Sublexical Translations for Low-Resource Language

Khan Md. Anwarus Salam[1] Setsuo Yamada[2] Tetsuro Nishino[1]
(1) The University of Electro-Communications, Tokyo, Japan.
(2) NTT Corporation, Tokyo, Japan.
salamkhan@uec.ac.jp, yamada.setsuo@lab.ntt.co.jp, nishino@uec.ac.jp

ABSTRACT

Machine Translation (MT) for low-resource language has low-coverage issues due to Out-Of-Vocabulary (OOV) Words. In this research we propose a method using sublexical translation to achieve wide-coverage in Example-Based Machine Translation (EBMT) for English to Bangla language. For sublexical translation we divide the OOV words into sublexical units for getting translation candidates. Previous methods without sublexical translation failed to find translation candidate for many joint words. In this research using WordNet and IPA transliteration algorithm we propose to translate OOV words with explanation. The proposed method is better than previous OOV words handling. Our proposal improved translation quality by 20 points in human evaluation.

KEYWORDS : Example-Based Machine Translation, Out-Of-Vocabulary Words, WordNet, Word Sense Disambiguation

1 Introduction

Since significant amount of the web contents are in English[1], it is very important to have a Machine Translation (MT) system for monolingual speakers of different languages. Bangla is the native language of around 230 million speakers worldwide, mostly from Bangladesh and West Bengal of India. To improve the information access to those Bangla speaking monolingual people, it is important to have good English to Bangla Machine Translation (MT) system. However, Bangla is a low-resource language due to the lack of language resources like Bangla WordNet and authorized parallel corpus, which makes the development of the MT system very challenging. More specifically, we are concerned about translating Out-Of-Vocabulary (OOV) words. Because MT systems for low-resource language has high probability of handling OOV words. English has rich language resources like automated parser, tokenizer and WordNet. WordNet is a large lexical database of English (Miller, 1995). On the other hand Bangla is a low-resource language due to the lack of language resources like Bangla WordNet and authorized parallel corpus. In this situation, to utilize the available language resources for English, we consider using English as source language (SL) and Bangla as target language (TL).

There were several attempts at building English-Bangla MT systems. The first available free MT system from Bangladesh was Akkhor Bangla Software[2]. The second available online MT system was apertium based Anubadok[3]. These systems used Rule-Based approach and did not handle OOV Words considering low-resource scenario. Most recently from June 2011, Google Translation[4] started offering MT service for Bangla language, having issues in translating OOV Words.

We considered Example-Based MT (EBMT) approach by improving the translation quality using WordNet. For using WordNet in to generalize the example-base we used chunk-string templates (CSTs) (Salam et. al, 2011a). CSTs consist of a chunk in the source language (English), a string in the target language (Bangla), and the word alignment information between them. CSTs are generated from the aligned parallel corpus and WordNet, by using English chunker. For clustering CSTs, we used <lexical filename> information for each words, provided by WordNet-Online[5]. Translaing OOV words using WordNet did not quantify the translation quality improvement (Salam et. al, 2012). In EBMT, it has been proposed to perform a fuzzy match on the corpus based on semantic distance (Sato and Nagao, 1990). Generalized templates proven to be useful for EBMT to achieve wide-coverage (Gangadharaiah, 2011). Using Chunks also helps for low-resource EBMT (Kim, 2010) .

However, previous approaches did not consider sublexical translation techniques together with WordNet to find the translation candidates of OOV words. In this paper, sublexical is part of the word which has independent meaning. For example, “bluebird” has two sublexical units: “blue” and “bird”.

In this research using WordNet and IPA transliteration algorithm we propose to translate OOV words with explanation. The proposed method is better than previous OOV words handling.

This method is effective to find translation candidates in Example-Based Machine Translation (EBMT) for English to Bangla language. To improve the translation quality, we implemented the proposed method in EBMT.

[1] http://www.netz-tipp.de/languages.html
[2] http://www.akkhorbangla.com
[3] anubadok.sourceforge.net
[4] http://translate.google.com/#en|bn|
[5] http://wordnetweb.princeton.edu/perl/webwn

To find semantically related English words from WordNet for the OOV word, we need to select the correct WordNet synset. In this research, in order to translate OOV words with better quality, we introduced word-sense-disambiguation technique to choose the semantically closest WordNet synset. Using the WordNet synset and English-Bangla dictionary, we proposed an improved mechanism to translate the OOV word. If no Bangla translation exists, the system uses IPA-based-transliteration. For proper nouns, the system uses the transliteration mechanism provided by Akkhor Bangla Software. Based on the above methods, we built an English-to-Bangla EBMT. Proposed solution improved translation quality by 20 points in human evaluation.

2 Background

We used EBMT approach for low-resource language using chunk-string templates (Salam et. al., 2011a). The Figure 1 shows the EBMT architecture. During the translation process, at first, the input sentence is parsed into chunks using OpenNLP Chunker. The output of Source Language Analysis step is the English chunks. Then the chunks are matched with the example base using the Matching algorithm as described in section IV. This process provides the CSTs candidates from the example-base. It also marks the OOV Words. In OOV Word Translation step, we try to choose the translation candidate for those OOV Words with the help of WordNet. Our improved WSD technique helps the system to choose the correct WordNet system for better translation of the OOV word. Finally in Generation process WordNet helps to translate determiners and prepositions correctly to improve MT performance (Salam et. al, 2011b). Finally using the generation rules we output the target-language strings. Based on the above MT system architecture, we built an English-to-Bangla MT system.

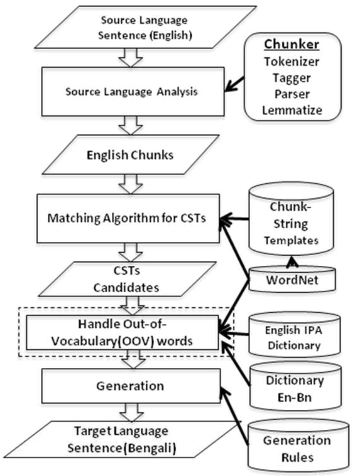

FIGURE 1 – EBMT Architecture

The dotted rectangle in Figure 1 identified the new contribution area of this research. Here we used EBMT based on chunk-string templates (CSTs), which is especially useful for developing a MT system for high-resource to low-resource language. CSTs consist of a chunk in the source language (English), a string in the target language (Bangla), and the word alignment information between them. From the English-Bangla aligned parallel corpus CSTs are generated automatically.

English	**Bangla**	**Align**
Bangla is the native language of 1 2 3 4 5 6 around 230 million people worldwide 7 8 9 10 11	বিশ্বব্যাপী বাংলা হচ্ছে প্রায় ২৩০ মিলিয়ন মানুষ –এর মাতৃভাষা	11 1 2 7 8 9 10 6 4

TABLE 1 – Example word-aligned parallel corpus.

Table 1 shows sample word-aligned parallel corpus. Here the alignment information contains English position number for each Bangla word. For example, the first Bangla word "বিশ্বব্যাপী" is aligned with the 11th word in the English sentence. That means "বিশ্বব্যাপী" is aligned with "worldwide".

The example-base of our EBMT is stored as CSTs. We produced CSTs from the parallel corpus. Table 2 shows the initial CSTs for the parallel sentence given in Table1. In Table 2, c is a chunk in the source language (English), s is a string in the target language (Bangla), and t is the alignment information calculated from the original word alignment.

CST#	**English Chunk (C)**	**Bangla (S)**	**T**
CST1	[NP Bangla/NNP]	বাংলা	1
CST2	[VP is/VBZ]	হচ্ছে	1
CST3	[NP the/DT native/JJ language/NN]	মাতৃভাষা	2
CST4	[PP of/IN]	–এর	1
CST5	[NP around/RB 230/CD million/CD people/NNS]	প্রায় ২৩০ মিলিয়ন মানুষ	1 2 3 4
CST6	[ADVP worldwide/RB]	বিশ্বব্যাপী	1

TABLE 2 – Example of initial CSTs.

In the next step CSTs are generalized by using WordNet to increase the EBMT coverage. To generalize we only consider nouns, proper nouns and cardinal number (NN, NNP, CD in OpenNLP tagset). For each proper nouns we search in WordNet. If available we replace that

NNP with <lexical filename> returned from the WordNet. For example WordNet return <noun.communication> for "Bangla". For cardinal number we simply CDs together to <noun.quantity>. We show example generalized CSTs produced using WordNet in Table 3.

CST#	English Chunk (C)	Generalized Chunk
CST1	[NP Bangla /NNP]	[NP<noun.communication>/NNP]
CST5	[NP around/RB 230/CD million/CD people/NNS]	[NP around/RB <noun.quantity> people/NNS]

TABLE 3 – Combined- CSTs examples.

Finally we get the CSTs database which has three tables: initial CSTs, generalized CSTs and Combined-CSTs. From the example word-aligned parallel sentence of Table 1, system generated 6 initial CSTs, 2 Generalized CSTs and 4 Combined-CSTs.

3 Handle Out-of-Vocabulary Problem

As in our assumption, the main users of this EBMT will be monolingual people; they cannot read or understand English words written in English alphabet. However, with related word translation using WordNet and Transliteration can give them some clues to understand the sentence meaning. As Bangla language accepts foreign words, transliterating an English word into Bangla alphabet, makes that a Bangla foreign word. For example, in Bangla there exist many words, which speakers can identify as foreign words.

Figure 2 shows the OOV or Unknown Words translation process in a flow chart. Proposed system first tries to find semantically related English words from WordNet for the OOV word. From these related words, we rank the translation candidates using WSD technique and English-Bangla dictionary. If no Bangla translation exists, the system uses IPA-based-transliteration. For proper nouns, the system uses transliteration mechanism provided by Akkhor Bangla Software.

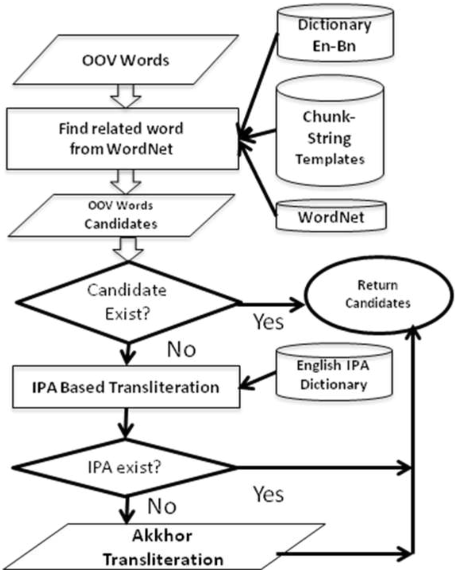

FIGURE 2 – Steps of handling OOV words

3.1 Find Sublexical Translations

For sublexical matching our system divide the OOV word into sublexical units and then find possible translation candidates from these sublexical units. For this the system use following steps:

(1) Find the possible sublexical units of the OOV. For example, the OOV "bluebird" gets divided into "blue" and "bird".

(2) Extract sublexical translations and restrain translation choices.

(3) Remove less probable sublexical translations

(4) Output translation candidates with the POS tags for the sublexical units of the OOV.

From the set of all CSTs we select the most suitable one, according to the following criteria:
1. The more exact CSTs matched, the better;
2. Linguistically match give priority by following these ranks, higher level is better:
•Level 4: Exact match.
•Level 3: Sublexical unit match, <lexical filename> of WordNet and POS tags match
•Level 2: Sublexical unit match, <lexical filename> of WordNet match
•Level 1: Only POS tags match.

•Level 0: No match found, all OOV words.

3.2 Find Candidates from WordNet

Due to small English-Bangla parallel corpus availability, there is high probability for the MT system to handle OOV words. Therefore, it is important to have a good method for translating OOV words. When the word has no match in the CSTs, it tries to translate using English WordNet and bilingual dictionary for English-Bangla.

Input of this step is OOV or unknown words. For example "canine" is a OOV in our system. Output of this process is the related OOV words translation.

3.2.1 Find Candidates from WordNet

The system first finds the synonyms for the OOV word from the WordNet synsets. Each synset member becomes the candidate word for OOV. WordNet provide related word for nouns, proper nouns, verbs, adjectives and adverbs. Synonymy is WordNet's basic relation, because WordNet uses sets of synonyms (synsets) to represent word senses. Synonymy is a symmetric relation between word forms. We can also use Entailment relations between verbs available in WordNet to find OOV candidate synonyms.

3.2.2 Find Candidates Using Antonyms

WordNet provide related word for nouns, proper Antonymy (opposing-name) is also a symmetric semantic relation between word forms, especially important in organizing the meanings of adjectives and adverbs. For some OOV we can get the antonyms from WordNet. If the antonym exists in the dictionary we can use the negation of that word to translate the OOV word. For example, "unfriendly" can be translated as "not friendly". In Bengali to negate such a word we can simply add "না" (na) at the end of the word. So, "unfriendly" can be translated as "বন্ধুত্বপূর্ণ না" (bondhuttopurno na). It helps to translate OOV words like "unfriendly", which improves the machine translation quality.

Hyponymy (sub-name) and its inverse, hypernymy (super-name), are transitive relations between synsets. Because there is usually only one hypernym, this semantic relation organizes the meanings of nouns into a hierarchical structure. We need to process the hypernyms to translate the OOV word.

3.2.3 Find Candidates Using Hypernyms

For nouns and verbs WordNet provide hypernyms, which is defined as follows:

Y is a hypernym of X if every X is a (kind of) Y.

For example "canine" is a hypernym of noun "carnivore", because every dog is a member of the larger category of canines. Verb example, "to perceive" is an hypernym of "to listen". However, WordNet only provides hypernym(s) of a synset, not the hypernym tree itself. As hypernyms can express the meaning, we can translate the hypernym of the unknown word. To do that, until any hypernym's Bangla translation found in the English-Bangla dictionary, we keep discovering upper level of hypernym's. Because, nouns and verbs are organized into hierarchies, defined by hypernyms or is-a-relationships in WordNet. So, we considered lower level synset words are generally more suitable then the higher level synset words.

This process discovers the hypernym tree from WordNet in step by step. For example, from the hypernym tree of "dog" from WordNet, we only had the "animal" entry in our English-

Bangla dictionary. Our system discovered the hypernym tree of “dog” from WordNet until “mammal”. Following is the discovered hypernym tree:

dog, domestic dog, Canis familiaris

=> canine, canid

=> carnivore => placental, placental mammal

=> mammal => vertebrate, craniate => chordate

=> animal => ...

This process search in English-Bangla dictionary, for each of the entry of this hypernym tree. So at first we used the IPA representation of the English word from our dictionary, then using transliterating that into Bengali. Then system produce “a kind of X” - এক ধরনের X [ek dhoroner X]. For the example of “canine” we only had the Bengali dictionary entry for “animal” from the whole hypernym tree. We translated “canine” as the translation of “canine, a kind of animal”, in Bangla which is “ক্যানাইন, এক ধরনের পশু” [kjanain, ek dhoroner poshu].

Similarly, for adjectives we try to find “similar to” words from WordNet. And for Adverbs we try to find “root adjectives”.

Finally, this step returns OOV words candidates from WordNet which exist in English-Bangla dictionary.

Using the same technique described above, we can use Troponyms and Meronyms to translate OOV words. Troponymy (manner-name) is for verbs what hyponymy is for nouns, although the resulting hierarchies are much shallower. Meronymy (part-name) and its inverse, holonymy (whole-name), are complex semantic relations. WordNet distinguishes component parts, substantive parts, and member parts.

3.3 Rank Candidates

To choose among the candidates for the OOV word, we need to rank all the candidates. Here we used following technique for this.

3.3.1 Selecting Adequate WordNet synset for OOV

Especially polysemous OOV words need to select the adequate WordNet synset to choose the right candidate. The system perform Google search with the input sentence as a query, by replacing the OOV word with each candidate words. We add quotation marks in the input sentence to perform phrase searches in Google, to find the number of in documents the sentence appear together. If the input sentence with quotation mark returns less than 10 results, we perform Google search with four and two neighbour chunks. Finally, the system ranks the candidate words using the Google search hits information.

For example, the input sentence in SL is: This dog is really cool. The system first adds double quotation with the input sentence: “This dog is really cool”, which returns 37,300 results in Google. Then the system replaces the OOV “dog” from discovered hypernym tree. Only for "This animal is really cool.", returned 1,560 results by Google. That is why “animal” is the second most suitable candidate for “dog”.

However, other options "This domestic dog is really cool." or "This canine is really cool." etc. returns no results or less than 10 results in Google. So in this case we search with neighbour chunks only.

For example, in Google:

"This mammal is" returns 527,000 results;

"This canid is" returns 503,000 results;

"This canine is" returns 110,000 results;

"This carnivore is" returns 58,600 results;

"This vertebrate is" returns 2,460 results;

"This placental is" returns 46 results;

"This craniate is" returns 27 results;

"This chordate is" returns 27 results;

"This placental mammal is" returns 6 result;

Finally the system returns the OOV candidates: mammal, canid, canine, carnivore, vertebrate, placental, craniates, chordate, placental mammal.

3.4 Final Candidate Generation

In this step, we choose one translation candidate. If any of the synonyms or candidate word exist in English-Bangla dictionary, the system translates the OOV word with that synonym meaning. If multiple synonyms exist then the entry with highest Google search hits get selected.

English-Bangla dictionary also contains multiple entries in target language. For WSD analysis in target language, we perform Google search with the produced translation by the system. The system chooses the entry with highest Google hits as final translation of the OOV word. For example, for OOV “dog”, animal get selected in our system.

However, if there were no candidates, we use IPA-Based-Transliteration.

3.4.1 IPA-Based-Transliteration

When OOV word is not even found in WordNet, we use IPA-Based transliteration using the English IPA Dictionary (Salam et. al., 2011b). Output for this step is the Bangla word transliterated from the IPA of the English word. In this step, we use English-Bangla Transliteration map to transliterate the IPA into Bangla alphabet.

Mouth narrower vertically	[iː] ই / ি sleep /sliːp/	[I] ই / ি slip /sl**I**p/	[ʊ] উ / ু book /bʊk/	[uː]ঊ/ ূ boot /buːt/
	[e] এ / ে ten /ten/	[ə] আ / া after /aːftə/	[ɜː] আ / া bird /bɜːd/	[ɔː] র্ bored /bɔːd/

Mouth wider vertically	[æ]এ্যা/◌্যা cat /kæt/	**[^] আ** /◌া cup / k^p/	[ɑː] **আ** / ◌া car / cɑːr/	[ɒ] অ hot /hɒt/

English-Bengali IPA mapping for vowels

[**I**ə] ইয়া/ি◌য়া beer /b**Iər/**	[e**I]** এই/ ে◌ই say /se**I**/	
[ʊə] উয়া/ ◌ুয়া fewer /fjʊər/	[ɔ**I**] অয়/য় boy /bɔ**I**/	[ə ʊ] ও / ে◌া no /nəʊ/
eə ঈয়া/ ◌ীয়া bear /beər/	[a**I]** ◌াই/ আই high /ha**I**/	[aʊ]আউ /◌াউ cow /kaʊ/

English-Bengali IPA mapping for diphthongs

[p] প pan /pæn/	[b] ব ban /bæn/	[t] ট tan /tæn/	[d] ড day /deI/	[ʧ] চ chat /ʧæt/	[ʤ] জ judge /ʤ^ʤ/	[k] ক key /kiː/	[g] গ get /get/
[f] ফ fan /fæn/	[v] ভ van / væn/	[θ] থ thin /θIn/	[ð] দ than /ðæn/	[s] স sip /sIp/	[z] জ zip / zIp/	[ʃ] শ ship /ʃIp/	[ʒ] স vision /vIʒ^n/
[m] ম might /maIt/	[n] ন night /naIt/	[ŋ]◌ং/ঙ thing /θIŋ/	[h] হ height /haIt/	[l] ল light /laIt/	[r] র right /raIt/	[w] য় white /hwaIt/	[j]ইয়ে yes /jes/

English-Bengali IPA mapping for consonants

FIGURE 3– English-Bengali IPA mapping

From English IPA dictionary the system can obtain the English words pronunciations in IPA format. Output for this step is the Bengali word transliterated from the IPA of the English word. In this step, we use following English-Bengali Transliteration map to transliterate the IPA into Bengali alphabet. Figure 3 shows our proposed English-Bengali IPA chart for vowels, diphthongs and consonants. Using rule-base we transliterate the English IPA into Bangla alphabets. The above IPA charts leaves out many IPA as we are considering about translating from English only. To translate from other language such as Japanese to Banglawe need to create Japanese specific IPA transliteration chart. Using the above English-Bangla IPA chart we produced transliteration from the English IPA dictionary. For examples: pan(pæn): প্যান; ban(bæn): ব্যান; might(maIt): মাইট.

However, when unknown word is not even found in the English IPA dictionary, we use transliteration mechanism of Akkhor Bangla Software. For example, for the word "Muhammod" which is a popular Bangla name, Akkhor transliterated into "মুহাম্মদ" in Bangla.

বাংলা	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
English	A	a/aa/`a	i/`i	I/ee/`I	u/`u	U/`U	ri/`ri	e/`e	oi/`oi	o/`o	ou ou

বাংলা	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
English	k	kh	g	gh	Ng	ch	Ch	j	jh	Y
বাংলা	ত	থ	দ	ধ	ন	ট	ঠ	ড	ঢ	ণ
English	t	th	d	dh	n	**T**	Th	D	Dh	N
বাংলা	প	ফ	ব	ভ	ম	য	র	ল	শ	ষ
English	p	f/ph	b	bh/v	m	**z**	r	l	sh	S
বাংলা	স	ক্ষ	হ	ড়	ঢ়	য়	ং	ঃ	ঁ	
English	S	k-S	h	R	rh	y	ng	:	~	
বাংলা	১	২	৩	৪	৫	৬	৭	৮	৯	০
English	1	2	3	4	5	6	7	8	9	1
বাংলা	কা	কে	কি	কু	কো	ক্র	ক্রে	ক্রি	ক্রু	ক্রূ
English	ka	ke	ki	ku	kO	kro	kre	kree	kru	krU
বাংলা	কী	চী	মী	কূ	মূ	বূ	ণূ	নূ	ক্য	ব্য
English	kI	chI	mI	kU	mU	bU	NU	nU	k-z	b-z

FIGURE 4– Akkhor phonetic mapping for Bengali alphabets

4 Exeriment

We did quality evaluations for the proposed EBMT with unknown words, by comparing with baseline EBMT system. Quality evaluation measures the translation quality through human evaluation.

Baseline system architecture has the same components as described in Fig. 1, except for the components inside dotted rectangles. Matching algorithm of baseline system is that not only match with exact translation examples, but it can also match with POS tags. The Baseline EBMT use the same training data: English-BANGLA parallel corpus and dictionary, but does not use CSTs, WordNet and unknown words translation solutions.

Currently from the training data set of 2000 word aligned English-Bangla parallel sentences, system generated 15356 initial CSTs, 543 Generalized CSTs and 12458 Combined-CSTs. As this research is focused for low-resource language, we trained our MT system with 2000 word aligned parallel corpus and small dictionary.

The development environment was in windows-OS using C Sharp language. Our test-set contained 336 sentences, which are not same as training data. The test-set includes simple and complex sentences, representing various grammatical phenomena. We have around 20,000 English-Bangla dictionary entries.

Translation Quality	Grade	Baseline EBMT %	EBMT with Sublexical %
Perfect	A	22.67	36.00
Good	B	25.33	32.00
Medium	C	19.67	20.00
Poor	D	32.33	12.00
Total		**100%**	**100%**

TABLE 4 – Human evaluation of EBMT system using same testset.

Quality evaluation measures the translation quality through human evaluation. Perfect Translation means there is no problem in the target sentence, and exact match with test-set translation. Good Translation means not exact match with test-set reference, but still understandable for human. Medium means there are several problems in the target sentence, like wrong word choice and wrong word order. Poor Translation means there are major problems in the target sentence, like non-translated words, wrong word choice and wrong word order. Our phonetic transcription component helped to improve such poor translation into medium translation quality. Table 4 shows the human evaluation of current system. Around 20 points of poor or medium translations produced by “Baseline EBMT” was improved using the proposed sublexical word translation mechanism.

#	OOV context	EBMT sublexical
1	WordNet is a..	শব্দজাল হচ্ছে.. (shobdojal hocche..) WordNet is a..(A)
2	Sublexical units of a word ..	শব্দের উপ-আভিধানিক অংশ.. (shobder upoabhidhanik onghsho) Sublexical units of a word .. (A)
3	This is a bluebird..	এটা নীলপাখি .. (eta nilpakhi..) This is a bluebird.. (A)

TABLE 5 – Comparison of produced translations of OOV words

Table 5 shows the sample produced translations of OOV words. The “OOV context” column shows the OOV word with context where the underlined word is OOV word. “EBMT sublexical” shows the OOV translation produced by our proposed mechanism. Column 3 shows the OOV translation produced in Bengali alphabet, then the transliteration in brackets, then the English meaning of the produced translation and finally the quality of translation using the above grades.

Conclusion

We proposed a method using sublexical translation to achieve wide-coverage in Example-Based Machine Translation (EBMT) for English to Bangla language. Our experiment showed the method is effective to handle the OOV problem in EBMT. Proposed solution improved translation quality by 20 points in human evaluation. In future, we want to experiment the proposed method with other low-resource Indian languages having larger training data.

References

Abney, Steven. 1991. Parsing by chunks. In Principle- Based Parsing, pages 257–278. Kluwer Academic Publishers.

Chung-Chi Huang, Ho-Ching Yen, Ping-Che Yang, Shih-Ting Huang, and Jason S. Chang. 2011. Using Sublexical Translations to Handle the OOV Problem in Machine Translation. 10, 3, Article 16 (September 2011), 20 pages. DOI=10.1145/2002980.2002986 http://doi.acm.org/10.1145/2002980.2002986

Diganta Saha, Sivaji Bandyopadhyay. 2006. A Semantics-based English-Bengali EBMT System for translating News Headlines. Proceedings of the MT Summit X, Second workshop on Example-Based Machine Translation Programme.

Diganta Saha, Sudip Kumar Naskar, Sivaji Bandyopadhyay. 2005. A Semantics-based English-Bengali EBMT System for translating News Headlines, MT Xummit X.

George A. Miller. 1995. WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

Jae Dong Kim, Ralf D. Brown, Jaime G. Carbonell. 2010. Chunk-Based EBMT. EAMT, St Raphael, France.

Khan Md. Anwarus Salam, Setsuo Yamada and Tetsuro Nishino. 2011a. Example-Based Machine Translation for Low-Resource Language Using Chunk-String Templates, 13th Machine Translation Summit, Xiamen, China.

Khan Md. Anwarus Salam, Yamada Setsuo and Tetsuro Nishino. 2011b. Translating Unknown Words Using WordNet and IPA-Based-Transliteration . 14th International Conference on Computer and Information Technology (ICCIT 2011), Dhaka, Bangladesh.

Khan Md. Anwarus Salam, Yamada Setsuo and Tetsuro Nishino. 2012. WordNet to Handle the OOV Problem in English to Bangla Machine Translation. Global WordNet Conference 2012. Matsue, Japan.

R. Gangadharaiah, R. D. Brown, and J. G. Carbonell. 2011. Phrasal equivalence classes for generalized corpus-based machine translation. In Alexander Gelbukh, editor, Computational Linguistics and Intelligent Text Processing, volume 6609 of Lecture Notes in-Computer Science, pages 13–28. Springer Berlin / Heidelberg, 2011.

Ralf D. Brown. 1999. Adding Linguistic Knowledge to a Lexical Example-Based Translation, Proceedings of the Eighth International Conference on Theoretical and Methodological Issues in Machine Translation System. Pp. 22-32. Chester, England.

Satoshi Sato, Makoto Nagao. 1990. Toward Memory-Based Translation. COLING 1990.

Sudip Kumar Naskar and Sivaji Bandyopadhyay. 2006b. Handling of Prepositions in English to Bengali Machine Translation. In the proceedings of Third ACL-SIGSEM Workshop on Prepositions, EACL 2006. Trento, Italy.

Zhanyi Liu, Haifeng Wang And Hua Wu. 2006. Example-Based Machine Translation Based on Tree-string Correspondence and Statistical Generation. Machine Translation, 20(1): 25-41