Manipuri-English Bidirectional Statistical Machine Translation Systems using Morphology and Dependency Relations

Thoudam Doren Singh
Department of Computer Science and
Engineering
Jadavpur University
thoudam.doren@gmail.com

Sivaji Bandyopadhyay
Department of Computer Science and
Engineering
Jadavpur University
sivaji_cse_ju@yahoo.com

Abstract

The present work reports the development of Manipuri-English bidirectional statistical machine translation systems. In the English-Manipuri statistical machine translation system, the role of the suffixes and dependency relations on the source side and case markers on the target side are identified as important translation factors. A parallel corpus of 10350 sentences from news domain is used for training and the system is tested with 500 sentences. Using the proposed translation factors, the output of the translation quality is improved as indicated by baseline BLEU score of 13.045 and factored BLEU score of 16.873 respectively. Similarly, for the Manipuri English system, the role of case markers and POS tags information at the source side and suffixes and dependency relations at the target side are identified as useful translation factors. The case markers and suffixes are not only responsible to determine the word classes but also to determine the dependency relations. Using these translation factors, the output of the translation quality is improved as indicated by baseline BLEU score of 13.452 and factored BLEU score of 17.573 respectively. Further, the subjective evaluation indicates the improvement in the fluency and adequacy of both the factored SMT outputs over the respective baseline systems.

1 Introduction

Manipuri has little resource for NLP related research and development activities. Manipuri is a less privileged Tibeto-Burman language spoken by approximately three million people mainly in the state of Manipur in India as well as its neighboring states and in the countries of Myanmar and Bangladesh. Some of the unique features of this language are tone, the agglutinative verb morphology and predominance of aspect than tense, lack of grammatical gender, number and person. Other features are verb final word order in a sentence i.e., Subject Object Verb (SOV) order, extensive suffix with more limited prefixation. In Manipuri, identification of most of the word classes and sentence types are based on the markers. All sentences, except interrogatives end with one of these mood markers, which may or may not be followed by an enclitic. Basic sentence types in Manipuri are determined through illocutionary mood markers, all of which are verbal inflectional suffixes, with the exception of the interrogatives that end with an enclitic. Two important problems in applying statistical machine translation (SMT) techniques to English-Manipuri bidirectional MT systems are: (a) the wide syntactic divergence between the language pairs, and (b) the richer morphology and case marking of Manipuri compared to English. The first problem manifests itself in poor word-order in the output translations, while the second one leads to incorrect inflections and case marking. The output Manipuri sentences in case of English-Manipuri system suffer badly when morphology and case markers are incorrect in this free word order and morphologically rich language.

The parallel corpora used is in news domain which have been collected, cleaned and aligned (Singh et al., 2010b) from the Sangai Express newspaper website www.thesangaiexpress.com available in both Manipuri and English. A daily basis collection was done covering the period from May 2008 to November 2008 since there is no repository.

2 Related Works

Koehn and Hoang (2007) developed a framework for statistical translation models that tightly integrates additional morphological, syntactic, or semantic information. Statistical Machine Translation with scarce resources using morpho-syntactic information is discussed in (Nieβen and Ney, 2004). It introduces sentence level restructuring transformations that aim at the assimilation of word order in related sentences and exploitation of the bilingual training data by explicitly taking into account the interdependencies of related inflected forms thereby improving the translation quality. Popovic and Ney (2006) discussed SMT with a small amount of bilingual training data. Case markers and morphology are used to address the crux of fluency in the English-Hindi SMT system (Ramanathan et al., 2009). Work on translating from rich to poor morphology using factored model is reported in (Avramidis and Koehn, 2008). In this method of enriching input, the case agreement for nouns, adjectives and articles are mainly defined by the syntactic role of each phrase. Resolution of verb conjugation is done by identifying the person of a verb and using the linguistic information tag. Manipuri to English Example Based Machine Translation system is reported in (Singh and Bandyopadhyay, 2010a) on news domain. For this, POS tagging, morphological analysis, NER and chunking are applied on the parallel corpus for phrase level alignment. Chunks are aligned using a dynamic programming "edit-distance style" alignment algorithm. The translation process initially looks for an exact match in the parallel example base and returns the retrieved target output. Otherwise, the maximal match source sentence is identified. For word level mismatch, the unmatched words in the input are either translated from the lexicon or transliterated. Unmatched phrases are looked into the phrase level parallel example base; the target phrase translations are identified and then recombined with the retrieved output. English-Manipuri SMT system using morpho-syntactic and semantic information is reported in (Singh and Bandyopadhyay, 2010c). In this system, the role of the suffixes and dependency relations on the source side and case markers on the target side are identified as important translation factors.

3 Syntactic Reordering

This is a preprocessing step applied to the input English sentences for English-Manipuri SMT system. The program for syntactic reordering uses the parse trees generated by Stanford parser[1] and applies a handful of reordering rules written using perl module Parse::RecDescent. By doing this, the SVO order of English is changed to SOV order for Manipuri, and post modifiers are converted to pre-modifiers. The basic difference of Manipuri phrase order compared to English is handled by reordering the input sentence following the rule (Rao et al., 2000):

$$SS_mV\ V_mOO_mC_m\ \rightarrow\ C'_mS'_mS'O'_mO'V'_mV'$$

    where,   S: Subject
    O: Object
    V : Verb
    C_m: Clause modifier
    X': Corresponding constituent in Manipuri, where X is S, O, or V
    X_m: modifier of X

There are two reasons why the syntactic reordering approach improves over the baseline phrase-based SMT system (Wang et al., 2007). One obvious benefit is that the word order of the transformed source sentence is much closer to the target sentence, which reduces the reliance on the distortion model to perform reordering during decoding. Another potential benefit is that the alignment between the two sides will be of higher quality because of fewer "distortions" between the source and the target, so that the resulting phrase table of the reordered system would be better. However, a counter argument is that the reordering is very error prone, so that the added noise in the reordered data actually hurts the alignments and hence the phrase tables.

---

[1] http://nlp.stanford.edu/software/lex-parser.shtml

4 Morphology

The affixes are the determining factor of the word class in Manipuri. In this agglutinative language the number of verbal suffixes is more than that of nominal suffixes. Works on Manipuri morphology are found in (Singh and Bandyopadhyay, 2006) and (Singh and Bandyopadhyay, 2008). In this language, a verb must minimally consist of a verb root and an inflectional suffix. A noun may be optionally affixed by derivational morphemes indicating gender, number and quantity. Further, a noun may be prefixed by a pronominal prefix which indicates its possessor. Words in Manipuri consist of stems or bound roots with suffixes (from one to ten suffixes), prefixes (only one per word) and/or enclitics.

    (a) ইবোমচা–না   বোল–দু   কাওই
       Ibomcha-na   *Ball-du*   *kao-i*
       Ibomcha-nom Ball-distal   kick
       Ibomcha kicks the ball.

    (b) বোল–দু   ইবোমচা–না   কাওই
       Ball-du   *Ibomcha-na*   *kao-i*
       Ball-distal   Ibomcha-nom kick
       Ibomcha kicks the ball.

The identification of subject and object in both the sentences are done by the suffixes না (*na*) and দু (*du*) as given by the examples (a) and (b). The case markers convey the right meaning during translation though the most acceptable order of Manipuri sentence is SOV. In order to produce a good translation output all the morphological forms of a word and its translations should be available in the training data and every word has to appear with every possible suffixes. This will require a large training data. By learning the general rules of morphology, the amount of training data could be reduced. Separating lemma and suffix allows the system to learn more about the different possible word formations.

Manipuri	Gloss	English Meaning
তোম্না	*Tom-na*	by Tom
তোমদগী	*Tom-dagi*	from Tom
তামসু	*Tom-su*	Tom also
তোমগী	*Tom-gi*	of Tom
তোমগা	*Tom-ga*	with Tom

Table 1: Some of the inflected forms of names in Manipuri and its corresponding English meaning

Table 1 gives some examples of the inflected forms of a person name and its corresponding English meaning. The Manipuri stemmer separates the case markers such as –না *(-na)*, -দগী *(-dagi)*, -সু *(-su)*, -গী *(-gi)*, -গা *(-ga)* etc. from surface forms so that 'তোম' (*Tom*) from Manipuri side matches with "Tom" at English side helping to overcome the data sparseness. Enclitics in Manipuri fall into six categories: determiners, case markers, the copula, mood markers, inclusive / exclusive and pragmatic peak markers and attitude markers. The role of the enclitics used and its meaning differs based on the context.

5 Factored Model of Translation

Using factored approach, a tighter integration of linguistic information into the translation model is done for two reasons[2]:

- Translation models that operate on more general representations, such as lemma instead of surface forms of words, can draw on richer statistics and overcome the data sparseness problem caused by limited training data.

- Many aspects of translation can be best explained at a morphological, syntactic or semantic level. Having such information available to the translation model allows the direct modeling of these aspects. For instance, reordering at the sentence level is mostly driven by general syntactic principles, local agreement constraints that show up in morphology, etc.

5.1 Combination of Components in Factored Model

Factored translation model is the combination of several components including language model, reordering model, translation steps and generation steps in a log-linear model[3]:

$$p(\mathbf{e}|\mathbf{f}) = \frac{1}{Z} \sum_{i=1}^{n} \lambda_i h_i(\mathbf{e}, \mathbf{f}) \tag{1}$$

Z is a normalization constant that is ignored in practice. To compute the probability of a translation **e** given an input sentence **f**, we have to evaluate each feature function h_i. The feature weight

λ_i in the log linear model is determined by using minimum error rate training method (Och, 2003).

For a translation step component, each feature function h_t is defined over the phrase pairs (f_j, e_j) given a scoring function τ:

$$h_t(\mathbf{e}, \mathbf{f}) = \sum_j \tau(f_j, e_j) \qquad (2)$$

For the generation step component, each feature function h_g given a scoring function γ is defined over the output words e_k only:

$$h_g(\mathbf{e}, \mathbf{f}) = \sum_k \gamma(e_k) \qquad (3)$$

5.2 Stanford Dependency Parser

The dependency relations used in the experiment are generated by the Stanford dependency parser (Marie-Catherine de Marneffe and Manning, 2008). This parser uses 55 relations to express the dependencies among the various words in a sentence. The dependencies are all binary relations: a grammatical relation holds between a governor and a dependent. These relations form a hierarchical structure with the most general relation at the root.
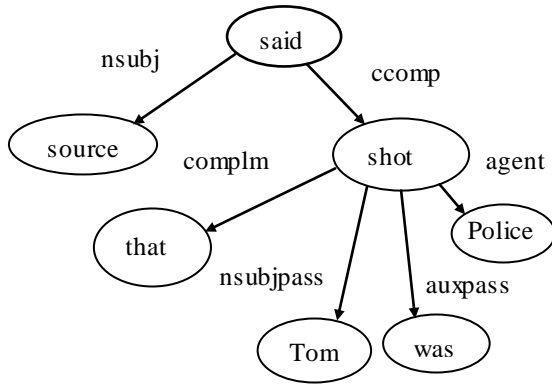

Figure 1. Dependency relation graph of the sentence "*Sources said that Tom was shot by police*" generated by Stanford Parser

There are various argument relations like subject, object, objects of prepositions and clausal complements, modifier relations like adjectival, adverbial, participial, infinitival modifiers and other relations like coordination, conjunct, expletive and punctuation. Let us consider an example "*Sources said that Tom was shot by police*". Stanford parser produces the dependency rela-

tions, nsubj(said, sources) and agent (shot, police) . Thus, *sources*|nsubj and *police*|agent are the factors used. "Tom was shot by police" forms the object of the verb "*said*". The Stanford parser represents these dependencies with the help of a clausal complement relation which links "*said*" with "*shot*" and uses the complementizer relation to introduce the subordination conjunction. Figure 1 shows the dependency relation graph of the sentence "*Sources said that Tom was shot by police*".

5.3 Factorization approach of English-Manipuri SMT system

Manipuri case markers are decided by dependency relation and aspect information of English. Figure 2 shows the translation factors used in the translation between English and Manipuri.

(i) Tomba drives the car.

তোম্বনা কারদু থোই
Tomba-na car-du thou-i

(Tomba) (the car) (drives)

Tomba|empty|nsubj drive|s|empty the|empty|det car|empty|dobj

A subject requires a case marker in a clause with a perfective form such as –না *(na)*. It can be represented as,

suffix+ dependency relation → case marker
s|empty + empty|dobj → না *(na)*

(ii) Birds are flying.

উচেকশিং পাইরি
ucheksing payri

(birds are) (flying)

Bird|s|nsubj are|empty|aux fly|ing|empty

Thus, English-Manipuri factorization consists of

- a lemma to lemma translation factor [i.e., Bird → উচেক *(uchek)*]
- a suffix + dependency relation → suffix [i.e., s + nsubj → শিং *(sing)*]
- a lemma + suffix → surface form generation factor
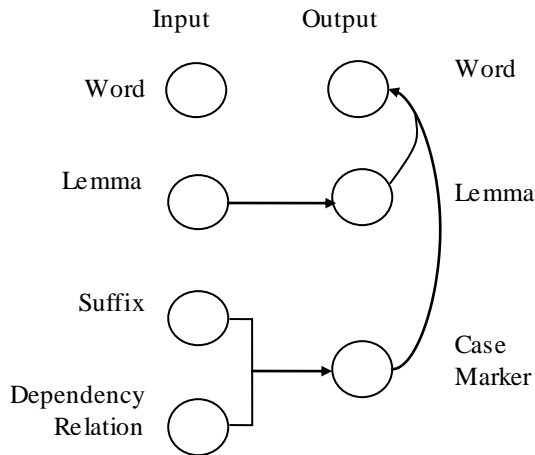 [i.e., উচেক *(uchek)* + শিং *(sing)* → উচেকশিং *(ucheksing)*]

Figure 2. English to Manipuri translation factors

5.4 Factorization approach of Manipuri-English SMT system

Manipuri case markers are responsible to identify dependency relation and aspect information of English. Figure 3 shows the translation factors used in the translation between Manipuri and English. The Manipuri- English factorization consists of:

- **Translation factor:** lemma to lemma
  [e.g., উচেক (*uchek*) → Bird]
- **Translation factor:** suffix + POS → dependency relation + POS + suffix
  [e.g., শিং (*sing*) + NN → nsubj + NN + s]
- **Generation factor:** lemma + POS + dependency Relation +suffix → surface form generation factor
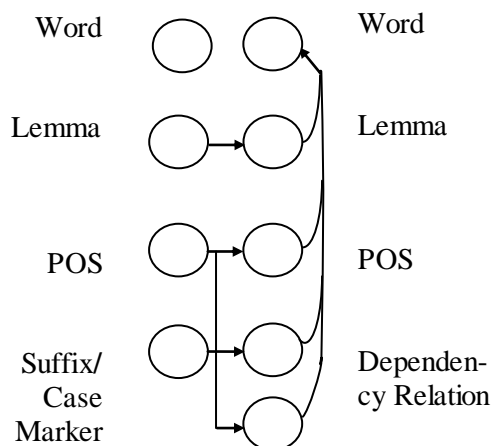 [e.g., উচেক (*uchek*) + NN + nsubj + শিং (*sing*) → উচেকশিং (*ucheksing*]

Figure 3. The Manipuri-English translation factors

5.5 Syntactically enriched output

High-order sequence models (just like n-gram language models over words) are used in order to support syntactic coherence of the output (Koehn and Hoang, 2007).
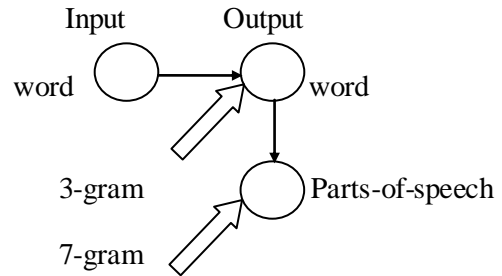

Figure 4. By generating additional linguistic factors on the output side, high-order sequence models over these factors support syntactical coherence of the output.

Adding part-of-speech factor on the output side and exploiting them with 7-gram sequence models (as shown in Figure 4) results in minor improvements in BLEU score.

6 Experimental Setup

A number of experiments have been carried out using factored translation framework and incorporating linguistic information. The toolkits used in the experiment are:

- *Stanford Dependency Parser*[4] was used to (i) generate the dependency relations and (ii) syntactic reordering of the input English sentences using Parse::RecDescent module.
- *Moses*[5] toolkit (Koehn, 2007) was used for training with GIZA++[6], decoding and minimum error rate training (Och, 2003) for tuning.
- *SRILM*[7] toolkit (Stolcke, 2002) was used to build language models with 10350 Manipuri sentences for English-Manipuri system and four and a half million English wordforms collected from the news domain for Manipuri-English system.
- English morphological analyzer *morpha*[8] (Minnen et al., 2001) was used and the

---

stemmer from Manipuri Morphological analyzer (Singh and Bandyopadhyay, 2006) was used for the Manipuri side.

- *Manipuri POS tagger* (Singh et. al., 2008) is used to tag the POS (Parts of speech) factors of the input Manipuri sentences.

7 Evaluation

7.1 English-Manipuri SMT System

The evaluation of the machine translation systems developed in the present work is done in two approaches using automatic scoring with reference translation and subjective evaluation as discussed in (Ramanathan et al., 2009).

Evaluation Metrics:

- *NIST* (Doddington, 2002): A high score means a better translation by measuring the precision of n-gram.
- *BLEU* (Papineni et al, 2002): This metric gives the precision of n-gram with respect to the reference translation but with a brevity penalty.

	No of sentences	No of words
Training	10350	296728
Development	600	16520
Test	500	15204

Table 2. Training, development and testing corpus statistics

Table 2 shows the corpus statistics used in the experiment. The corpus is annotated with the proposed factors. The following models are developed for the experiment.

Baseline:

The model is developed using the default setting values in MOSES.

Lemma +Suffix:

It uses lemma and suffix factors on the source side, lemma and suffix on the target side for lemma to lemma and suffix to suffix translations with generation step of lemma plus suffix to surface form.

Lemma + Suffix + Dependency Relation:

Lemma, suffix and dependency relations are used on the source side. The translation steps are (a) lemma to lemma (b) suffix + dependency relation to suffix and generation step is lemma + suffix to surface form. Table 3 shows the BLEU and NIST scores of the system using these factors.

Table 4 shows the BLEU and NIST scores of the English-Manipuri SMT systems using lexicalized and syntactic reordering.

Model	BLEU	NIST
Baseline (surface)	13.045	4.25
Lemma + Suffix	15.237	4.79
Lemma + Suffix + Dependency Relation	16.873	5.10

Table 3. Evaluation Scores of English - Manipuri SMT System using various translation factors

Model	Reordering	BLEU	NIST
Baseline (surface)		13.045	4.25
Surface	Lexicalized	13.501	4.32
Surface	Syntactic	14.142	4.47

Table 4. Evaluation Scores of English-Manipuri SMT system using Lexicalized and Syntactic Reordering

Input/Output of English-Manipuri SMT:

(1a) **Input:** Going to school is obligatory for students.

স্কুল চৎপা ছাত্রশিংগী তৌদ য়াড্রবা মখোনি ।
School chatpa shatra-sing-gi touda ya

draba mathouni.

Baseline output: স্কুল মখৌ চৎপা ওই ছাত্র
school mathou chatpa oy shatra
gloss : school duty going is student.

Syntactic Reorder output: ছাত্র স্কুল চৎপা তৌদ য়াড্রবা
shatra school chatpa touda yadraba
gloss: Student school going compulsory.

Dependency output: ছাত্রশিং স্কুল চৎপা মখোনি
shatrasing schoolda chatpa mathouni
gloss: Students going to the school is duty.

(1b) **Input:** Krishna has a flute in his hand.
কৃষ্ণগী   খুত্তা   তৌদ্রি   অমা লৈ ।
Krishna-gi khut-ta toudri ama lei.

Syntactic Reorder output: কৃষ্ণ লৈ খুত্ত অমা তৌদ্রি
Krishna lei khut ama toudri
gloss : Krishna has a hand flute

Dependency output: কৃষ্ণগী লৈ তৌদ্রি অমা খুত্তা
krishnagi lei toudri ama khutta
gloss : Krishna has a flute in his hand

One of the main aspects required for the fluency of a sentence is agreement. Certain words have to match in gender, case, number, person etc. within a sentence. The rules of agreement are language dependent and are closely linked to the morphological structure of language. Subjective evaluations on 100 sentences have been performed for fluency and adequacy by two judges. The fluency measures how well formed the sentences are at the output and adequacy measures the closeness of the output sentence with the reference translation. The Table 5 and Table 6 show the adequacy and fluency scales used for evaluation and Table 7 shows the scores of the evaluation.

Level	Interpretation
4	Full meaning is conveyed
3	Most of the meaning is conveyed
2	Poor meaning is conveyed
1	No meaning is conveyed

Table 5. Adequacy scale

Level	Interpretation
4	Flawless with no grammatical error
3	Good output with minor errors
2	Disfluent ungrammatical with correct phrase
1	Incomprehensible

Table 6. Fluency scale

	Sentence length	Fluency	Adequacy
Baseline	<=15 words	1.95	2.24
	>15 words	1.49	1.75
Reordered	<=15 words	2.58	2.75
	>15 words	1.82	1.96
Dependency Relation	<=15 words	2.83	2.91
	>15 words	1.94	2.10

Table 7. Scale of Fluency and Adequacy on sentence length basis of English-Manipuri SMT system

7.2 Manipuri-English SMT System

The system uses the corpus statistics shown in Table 2. The corpus is annotated with the proposed factors. The following models are developed for the experiment. The *baseline* and *lemma+suffix* systems follow same factors as English-Manipuri.

Lemma + Suffix + POS:
Lemma, suffix and POS are used on the source side. The translation steps are (a) lemma to lemma (b) suffix + POS to POS + suffix + dependency relation and generation step is lemma + suffix + POS + dependency relation to surface form.

Model	BLUE	NIST
Baseline (surface)	13.452	4.31
Lemma + Suffix	16.137	4.89
Lemma + Suffix + POS	17.573	5.15

Table 8. Evaluation Scores of Manipuri-English SMT system using various translation factors

Table 8 shows the BLEU and NIST scores of the Manipuri-English systems using the different factors. Table 9 shows the scores of using lexicalized reordering and POS language model.

Model	BLUE	NIST
Baseline + POS LM	14.341	4.52
Baseline + Lexicalized	13.743	4.46
Baseline + Lexicalized +POS LM	14.843	4.71

Table 9. Evaluation Scores of Manipuri-English SMT system using Lexicalized reordering and POS Language Model

Input/Output of Manipuri-English SMT:

(2a) **Input:** স্কুল চৎপা ছাত্রশিংগী তৌদ য়াদ্রবা মথৌনি ।
   *gloss***:** *School chatpa shatra-sing-gi touda yadraba mathouni.*
   Going to school is obligatory for students.
Baseline output: school going to the students important
Lexicalized Reordered output: school going important to the students
Lemma+Suffix+POS+lexicalized reordered output: School going important to the students

(2b) **Input:** কৃষ্ণগী খুত্রা তৌদ্রি অমা লৈ ।
   gloss: Krishna-gi khut-ta toudri ama lei.
   Krishna has a flute in his hand.
Baseline output: Krishna is flute and hand
Lexicalized Reordered output: Krishna flute has his hand

Lemma+Suffix+POS+lexicalized reordered output: Krishna has flute his hand

By considering the lemma along with suffix and POS factors, the fluency and adequacy of the output is better addressed as given by the sample input and output (2a) and (2b) over the baseline system. Using the Manipuri stemmer, the case markers and suffixes are taken into account for different possible word forms thereby helping to overcome the data sparseness problem. Table 10 shows the scores of adequacy and fluency of the evaluation.

	Sentence length	Fluency	Adequacy
Baseline	<=15 words	1.93	2.31
	>15 words	1.51	1.76
Reordered	<=15 words	2.48	2.85
	>15 words	1.83	1.97
Lemma + Suffix + POS	<=15 words	2.86	2.92
	>15 words	2.01	2.11

Table 10. Scale of Fluency and Adequacy on sentence length basis of Manipuri-English SMT system

Subjective evaluations on 100 sentences have been performed for fluency and adequacy. In the process of subjective evaluation, sentences were judged on fluency, adequacy and the number of errors in case marking/morphology. It is observed that poor word-order makes the baseline output almost incomprehensible, while lexicalized reordering solves the problem correctly along with parts-of-speech language model (POS LM). Statistical significant test is performed to judge if a change in score that comes from a change in the system reflects a change in overall translation quality. It is found that all the differences are significant at the 99% level.

8 Discussion

The factored approach using the proposed factors show improved fluency and adequacy at the Manipuri output for English-Manipuri system as shown in the Table 6. Using the Stanford generated relations shows an improvement in terms of fluency and adequacy for shorter sentences than the longer ones.

Input : Khamba pushed the stone with a lever.
থম্বনা জম্ফম্না নুং অদু ইল্লম্মী ।
Outputs:
Syntactic Reordered: থম্ব নুং জম্ফত অদু ইল্লি ।
Khamba nung jamfat adu illi
gloss: Khamba stone the lever push
Dependency: থম্বনা নুং অদু জম্ফম্না ইল্লি ।
Khambana nung adu jamfatna illi
gloss: Khamba the stone pushed with lever

By the use of semantic relation, না (*na*) is attached to থম্ব (*Khamba*), which makes the meaning থম্বনা "by Khamba" instead of just থম্ব "Khamba".
Input : Suddenly the woman burst into tears.
থঙহৌদনা মৌ অদুনা মপি শিল্লরকই ।
Outputs:
Syntactic Reordered: নুপী থুনা পিরাংগা কপ্পী ।
Nupi thuna pirang-ga kappi
gloss: woman soon tears cry
Dependency: অথুবদা নুপীদু কপ্লম্মী ।
Athubada nupidu kaplammi
gloss: suddenly the woman cried

Here, in this example, the নুপী (nupi) is suffixed by the দু (du), to produce নুপীদু "the woman" instead of just নুপী "woman".

The factored approach of Manipuri-English SMT system also shows improved BLEU and NIST scores using the proposed factors as shown in Table 8 not only gain in fluency and adequacy scores as shown in Table 10.

9 Conclusion

A framework for Manipuri and English bidirectional SMT system using factored model is experimented with a goal to improve the translation output and reduce the amount of training data. The output of the translation is improved by incorporating morphological information and semantic relations by tighter integration. The systems are evaluated using automatic scoring techniques BLEU and NIST. The subjective evaluation of the systems is done to find out the fluency and adequacy. The fluency and adequacy are also addressed better for the shorter sentences than the longer ones using semantic relations. The improvement is statistically significant.

References

Avramidis, E. and Koehn, P. 2008. Enriching morphologically poor languages for Statistical Machine Translation, *Proceedings of ACL-08: HLT*

Callison-Burch, Chris., Osborne, M. and Koehn, P. 2006. Re-evaluating the Role of Bleu in Machine Translation Research" *In Proceedings of EACL-2006*

Doddington, G. 2002. Automatic evaluation of Machine Translation quality using n-gram co-occurrence statistics. *In Proceedings of HLT 2002*, San Diego, CA.

Koehn. P., and Hoang, H. 2007. Factored Translation Models, *In Proceedings of EMNLP-2007*

Koehn, P., Hieu, H., Alexandra, B., Chris, C., Marcello, F., Nicola, B., Brooke, C., Wade, S., Christine, M., Richard, Z., Chris, D., Ondrej, B., Alexandra, C., Evan, H. 2007. Moses: Open Source Toolkit for Statistical Machine Translation, *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 177–180, Prague.

Marie-Catherine de Marneffe and Manning, C. 2008. Stanford Typed Dependency Manual

Minnen, G., Carroll, J., and Pearce, D. 2001. Applied Morphological Processing of English, *Natural Language Engineering*, 7(3), pages 207-223

Nieβen, S., and Ney, H. 2004. Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information, *Computational Linguistics*, 30(2), pages 181-204

Och, F. 2003. Minimum error rate training in Statistical Machine Translation , *Proceedings of ACL*

Papineni, K., Roukos, S., Ward, T., and Zhu, W. 2002. BLEU: a method for automatic evaluation of machine translation. *In Proceedings of 40th ACL*, Philadelphia, PA

Popovic, M., and Ney, H. 2006. Statistical Machine Translation with a small amount of bilingual training data, *5th LREC SALTMIL Workshop on Minority Languages*

Ramanathan, A., Choudhury, H., Ghosh, A., and Bhattacharyya, P. 2009. Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*: Volume 2, pages: 800-808

Rao, D., Mohanraj, K., Hegde, J., Mehta, V. and Mahadane, P. 2000. A practical framework for syntactic transfer of compound-complex sentences for English-Hindi Machine Translation, *Proceedings of KBCS* 2000, pages 343-354

Singh, Thoudam D., and Bandyopadhyay, S. 2006. Word Class and Sentence Type Identification in Manipuri Morphological Analyzer, *Proceeding of Modeling and Shallow Parsing of Indian Languages(MSPIL) 2006*, IIT Bombay, pages 11-17, Mumbai, India

Singh, Thoudam D., and Bandyopadhyay, S. 2008. Morphology Driven Manipuri POS Tagger, *In proceedings International Joint Conference on Natural Language Processing (IJCNLP-08) Workshop on Natural Language Processing of Less Privileged Languages (NLPLPL) 2008*, pages 91-98, Hyderabad, India

Singh, Thoudam D., and Bandyopadhyay, S. 2010a. Manipuri-English Example Based Machine Translation System, *International Journal of Computational Linguistics and Applications* (IJCLA), ISSN 0976-0962, pages 147-158

Singh, Thoudam D., Singh, Yengkhom R. and Bandyopadhyay, S., 2010b. Manipuri-English Semi Automatic Parallel Corpora Extraction from Web, *In proceedings of 23rd International Conference on the Computer Processing of Oriental Languages (ICCPOL 2010) - New Generation in Asian Information Processing* , San Francisco Bay, CA, USA, Pages 45-48

Singh, Thoudam D. and Bandyopadhyay, S., 2010c. Statistical Machine Translation of English-Manipuri using Morpho-Syntactic and Semantic Information, *In the proceedings of Ninth Conference of the Association for Machine Translation in Americas (AMTA 2010)*, Denver, Colorado, USA. (To appear)

Stolcke. A. 2002. SRILM - An Extensible Language Modeling Toolkit. *In Proc. Intl. Conf. Spoken Language Processing*, Denver, Colorado, September.

Wang, C., Collin, M., and Koehn, P. 2007. Chinese syntactic reordering for statistical machine translation, *Proceedings of EMNLP-CoNLL*