largely on how strong the WordNet is. Same word can have drastically different meaning in two languages. Some words are polysemous (eg bank has different meanings in Bangla) in one language, but not in other. Words which have subtly different meaning in two languages can be misunderstood to have same meaning.

Chapter 8. Conclusion

We tried to improve English to Bengali EBMT. Right now our system can translate English to Bengali sentence. But it has limited knowledge base. By increasing the knowledge base we can improve its efficiency for general purpose use.

Chapter 9: Reference

[1]. Book: Machine Translation: An Introductory Guide , *By Doug Arnold, Lorna Balkan, Siety Meijer, R.Lee Humphreys, Louisa Sadler; Colchester, August 1993*

[2]. Contributions To English To Hindi Machine Translation Using Example-Based Approach, Phd Theses in January 2005, Deepa Gupta, IIT Delhi.

[3]. D. Gupta and N. Chatterje., Study of Divergence for Example Based English-Hindi Machine Translation. STRANS-2001, IIT Kanpur, 2001 pp. 43-51.

[4]. Balanced Bengali Language Corpus: A Proposal, *By Khan Md. Anwarus Salam, S M Murtoza Habib and Dr. Mumit Khan*, Research work in BRAC University in 2008.

[5]. H.A. Guvenir and I. Cicekli., Learning Translation Templates from Examples. Elsevier Science Ltd., 1998

[6]. R. Jain, R.M.K Sinha and A. Jain., ANUBHATRI: Using Hybrid Example-Based Approach for Machine Translation.. STRANS-2001, IIT Kanpur, 2001 pp. 20-32.

[7]. Verb Transfer For English To Urdu Machine Translation, Independent Study by Nayyara Karamat, FAST-Lahore, 2006

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

[9]. Shah Asaduzzaman and Muhammad Masroor Ali, “Transfer Machine Translation-An Experience with Bangla English Machine Translation System”. In the Proceedings of the International Conference on Computer and Information Technology (ICCIT), Bangladesh, 2003.

[10]. Resolving Pattern Ambiguity for English to Hind Machine Translation Using WordNet, Niladri Chatterjee Shailly Goyal Anjali Naithani, Department of Mathematics, Indian Institute of Technology Delhi, Published in Workshop Modern Approaches in Translation Technologies 2005 - Borovets, Bulgaria

[11]. “Example Based English to Bengali Machine Translation” Thesis work of Khan Md. Anwarus Salam completed in August 2009, BRAC University