Coupling Statistical Machine Translation with Rule-based Transfer and Generation

Arafat Ahsan
LTRC, IIIT-Hyderabad
{arafat}@research.iiit.ac.in

Prasanth Kolachina
LTRC, IIIT-Hyderabad
{prasanth_k}@research.iiit.ac.in

Sudheer Kolachina
LTRC, IIIT-Hyderabad
{sudheer.kpg08}@research.iiit.ac.in

Dipti Misra Sharma
LTRC, IIIT-Hyderabad
{dipti}@iiit.ac.in

Rajeev Sangal
LTRC, IIIT-Hyderabad
{sangal}@iiit.ac.in

Abstract

In this paper, we present the insights gained from a detailed study of coupling a highly modular English-Hindi RBMT system with a standard phrase-based SMT system. Coupling the RBMT and SMT systems at various stages in the RBMT pipeline, we observe the effects of the source transformations at each stage on the performance of the coupled MT system. We propose an architecture that systematically exploits the structural transfer and robust generation capabilities of the RBMT system. Working with the English-Hindi language pair, we show that the coupling configurations explored in our experiments help address different aspects of the typological divergence between these languages. In spite of working with very small datasets, we report significant improvements both in terms of BLEU (7.14 and 0.87 over the RBMT and the SMT baselines respectively) and subjective evaluation (relative decrease of 17% in SSER).

1 Introduction

Integrating systems with complementary strengths has been a focal theme in core areas of NLP research recently. In machine translation, quite a few efforts to integrate rule-based and data-driven approaches have been reported in recent literature.

Simard et al. (2007) present Automatic Post-editing using phrase-based SMT as a simple yet effective strategy to combine rule-based and statistical MT technologies. Their approach is quite simple - the source language part of a bilingual corpus (English-French and French-English in this case) is given as input to an RBMT system. The output of this RBMT system along with its manually post-edited counterpart (the reference translation) is then treated as a bilingual corpus, over which a phrase-based SMT system is trained to translate from the former to the latter. Experimenting with the English-French language pair, Simard et al. (2007) report that their automatic post-editing system outperforms both the RBMT and standalone phrase-based SMT systems even when trained on small datasets. Thus, coupled systems are presented as a strategy to counter the effect of small training corpora on MT performance, and to facilitate domain adaptation of MT systems.

Dugast et al. (2007) use a similar combination of an RBMT system and a phrase-based SMT system and demonstrate improvements in translation quality for four other European language pairs. They also present a qualitative analysis of the improvement in the RBMT output achieved using a statistical post-editor. Voss et al. (2008) alternatively combine their in-house lexicon-based MT (LBMT) and RBMT systems with automatic post-editors and report improved performance of both the coupled systems for the syntactically divergent language pair of English and Urdu. Ueffing et al. (2008) explore an integration of the RBMT and SMT approaches using a similar serial system combination for Chinese-English MT. Instead of simply taking the 1-best translation from the RBMT system and giving it as input to the SMT system, they break up the RBMT output into annotated chunks which have confidence values assigned to them. These chunks are then fed into the SMT system.

However, all these above mentioned approaches, essentially treat the RBMT system as a black box, seeking to improve upon its final output. In this paper, we build on this recent work in MT hybridization by exploring different ways of combining RBMT and phrase-based SMT approaches for the typologically divergent English-Hindi language pair. We use a highly modular RBMT system which enables us to combine the two approaches at different stages in the RBMT system's pipeline. This allows us to closely observe the effects of the RBMT's transfer and generation capabilities on the performance of the hybrid system. This mode of combination, we show, helps address the typological differences between English and Hindi.

The rest of this paper is organized as follows. We briefly discuss some issues pertinent to English-Hindi MT in Section 2. In Section 3 we provide an overview of the systems used in our experiments and their possible combinations. In Section 4, we describe the experimental setup followed by results in section 5. Finally, we present our conclusions in section 6.

2 English-Hindi MT

English to Hindi machine translation, in addition to the lack of large-scale training corpora, also grapples with a number of issues owing to the typological divergence between the two languages. In particular, Hindi's rich morphology, extensive use of case-markers, and relatively free word order add to the complexity of English-Hindi MT. We take a closer look at some of these issues below.

2.1 Morphology and Case-Marking

Compared to English, Hindi exhibits highly inflected morphology. Hindi noun phrases not only inflect for gender, number, person, but also take different case-markers depending on their role in a sentence. Similarly, Hindi verb morphology carries agreement features corresponding to the above mentioned nominal inflections; this is in addition to the accompanying auxiliaries which convey tense, aspect and modality (TAM) information. The correct mapping and generation of these features is crucial for the transfer of information across the two languages. For a more detailed discussion on morphology and its role in MT, see Bharati et al. (1995). In recent SMT research, translating from poor to rich morphology has been recognized to be a challenging problem (Avramidis and Koehn, 2008).

In this context, Ramanathan et al. (2008) initially proposed the use of stemmers for separating words from their suffixes. They show this method to improve translation performance and help address the data-sparsity problem of this language pair. In a later work, Ramanathan et al. (2009) refine their method and make use of factored models to generate Hindi suffixes and case-markers from English suffixes and semantic relations. They report an improvement in both BLEU and subjective evaluation scores.

2.2 Local and Long distance Reordering

Owing to the syntactic divergence between English and Hindi, reordering is arguably the most crucial aspect of English-Hindi MT. To address this, a number of recent approaches in phrase-based SMT use various pre-processing techniques that reorder the source side in accordance with the syntactic properties of the target language. Results have shown this approach - of reordering the words in the source sentences to match the target word order and then decoding monotone - to significantly improve the translation quality, especially between syntactically divergent languages. While some of these approaches try to automatically infer local syntactic transformations using POS-tagged bilingual corpora (either on the source, or on both the source and target sides) (Rottmann and Vogel, 2007; Popovic and Ney, 2006); others have proposed more global (long distance/non-local) transformations of the source structure (Collins et al., 2005). For the English-Hindi language pair, Ramanathan et al. (2008) show that the use of hand-crafted rules to reorder English sentences improves translation accuracy.

In our experiments on coupled MT, we use a modular transfer-based MT system working at the level of chunks to tease apart the influence of local and long-distance reordering transformations. This allows us to study the effects of local and long-distance reordering rules independently and identify those reorderings that are most crucial towards improving system performance.

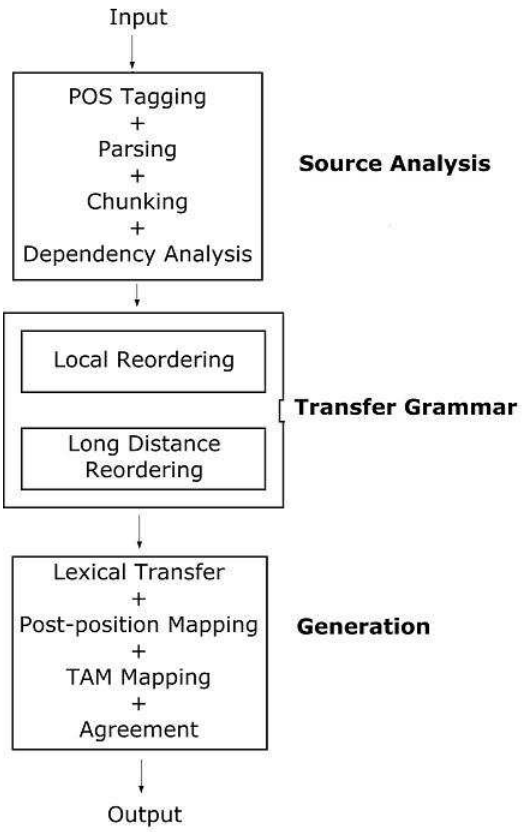

Figure 1: RBMT system architecture

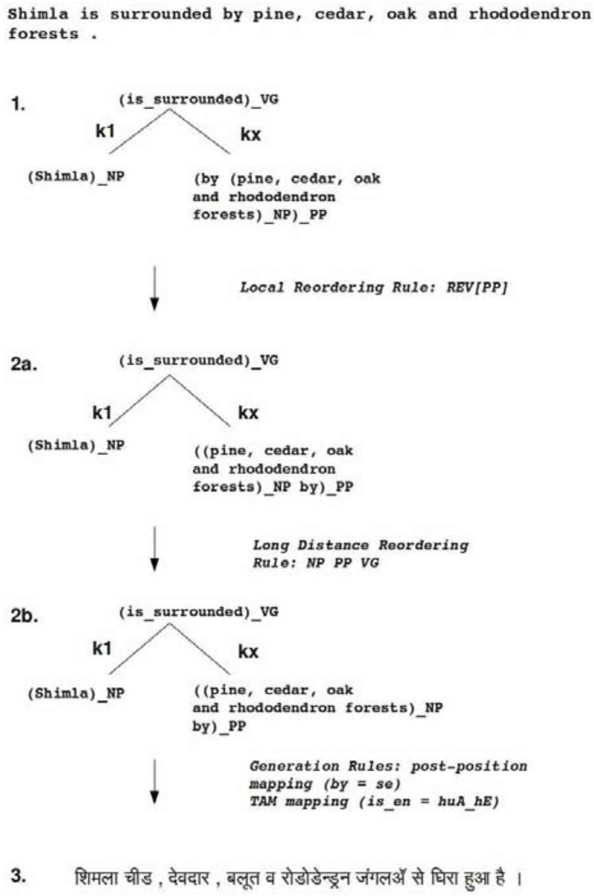

Figure 2: Transformations of input in RBMT phases

3 Description of Systems

3.1 RBMT System

The in-house RBMT system used in our experiments is a highly modular transfer-based English to Hindi MT system which breaks the translation task into a three-stage process (Figure 1).

In the first stage, the source analyzer performs extensive linguistic analysis by running Brill's POS tagger (Brill, 1992) and the Stanford dependency parser (de Marneffe and Manning, 2008) on the input sentence. It then converts the source into a chunk-based unordered dependency tree. This is a labeled tree where the dependency labels are based on the HyDT annotation scheme (Begum et al., 2008) and are derived by mapping them to the labels assigned by the Stanford parser.

In the next stage, the Transfer Grammar performs local and long-distance reorderings. By chunking the source sentences and converting them into a dependency structure, the RBMT system separates local (intra-chunk) reordering decisions from global (inter-chunk) reorderings. This allows for separate specifications of local and long-distance rules; thus, greatly reducing the number of rules that must be written into the grammar. Figure 2 shows the transformations a source sentence undergoes at each stage of the RBMT system and the rules that govern these transformations.

An example of a local transformation is the prepositional phrase (PP) inversion rule of the form REV[PP], which moves a preposition from the initial to the final position inside the PP chunk in accordance with the syntactic structure of Hindi. On the other hand, a long-distance reordering rule works using the dependency relations between different chunks of an unordered source tree. For each matching rule, the tree is mapped to Hindi word-order by explicitly specifying the linear order of the chunks on the target side. In essence, this part of the trans-

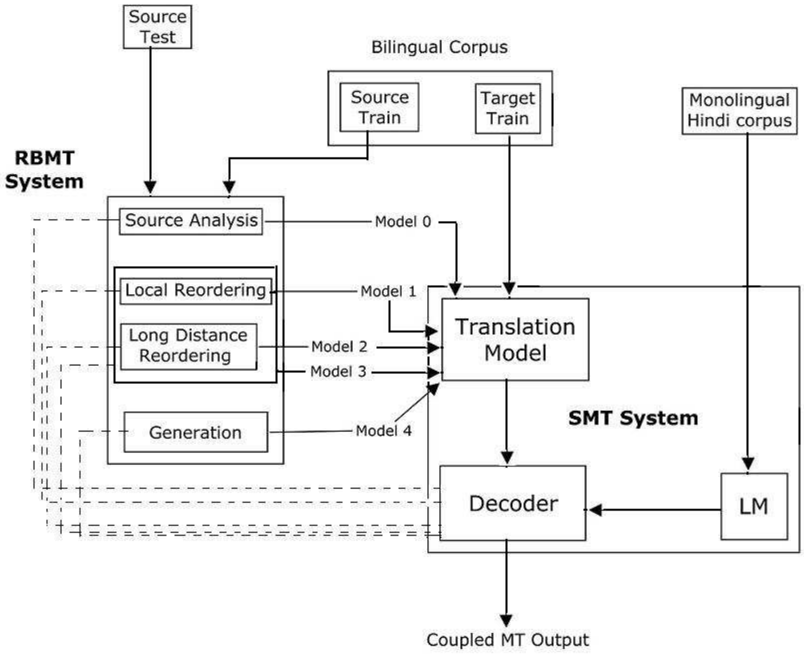

Figure 3: Architecture of the Coupled System

fer grammar can be viewed as tree-to-string transformations. For example, the reordering rule for the source sentence discussed in Figure 2 is as follows:

```
0 ROOT  VG
1 k1    NP
l kx    PP
r NP PP VG
```

The RHS in the above rule (indicated using `r` in the first column) specifies the linear order in Hindi of the three chunks present in the source dependency tree. The transfer grammar component used in our experiments contains around 30 different rules for long-distance reorderings.

In the third and final stage, the generation component of the RBMT system handles the tasks of lexical transfer, agreement, and the mapping and insertion of correct TAMs and post-positions into the target string. In this, it is aided by a number of handcrafted rules and dictionaries. For example, as shown in Figure 2 the English TAM *is_en* is mapped to its Hindi equivalent *huA_hE* via a TAM dictionary and then appended to the verbal root. Post-positions are mapped in a similar manner, with additional rules to disambiguate between them. The modularity of this architecture makes it possible to access the output of the RBMT system at each of these stages.

Building such resource-heavy systems, however, is a tedious and time-intensive task. Moreover, they do not offer the robustness and versatility of SMT systems, which are more easily and readily adaptable to domain specific demands. Therefore, in our approach, we try to combine the respective strengths of both rule-based and statistical systems in a coupled machine translation architecture.

3.2 Phrase-based SMT System

We use the standard phrase-based SMT system as described in Koehn et al. (2003) in our experiments.

3.3 Coupled MT System

Given the highly modular nature of the RBMT system as described in section 3.1, it is possible to couple the RBMT and the phrase-based SMT systems at different stages of processing. Figure 3 shows the architecture of such a coupled system at each stage of the RBMT pipeline.

The following is the nomenclature used for the standalone and coupled systems in our experiments:

- RBMT-baseline: Standalone RBMT system

- SMT-baseline: Standalone SMT system

- Model-0: Coupled baseline model with input from RBMT source analysis phase

- Model-1: Coupled reordering model with local reordering only

- Model-2: Coupled reordering model with long distance reordering only

- Model-3: Coupled reordering model with both local and long distance reordering

- Model-4: Coupled model with input from RBMT generation phase (similar to Automatic/Statistical Post-editing serial combination)

Within this architecture, the source side of the training corpus (S, T) undergoes transformations depending upon the RBMT stage at which the two systems are coupled together. Table 1 lists the training corpora derived at each stage in the RBMT pipeline.

4 Experimental Setup

4.1 Datasets

The corpus used in our experiments is a tourism domain corpus, part of which was released during the ICON-2008 NLP tools contest for SMT[1]. Table 2 lists the corpus details[2].

4.2 Description of Experiments

We begin by establishing the baselines for the standalone SMT and RBMT systems. Our baseline SMT model is a standard phrase-based model trained on the parallel corpus (S, T). We use the Moses toolkit (Koehn et al., 2007) to build all our phrase-based models. The reordering feature used for the baseline system is *msd-bidirectional-fe*, which allows for all reorderings over a specified distortion limit. In this case, the distortion limit was fixed at 6 after a few initial experiments.

[1] `http://ltrc.iiit.ac.in/icon2008/nlptools.php`

[2] The original test set in this corpus contained 500 sentences. However, after checking the training and test sets we found a large number of sentences in the test and training sets to be duplicates. The sizes we report here are after elimination of all duplicates.

Model	Training Corpus	RBMT phase
RBMT baseline	none	None
SMT baseline	(S, T)	None
Model-0	(S_0, T)	Source analysis
Model-1	(S_1, T)	Local reordering
Model-2	(S_2, T)	Long-distance reordering
Model-3	(S_3, T)	Local+long-distance reordering
Model-4	(S_4, T)	Generation

Table 1: Training corpus transformations.

While establishing the baseline for the standalone SMT system, we initially experimented with factored models using factors such as lemma, lexical category and POS tag to obtain the best possible alignments. However our observation is that the use of these factors (in all possible combinations) does not help improve upon the baseline achieved by using surface to surface alignments. Although the use of factors has been reported to improve the performance of English-Hindi SMT (Ramanathan et al., 2009), the gains seem to be limited to factored generation. Since in our experiments, we were looking to test the effect of the RBMT system's generation capabilities, we chose to work with unfactored models after these initial tests.

The goal of all our experiments on coupling the RBMT and SMT approaches is to study the effects of this serial combination at various stages in the RBMT pipeline.

In the first experiment on coupled systems, the source part of the parallel corpus (S, T) is passed through the source analysis phase in the RBMT pipeline to obtain the transformed source S_0 (see Fig. 3). This first coupled model (Model-0) is now trained over (S_0, T). Comparing the performance of Model-0 against the SMT baseline, allows us to infer the cost of coupling the RBMT and SMT approaches. The setup in all the experiments is essentially the same with variations only on the source side of the parallel corpus, depending on the stage at which the systems are combined.

The next set of experiments are to study the effect of the RBMT system's transfer grammar on the coupled models. Model-1 in which, the source is reordered using only local reordering rules from the

Corpus	sentences	source words	target words
training	8169	0.17M	0.18M
tuning	358	7741	7992
test	241	5439	5552
Monolingual Hindi(LM)	11300	n.a	0.30M

Table 2: Corpus Statistics

4	Perfect
3	Comprehensible with occasional errors
2	Comprehensible with quite a few errors
1	Some parts make sense but not comprehensible overall
0	Nonsense

Table 3: Grading scale for subjective evaluation

baseline systems	BLEU	SSER
RBMT baseline	4.30	83.2
SMT baseline	10.57	80.8

Table 4: Standalone baseline systems

grammar, is trained on the (S_1, T) pairing of the parallel corpus. Similarly, the effect of transforming the source using long-distance transfer rules is studied by training Model-2 on (S_2, T). Model-3 trained over (S_3, T) is meant to study the cumulative effect of both local and long-distance reordering rules applied together.

In our final experiment on coupling, we train Model-4 over (S_4, T). Here S_4 is the output of the RBMT system from the generation phase. Note that this is also the *final* output of the RBMT system. Model-4 is very similar to the Statistical post-editor/Serial System Combination proposed in previous work (Simard et al., 2007; Ueffing et al., 2008). The main aim of this experiment is to study the effect of the generation phase of the rule-based system. Since the RBMT system has a robust paradigm-based generation component that efficiently handles morphological transformations, our intuition is that it will help address the problem of translating from the poorer morphology of English to Hindi.

During tuning and decoding, the source side data undergo the same pre-processing (transformations) as the training data for each model. Decoding on all the coupled models is performed *monotone*. A trigram Language Model, common to all the SMT systems described in the experimental setup, is built using the SRILM toolkit. The Kneser-Ney method (Goodman and Chen, 1999) is used for smoothing.

5 Results and Discussion

We evaluate the output of the MT models in our experiments using the BLEU automatic evaluation metric (Papineni et al., 2002) as well as human subjective evaluation. For subjective evaluation, a subset[3] of the MT output was assessed by 10 native speakers of Hindi using a grading scale of 0-4 shown in Table 3. Subjective Sentence Error Rate (SSER) was estimated using the method described in Nießen et al. (2000). The use of subjective evaluation metrics like SSER to evaluate the output of machine translation systems is all the more pertinent in case of hybrid systems, since BLEU seems to consistently underestimate the gains obtained through such system combinations (Ueffing et al., 2008).

In Table 4, the BLEU scores of the two baseline standalone systems are shown along with their corresponding error rates. Note that the performance of the RBMT system in terms of BLEU is quite low when compared to the SMT baseline. This is understandable as the RBMT output is generated *independently* using built-in bilingual dictionaries for lexical transfer. As a result, the lexical choices in the RBMT and SMT output tend to be radically different, resulting in lower BLEU scores when the RBMT output is compared against reference gold translations. The SSER scores on the other hand, estimate the performance of the two systems to be much closer than predicted by the automatic evaluation metric.

The first coupled model, Model-0 registers a drop of about 0.48 BLEU compared to the standalone SMT baseline (See Table 5). However, according to the subjective evaluation metric, the system in fact seems to have improved. The possible reason for this discrepancy between the two metrics could be, that the RBMT system processes the source through

[3] a randomly selected set of 100 sentences

Model	BLEU	SSER
RBMT baseline	4.30	83.2
SMT baseline	10.57	80.8
Model-0	10.09	80.0
Model-1	10.57	79.0
Model-2	10.44	74.8
Model-3	11.01	71.4
Model-4	11.44	68.2

Table 5: Experiment Results: Coupled systems

a number of different modules. For example, in addition to modules such as a POS-tagger, a chunker and a parser, as described in section 3.1, the source analysis phase also consists of a number of other modules, such as a *collocations-identifier*, which on having identified a collocation replaces it with its Hindi equivalent in this very phase. This is to prevent it from being split at a later stage. However for Model-0 trained over the parallel corpus (S_0, T) built from this phase, these substitutions lead to a drop in BLEU score when compared to the baseline system, trained over the untransformed corpus (S, T). This drop can be treated as an initial cost of coupling the two systems together.

The next three coupled models (Model-1, Model-2, Model-3) capture the effect of the structural transfer component of the RBMT on the coupled SMT system. Model-1 trained over (S_1, T) registers an improvement over Model-0 by almost 0.48 BLEU, drawing level with the standalone SMT baseline. According to the automatic evaluation metric these gains are substantial. However, in terms of subjective evaluation, local reorderings (the structural transformations this model aims to capture) do not seem to have that significant an effect on fluency. This is also quite apparent from the sample output of the model given in Table 6. Although the preposition has moved to the end of the noun phrase, this alone, does not help improve the readability of the sentence.

The BLEU and SSER scores obtained for Model-2 almost imply the opposite. While the error-rate decreases by 5.3% (relative to Model-1), this model scores lower than Model-1 in terms of BLEU. The failure of the RBMT system to perform any reorderings in some cases (when no long-distance rules are matched) while building S_2, and also during preprocessing of the test data before decoding, might have affected the BLEU score for this model. However, in the cases where it does reorder, it scores high on the subjective evaluation metric. This seems to suggest, that a combination of both local and long-distance rules is necessary to achieve overall improved performance.

In Model-3, where local and long-distance reordering rules in the transfer grammar are both applied to transform the source side, the performance of the coupled MT system betters the SMT baseline by around 0.4 BLEU. This observation is in agreement with conclusions of previous work (Ramanathan et al., 2008; Ramanathan et al., 2009) that reordering (or restructuring) the source side using structural information has a positive effect on the performance of the system. The corresponding decrease in SSER further adds weight to this observation. However, we are unable to directly compare our scores with the previous work on this language pair owing to the differences in corpus domains and sizes.

In the case of Model-4, the SMT system was trained over (S_4, T), where S_4 was the output of the RBMT system. This combination of RBMT and SMT approaches has been discussed in literature as an Automatic Post-editing System. This serial system combination gives the best performance compared to all the other hybrid configurations. Model-4 outperforms the RBMT and SMT baselines by 7.14 and 0.87 BLEU respectively. This improvement seems significant given the small size of the datasets available to us. In terms of SSER, a relative decrease of 17% in the error rate is also substantial. This improved performance of Model-4 over Model-3 can be attributed only to the sophistication of the generation phase in the RBMT pipeline. During the generation phase, the information obtained during source analysis is used to generate the appropriate morphological forms. Note the correct mapping of the TAM and the post-position in the sample output of this model in Table 6. This alone, leads to a dramatic improvement in the quality of the output as indicated by the SSER scores. Thus, although the source analysis phase in itself does not contribute much towards improving the system performance, the information obtained by the RBMT system during this phase is capitalized upon during generation.

Model	Sample output sentence
Source Input	shimla is surrounded by pine , cedar , oak and rhododendron forests .
Reference	शिमला चीड , देवदार , बलूत व रोडोडेन्ड्रन जंगलों से घिरा है । SimalA cIda , xevaxAra , balUwa va rododendrana jaMgaloM se GirA hE . Shimla pine , cedar , oak and rhododendron forests by surrounded is .
RBMT baseline	शिमला चीड का पेड , देवदार , शाहबलूत और रोडोडेन्ड्रन जंगलों के द्वारा shimla cIda kA peda , xevaxAra , SAhabalUwa Ora rhododendron jaMgaloM ke xvArA Shimla pine of tree , cedar , oak and rhododendron forests by चारों ओर से घेरा गया है । cAroM ora se GerA gayA hE . four sides on is surrounded .
SMT baseline	शिमला से घिरा हुआ है चीड , देवदार ,ओक और बुरुंश जंगलों है । SimalA se GirA huA hE cIda , xevaxAra , oka Ora buruMSa jaMgaloM hE . Shimla by surrounded is pine , cedar , oak and rhododendron forests is .
Model-0	शिमला से घिरा हुआ है चीड , देवदार , सुंदर शाहबलूत और बुरुंश जंगलों है । SimalA se GirA huA hE cIda , xevaxAra , suMxara SAhabalUwa Ora buruMSa jaMgaloM hE . Shimla by surrounded is pine , cedar , beautiful oak and rhododendron forests is .
Model-1	शिमला से घिरा है चीड , देवदार , सुंदर शाहबलूत और बुरुंश जंगलों से है । SimalA se GirA hE cIda , xevaxAra , suMxara SAhabalUwa Ora buruMSa jaMgaloM se hE . Shimla by surrounded is pine , cedar , beautiful oak and rhododendron forests by is .
Model-2	शिमला द्वारा चीड , देवदार ,ओक और बुरुंश जंगलों से घिरा है । SimalA xvArA cIda , xevaxAra , oka Ora buruMSa jaMgaloM se GirA hE . Shimla by pine , cedar , oak and rhododendron forests by surrounded is .
Model-3	शिमला चीड , देवदार ओक और बुरुंश जंगलों से घिरा है । SimalA cIda , xevaxAra oka Ora buruMSa jaMgaloM se GirA hE . Shimla pine , cedar oak and rhododendron forests by surrounded is .
Model-4	शिमला चीड , देवदार ओक और बुरुंश जंगलों से घिरा हुआ है । SimalA cIda , xevaxAra oka Ora buruMSa jaMgaloM se GirA huA hE . Shimla pine , cedar oak and rhododendron forests by surrounded is .

Table 6: Sample Output of different models

To conclude, the results of the various models in coupled combinations show the incremental gains made by the systems at each stage of the source processing. Local reorderings alone help draw the system level with the baseline SMT, while combining both local and long-distance rules further improves the performance of the coupled system. We observe that although SMT systems are able to handle local reorderings by themselves, in case of long-distance transformations, they benefit considerably from external guidance (the RBMT system in this case). These results are indicative of the importance of structural transfer and generation in English-Hindi machine translation.

Table 6 shows the output of the various coupled MT models studied in our experiments on a sample input sentence. The output translations, in addition to the devanagari text, are also transcribed using a roman script[4] and the word gloss is provided below each translation.

6 Conclusions and Future work

In this paper we proposed a coupled machine translation architecture and observed incremental gains in the performance of the coupled systems at each stage of the coupling. There were significant improvements in terms of both automatic evaluation and subjective evaluation metrics. This tight combination of a rule-based and statistical phrase-based system also helped bring into focus the areas that are most crucial towards further improving English-

[4]Notation for Hindi transcription can be found at `http://ltrc.iiit.ac.in/MachineTrans/research/tb/map.pdf`

Hindi machine translation.

Our future efforts would focus on the possibility of adapting this combination to the tree-based SMT framework since our source side is richly annotated with dependency annotation, which is currently being indirectly used by the RBMT system for reordering purposes. This course of action can prove to be particularly useful in further improving reorderings in an SMT model.

References

E. Avramidis and P. Koehn. 2008. Enriching morphologically poor languages for statistical machine translation. In *Proc. of ACL*.

R. Begum, S. Husain, A. Dhwaj, D.M. Sharma, L. Bai, and R. Sangal. 2008. Dependency annotation scheme for Indian languages. *Proceedings of International Joint Conference on Natural Language Processing*.

Akshar Bharati, Vineet Chaitanya, and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. Prentice Hall of India, Delhi.

E. Brill. 1992. A simple rule-based part of speech tagger. In *Proceedings of the third conference on applied natural language processing*, pages 152–155. Association for Computational Linguistics.

M. Collins, P. Koehn, and I. Kucerova. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*, volume 43, page 531.

M.C. de Marneffe and C.D. Manning. 2008. Stanford typed dependencies manual.

L. Dugast, J. Senellart, and P. Koehn. 2007. Statistical post-editing on SYSTRAN's rule-based translation system. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 220–223. Association for Computational Linguistics.

J. Goodman and S. Chen. 1999. An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 13(4):359–393.

P. Koehn, F.J. Och, D. Marcu, et al. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, volume 1, pages 48–54.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Annual meeting-association for computational linguistics*, volume 45, page 2.

S. Nießen, F.J. Och, G. Leusch, and H. Ney. 2000. An evaluation tool for machine translation: Fast evaluation for MT research. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 39–45.

K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

M. Popovic and H. Ney. 2006. POS-based word reorderings for statistical machine translation. In *Proc. of the Fifth Int. Conf. on Language Resources and Evaluation (LREC)*.

A. Ramanathan, P. Bhattacharyya, J. Hegde, R.M. Shah, M. Sasikumar, and C. Mumbai. 2008. Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing*. Citeseer.

A. Ramanathan, H. Choudhary, A. Ghosh, and P. Bhattacharyya. 2009. Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT. In *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-Volume 2*, pages 800–808. Association for Computational Linguistics.

K. Rottmann and S. Vogel. 2007. Word reordering in statistical machine translation with a POS-based distortion model. *TMI. Sk
"ovde, Sweden*.

M. Simard, C. Goutte, and P. Isabelle. 2007. Statistical phrase-based post-editing. In *Proceedings of NAACL HLT*, pages 508–515.

N. Ueffing, J. Stephan, E. Matusov, L. Dugast, G. Foster, R. Kuhn, J. Senellart, and J. Yang. 2008. Tighter integration of rule-based and statistical MT in serial system combination. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 913–919. Association for Computational Linguistics.

C.R. Voss, M. Aguirre, J. Micher, R. Chang, J. Laoudi, and R. Hobbs. 2008. Boosting Performance of Weak MT Engines Automatically: Using MT Output to Align Segments & Build Statistical Post-Editors. In *Proceedings of EAMT 2008: 12th annual conference of the European Association for Machine Translation*, pages 192–201.